Indira Gandhi National Open University
School of Sciences

MST-016
STATISTICAL INFERENCE

Block
2

PROPERTIES OF GOOD ESTIMATOR

UNIT 6
Unbiasedness 163

UNIT 7
Consistency 181

UNIT 8
Efficiency and Mean Squared Error 201

UNIT 9
Sufficiency and Minimal Sufficiency 221

BLOCK 1: Sampling Distributions

Unit 1: Basic Concepts of Sampling Distribution

Unit 2: Sampling Distributions of Sample Means

Unit 3: Sampling Distributions of Sample Proportions and Variances

Unit 4: Sampling Distributions Associated with Normal Populations-I

Unit 5: Sampling Distributions Associated with Normal Populations-II

BLOCK 2: Properties of Good Estimator

Unit 6: Unbiasedness

Unit 7: Consistency

Unit 8: Efficiency and Mean Squared Error

Unit 9: Sufficiency and Minimal Sufficiency

BLOCK 3: Methods of Estimation

Unit 10: Method of Maximum Likelihood Estimation

Unit 11: Other Methods of Point Estimation

Unit 12: Interval Estimation for Means

Unit 13: Interval Estimation for Proportions

Unit 14: Interval Estimation for Variances

BLOCK 4: Testing of Hypothesis: Parametric Tests

Unit 15: Basic Concepts of Testing of Hypothesis

Unit 16: Tests for Means

Unit 17: Tests for Proportions

Unit 18: Tests for Variances

BLOCK 2: PROPERTIES OF GOOD ESTIMATOR

In Block 1 of this course, you have studied the sampling distributions of various statistics such as sample mean, difference of two sample means, sample proportion, difference of two sample proportions, sample variance, and ratio of two sample variances. Also, have studied some standard sampling distributions such as chi-square, t, and F-distributions that provide a platform to draw inferences about the population parameters on the basis of the samples.

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 all possible estimators and the second is to derive some methods or techniques through which we obtain an estimator which possesses such properties. This block is devoted to explaining the criteria of a good estimator.

This block comprises four units.

Unit 6: Unbiasedness is devoted to explaining the concept of estimation (point and interval estimation) with the first property of a good estimator, i.e. unbiasedness. The properties of an unbiased estimator are also described in this unit.

Unbiasedness property is defined for a fixed sample size. In **Unit 7: Consistency,** you will learn about consistency which is defined for increasing sample size. Here, we describe the concept of consistency and its asymptotic distribution with a suitable normalisation with examples. We also explain the properties of the consistent estimator.

There may exist more than one unbiased estimator of a parameter, therefore, to check which one is better, we explain the concept of efficiency. **Unit 8: Efficiency and Mean Squared Error** is devoted to explaining the concept of efficiency, mean squared error and minimum variance unbiased estimator which help us to compare estimators and make the decision which one is better.

In the continuation of finding the best estimator, we introduce the concept of sufficiency in **Unit 9: Sufficiency and Minimal Sufficiency**. In this unit, you will study the concept of sufficient and minimal sufficient estimators with examples. The Fisher-Nayman Factorization theorem for finding sufficient estimators is also explained in this unit.

Expected Learning Outcomes

After completing this block, you should be able to:

- define the parameter space and describe the properties of a good estimator;
- explain the unbiasedness, consistency, efficiency and sufficiency properties of a good estimator;
- check whether an estimator is unbiased, consistent, efficient or sufficient ;
- describe the properties of unbiased, consistent, efficient and sufficient estimators;
- describe the Fisher-Nayman Factorization theorem and how to use it to find sufficient statistic; and
- explain the concept of a consistent asymptotically normal estimator; mean squared error, most efficient estimator, minimum variance unbiased estimator, and minimal sufficient statistic.

Notations and Symbols

SAQ/TQ	:	Self Assessment Question/Terminal Question
Fig./Figs.	:	Figure/Figures
$X_1, X_2, \ldots, X_n$	:	A random sample of size n
$\bar{X}$	:	Sample mean
S^2	:	Sample variance
μ and σ^2	:	Mean and variance of a population
E(X) and Var(X)	:	Mathematical expectation and variance of X
$Z \sim N(0, 1)$	:	Standard normal variate
P and p	:	Population and sample proportion
$B(a,b) = \frac{\Gamma a \, \Gamma b}{\Gamma a+b}$	:	Beta function
Γa	:	Gamma function
Θ		Parametric Space
$T = t(X_1, X_2, \ldots, X_n)$	:	Estimator
e and MSE	:	Efficiency and mean squared error
$g[t(x), \theta]$	:	Non-negative function of the parameter θ and observed sample values
$h(x_1, x_2, \ldots, x_n)$	:	Non-negative function of observed sample values

UNIT 6

UNBIASEDNESS

Structure

6.1 Introduction

Expected Learning Outcomes

6.2 Basic Terminology

6.3 Properties of Good Estimator

6.4 Unbiasedness

6.5 Properties of Unbiased Estimator

6.6 Summary

6.7 Terminal Questions

6.8 Solutions /Answers

6.1 INTRODUCTION

In many real-life problems, the population parameter (characteristic of the population) is not known and someone is interested in obtaining the value of the parameter. But, if

- the whole population is too large to study,
- the units of the population are destructive in nature,
- there are limited resources and manpower available, etc.

then it is not practically convenient to examine each unit of the population to find the value of the parameter. For example, as you know many of us use Facebook and you are interested to know the average age of the people who use Facebook. However, the true value (average age) of Facebook users is not known. The only way to know the true average age of Facebook users is to survey each and every person in the world who uses Facebook. But it is not possible to survey everyone in the world. In such a situation, one can select randomly some persons who use Facebook and note their age. Suppose we randomly selected 20 Facebook users and obtained the following data of their age (in years):

20	42	36	30	20	52	32	18	70	22
45	18	40	16	18	20	30	19	41	20

If we use the sample average age to estimate the unknown average age of the Facebook users, then we get an estimate of the same as

Tools You Will Need

The following terms are considered essential background material for this Unit. If you doubt your knowledge of any of these terms, you should review the appropriate Unit or section before proceeding:

- Sampling distributions (Units 2,3, 4 and 5).
- Probability distributions (MST-012).

Unit Writer- Dr. Prabhat Kumar Sangal, School of Sciences, IGNOU, New Delhi

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{607}{20} = 30.25$$

Estimating is not something new to us. Every one of us uses an estimate in our day-to-day life. Some situations are as follows:

- At the metro station of Delhi, a guard may estimate the height of a child to be 3 feet or longer.
- Lavnik estimates the time to reach his school from home is about 20 minutes.
- A family estimates the monthly expenditure on the basis of particular needs.
- The distance between New Delhi and Gujrat is approximately 1112 km.

So, the question is, what is an estimation?

The technique of finding an estimator to produce an estimate (approximate value) of the unknown parameter of the population on the basis of a sample is called estimation.

Any statistic used to estimate an unknown population parameter is known as **estimator** and the particular value of the estimator is known as **estimate** of parameter. The estimated value of sample mean and sample variance are denoted by $\bar{X}$ and S^2, respectively.

As its name suggests, the objective of estimation is to determine the approximate value of a population parameter on the basis of a sample statistic.

There are two methods of estimation:

1. Point Estimation
2. Interval Estimation

In point estimation, we determine an appropriate single statistic whose value is used to estimate the unknown parameter whereas, in interval estimation, we determine an interval that contains the true value of the unknown parameter with a certain confidence. For example, in the case of Facebook users, we get the point estimate as 30.25 years because we estimated it by only one value (30.25 years) whereas if we estimate the same as the age group (18, 34) uses Facebook then it is an interval estimation because we estimated it by using an interval (18, 34) age. The point estimation and interval estimation are briefly described in Units 10-11 and Units 12-14, respectively.

Estimation admits two problems:

- The first is to select some criteria or properties such that if an estimator possesses these properties then it is called the best estimator among possible estimators, that is, properties of a good estimator, and
- The second is to derive some methods or techniques through which we obtain an estimator which possesses such properties, that is, methods of estimation.

Units 6, 7, 8 and 9 are devoted to describing the properties of a good estimator in detail, however, Units 10, 11, 12, 13 and 14 explain the methods of estimation.

This unit is divided into nine sections. Section 6.1 is introductory in nature. The basic terms used in estimation are defined in Section 6.2. Section 6.3 is devoted to explaining the criteria of a good estimator. Section 6.4 explores the concept of unbiasedness with examples. The properties of an unbiased estimator are described in Section 6.5. The unit ends by providing a summary of what we have discussed in this unit in Section 6.6. The terminal questions and the solution of the SAQs/TQs are given in Sections 6.7 and 6.8, respectively.

In the next unit, we shall discuss the second characteristic of a good estimator, that is, consistency.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ define the parameter space;
- ❖ describe the properties of a good estimator;
- ❖ explain the unbiasedness characteristic of an estimator;
- ❖ check whether an estimator is unbiased or not; and
- ❖ describe the properties of an unbiased estimator.

6.2 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 the theory of estimation discussed.

Discrete and Continuous Distributions

In Units 9 to 16 of MST-012, we have discussed standard discrete and continuous distributions as binomial, Poisson, normal, exponential, etc. We know that a population can be described with the help of a distribution, therefore, standard discrete and continuous distributions are also used in statistical inference. Here, we discuss some standard discrete and continuous distributions in brief as in tabular form and you have to learn at least the mean and variance of these distributions which will pay you to do the estimation questions easily.

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

6	Geometric (discrete) $P[X=x]=p(1-p)^x;\quad x=0,1,2,...$	p	$\frac{p}{(1-p)}$	$\frac{p}{(1-p)^2}$
7	Negative Binomial (discrete) $P[X=x]=\binom{x+r-1}{r-1}p^r(1-p)^x;\quad x=0,1,2,\ldots$	r & p	$\frac{rp}{(1-p)}$	$\frac{rp}{(1-p)^2}$
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) $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2};\quad -\infty<x<\infty$	--	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) $f(x)=\theta e^{-\theta x};\quad x\geq 0;\ \theta>0$ Negative Exponential or simply exponential (continuous) $f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}};\quad x\geq 0;\ \theta>0$	θ θ	$\frac{1}{\theta}$ θ	$\frac{1}{\theta^2}$ θ^2
12	Gamma (continuous) $f(x)=\frac{b^a}{\Gamma a}e^{-bx}x^{a-1};\ x>0; a,b>0$	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) $f(x)=\frac{1}{B(a,b)}\frac{x^{a-1}}{(1+x)^{a+b}};\ x>0;\ a,b>0$	a & b	$\frac{a}{b-1}$	$\frac{a(a+b+1)}{(b-1)^2(b-2)}$
15	Standard Cauchy $f(x)=\frac{1}{\pi(1+x^2)};\ -\infty<x<\infty$	---	Does not exist	Does not exist
16	Laplace $f(x)=\frac{1}{2b}e^{-\left\lvert\frac{x-\mu}{b}\right\rvert};\ -\infty<x<\infty$	μ & b	μ	$2b^2$

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 finding the parameter space of a parameter, we have to think all possible values of the parameter yet the chance of these is very very small. For example, suppose the parameter θ represents the average life of electric bulbs manufactured by a company. Since the bulb can be fused at the initial time 0 or at 1, 2, 2.3, 3 hours, and so on, therefore, it lies from 0 to ∞. Hence, the parameter space of the average life of the bulbs, that is, θ is $\Theta = \{\theta : \theta \geq 0\}$.

It means that the parameter average life θ can take all possible values greater than or equal to 0, Similarly, in a normal distribution (μ, σ^2), the parameter space of parameters μ and σ^2 is $\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty;\ \ 0 < \sigma < \infty\}$.

Mathematical Expectation

If X is a continuous random variable having the probability density function $f(x)$, then the expected value of X (mean) is defined as

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx \text{ and } E(X^r) = \int_{-\infty}^{\infty} x^r\, f(x)\, dx$$

If X is a discrete random variable having the probability mass function $p(x)$, then the expected value of X is defined as

$$E(X) = \sum_{i=1}^{n} x_i\, p(x_i) \text{ and } E(X^r) = \sum_{i=1}^{n} x_i^r\, p(x_i)$$

Some properties of mathematical expectation are:

- $E(a) = a$ where 'a' is a constant
- $E(aX) = aE(X)$
- $E(aX \pm bY) = aE(X) \pm bE(Y)$

Variance

If X is a random variable then the variance of X in terms of expectation is defined as

$$Var(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

Some properties of variance are:

- $Var(a) = 0$
- $Var(aX) = a^2 Var(X)$
- If random variables X and Y are independent, then

  $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y)$

Now, try the following Self Assessment Question.

SAQ 1

If θ represents the average marks (out of 50) of the learner in the Term-End-Exam paper of the MST-016 course, then find the parameter space of θ.

After understanding the basic definition and terminology which will help you to understand the properties of a good estimator. We now finally discuss the properties of a good estimator in the next section.

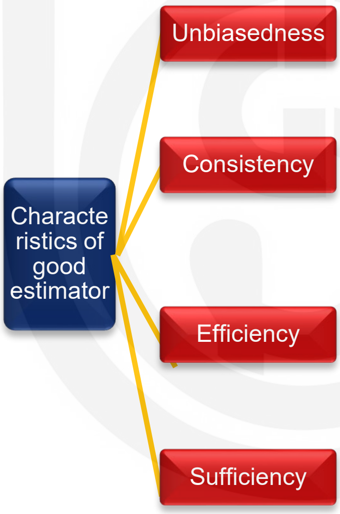
6.3 PROPERTIES OF GOOD ESTIMATOR

It is to be noted that a large number of estimators can be proposed for an unknown parameter. For example, in our case of estimating the average age of Facebook users, some possible estimators are:

- Sample mean $\overline{X} = \dfrac{607}{20} = 30.25$
- Sample median $\tilde{X} = \dfrac{22+30}{2} = 26$
- Sample mode $X_0 = 20$
- Average of extreme users $= \dfrac{\max + \min}{2} = \dfrac{70+14}{2} = 42$

Now, the questions arise,

- Which estimator should you use, that is, which is likely to give estimates closer to the true (but unknown) population value?
- Are some of the possible estimators better, in some sense, than the others?”



In general, an estimator whose sampling distribution concentrates as closely as possible near the true value of the parameter may be regarded as a good estimator. To give the answer to the above questions, Prof. Ronald A. Fisher gave some properties of a good estimator which are as follows:

- Unbiasedness
- Consistency
- Efficiency
- Sufficiency

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

Now, answer the following Self Assessment Question.

SAQ 2

Write properties of a good estimator.

We now discuss the first characteristic of a good estimator in the next section.

6.4 UNBIASEDNESS

Any statistic which is used to estimate an unknown population parameter is known as **estimator.**

In the previous units, you have studied that any statistic such as sample mean, sample variance, sample proportion, etc. which is used to estimate an unknown population parameter is known as an estimator. You also saw that the value of any estimator changes from sample to sample, therefore, we consider the estimator as a random variable and we can find the mean and variance of the estimator. So we can define an estimator as an unbiased estimator as:

An estimator is said to be unbiased for a population parameter if and only if the average or mean of the sampling distribution of the estimator is equal to the true value of the parameter. This property of the estimator is called unbiasedness.

Let us see some examples,

- In Unit 1, you have seen that the mean of the sampling distribution of the sample mean of monthly salary of the employees is equal to the mean salary of all employees of the industry. So sample mean is an unbiased estimate of the population mean.
- Similarly, in Unit 3, we saw that the mean of the sample proportions of the children who like to dance is equal to the population proportion. Therefore, sample proportion is an unbiased estimate of the population proportion.

In general, we denote any population parameter such as a population mean, population standard deviation, population proportion, and so on by the Greek letter theta θ, and its estimator such as the sample mean, sample standard deviation, and sample proportion by T or $\hat{\theta}$ (pronounced as **"theta-hat"**).

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 an unbiased estimator of the parameter θ if and only if

$$E(T) = \theta$$

for all possible values of the parameter θ.

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.

However, if the expected value of the estimator does not equal to the true value of the parameter, then the estimator is said to be a **"biased estimator"**, that is, if

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

then the estimator T is called the biased estimator of θ.

We can also define bias as

The distance between the estimate obtained from a sample and the actual value of the population parameter from which the sample was taken is called bias.

The amount of biases 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 the parameter θ.
- If b(θ) < 0 or E(T) < θ, then the estimator T is said to be negatively biased for the parameter θ.
- If $E(T) \to \theta$ as $n \to \infty$, that is, if an estimator T is unbiased for a large sample only then the estimator T is said to be asymptotically unbiased for θ. For example, suppose $E(T) = \theta + \frac{1}{n}$ then as $n \to \infty, E(T) \to \theta$.

An unbiased estimator is generally preferred in comparison to a biased estimator.

Now, we explain the procedure to show whether an estimator is unbiased or not for a parameter with the help of some examples.

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

Solution: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from any population with mean μ. We have to show that the sample mean $\bar{X}$ is an unbiased estimator for μ, therefore, we have to find $E(\bar{X})$ and check whether it is equal to μ or not. That is,

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

Consider,

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

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

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from the same population with mean μ and variance σ^2, therefore,

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

Thus,

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

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

Example 2: Suppose the speed of lightweight vehicles on a particular stretch of roadway is normally distributed with a known standard deviation of 5 kph. A researcher measured the speed of 10 lightweight vehicles randomly and obtained the following results:

Vehicle	1	2	3	4	5	6	7	8	9	10
Speed (in kph)	62	70	65	68	64	65	70	64	55	60

(i) Find the point estimate of the average speed.

(ii) Show that the sample average speed is an unbiased estimator of the average speed of all the lightweight vehicles on the roadway.

Solution: Generally, to draw the inference about the population mean, we use the sample mean, therefore, to find the point estimate of the average speed, we use the sample mean.

We can obtain the point estimate of average speed as

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

$$= \frac{62+70+65+68+64+65+70+64+55+60}{10} = 64.3$$

Now, we have to show that the sample average speed is an unbiased estimate of the average speed of all lightweight vehicles on the roadway.

Since the speed of the vehicles is normally distributed and standard deviation (σ) is known, therefore, the sample average speed also follows a normal distribution with mean μ and variance σ^2/n.

Thus, $E\left(\bar{X}\right) = \mu$

Hence the sample average speed is an unbiased estimate of the average speed of all lightweight vehicles on the roadway.

Example 3: A machine produces a large number of water bottles. A quality inspector selected 40 water bottles randomly and found 2 defective water bottles. Find the point estimate of the proportion of all defective water bottles.

Solution: To draw the inference about the population proportion, we use the sample proportion, therefore, to find the point estimate of the proportion of all defective water bottles, we use sample proportion defectives.

Therefore, we can obtain the point estimate of the proportion of all defective water bottles as

$$p = \frac{X}{n} = \frac{2}{40} = 0.05$$

Hence, the point estimate of the proportion of all defective water bottles is 0.05.

Example 4: A furniture company manufacturing square tables of a side length μ. Thus, the area of the table will be μ^2 (unknown). Based on n independent measurements $X_1, \ldots, X_n$ of the length, estimate area of the table. Assume that the measurements of the length have mean μ and variance σ^2.

(i) Show that $\bar{X}^2$ is not an unbiased estimator for μ^2.

(ii) For what value of k, is the estimator $\bar{X}^2 - kS^2$ unbiased for μ^2?

Solution: Since the measurements of the length have mean μ and variance σ^2, therefore, the sampling distribution of mean has mean and variance as follows:

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

We have to show that $\bar{X}^2$ is not an unbiased estimator for μ^2, therefore, we have to find $E(\bar{X}^2)$ and check whether it is equal to μ^2 or not. But the question is how we find the $E(\bar{X}^2)$ without knowing the sampling distribution of $\bar{X}^2$. However, we know that

$\operatorname{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n}$ and by the formula of variance, we have

$$\text{Var}\left(\bar{X}\right) = E\left(\bar{X}^2\right) - \left[E\left(\bar{X}\right)\right]^2$$

Therefore,

$$E\left(\bar{X}^2\right) = \text{Var}\left(\bar{X}\right) + \left[E\left(\bar{X}\right)\right]^2 = \frac{\sigma^2}{n} + \mu^2 \neq \mu^2$$

Hence, $\bar{X}^2$ is not an unbiased estimator of the area of the table, that is, μ^2 and we can calculate the bias of the estimator as

$$E\left(\bar{X}^2\right) - \mu^2 = \frac{\sigma^2}{n} \Rightarrow E\left(\bar{X}^2 - \mu^2\right) = \frac{\sigma^2}{n} \qquad \left[\because E(a) = a\right]$$

We now find the value of k such that the estimator $\bar{X}^2 - kS^2$ is unbiased for μ^2.

Since the estimator $\bar{X}^2 - kS^2$ is unbiased for μ^2, therefore,

$$E\left(\bar{X}^2 - kS^2\right) = \mu^2$$

$$E\left(\bar{X}^2\right) - kE\left(S^2\right) = \mu^2 \qquad \left[\because E\left(aX \pm bY\right) = aE\left(X\right) \pm bE\left(Y\right)\right]$$

$$\mu^2 + \frac{\sigma^2}{n} - k\sigma^2 = \mu^2$$ [since S^2 is an unbiased estimator of σ^2,i.e. $E(S^2)= \sigma^2$]

Therefore,

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

Hence, for $k = \frac{1}{n}$, the estimator $\bar{X}^2 - kS^2$ is unbiased for μ^2.

Example 5: If $X_1, X_2, ..., X_n$ is a random sample taken from a population with a mean μ and variance σ^2, then

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$ is a biased estimator of σ^2.

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

Solution: We have to show that S'^2 is not an unbiased estimator for σ^2, therefore, we have to find $E(S'^2)$ and check whether it is equal to σ^2 or not. Therefore, we consider:

$$E\left(S'^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) \quad \left[\because E\left(aX\right) = aE\left(X\right)\right]$$

$$= \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i^2 + \bar{X}^2 - 2X_i\bar{X}\right)\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2 + \sum_{i=1}^{n}\bar{X}^2 - -2\bar{X}\sum_{i=1}^{n}X_i\right)$$

$$= \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2 + n\bar{X}^2 - 2\bar{X}n\bar{X}\right) \quad \left[\because \bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i \Rightarrow \sum_{i=1}^{n}X_i = n\bar{X}\right]$$

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

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

$$=\frac{1}{n}\left[\sum_{i=1}^{n}\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\frac{\sigma^2}{n}\right)\right]\quad \text{[As discussed in Example 4]}$$

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

$$=\mu^2+\sigma^2-\mu^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

$$E\left(S'^2\right)=\frac{n-1}{n}\sigma^2\neq\sigma^2$$

Hence, S'^2 is not an unbiased estimator for σ^2.

Now, we check whether S^2 is unbiased of σ^2 or not, therefore, we consider:

$$E\left(S^2\right)=\frac{n}{n-1}E\left(S'^2\right)=\frac{n}{n-1}\frac{n-1}{n}\sigma^2=\sigma^2$$

Hence, estimator S^2 is an unbiased estimator for σ^2.

This is the reason why we consider S^2 in place of S'^2 for estimating the sample variance.

Now, you can assess your understanding by answering the following Self Assessment Question.

SAQ 3

a. A company produces batteries for laptops and wants to estimate the average life of the battery. For that, the statistician of the company selected 5 batteries from the production and measured their lives. He suggests two estimators for estimating the average life of the battery:

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

where X_1, X_2, X_3, X_4 and X_5 repersent life of the selected batteries. It is known that the life of batteries has mean μ and variance σ^2. Are both estimators unbiased?

b. Show that the sample mean and sample median are both unbiased estimators for mean (μ) of a normal distribution.

6.5 PROPERTIES OF UNBIASED ESTIMATOR

After understanding the concept of unbiasedness and how to check whether an estimator is unbiased or not, we now discuss some properties of the unbiased estimator as follows:

1. Unbiased estimators may not be unique. For example, the sample mean and sample median are unbiased estimators of the population mean of a normal population.

2. Unbiased estimators do not always exist for all 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, the sample mean is an absolutely unbiased estimator of the population mean, if the population mean exists.

4. If T and T* are two unbiased estimators of the parameter θ then

   $$aT + (1-a)\,T^*$$

   is also an unbiased estimator of θ where 'a' $(0 \le a \le 1)$ is any constant.

For a better understanding of the unbiasedness try Self Assessment Questions.

SAQ 4

Some patients with high blood pressure are randomly assigned to a placebo group and a treatment group. The placebo patients receive an inactive pill, and the treatment patients receive a new drug that is expected to lower blood pressure. After the patients are treated for two months, the high blood pressures of the patients of both groups are measured and given as follows:

Placebo Group (X)	140	165	170	140	135	170	165	150	140
Treatment Group (Y)	130	135	140	130	120	120	132	118	120

If $E(X) = \mu_1, \mathrm{Var}(X) = \sigma_1^2$ and $E(Y) = \mu_2, \mathrm{Var}(Y) = \sigma_2^2$, then

(i) Show that the statistic $(\overline{X} - \overline{Y})$ is an unbiased estimator of the parameter $(\mu_1 - \mu_2)$. Also, find the estimate of the same using the given data.

(ii) Calculate the variance and standard deviation of the estimator in Part (i). Also, find the estimate of standard error.

(iii) Calculate an estimate of the ratio σ_1 / σ_2.

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

6.6 SUMMARY

In this unit, we have covered the following points:

- If we estimate an unknown parameter by a single statistic then this technique is known as point estimation whereas if we determine an interval (using sample values) that contains the true value of the unknown parameter with a certain confidence then it is known as interval estimation.

- 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.
- The properties of a good estimator are unbiasedness, consistency, efficiency and sufficiency.
- 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.
- The properties of an unbiased estimator.

6.7 TERMINAL QUESTIONS

1. A random sample $X_1, X_2, \ldots, X_n$ 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.

2. Consider a population comprising three LED televisions of a certain company. If the lives of the LED televisions are 8, 6 and 10 years then construct the sampling distribution of the average life of the LED televisions by taking samples of size 2 and show that the sample mean is an unbiased estimator of the 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

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

6.8 SOLUTIONS / ANSWERS

Self Assessment Questions (SAQs)

1. Since parameter θ represents the average marks in the paper of the MST-016 course of 50 marks and there is no negative marking, therefore, a learner can take a minimum 0 marks and a maximum 50 marks. Thus, the parameter space of θ is $\Theta = \{\theta : 0 \le \theta \le 50\}$.

2. Refer to Section 6.3.

3(a) We have to check whether estimator T_1 and T_2 are unbiased or not. Therefore, we have to find $E(T_1)$ and $E(T_2)$ and check whether it is equal to μ or not. Since X_1, X_2, X_3, X_4 and X_5 are independent and taken from the same population with a mean μ and variance σ^2, therefore,

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

We now consider

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

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

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

$$E(T_1) = \mu$$

Similarly,

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

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

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

$$E(T_2) = \mu$$

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

3(b) Let $\bar{X}$ and $\tilde{X}$ be the sample mean and sample median respectively. We have seen in Unit 2 that if we draw samples from the population whose mean is μ and variance σ^2 then the sampling distribution of mean has mean μ and variance σ^2/n.

Therefore,

$$E(\bar{X}) = \mu \text{ and } Var(\bar{X}) = \frac{\sigma^2}{n}$$

Hence, the sample mean is an unbiased estimator of the population mean.

Similarly, the sampling distribution of the median has mean μ and variance $\frac{\pi}{2}\frac{\sigma^2}{n}$. Therefore,

$$E(\tilde{X}) = \mu \text{ and } Var(\tilde{X}) = \frac{\pi\sigma^2}{2n}$$

Hence, the sample median is also an unbiased estimator of the population mean.

4. We have to show that $(\bar{X} - \bar{Y})$ is an unbiased estimator for $(\mu_1 - \mu_2)$, therefore, we have to find $E(\bar{X} - \bar{Y})$ and check whether it is equal to $(\mu_1 - \mu_2)$ or not. Thus, we consider

$$E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = \mu_1 - \mu_2 \quad \left[\because E(aX - bY) = aE(X) - bE(Y)\right]$$

Hence, the estimator $(\bar{X} - \bar{Y})$ is an unbiased estimator for $(\mu_1 - \mu_2)$.

We now find the estimate of the same using the given data. Since estimate is the value of the estimator, therefore, we find the mean of each group as

Placebo Group (X)	Treatment Group (Y)	X^2	Y^2
140	130	19600	16900
165	135	27225	18225
170	140	28900	19600
140	130	19600	16900
135	120	18225	14400
170	120	28900	14400
165	132	27225	17424
150	118	22500	13924
140	120	19600	14400
1375	**1145**	**211775**	**146173**

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i = \frac{1375}{9} = 152.78$$

$$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i = \frac{1145}{9} = 127.22$$

Thus, we find the estimate of the parameter $(\mu_1 - \mu_2)$ as

$$\bar{X} - \bar{Y} = 152.78 - 127.22 = 25.56.$$

We now find the variance and standard deviation of $(\bar{X} - \bar{Y})$ as

$$\operatorname{Var}(\bar{X} - \bar{Y}) = \operatorname{Var}(\bar{X}) + \operatorname{Var}(\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

And standard deviation

$$SD(\bar{X} - \bar{Y}) = \sqrt{\operatorname{Var}(\bar{X} - \bar{Y})} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

By the definition of standard error, we can find it as

$$SE(\bar{X} - \bar{Y}) = SD(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

We can compute the estimate of standard error as

$$SE(\bar{X} - \bar{Y}) = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Therefore, we first calculate S_1^2 and S_2^2 as

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 = \frac{1}{n_1 - 1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

$$= \frac{1}{8}(211775 - 9 \times 152.78 \times 152.78) = 212.43$$

$$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2 = \frac{1}{n_2 - 1}\left(\sum_{i=1}^{n_2} Y_i^2 - n\bar{Y}^2\right)$$

$$= \frac{1}{8}\left(146173 - 9 \times 127.22 \times 127.22\right) = 63.58$$

Finally, we can compute the estimate of the standard error as

$$SE\left(\bar{X} - \bar{Y}\right) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{212.43}{9} + \frac{63.58}{9}} = \sqrt{30.67} = 5.54$$

Note that S^2 is an unbiased estimator of σ^2 however, S is not an unbiased estimator of σ. Similarly, S_1/S_2 is not an unbiased estimator of σ_1/σ_2.

An estimate of σ_1/σ_2 is S_1/S_2 (this is a biased estimate)

$$\frac{S_1}{S_2} = \sqrt{\frac{S_1^2}{S_2^2}} = \sqrt{\frac{212.43}{63.58}} = 1.83$$

Terminal Questions (TQs)

1. We have to show that $\bar{X}$ is an unbiased estimator for θ, therefore, we have to find $E(\bar{X})$ and check whether it is equal to θ or not. Here, we are given that

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

It is negative exponential distribution with parameter θ and know the mean of this distribution as

$$E(X) = \theta$$

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

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

Therefore, we consider

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

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

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

Thus, the sample mean is an unbiased estimator of the parameter θ.

2. Here, the population consists of three LED 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

N = 3 and n = 2

Therefore, the 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 the necessary calculations for these results are shown in the following table:

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

Sample	Sample Observations	$\bar{X}$	$\sum_{i=1}^{2}(X_i - \bar{X})^2$	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

$$\bar{X}_1 = \frac{1}{2}(8+8) = 8,$$

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

$$\bar{X}_9 = \frac{1}{2}(10+10) = 10$$

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

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

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

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

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

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

From the above table, we have

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

Hence, the sample mean is an unbiased estimator of the population mean.

Also

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

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

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

Hence, the estimator S^2 is an unbiased estimator of parameter σ^2.

UNIT 7

CONSISTENCY

Structure

7.1 Introduction

Expected Learning Outcomes

7.2 Consistency

7.3 Properties of Consistent Estimator

7.4 Consistent Asymptotically Normal Estimator

7.5 Summary

7.6 Terminal Questions

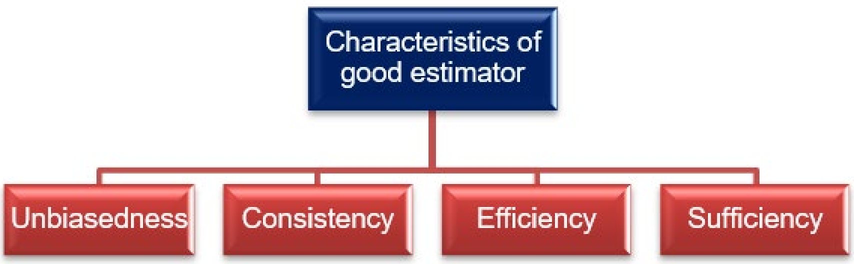
7.7 Solutions /Answers

7.1 INTRODUCTION

In the previous unit, you have seen that there exists more than one estimator for an unknown parameter. For example, for estimating unknown population mean, we may use sample mean, sample median, sample mode, average of extreme observations, etc. Now, the questions may arise:

- Which estimator should we use, that is, which is likely to give estimates closer to the true (but unknown) population value?
- Are some of the possible estimators better, in some sense, than the others?"

To answer the above questions, Prof. Ronald A. Fisher gave some characteristics of a good estimator which are as follows:



In the previous unit, you studied one of the characteristics of a good estimator, that is, unbiasedness.

An estimator is said to be unbiased for the population parameter if and only if the average or mean of the sampling distribution of the estimator is equal to the true value of the parameter. In other words, an estimator is said to be

Tools You Will Need

The following terms are considered essential background material for this Unit. If you doubt your knowledge of any of these terms, you should review the appropriate Unit or section before proceeding:

- Sampling distributions (Units 2,3, 4 and 5).
- Basic terms of estimation (Unit 6).
- Unbiased (Unit 6).
- Probability distributions (MST-012).

Unit Writer- Dr. Prabhat Kumar Sangal, School of Sciences, IGNOU, New Delhi

unbiased **if the expected value of the estimator is equal to the true value of the parameter being estimated**, that is,

$$E(T) = \theta$$

This concept was defined for a fixed sample size. In this unit, you will learn about consistency which is defined for increasing sample size.

This unit is divided into seven sections. Section 7.1 is introductory in nature. Consistency is described with examples in Section 7.2. Section 7.3 is devoted to describing various properties of the consistent estimator. Section 7.4 is devoted to the study of an additional property of a consistent estimator, which involves its asymptotic distribution. The unit ends by providing a summary of what we have discussed in this unit in Section 7.5. The terminal questions and the solution of the SAQs/TQs are given in Sections 7.6 and 7.7, respectively.

In the next unit, we shall discuss the third characteristic of a good estimator, that is, efficiency.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ comprehend the concept of consistency of an estimator;
- ❖ describe various properties of a consistent estimator;
- ❖ explain the concept of a consistent asymptotically normal estimator; and
- ❖ define consistent asymptotically normal estimator.

7.2 CONSISTENCY

In the previous unit, we have learnt about unbiasedness. An estimator T is said to be an unbiased estimator of a parameter, say, θ if the mean of the sampling distribution of estimator T is equal to the true value of the parameter θ, that is,

$$E(T) = \theta$$

This concept was defined for a fixed sample size. In this section, we will learn about consistency which is defined for increasing sample size. In general, we construct an estimator as a function of an available sample of size n for a parameter. Suppose we are able to keep collecting data and expanding the sample. In this way, we would obtain a sequence of estimates such as $T_1 = t_1(X_1)$, $T_2 = t_2(X_1, X_2)$, $T_3 = t_3(X_1, X_2, X_3)$,…, $T_n = t_n(X_1, X_2, ..., X_n)$,…. Here, we denote an estimator as T_n (indexed by n) to represent the estimator based on sample size n, instead of T as used in the previous unit. The consistency is a property of what occurs as the sample size **"grows to infinity"**.

> In statistics, a consistent estimator is an estimator that converges to the true value of the parameter as the sample size increases. It means that the estimation becomes more and more accurate as more data is collected.

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 consider a sequence of estimators, say, $\{T_1, T_2, \ldots, T_n\}$. A sequence of estimators $\{T_1, T_2, \ldots, T_n\}$ is said to be a consistent estimator for a parameter θ if the deviation/difference of the values of an estimator from the parameter tends to zero as the sample size increases. This indicates that as sample size increases, the estimator values tend to approach the parameter.

In other words, we can say that as the sample size approaches infinity, the sampling distribution of a consistent estimator becomes concentrated on the value of the parameter. It means that the standard error of the estimator declines to 0 and the sampling distribution concentrates around the population parameter.

For example, suppose $\{T_1, T_2, T_3, \ldots\}$ is a sequence of estimators for parameter θ whose true value is 5. As the sample size increases, the sampling distributions of these estimators (as shown in Fig. 7.1) are getting more and more concentrated near the true value θ = 5 (even the estimators are biased) and the density is more tightly distributed around the true value. As the sample size becomes infinite, the sampling distribution of the sequence collapses to a spike at the true value. Therefore, we can say that this sequence is consistent.

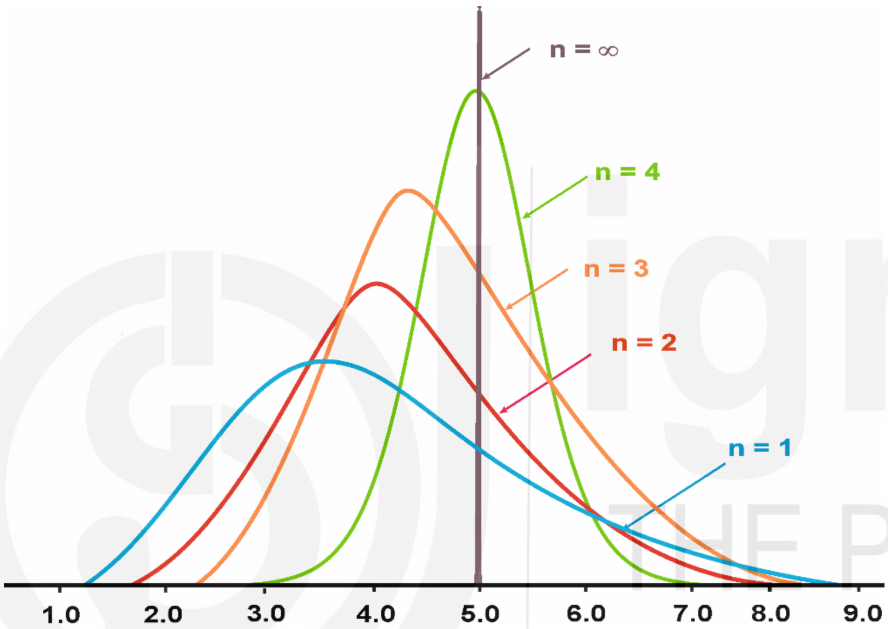


Fig. 7.1: Sampling distribution of a consistent estimator T_n with increasing sample size.

Formally, we can define a consistent estimator as

A sequence $\{T_n\}$ of estimators or simply an estimator T_n is said to be a consistent estimator of θ if T_n converges to θ in probability, that is

$$T_n \xrightarrow{p} \theta \quad \text{as } n \to \infty \text{ for every } \theta$$

i.e. for every $\varepsilon > 0$

$$\lim_{n\to\infty} P\left[\left|T_n - \theta\right| < \varepsilon\right] \to 1$$

i.e. for every $\varepsilon > 0$ and $\eta > 0$, there exist $n \geq m$ such that

$$P\left[\left|T_n - \theta\right| < \varepsilon\right] > 1 - \eta; \quad n \geq m$$

where m is some very large value of n. The above expressions are to mean the same thing.

The above definition says that an estimator is said to be a consistent estimator if the probability of accurate estimates (estimates close to the value of the population parameter) increases as sample size increases.

Consistency as defined above is sometimes called **weak consistency**. If we replace convergence in probability with almost sure convergence, then the estimator is said to be strongly consistent. Therefore,

A sequence of random $\{X_n\}$ variables is said to be almost sure convergence to a random variable X if $\{X_n\}$ converges to X with probability 1. That is,

$\lim_{n\to\infty} P\left[\left|X_n - X\right|\right] = 1$

An estimator T_n of parameter θ is said to be strongly consistent, if it converges almost surely to the true value of the parameter, that is,

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

The definition of the consistent estimator could appear a little difficult to understand. Furthermore, it is not always simple to verify whether an estimator is consistent or not using this definition. For this reason, we will present a more straightforward applied criteria (known as sufficient conditions for consistency) of consistency through which you can easily check the consistency of an estimator.

Sufficient conditions for consistency

If $\{T_n\}$ is a sequence of estimators or simply an estimator T_n is such that for all $\theta \in \Theta$ then an estimator T_n is a consistent estimator of θ if

(i) The estimator T_n is an asymptotically unbiased or simply unbiased estimator of θ, that is,

$$E(T_n) \to \theta \text{ as } n \to \infty, \text{ and}$$

(ii) The variance of estimator T_n decreases with increasing sample size. In other words, we can say that the variance of the estimator approaches zero as $n \to \infty$, that is,

$$Var(T_n) \to 0 \text{ as } n \to \infty.$$

Formally, an unbiased estimator T_n for parameter μ is said to be consistent if Var (T_n) approaches zero as $n \to \infty$.

Note: The concept of consistency relates to a sequence of estimators $\{T_n\}_{n\to\infty}$ but we usually say consistency of an estimator T_n for simplicity. Further, consistency is a large sample property of an estimator.

Since consistency is a large sample property of an estimator, some statisticians suggest that consistency should not be used alone for judging the goodness of an estimator; rather it should be used along with other criteria.

After understanding the concept of consistency, let us take some examples to understand how the definition and sufficient conditions for a consistent estimator are used.

Example 1: Prove that the sample mean is a consistent estimator of the population mean (μ) provided that the population has 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 consider $\lim_{n\to\infty} P\left[\left|T_n - \theta\right| < \varepsilon\right]$ and check whether it converges to 0 or not.

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

To find this probability, we convert it to the standard form. Recall from Unit 1 that the sample mean ($\bar{X}$) has mean μ and finite variance σ^2/n, therefore, we convert $\bar{X} - \mu$ in the standard form by dividing it by $\sigma / \sqrt{n}$ as

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

By the central limit theorem (described in Unit 1 of this course), you 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 the 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, the sample mean is a consistent estimator of the population mean.

I think that you may feel this process is a little difficult to show whether an estimator is consistent or not. Furthermore, it is not always simple to verify whether an estimator is consistent or not using this definition. For this reason, you can use sufficient conditions for consistency. Let us solve this example with the help of sufficient conditions for consistency.

First, we have to show that the sample mean is asymptotically unbiased or simply an unbiased estimator for parameter μ. Therefore, we have to find $E\left(\bar{X}\right)$ and check whether it is equal to μ or not as $n \to \infty$. Consider,

$$E\left(\bar{X}\right) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \qquad \text{[by definition of the sample mean]}$$

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

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from the same population, therefore, they also have same mean and variance. Therefore,

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

Thus,

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

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

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

Now, we consider the variance of the sample mean $(\bar{X})$ and check whether it converges to 0 or not as $n \to \infty$.

Thus, we consider,

If X and Y are two independent random variables, then

Var(aX±bY) = a^2 Var(X) + b^2 Var(Y).

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

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

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

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

Hence, by the sufficient conditions of consistency, we can say that the sample mean $(\bar{X})$ is a consistent estimator of the population mean.

We prove this result in general, so the sample mean is always an unbiased and consistent estimator of the population mean of all populations which follows normal, Poisson, binomial, exponential, etc. Let us consider the next example.

Example 2: If the number of weekly accidents occurring on a mile stretch of a particular road follows a Poisson distribution with parameter λ then show that the sample mean $(\bar{X})$ is a consistent estimator of λ. Also, find the estimate of parameter λ on the basis of the following data:

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

Solution: Here, the number of weekly accidents occurring on a mile stretch of a particular road follows a Poisson distribution with parameter λ. You also know that the mean and variance of Poisson distribution (λ) are

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

Since $X_1, X_2, \ldots, X_n$ are independent and come from the same Poisson distribution, therefore,

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

We now consider

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

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

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

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

Thus, the sample mean is an unbiased estimator of the parameter λ.

Now we consider the variance of the sample mean and check whether it converges to 0 or not as $n \to \infty$.

Therefore, we consider,

If X and Y are two independent random variables, then

$Var(aX \pm bY) = a^2\,Var(X) + b^2\,Var(Y)$.

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

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

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

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

Hence, by the sufficient conditions of consistency, we can say that the sample mean is a consistent estimator of the parameter λ of the Poisson distribution.

Since the mean of the Poisson distribution is λ, therefore, we estimate the population mean by the sample mean. Thus, we calculate the sample mean of the given data as follows:

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

The formula for calculating the mean is

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

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

Hence, the estimate of the parameter λ is 2.

Example 3: A company produces ball bearings. The quality control inspector found that there is a variation in the diameters of the steel ball bearings. To estimate the variation in the diameter, he randomly selected n ball bearings from the production line and measured the diameter of each selected ball bearing. Suppose the measured diameters of the ball bearings are $X_1, X_2, \ldots, X_n$. He proposed an estimator for the variance of all ball bearings which is given as follows:

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

If the actual variation in the diameter of the ball bearing is σ^2 then show that S'^2 is a consistent estimator.

Solution: To show that the proposed estimator S'^2 is consistent, we have to show that

- It is asymptotically unbiased, that is

  $E(S'^2) \to \sigma^2$ as $n \to \infty$ and

- Its variance tends to zero as n tends to infinity, that is,

  $Var(S'^2) \to 0$ as $n \to \infty$

Therefore, we consider

$$E\left(S'^2\right) = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] \quad \left[\because E(aX) = aE(X)\right]$$

$$= \frac{1}{n}E\left[(n-1)S^2\right] \quad \left[\because S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \Rightarrow \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = (n-1)S^2\right]$$

$$= \frac{\sigma^2}{n}E\left[\frac{(n-1)S^2}{\sigma^2}\right] \quad \text{[multiplying and dividing by } \sigma^2]$$

Recall from Units 3 and 4 that $\frac{(n-1)S^2}{\sigma^2}$ follows a chi-square distribution with (n − 1) degrees of freedom and from the properties of the chi-square distribution, we have the mean and variance of the chi-square distribution with (n − 1) degrees of freedom as (n − 1) and 2(n − 1), therefore, we have

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = E\left[\chi^2_{(n-1)}\right] = n-1$$

And $Var\left[\frac{(n-1)S^2}{\sigma^2}\right] = Var\left[\chi^2_{(n-1)}\right] = 2(n-1)$

Therefore, we have

$$E\left(S'^2\right) = \frac{\sigma^2}{n}E\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n}(n-1) = \sigma^2 - \frac{\sigma^2}{n}$$

Therefore,

$$\lim_{n\to\infty} E\left(S'^2\right) = \lim_{n\to\infty}\left(\sigma^2 - \frac{\sigma^2}{n}\right) = \sigma^2 \quad \left(\because \lim_{n\to\infty}\frac{\sigma^2}{n} = 0\right)$$

Hence S'^2 is an asymptotically unbiased estimator for σ^2.

We now consider the variance of S'^2 as

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

$$= \frac{1}{n^2}\text{Var}\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] \quad \left[\because \text{Var}(aX) = a^2\text{Var}(X)\right]$$

$$= \frac{1}{n^2}\text{Var}\left[\sigma^2\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\left(\sigma^2\right)^2}{n^2}\text{Var}\left[\chi^2_{n-1}\right] = \frac{\left(\sigma^2\right)^2}{n^2}\times 2(n-1)$$

$$\lim_{n\to\infty}\text{Var}\left(S'^2\right) = 2\sigma^4\lim_{n\to\infty}\left(\frac{1}{n} - \frac{1}{n^2}\right) = 2\sigma^4 \times 0 = 0$$

Hence, the estimator $S'^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ is a consistent estimator for population variance σ^2.

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

SAQ 1

(i) If the grades of the students in a paper of the MSCAST programme follow a normal distribution with mean μ and variance σ^2 then show that the sample median is a consistent estimator of the population mean (μ).

(ii) The magnitude of earthquakes recorded in a region modelled as an exponential distribution with an unknown parameter θ whose pdf is given by

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

A researcher considered the following two estimators for the parameter θ:

$$T_1 = \frac{1}{n}\sum_{i=1}^{n}X_i \text{ and } T_2 = \frac{1}{n+1}\sum_{i=1}^{n}X_i$$

Check whether both estimators are unbiased and consistent.

(iii) The average weight gained by a person over the winter months is uniformly distributed and ranges from θ to θ+6 lbs, then show that the sample mean is an unbiased as well as a consistent estimator of $(\theta + 3)$.

I think you have understood what a consistent estimator is and how to check whether an estimator is consistent or not. Let us study the properties of a consistent estimator in the next session.

7.3 PROPERTIES OF CONSISTENT ESTIMATOR

After understanding the concept of consistency and how to check whether an estimator is a consistent estimator or not, we now discuss some important properties of a consistent estimator as follows:

1. Consistent estimators may not be unique. For example, the sample mean and the sample median both are consistent estimators of the population mean of a normal population (see Example 1 and SAQ 1(i)). Also, the sample variances S'^2 and S^2, respectively are consistent estimators for population variance (see Example 3 and TQ 2).

2. Consistent estimators need not be unbiased. For example, the sample variance $S'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is not an unbiased estimator but consistent (see Example 3).

3. An unbiased estimator need not be consistent (see Example 4).

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

5. If T_n and S_n are consistent estimators of the parameters α and β, respectively, then

(i) $T_n + S_n$ is a consistent estimator of α + β

(ii) T_nS_n is a consistent estimator of αβ

(iii) If $\beta \neq 0$ then T_n/S_n is a consistent estimator of α/β

A function f(x) is said to be continuous at x = a if $\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = f(a)$, In other words, we can say that a function is continuous at a point x = a, if $L.H.L._{at\,x=a} = R.H.L._{at\,x=a} =$ = value of the function at x =a

Let us study the use of the properties of a consistent estimator with the help of an example.

Example 4: The height of the person living on hills follows a normal distribution with mean μ inches and variance σ^2 square inches. To estimate the average height, a researcher suggested an estimator $T = X_1$ (the first observation of a sample). Check whether it is unbiased and consistent.

Solution: To check whether the suggested estimator T is unbiased or not, we have to find E(T). Since X_1 was taken from the same population with a mean μ and variance σ^2, therefore,

$$E(X_1) = \mu \text{ and } Var(X_1) = \sigma^2$$

Thus,

$$E(T) = E(X_1) = \mu$$

Therefore, the estimator $T= X_1$ is an unbiased estimator.

Since $Var(T) = Var(X_1) = \sigma^2$ but it does not converge to zero as n tends to infinity, therefore, by the sufficient conditions of the consistency, it is not a consistent estimator.

Example 5: Consider Example 2. Find the consistent estimator of $\lambda(\lambda - 1)$. Also, find the estimate of it.

Solution: In such questions where we have to find out consistent estimator for a function of a parameter, we solve the question as follows:

(i) First, we find the consistent estimator of the parameter (here λ).

(ii) After that, we check whether the given function is continuous or not.

(iii) Finally, we use the invariance property of the consistent estimator to find the consistent estimator of the given function.

In Example 2, we showed that the sample mean $(\bar{X})$ is a consistent estimator of the parameter λ. Therefore, we move the second point and check whether $\lambda(\lambda-1)$ is a continuous function or not. Since $\lambda(\lambda-1)$ is a polynomial and we know that each polynomial is a continuous function, therefore, $\lambda(\lambda-1)$ is a continuous function. Since the sample mean is a consistent estimator of λ and $\lambda(\lambda-1)$ is a continuous function of λ, therefore, by the invariance property of consistency $\bar{X}(\bar{X}-1)$ will be the consistent estimator of $\lambda(\lambda-1)$.

In Example 2, we also calculated the values of the sample mean as

$$\bar{X} = 2$$

Therefore, the estimate of $\lambda(\lambda-1)$ is $\bar{X}(\bar{X}-1) = 2\times(2-1) = 2$.

It is now time for you to try the following Self Assessment Question.

SAQ 2

Suppose it is known that the probability that a certain company experiences a network failure in a given week is θ and the distribution of the number of weeks the company does not experience a network failure follows a geometric distribution with parameter θ, then show that the sample mean $\bar{X}$ is a consistent estimator of $1/\theta$. Also, find a consistent estimator of $e^{1/\theta}$.

After studying the properties of a consistent estimator, the next section is devoted to a study of an additional property of a consistent estimator, which involves its asymptotic distribution with a suitable normalisation. These play a key role in large sample inference theory.

7.4 CONSISTENT ASYMPTOTICALLY NORMAL ESTIMATOR

In the previous sections, you have studied that a sequence of estimators $\{T_1, T_2, \ldots, T_n\}$ or simply an estimator T_n is said to be a consistent estimator for a parameter θ if the sampling distribution of the estimator T_n becomes concentrated on the value of the parameter as the sample size approaches infinity. Everything is OK with a consistent estimator, but we face mainly two problems with the consistent estimator which are given as follows:

- The shape of the sampling distribution of most of the estimators (consistent) is not known because the estimators are complicated non-linear functions of random samples. If we know the sampling distribution of our estimator for every sample size, we could use it to draw inferences using this finite-sample distribution.
- If the sampling distribution of the consistent estimator is known then its shape changes with the sample size.

For example, in Example 1, you have seen that the sample mean is a consistent estimator for the population mean. To show the impact of the sample size on the sampling distribution of the consistent estimator, we plot the sampling distribution of the estimator (the sample mean for the population mean), say, $\theta = 5$ for sample sizes 100, 300, 500, 1000… as shown in Fig. 7.2.

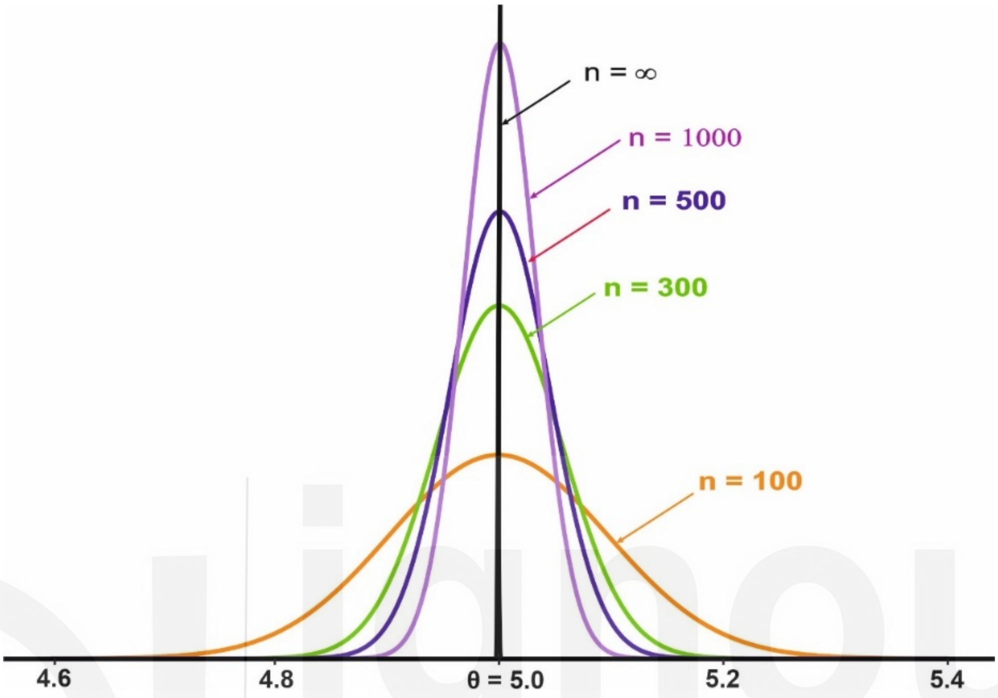


Fig. 7.2: Sampling distribution of sample mean with increasing sample size.

From Fig. 7.2, you can see that as the sample size increases the shape of the sampling distribution changes and its average becomes increasingly tight around the true value of the population mean. Also, in Unit 15 of the course MST-012: Probability and Probability Distributions, you have studied that convergence in probability implies convergence in law (distribution), therefore, the estimator T_n converges to the parameter $\theta = 5$ in distribution as sample size approaches infinity. It means that the asymptotic distribution of the consistent estimator T_n is degenerate at θ (in our case $\theta = 5$, as shown in Fig. 7.2). It indicates that the estimator will take one value θ with probability 1. Such a degenerate distribution is not helpful to find the rate of convergence or to find an interval estimator of θ. If we know the sampling distribution of our estimator for every sample size, we could use it to draw inferences using this finite-sample distribution. Hence, we aim to find an estimator whose sampling distribution does not change for a large sample size. For that, we use the concept of consistent asymptotically normal distribution.

It is observed that if we re-centre and re-scale the estimator, then the form of the sampling distributions of the new version of the estimator does not change with sample size and non-degenerate as the sample size tends to infinity. Also, the shape of the sampling distribution gets arbitrarily close to a normal distribution as the sample size increases. For illustration purposes, instead of looking at the distribution of the estimator $T_n = \bar{X}$ for sample size n, let's look at the distribution of $\sqrt{n}\left(T_n - \theta_0\right)$, where θ_0 is the true value of the population mean (parameter) for which the estimator T_n is consistent. We plot again the sampling distribution of $\sqrt{n}\left(T_n - \theta_0\right) = \sqrt{n}(\bar{X} - 5.0)$ instead of the sample mean with the sample sizes 100, 300, 500 and 1000 in Fig. 7.3.

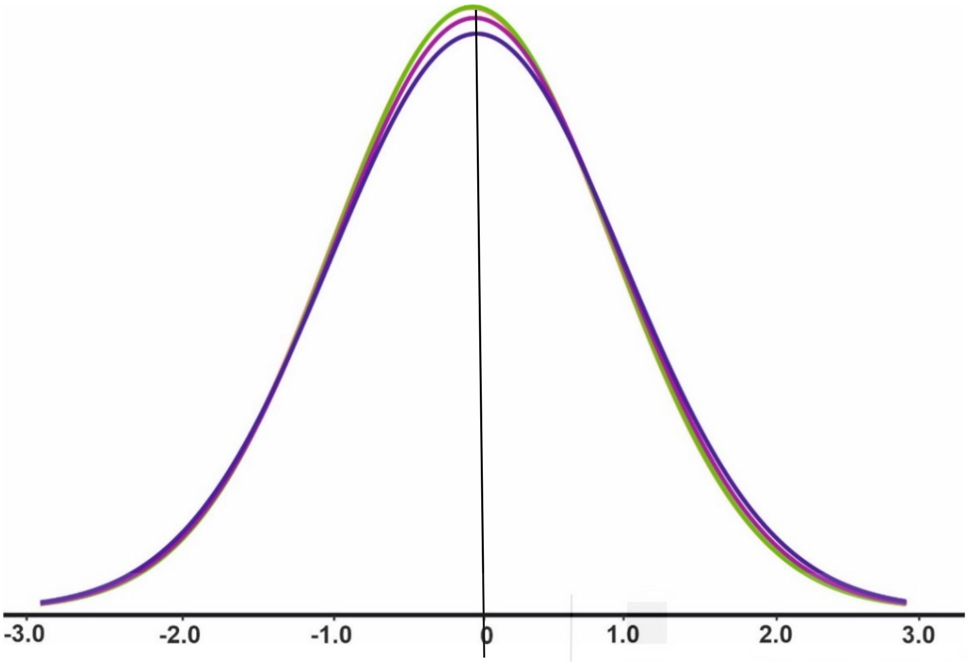


Fig. 7.3: Sampling distribution of a re-centre and re-scale sample mean with increasing sample size

From the above figure, you can observe that the sampling distributions of $\sqrt{n}\left(T_n - \theta_0\right) = \sqrt{n}(\bar{X} - 5.0)$ (for sample size 100, 300, 500 and 1000) are indistinguishable from each other and look closely. They also follow the normal distributions with mean 0 and a constant variance σ^2 (population variance). In other words, we can say that the distribution of $\sqrt{n}\left(T_n - \theta_0\right) = \sqrt{n}(\bar{X} - 5.0)$ gets arbitrarily close to a N(0, σ^2) distribution as $n \to \infty$. Therefore, we can define the consistent asymptotically normal estimator as follows:

An estimator T_n is said to be a consistent asymptotically normal estimator for the parameter θ if the sampling distribution of $\sqrt{n}\left(T_n - \theta_0\right)$ follows a normal distribution with mean 0 and constant variance σ^2.

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

7.5 SUMMARY

In this unit, we have covered the following points:

- A consistent estimator is an estimator that converges to the true value of the parameter as the sample size increases. This means that the estimate becomes more and more accurate as more data is collected.
- A sequence $\{T_n\}$ of estimators or simply an estimator T_n is said to be consistent for a parameter θ, if T_n converges to θ in probability, that is,

$$T_n \xrightarrow{p} \theta \text{ as } n \to \infty \quad \text{for every } \theta$$

- If $\{T_n\}$ is a sequence of estimators such that for all $\theta \in \Theta$ then an estimator T_n is a consistent estimator of θ if

- The estimator T_n is an asymptotically unbiased or simply unbiased estimator of θ, that is,

  $E(T_n) \to \theta$ as $n \to \infty$, and

- The variance of estimator T_n decreases with increasing sample size, or we can say that variance of the estimator approaches zero as $n \to \infty$, that is,

  $Var(T_n) \to 0$ as $n \to \infty$.

- If T_n is a consistent estimator of θ and $f(\theta)$ is a continuous function of θ then $f(T_n)$ will be the consistent estimator of $f(\theta)$. This property is known as **invariance property**.
- An estimator T_n is said to be a consistent asymptotically normal estimator for the parameter θ if the sampling distribution of $\sqrt{n}(T_n - \theta_0)$ follows a normal distribution with mean 0 and constant variance σ^2.

7.6 TERMINAL QUESTIONS

1. Define consistency.

2. Consider Example 2 and suppose the quality control inspector proposed the estimator $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ for the population variance σ^2 then show that it is also consistent.

7.7 SOLUTIONS / ANSWERS

Self Assessment Questions (SAQs)

1. (i) To check the consistency, first, we have to show that the sample median ($\tilde{X}$) is asymptotically unbiased or simply an unbiased estimator for population mean μ. Therefore, we have to find $E(\tilde{X})$ and check whether it is equal to μ or not as $n \to \infty$.

Recall from Unit 2 that if we draw the samples from a normally/non-normally distributed population with mean μ and variance σ^2, then the sampling distribution of median will be normally distributed, only for the large sample, with mean

$E[\tilde{X}] = \mu$ and standard deviation

$$SD[\tilde{X}] = 1.253\frac{\sigma}{\sqrt{n}}$$

Since $E[\tilde{X}] = \mu$, therefore, the sample median is an unbiased estimator of the population mean μ.

We now consider the variance of the sample median ($\tilde{X}$) and check whether it converges to 0 or not as $n \to \infty$. Therefore, we consider,

$$\mathrm{Var}\left(\tilde{X}\right) = (SD)^2 = \left(1.253\frac{\sigma}{\sqrt{n}}\right)^2 = 1.57\frac{\sigma^2}{n}$$

$$\mathrm{Var}\left(\tilde{X}\right) = \rightarrow 0 \text{ as } n \rightarrow \infty$$

Hence, by the sufficient conditions of consistency, we can say that the sample median ($\tilde{X}$) is a consistent estimator of the population mean.

(ii) Since the magnitude of earthquakes recorded in a region is modelled as the exponential distribution, therefore, it has mean as θ and variance θ^2.

We have to check whether estimators T_1 are T_2 are unbiased and consistent. Therefore, first, we check whether they are unbiased. For that, we have to find $E(T_1)$ and $E(T_2)$ and check whether these are equal to μ or not. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from exponential distribution. Since the sample observations are independent and taken from the same population with a mean θ and variance θ^2, therefore,

$$E(X_i) = \theta \text{ and } \mathrm{Var}(X_i) = \theta^2 \quad \text{for all i = 1, 2, …,n}$$

Consider,

$$E(T_1) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$$

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

$$E(T_1) = \theta$$

Hence, the estimator T_1 is an unbiased estimator of θ.

Similarly,

$$E(T_2) = E\left[\frac{1}{n+1}\sum_{i=1}^{n} X_i\right] = \frac{1}{n+1}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$$

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

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

Since $E(T_2) = \left(\frac{1}{n} + 1\right)\theta \neq \theta$

Hence, the estimator T_2 is not an unbiased estimator of the parameter θ.

We now check the consistency of both estimators. For that, we check whether the variances of these estimators converge to 0 or not as $n \rightarrow \infty$. Therefore, we consider

$$\mathrm{Var}(T_1) = \mathrm{Var}\left[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right]$$

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

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

$$Var(T_{1'}) = \frac{\theta^2}{n} \to 0 \text{ as } n \to \infty$$

Hence, $E(T_1) = \theta$ and $Var(T_{1'}) \to 0$ as $n \to \infty$, therefore, by the sufficient conditions of consistency, the estimator T_1 is a consistent estimator of the parameter θ.

We consider

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

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

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

$$Var(T_{2'}) == \frac{n\theta^2}{(n+1)^2} \to 0 \text{ as } n \to \infty$$

Also, $E(T_2) = \dfrac{1}{\left(1 + \frac{1}{n}\right)}\theta \to \theta$ as $n \to \infty$

Hence, $E(T_2) \to \theta$ and $Var(T_{2'}) \to 0$ as $n \to \infty$, therefore, by the sufficient conditions of consistency, the estimator T_2 is also a consistent estimator of the parameter θ.

Hence, both estimators T_1 and T_2 are consistent but not unbiased estimators of the parameter θ.

(iii) As we know the pdf of the uniform distribution with parameters 'a' and 'b' as follows:

$$f(x) = \frac{1}{b-a}; \quad a \le x \le b$$

The mean and variance of the uniform distribution are given as follows:

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

In our case, a = θ and b = θ + 6, therefore, the pdf, mean and variance are given as follows:

$$f(x)=\frac{1}{\theta+6-\theta}=\frac{1}{6}; \quad \theta \le x \le \theta+6$$

$$E(X)=\frac{\theta+\theta+6}{2}=\theta+3 \text{ and } Var(X)=\frac{(\theta+6-\theta)^2}{12}=3$$

Let $X_1, X_2, ..., X_n$ denote the weight of the person on n random winter days, therefore,

$$E(X_i)=E(X)=\theta+3 \text{ and } Var(X_i)=Var(X)=3 \text{ for all } i=1, 2, ..., n$$

To show that $\bar{X}$ is unbiased estimator of $(\theta+3)$, we consider

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

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

$$=\frac{1}{n}\left[\underbrace{(\theta+3)+(\theta+3)+...+(\theta+3)}_{\text{n-times}}\right]$$

$$=\frac{1}{n}[n(\theta+3)]=\theta+3$$

Therefore, the sample mean is an unbiased estimator of $(\theta+3)$.

For consistency, we have to show that

$$E(\bar{X})\to\theta+3 \text{ and } Var(\bar{X})\to 0 \text{ as } n\to\infty$$

Therefore, we consider,

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

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

$$=\frac{1}{n^2}\left(\underbrace{3+3+...+3}_{\text{n-times}}\right)$$

$$=\frac{1}{n^2}(3n)=\frac{3}{n}$$

$$Var(\bar{X})=\frac{3}{n}\to 0 \text{ as } n\to\infty$$

Thus, $E(\bar{X})=\theta+3$ and $Var(\bar{X})\to 0$ as $n\to\infty$

Hence, by the sufficient conditions of the consistency, the sample mean is also a consistent estimator of $(\theta+3)$.

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

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

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

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

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

Therefore, we consider,

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

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

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

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

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

Thus, the sample mean is an unbiased estimator of $\frac{1}{\theta}$.

We now consider

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

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

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

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

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

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

Hence, by the sufficient conditions of the consistency, the sample mean $\bar{X}$ is a consistent estimator of $1/\theta$.

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

Terminal Questions(TQs)

1. Refer to Section 7.2.

2. To show that the proposed estimator S^2 is a consistent estimator, we have to show that

- S^2 is asymptotically unbiased, that is

  $E(S^2) \to \sigma^2$ as $n \to \infty$ and

- The variance of S^2 tends to zero as n tends to infinity, that is,

  $Var(S^2) \to 0$ as $n \to \infty$

Therefore, we consider

$$E\left(S^2\right) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]$$

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

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

$$\left[\because S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \Rightarrow \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \left(n-1\right)S^2\right]$$

$$= \frac{\sigma^2}{n-1}E\left[\frac{\left(n-1\right)S^2}{\sigma^2}\right] \quad \text{[multiplying and dividing by } \sigma^2\text{]}$$

Recall from Units 3 and 4 that $\frac{\left(n-1\right)S^2}{\sigma^2}$ follows the chi-square distribution with (n − 1) degrees of freedom and from the properties of the chi-square distribution, we have the mean and variance of the chi-square distribution with (n − 1) degrees of freedom as (n − 1) and 2(n − 1), respectively, therefore, we have

$$E\left[\frac{\left(n-1\right)S^2}{\sigma^2}\right] = E\left[\chi^2_{(n-1)}\right] = n-1$$

and

$$Var\left[\frac{\left(n-1\right)S^2}{\sigma^2}\right] = Var\left[\chi^2_{(n-1)}\right]$$

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

Therefore, we have

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

Hence S^2 is an unbiased estimator for σ^2.

We now consider the variance of S^2 as

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

$$= \frac{1}{(n-1)^2} Var\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) \quad [\because Var(aX) = a^2 Var(X)]$$

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

$$= \frac{(\sigma^2)^2}{(n-1)^2} Var\left[\chi^2_{n-1}\right] = \frac{(\sigma^2)^2}{(n-1)^2} \times 2(n-1)$$

$$\lim_{n\to\infty} Var(S^2) = 2\sigma^4 \lim_{n\to\infty}\left(\frac{1}{n-1}\right) = 2\sigma^4 \times 0 = 0$$

Hence, by the sufficient conditions of consistency, the estimator $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a consistent estimator for the population variance σ^2.

UNIT 8

EFFICIENCY AND MEAN SQUARED ERROR

Structure

8.1 Introduction

Expected Learning Outcomes

8.2 Concept of Efficiency

8.3 Most Efficient Estimator

8.4 Properties of Efficient Estimator

8.5 Mean Squared Error

8.6 Minimum Variance Unbiased Estimator

8.7 Summary

8.8 Terminal Questions

8.9 Solutions /Answers

8.1 INTRODUCTION

In the previous Units 6 and 7, we discussed two characteristics: unbiasedness and consistency of a good estimator with various examples. I hope you understand both properties. You have also seen that the sample mean and sample median both are unbiased and consistent for the population mean μ when sampling is done from a normal population with mean μ and variance σ^2. Now, the question may arise: Are they as “good” as one another, or is there some reason to prefer one over another? This means that we need to consider other characteristics of a good estimator to check which one is better in comparison to another. Thus, this unit is devoted to explaining the concept of efficiency, mean squared error and minimum variance unbiased estimator which help us to compare estimators and make the decision which one is better.

This unit is divided into nine sections. Section 8.1 is introductory in nature. There may exist more than one unbiased estimator of a parameter, therefore, to check which one is better, we explain the concept of efficiency in Section 8.2. If we have a class of unbiased estimators of a parameter, then to compare them, we use the concept of the most efficient estimator which is explained in Section 8.3. Section 8.4 is devoted to discussing the properties of efficient estimators. Section 8.5 explains the concept of the mean squared error. Section 8.6 describes the minimum variance unbiased estimator. The unit ends by providing a summary of what we have discussed in this unit in Section 8.7. The terminal questions and the solution of the SAQ/TQ are given in

Tools You Will Need

The following terms are considered essential background material for this Unit. If you doubt your knowledge of any of these terms, you should review the appropriate Unit or section before proceeding:

- Sampling distributions (Units 2,3, 4 and 5).
- Basic terms of estimation (Unit 6).
- Unbiased and consistency (Units 6 and 7).
- Probability distributions (MST-012).

Unit Writer- Dr. Prabhat Kumar Sangal, School of Sciences, IGNOU, New Delhi

Sections 8.8 and 8.9, respectively.

Expected Learning Outcomes

After studying this unit, you should be able to:

- comprehend the concept of efficiency of an estimator;
- explain the concept of the most efficient estimator;
- describe various properties of an efficient estimator;
- define the mean squared error of an estimator; and
- describe the concept of minimum variance unbiased estimator.

8.2 CONCEPT OF EFFICIENCY

In some situations, we may see that there is more than one estimator of the same parameter which are unbiased. For example, the sample mean and the sample median both are unbiased estimators for the population parameter mean μ when the sampling is done from a normal population with mean μ and variance σ^2. Now, the question may arise: Are they all as "good" as one another, or is there some reason to prefer one over another? Let us assume that we have two unbiased estimators say, T_1 and T_2 (based on the same sample size) for the same parameter θ and they have variances $Var(T_1)$ and $Var(T_2)$, respectively. Suppose the shapes of the sampling distribution of both estimators are as shown in Fig. 8.1.

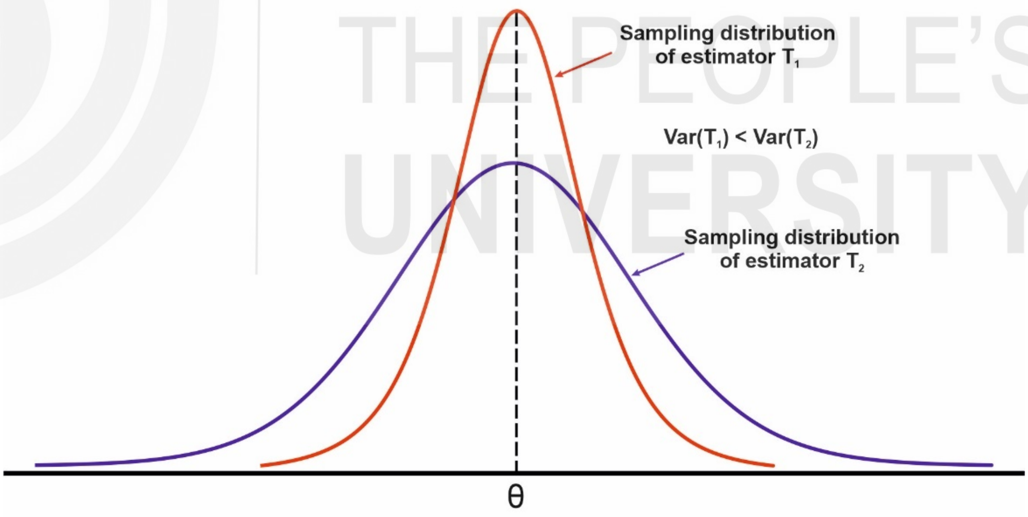


Fig. 8.1: Sampling distributions of estimators T_1 and T_2.

From Fig. 8.1, you can observe that the centre of both sampling distributions is θ so both estimators are unbiased, however, the sampling distribution of the estimator T_2 is more spread than the estimator T_1, therefore, we can conclude that the variance (spread) of the estimator T_1 is smaller than the estimator T_2. However, it is clear that we would also desire the estimator whose sampling distribution not be too spread out around the true value of the parameter because if it is too spread then there will be a high probability that an estimate could be generated will have a significant distance from the true value of the parameter. Therefore, there is a necessity for some further criterion which will enable us to choose between the estimators with the common property of unbiasedness. One way to compare estimators is by looking at their variance. If one unbiased estimator has a lower variance than another unbiased estimator, we say that the one with a lower variance is more efficient than the one with a higher variance. Such a criterion which is based on the variances of the sampling distributions of the estimators is usually known as efficiency. We can define it as follows:

There are some textbooks in which equal sample sizes are not mentioned. But this seems a bit unfair because as you know that the variance of an estimator decreases by increasing the sample size. In practice the sample size is fixed. It is hard to imagine a situation where you would select an estimator that is more efficient at a larger sample size than sample size of your data.

If T_1 and T_2 are two unbiased estimators of a parameter θ with the same size, then the estimator T_1 is said to be more efficient than the estimator T_2 if

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

It means that if we want to compare (which one is better) two unbiased estimators of the same size of a parameter then we can compare their variances, and which one has the less variance is said to be more efficient. An estimator with a smaller variance is relatively more efficient because its values are concentrated more closely on the true value of the parameter. Efficiency in statistical inference is important in comparing the performance of various estimators. The efficiency of an estimator can also be treated as the **precision** of the estimate. If an estimator is more efficient then we can say that it is the more precise estimator of the parameter.

Let us look at an example to see how this definition works.

Example 1: A company produces batteries for laptops and wants to estimate the average life of the batteries. For that, the statistician of the company selected 5 batteries from the production and measured their lives. He suggests two unbiased estimators for estimating the average life of the batteries:

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

where X_1, X_2, X_3, X_4 and X_5 represent the life of the selected batteries. If it is known that the life of batteries has mean μ and variance $σ^2$ then which one is more efficient?

Solution: We have to check which one of these proposed unbiased estimators T_1 and T_2 is more efficient. Therefore, we have to find the variances of both estimators and check which one is smaller. Since X_1, X_2, X_3, X_4 and X_5 are independent and taken from the same population with a mean μ and variance $σ^2$, therefore,

$$E(X_i) = \mu \text{ and } Var(X_i) = \sigma^2 \text{ for all } i = 1,2,...,5$$

So we consider,

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

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

$$= \frac{1}{25}\left[\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2\right] \qquad \left[\because Var(X_i) = \sigma^2\right]$$

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

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

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

Similarly,

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

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

$$= \frac{1}{225}\left[\begin{array}{l}\text{Var}\left(X_1\right) + 4\text{Var}\left(X_2\right) + 9\text{Var}\left(X_3\right) \\ \qquad\qquad + 16\text{Var}\left(X_4\right) + 25\text{Var}\left(X_5\right)\end{array}\right]$$

$$= \frac{1}{225}\left(\sigma^2 + 4\sigma^2 + 9\sigma^2 + 16\sigma^2 + 25\sigma^2\right) \qquad \left[\because \text{Var}\left(X_i\right) = \sigma^2\right]$$

$$= \frac{55\sigma^2}{225}$$

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

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

Example 2: Show that the sample mean is a more efficient estimator than the sample median for estimating the mean of the normal population.

Solution: To show that the sample mean is a more efficient estimator than the sample median for estimating the mean of the normal population, we have to compare the variance of the sample mean with the variance of the sample median.

Let $X_1, X_2, \ldots, X_n$ be a random sample taken from a 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 mean from a normal population follows a normal distribution with means μ and variance σ^2/n. Similarly, the sampling distribution of the median from a normal population also follows a normal distribution with mean μ and variance $\frac{\pi}{2}\frac{\sigma^2}{n}$.

Therefore,

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

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

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

I hope you understood the concept of efficiency and how to check which one is more efficient between the two estimators. Therefore, before going to the next section, you should assess yourself by answering the following Self Assessment Question.

SAQ 1

A company manufactures fruit juice packets. Suppose the weight of juice packets follows a normal distribution with mean weight μ ml and standard deviation σ ml. To estimate the average weight of the fruit juice packets, the quality control inspector measured the weight of three selected fruit juice packets X_1, X_2, and X_3 ml and proposed two estimators for estimating the average weight of fruit juice packets μ as follows:

$$T_1 = \frac{X_1 + X_2 + X_3}{3} \quad \text{and} \quad T_2 = \frac{X_1 + X_2}{4} + \frac{X_3}{2}$$

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

8.3 MOST EFFICIENT ESTIMATOR

In the previous section, you studied the concept of efficiency. According to this, if one unbiased estimator, say, T_1 has lower variance than another unbiased estimator, say, T_2, then we say that the estimator T_1 is more efficient than the estimator T_2 for all the same sample sizes. This concept is used when we have to compare two unbiased estimators. Sometimes, we have a class of unbiased estimators for a parameter then to compare the efficiency of the estimator, we use the concept of the most efficient estimator. We can define the most efficient estimator as follows:

In a class of unbiased estimators (based on the same sample size) of a parameter, if there exists one estimator whose variance is minimum (least) among the class, then it is said to be the most efficient estimator of that parameter.

For example, suppose T_1, T_2 and T_3 are three unbiased estimators of parameter θ having variance 1/n, 1/(n+1) and 5/n, respectively. Since the variance of estimator T_2 is minimum, therefore, estimator T_2 is the most efficient estimator in that class.

Efficiency

The efficiency of an unbiased estimator is measured by concerning the most efficient estimator is called **"Absolute Efficiency"**. If T^* is the most efficient estimator having variance $Var(T^*)$ and T is any other unbiased estimator having variance Var(T), then the efficiency of T is defined as

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

Since the variance of the most efficient estimator is minimum, therefore,

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

Let us take an example for illustration purposes.

Example 3: Suppose a market researcher proposed three unbiased estimators for estimating the average life of LED bulbs produced by a company on the basis of a sample of size 4 which are given as follows:

$$T_1 = \frac{X_1 + X_2 + X_3 + X_4}{4},\ T_2 = \frac{2X_1 + 3X_2 + \alpha X_4}{10}, T_3 = \frac{X_1 + X_2 + \beta X_3}{5}$$

where X_1, X_2, X_3, and X_4 represent the life of the selected LED bulbs in the random sample. It is known that the life of the LED bulbs has mean μ and variance σ^2.

(i) Find the values of α and β.

(ii) Which one is the most efficient estimator?

(iii) Calculate the efficiency of the remaining estimators.

Solution: Since it is given that the estimators are unbiased, therefore, by the definition of the unbiased estimator, their expected values equal to the average life of the LED bulbs, therefore,

$$E(T_1) = E(T_2) = E(T_3) = \mu$$

To find the value of α, we consider

If X and Y are two independent random variables and a & b are two constants, then E(aX ± bY) = aE(X) ± bE(Y) and

$$E(T_2) = \mu$$

$$E\left(\frac{2X_1 + 3X_2 + \alpha X_4}{10}\right) = \mu \Rightarrow \frac{2E(X_1) + 3E(X_2) + \alpha E(X_4)}{10} = \mu$$

Since X_1, X_2, X_3, and X_4 are independent and taken from the same group of the LED bulbs (population) with a mean μ and variance σ^2, therefore,

$$E(X_i) = \mu \text{ and } Var(X_i) = \sigma^2 \text{ for all } i = 1, 2, ..., 4$$

Therefore,

$$\frac{2\mu + 3\mu + \alpha\mu}{10} = \mu \Rightarrow \alpha = 5$$

Similarly, to find the value of β, we consider

$$E(T_3) = \mu$$

$$E\left(\frac{X_1 + X_2 + \beta X_3}{5}\right) = \mu \Rightarrow \frac{E(X_1) + E(X_2) + \beta E(X_3)}{5} = \mu$$

$$\frac{\mu + \mu + \beta\mu}{5} = \mu \Rightarrow \beta = 3$$

To check which one of these proposed estimators T_1, T_2 and T_3 is most efficient, we have to find variances of the estimators and check which one is the smallest. Therefore, we consider

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

$$Var(T_1) = Var\left[\frac{X_1 + X_2 + X_3 + X_4}{4}\right]$$

$$= \frac{1}{16}\left[Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4)\right]$$

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

$$Var(T_1) = \frac{\sigma^2}{4}$$

Similarly,

$$\text{Var}(T_2) = \text{Var}\left[\frac{2X_1 + 3X_2 + 5X_4}{10}\right]$$

$$= \frac{1}{100}\left[4\text{Var}(X_1) + 9\text{Var}(X_2) + 25\text{Var}(X_4)\right]$$

$$= \frac{1}{100}\left[4\sigma^2 + 9\sigma^2 + 25\sigma^2\right] = \frac{38}{100}\sigma^2$$

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

Similarly,

$$\text{Var}(T_3) = \text{Var}\left[\frac{X_1 + X_2 + 3X_3}{5}\right] = \frac{1}{25}\left[\text{Var}(X_1) + \text{Var}(X_2) + 9\text{Var}(X_3)\right]$$

$$= \frac{1}{25}\left[\sigma^2 + \sigma^2 + 9\sigma^2\right] = \frac{11}{25}\sigma^2$$

Since the variance of the estimator T_1 is minimum, therefore, by the definition of the most efficient estimator, we conclude that the estimator T_1 is the most efficient estimator in the class of three unbiased estimators.

We now come to part (iii). We can calculate the efficiency of an unbiased estimator as

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

where T^* is the most efficient estimator.

Since the estimator T_1 is the most efficient estimator in the class of three unbiased estimators, therefore, for computing the efficiency of estimator T_2, we take estimator T_1 in place of T^*

$$e = \frac{\text{Var}(T_1)}{\text{Var}(T_2)} = \frac{\sigma^2/4}{38\sigma^2/100}$$

$$= \frac{100}{38 \times 4} = 0.658$$

Similarly, we can compute the efficiency of estimator T_3 as follows:

$$e = \frac{\text{Var}(T_1)}{\text{Var}(T_3)} = \frac{\sigma^2/4}{11\sigma^2/25}$$

$$= \frac{25}{11 \times 4} = 0.568$$

Hence, we conclude that estimator T_2 is more efficient in the comparison of estimator T_3.

Note 1: Although an unbiased estimator is usually preferred over a biased one. But, there are situations in which a biased estimator with higher efficiency can be more valuable than an unbiased estimator with lower efficiency.

Note 2: The relative efficiency of two estimators may depend on the distribution involved. For example, the mean is more efficient than the median for normal distribution, however, this is not the case for highly skewed distribution.

I think you have a curiosity to find the efficiency of an estimator. Therefore, you can try the following Self Assessment Question.

SAQ 2

Consider the question of the manufacturing fruit juice packets discussed in SAQ 1. Suppose the quality control inspector proposed third estimator for estimating the average weight of fruit juice packets μ as follows:

$$T_3 = \frac{X_1 + 2X_2 + 3X_3}{6}$$

(i) Is estimator T_3 unbiased of μ?

(ii) Which one is the most efficient estimator among the three?

(iii) Calculate the efficiency of the remaining estimators.

After understanding the concept of the efficient estimator, we now discuss some important properties of the same in the next section.

8.4 PROPERTIES OF EFFICIENT ESTIMATOR

After understanding the concept of efficiency and how to calculate it, we now discuss some properties of the efficient estimator as follows:

1. Efficient estimators are not necessarily unbiased or consistent (see Example 4).
2. The most efficient estimator is unique.

After understanding the concept of efficiency and how to check whether an unbiased estimator is more efficient or not, we would like to indicate one weakness of an efficient estimator. The efficiency is restricted to unbiased estimators and excludes biased estimators. Although an unbiased estimator is usually preferred over a biased one, however, there are situations in which a biased estimator with higher efficiency can be more valuable than an unbiased estimator with lower efficiency. Therefore, in such cases, we require some other characteristics of a good estimator which compares the estimators. In the next section, we introduce the concept of such a tool known as mean squared error.

8.5 MEAN SQUARED ERROR

In the previous sections, you studied the concept of efficiency and the most efficient estimators. With the help of efficiency, we can compare unbiased estimators and judge which one is better or most efficient in the class of unbiased estimators. You also noticed that the concept of efficiency is restricted to unbiased estimators and excludes biased estimators. But there exist so many situations where a biased estimator has a smaller variance in comparison to the unbiased estimator for a parameter. For example, if an investigator proposed two estimators for the average height of the young males in a city as follows:

(i) T_1 = Sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(ii) T_2 = a constant = 165 cm

The first estimator T_1 is the sample mean which is unbiased and its value changes with the change of the samples. Therefore, it has a certain variance greater than zero. But the second estimator T_2 does not change with the samples and always takes a single value so its variance is zero but it is highly biased because not all young males may have the same height of 165 cm. Similarly, an estimator that multiplies the sample mean by [n/(n+1)] will underestimate the population mean (biased estimator) but have a smaller variance. Therefore, the question may arise:

(i) Is a biased estimator with a smaller variance better than an unbiased estimator with a larger variance?

(ii) How can we compare such estimators?

Think about that. To compare these estimators, we require a measuring device that explicitly trades off biasedness with the variance of an estimator. A simple approach is to compare estimators based on their mean squared error. It permits us to compare biased and unbiased estimators.

In statistics, the mean squared error is an essential measure which is used to assess the performance of a point estimator (biased or unbiased). It is also necessary for relating the concepts of **precision**, **bias** and **accuracy** during the statistical estimation. It is abbreviated as MSE. The mean squared error measures the average squared difference between the estimator and the parameter.

The mean square error may be called a **risk function** which agrees with the expected value of the loss of squared error. This difference or the loss could be developed due to the randomness or due to the estimator is not representing the true unknown parameter.

Therefore, we can define the mean squared error of an estimator T of a parameter θ as

$$\text{MSE} = E\left[T-\theta\right]^2$$

It is a function of parameter θ.

We can also express the mean squared error as

$$\text{MSE} = E\left[T-\theta\right]^2 = E\left[\underline{T-E(T)}+\underline{E(T)-\theta}\right]^2 \qquad \text{[add and subtract E(T)]}$$

$$= E\left[\left\{T-E(T)\right\}^2 + 2\left\{T-E(T)\right\}\left\{E(T)-\theta\right\} + \left\{E(T)-\theta\right\}^2\right]$$

$$= E\left\{T-E(T)\right\}^2 + 2E\left\{T-E(T)\right\}\left\{E(T)-\theta\right\} + E\left\{E(T)-\theta\right\}^2$$

$$= \text{Var}(T) + 2\left\{E(T)-E(T)\right\}\left\{E(T)-\theta\right\} + E\left[\text{Bias}(T,\theta)\right]^2 \left[\because E\left(E(T)\right)=E(T)\right]$$

$$\text{MSE} = \text{Var}(T) + \left\{\text{Bias}(T,\theta)\right\}^2 \quad \left[\because E\left\{\text{Bias}(T,\theta)\right\}^2 = \left\{\text{Bias}(T,\theta)\right\}^2\right]$$

Thus, the mean squared error incorporates two components, one measuring the variability of the estimator (precision) and the other measuring its bias (accuracy). It means that an estimator will be an efficient estimator if its variance and bias should be minimum.

Therefore, a desirable property of a good estimator is not only unbiased but also has a small variance. An estimator which has a smaller mean squared error is said to be better than the other, regardless of whether they are biased or unbiased. Therefore, we can say that an estimator will be an efficient estimator if its variance as well as its bias should be minimum. You can easily understand the same using an example of a dart board on which there are several situations of hits which are shown in Fig. 8.2.

From Fig. 8.2, you can observe that the hitting of the target is too good if the bullets have less variation (closely packed) and are at the centre or near the centre.

In a similar way, we can say that an estimator is too good if its variance is small and it is unbiased or has a small bias.

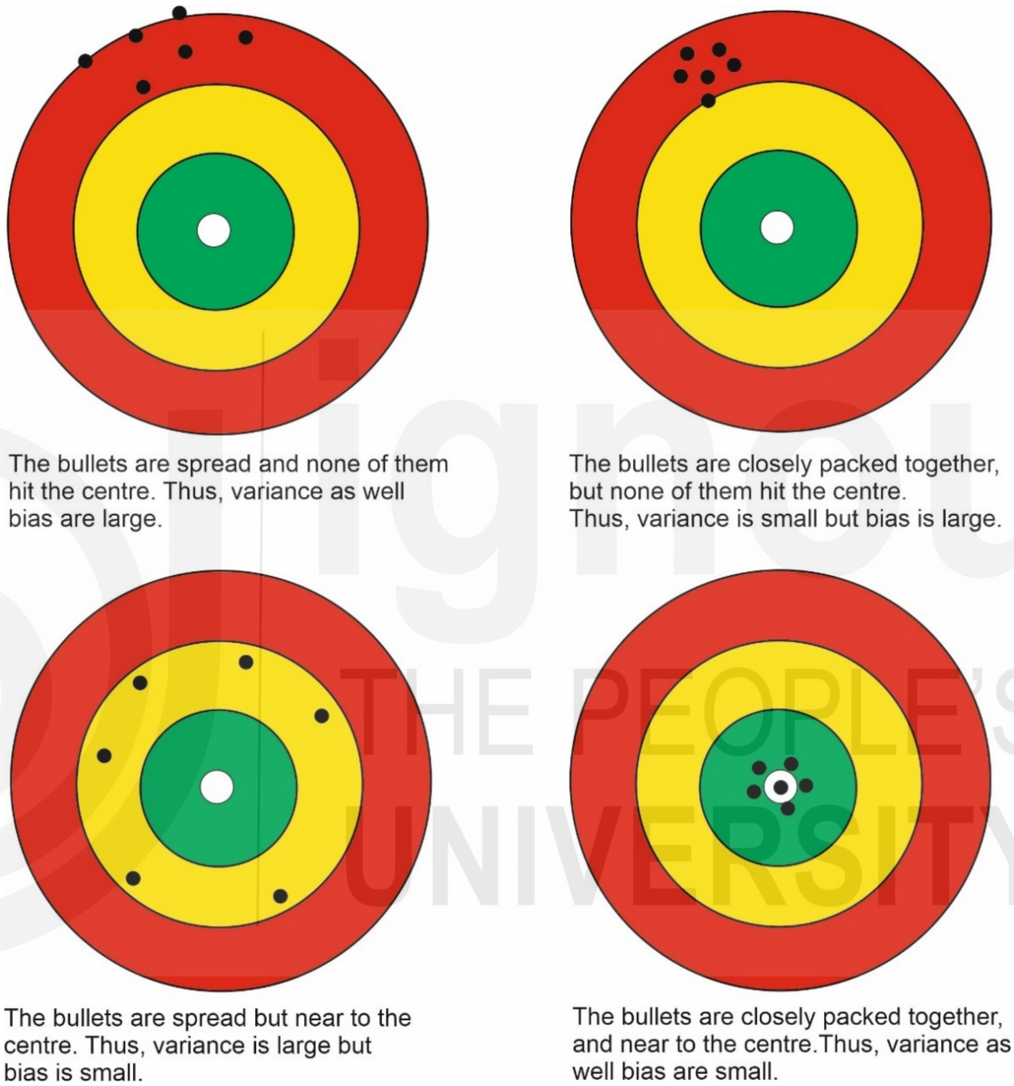


Fig. 8.2: Several situations of hits on a dart board.

If the estimator is unbiased then the mean squared error is equal to the variance of the estimator.

$$\text{MSE} = \text{Var}(T) + \{0\}^2 = \text{Var}(T)$$

For the unbiased estimator, the mean squared error is equal to the variance. Therefore, for comparing the estimators, we compare the mean squared error regardless of whether they are biased or unbiased. If T_1 and T_2 are two estimators (biased or unbiased) of a parameter θ with the same size, then the estimator T_1 is said to be more efficient than the estimator T_2 for all the same sample sizes if

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

We can also compare the mean squared errors of two estimators by using relative efficiency. If T_1 and T_2 are two estimators, then the efficiency of T_1 relative to T_2 is

$$e(T_1, T_2) = \frac{MSE(T_2)}{MSE(T_1)}$$

Sometimes the mean square error of an unbiased estimator is greater than that of a biased estimator. In such a situation, we prefer the biased estimator.

For a better understanding of the concept of the mean squared error, we take an example.

Example 4: Suppose the market researcher of Example 3 proposed the following estimators for estimating the average life of LED bulbs produced by the company as follows:

$$T_1 = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + 2,\ T_2 = \frac{1}{2}X_1 - X_2 + X_3 + \frac{1}{2}X_4$$

where X_1, X_2, X_3, and X_4 represent the life of the selected LED bulbs in the random sample. It is known that the life of the LED bulbs has mean μ and variance 2.

(i) Check whether the estimators are unbiased or not.

(ii) Find the bias, variance and mean squared error.

(iii) Which one is the more efficient estimator?

Solution: We have to check whether the estimators T_1 and T_2 are unbiased or not. Therefore, we have to find $E(T_1)$ and $E(T_2)$ of both estimators and check whether they are equal to μ or not. Since X_1, X_2, X_3 and X_4 are independent and taken from the same population with a mean μ and variance σ^2, therefore,

$$E(X_i) = \mu \text{ and } Var(X_i) = \sigma^2 \text{ for all } i = 1, 2, 3, 4$$

So we consider,

$$E(T_1) = E\left[\frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + 2\right]$$

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

$$= \frac{1}{2}\mu + \frac{1}{4}\mu + \frac{1}{4}\mu + 2 = \mu + 2 \qquad [\because E(a) = a]$$

If X and Y are two independent random variables and a & b are two constants, then
$E(aX \pm bY)$
$= aE(X) \pm bE(Y)$ and

Since $E(T_1) = \mu + 2 \neq \mu$ so the estimator T_1 is not an unbiased estimator of the parameter μ.

Similarly, we consider

$$E(T_2) = E\left[\frac{1}{2}X_1 - X_2 + X_3 + \frac{1}{2}X_4\right]$$

$$= \frac{1}{2}E(X_1) - E(X_2) + E(X_3) + \frac{1}{2}E(X_4)$$

$$= \frac{1}{2}\mu - \mu + \mu + \frac{1}{2}\mu = \mu$$

$$E(T_2) = \mu$$

Since $E(T_2) = \mu$ so the estimator T_2 is an unbiased for the parameter μ.

We now find the bias of the estimator which is not unbiased as

The bias of the estimator T_1 = $E(T_1) - \mu = \mu + 2 - \mu = 2$

We now find the variance of both estimators as

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

$$\text{Var}(T_1) = \text{Var}\left[\frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + 2\right]$$

$$= \frac{1}{4}\text{Var}(X_1) + \frac{1}{16}\text{Var}(X_2) + \frac{1}{16}\text{Var}(X_3) + \text{Var}(2)$$

$$= \frac{1}{4}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 + 0 = \frac{6}{16}\sigma^2 \quad [\because \text{Var}(a) = 0]$$

$$\text{Var}(T_1) = \frac{6}{16} \times 2 = 0.75$$

Similarly,

$$\text{Var}(T_2) = \text{Var}\left[\frac{1}{2}X_1 - X_2 + X_3 + \frac{1}{2}X_4\right]$$

$$= \frac{1}{4}\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \frac{1}{4}\text{Var}(X_4)$$

$$= \frac{1}{4}\sigma^2 + \sigma^2 + \sigma^2 + \frac{1}{4}\sigma^2 = \frac{5}{2}\sigma^2$$

$$\text{Var}(T_2) = \frac{5}{2} \times 2 = 5$$

We can calculate the mean squared error of both estimators as

$$\text{MSE}(T_1) = \text{Var}(T_1) + \{\text{Bias}(T_1, \theta)\}^2 = 0.75 + 4 = 4.75$$

$$\text{MSE}(T_2) = \text{Var}(T_2) + \{\text{Bias}(T_2, \theta)\}^2 = 5 + 0 = 5$$

Since $\text{MSE}(T_1) = 4.75 < \text{MSE}(T_2) = 5$

Thus, we conclude that the estimator T_1 is more efficient than the estimator T_2.

Example 5: Suppose the counsellor of the MST-016 course of the MSCAST programme gave the problem of estimating the variation in the marks of the cute play school children discussed in Unit 2 to the two groups of learners. Suppose the first group of learners estimates the same using the sample variance S^2 whereas the second by the sample variance S'^2 then

(i) Check whether both estimators are unbiased.

(ii) Find the mean squared error of the estimators.

Solution: As we know that

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

We also know (from Unit 3) that the sample variance has a mean

$$E\left(S^2\right) = \sigma^2$$

and variance

$$Var(S^2) = \frac{2\sigma^4}{n-1}$$

Since $E\left(S^2\right) = \sigma^2$ so S^2 is unbiased whereas

$$E\left(S'^2\right) = \frac{n-1}{n}E\left(S^2\right) = \frac{n-1}{n}\sigma^2 \quad \left[\because S'^2 = \frac{n-1}{n}S^2\right]$$

Since $E\left(S'^2\right) \neq \sigma^2$ so S'^2 is a biased estimator.

We can also find the variance of S'^2 as

$$Var(S'^2) = Var\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2 Var(S^2)$$

$$= \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2}$$

If X and Y are two independent random variables and a & b are two constants, then
E(aX ± bY)
= aE(X) ± bE(Y) and
Var(aX ± bY)
= a^2Var(X) + b^2Var(Y)

We now calculate the mean square errors as

$$MSE\left(S^2\right) = Var\left(S^2\right) + \left(Bias\right)^2 = \frac{2\sigma^4}{n-1} + 0 = \frac{2\sigma^4}{n-1}$$

$$MSE\left(S'^2\right) = Var\left(S'^2\right) + \left(Bias\right)^2 = \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2$$

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

$$= \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2}$$

$$MSE\left(S'^2\right) = \frac{(2n-1)\sigma^4}{n^2}$$

We now consider

$$MSE\left(S'^2\right) - MSE\left(S^2\right) = \left(\frac{2n-1}{n^2} - \frac{2}{n-1}\right)\sigma^4$$

$$= \left(\frac{(2n-1)\times(n-1) - 2n^2}{n^2(n-1)}\right)\sigma^4$$

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

$$MSE\left(S'^2\right) - MSE\left(S^2\right) = \left(\frac{1-3n}{n^2(n-1)}\right)\sigma^4 < 0$$

Therefore, $MSE\left(S'^2\right) < MSE\left(S^2\right)$

Thus, we conclude that the sample variance S'^2 has less mean squared error than the sample variance S^2 even S'^2 is a biased estimator.

The above example does not suggest that S^2 should not be used as an estimator of σ^2. The reasons are discussed in the remarks as follows:

Important Remarks

- From Example 5, you may have the curiosity to know why we take S^2 in place of S'^2 even mean squared error of S'^2 is less than S^2. One reason is that the concept of the mean square error is a fair criterion for location parameters, but it is not appropriate for scale parameters because the mean squared error penalizes equally for overestimation and underestimation, which is fine in the location case but in the scale case, the lower limit of the scale parameter is 0, so the estimation problem is not symmetric.
- The second reason is that if we use the mean squared error as a measure, then on average, S'^2 will be closer to σ^2 than S^2. However, S'^2 is biased and will, on average, underestimate σ^2. This fact alone may make us uncomfortable using S'^2 an estimator for σ^2. In general, the mean squared error is a function of the parameter, therefore, for some parameter values, one is better, and for other values, the other is better. Suppose we have two estimators, say, T_1 and T_2 and their respective mean squared errors are $MSE_{T_1=t_1}(\theta)$ and $MSE_{T_2=t_2}(\theta)$ which are function of the parameter θ and are likely cross to each other. For some values of θ estimator T_1 has a smaller mean squared error whereas for other values of θ, estimator T_2 has a smaller mean squared error as shown in Fig. 8.3.

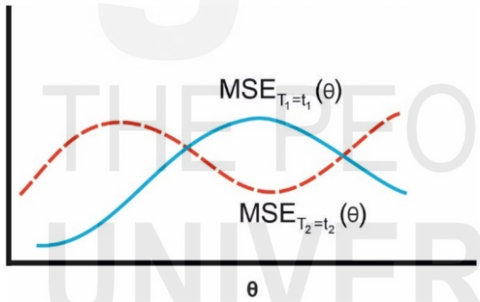


Fig. 8.3: Mean squared error of two estimators for various values of the parameter θ.

Therefore, we would have no basis for preferring one of the estimators over the other on the basis of mean squared error.

It is now time for you to try the following Self Assessment Question to make sure that you have understood the concept of mean squared error.

SAQ 3

The magnitude of earthquakes recorded in a region modelled as an exponential distribution with an unknown parameter θ whose pdf is given by

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

A researcher considered the following two estimators for estimating the parameter θ:

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } T_2 = \frac{1}{n+1}\sum_{i=1}^{n} X_i$$

Check which one is more efficient for θ.

8.6 MINIMUM VARIANCE UNBIASED ESTIMATOR

In the previous section, you studied the concept of the mean squared error, and we can define the mean squared error of an estimator T of a parameter θ as

$$MSE = E\left[T - \theta\right]^2$$

It is a function of parameter θ. Due to this, for some values of θ estimator T_1 has a smaller mean squared error whereas for other values of θ, estimator T_2 has a smaller mean squared error.

Also, you have studied that the unbiasedness criterion ensures only the average or mean of the sampling distribution of the estimator is equal to the true value of the parameter. However, it does not tell us the scatteredness (variance) of the sampling distribution of the estimator. Graphically, we show the sampling distribution of the two estimators T and T' of the parameter θ in Fig. 8.4.

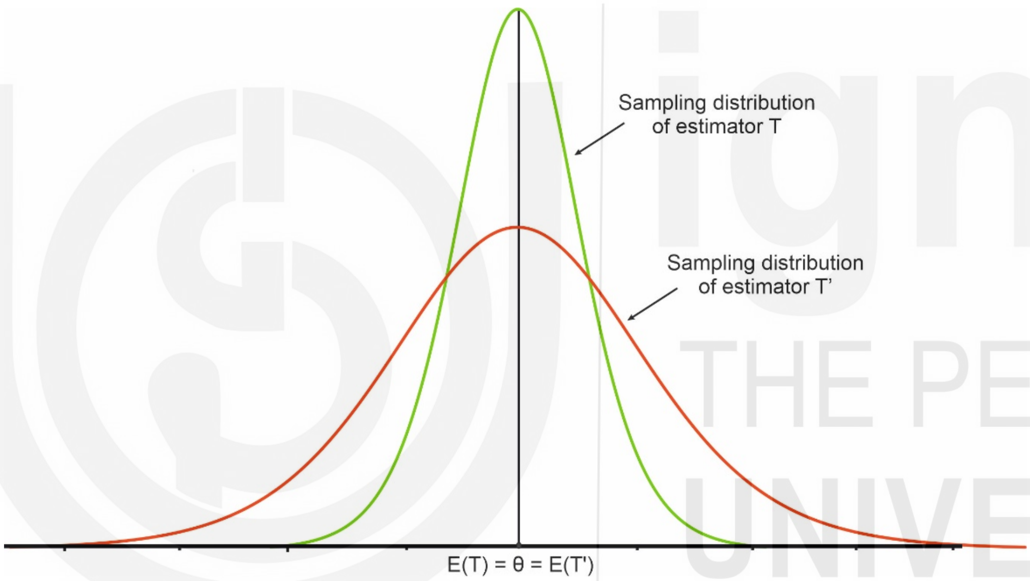


Fig. 8.4: Sampling distributions of estimator T and T`of parameter θ

From Fig. 8.4, you can observe that both estimators are unbiased however, the variance (spread) of the estimator T is smaller than the estimator T' estimator. It is clear that we would also desire the estimator whose sampling distribution not be too spread out around the true value of the parameter because if it is too spread then there will be a high probability that an estimate could be generated that will have a significant distance from the true value of the parameter. The foregoing considerations motivate that if one wishes to use an unbiased estimator of the parameter θ, one should use the unbiased estimator that also has minimum variance among all unbiased estimators of θ. Such an estimator is called a minimum variance unbiased estimator (MVUE). We can define it as follows:

An estimator T of the parameter θ is said to be a minimum variance unbiased estimator of θ if and only if

(i) $E(T) = \theta$, that is the estimator T is an unbiased estimator of the parameter θ; and

(ii) $Var(T) \leq Var(T')$ where T' is any other unbiased estimator of parameter θ.

The above definition implies that an estimator is a minimum variance unbiased estimator (MVUE) if and only if the estimator is unbiased and if there is no other unbiased estimator that has a smaller variance for any value of θ. Since it is for all values of the parameter θ, therefore it is also called a **uniformly minimum variance unbiased estimator** (UMVUE).

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

8.7 SUMMARY

In this unit, we have covered the following points:

- If T_1 and T_2 are two unbiased estimators of a parameter θ with the same size, then the estimator T_1 is said to be more efficient than the estimator T_2 if

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

- In a class of estimators of a parameter, if there exists one estimator whose variance is minimum (least) among the class, then it is said to be the most efficient estimator of that parameter.
- The efficiency of an estimator T is defined as

  $$e = \frac{Var(T^*)}{Var(T)}$$ where T^* is the most efficient estimator.

- The mean square error is defined as the average squared difference between the estimator and the parameter.

  $$MSE = E[T-\theta]^2 = Var(T) + \{Bias(T,\theta)\}^2$$

- An estimator T of the parameter θ is said to be a minimum variance unbiased estimator of θ if and only if

  (i) $E(T) = \theta$, that is, the estimator T is an unbiased estimator of the parameter θ.

  (ii) $Var(T) \leq Var(T')$ where T' is any other unbiased estimator of the parameter θ.

8.8 TERMINAL QUESTIONS

1. Describe efficiency and mean squared error.

2. Define the most efficient estimator and minimum variance unbiased estimator.

8.9 SOLUTIONS / ANSWERS

Self Assessment Questions (SAQs)

1. Since X_1, X_2, X_3 are the weight of juice packets which are taken randomly and independently from a normal population with a mean μ and variance σ^2, therefore,

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

To check whether the estimators T_1 and T_2 are unbiased or not, we find expectations of T_1 and T_2 as

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

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

If X and Y are two independent random variables and a & b are two constants, then
$E(aX \pm bY)$
$= aE(X) \pm bE(Y)$ and

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

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

Since $E(T_1) = \mu$ so it is unbiased.

We now consider

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

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

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

Since $E(T_2) = \mu$ so it is also unbiased.

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

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

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

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

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

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

We now consider

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

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

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

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

$$\text{Var}(T_2) = \frac{3\sigma^2}{8}$$

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

2. To check whether the estimator T_3 is unbiased or not, we find the expectation of T_3 as

$$E(T_3) = E\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right)$$

$$= \frac{1}{6}\left[E(X_1) + 2E(X_2) + 3E(X_3)\right]$$

$$E(T_3) = \frac{1}{6}\left[\mu + 2\mu + 3\mu\right]$$

$$= \frac{1}{6}(6\mu) = \mu$$

If X and Y are two independent random variables and a & b are two constants, then
E(aX ± bY)
= aE(X) ± bE(Y) and
Var(aX ± bY)
= a^2Var(X) + b^2Var(Y)

Since $E(T_3) = \mu$ so it is also unbiased.

We now find the variance of the estimator T_3 as

$$\text{Var}(T_{.3}) = \text{Var}\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right)$$

$$= \frac{1}{36}\left[\text{Var}(X_1) + 4\text{Var}(X_2) + 9\text{Var}(X_3)\right]$$

$$= \frac{1}{36}\left[\sigma^2 + 4\sigma^2 + 9\sigma^2\right] = \frac{14\sigma^2}{36}$$

$$\text{Var}(T_3) = \frac{7\sigma^2}{18}$$

To find the most efficient estimator among the three unbiased estimators, we compare their variances. We have

$$\text{Var}(T_1) = \frac{\sigma^2}{3} = 0.333\sigma^2,$$

$$\text{Var}(T_2) = \frac{3\sigma^2}{8} = 0.375\sigma^2 \text{ and}$$

$$\text{Var}(T_3) = \frac{7\sigma^2}{18} = 0.389\sigma^2$$

Since, the variance of the estimator T_1 is minimum, therefore, by the definition of the most efficient estimator, we conclude that the estimator T_1 is the most efficient estimator in the class of three unbiased estimators.

We can calculate the efficiency of the unbiased estimator as

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

where T^* is the most efficient estimator.

Since the estimator T_1 is the most efficient estimator in the class of three unbiased estimators, therefore, for computing the efficiency of estimator T_2, we take estimator T_1 in place of T^*

$$e = \frac{Var(T_1)}{Var(T_2)} = \frac{0.333\sigma^2}{0.375\sigma^2} = 0.88$$

Similarly, we can compute the efficiency of estimator T_3 as follows:

$$e = \frac{Var(T_1)}{Var(T_3)} = \frac{0.333\sigma^2}{0.389\sigma^2} = 0.85$$

Hence, we conclude that estimator T_2 is more efficient in comparison of estimator T_3.

3. Since the magnitude of the earthquakes recorded in the region is modelled as the exponential distribution. Therefore, it has mean θ and variance θ^2. To check which estimator is more efficient, first, we check whether estimators T_1 are T_2 are unbiased or not. For that, we have to find $E(T_1)$ and $E(T_2)$ and check whether it is equal to μ or not. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from exponential distribution. Since the sample observations are independent and taken from the same population with a mean θ and variance θ^2, therefore,

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

We consider,

$$E(T_1) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$$

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

$$E(T_1) = \theta$$

If X and Y are two independent random variables and a & b are two constants, then
$E(aX \pm bY)$
$= aE(X) \pm bE(Y)$ and

Hence, the estimator T_1 is an unbiased estimator of θ.

Similarly,

$$E(T_2) = E\left[\frac{1}{n+1}\sum_{i=1}^{n} X_i\right] = \frac{1}{n+1}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$$

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

Since $E(T_2) = \left(\frac{n}{n+1}\right)\theta \neq \theta$

Hence, the estimator T_2 is not an unbiased estimator of the parameter θ.

Since both estimators are not unbiased estimators so we use the concept of the mean squared error for judging which one is more efficient. Therefore, we have to find the variance of these unbiased estimators. We now consider

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

$$\text{Var}(T_1) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \text{Var}\left[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right]$$

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

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

$$\text{Var}(T_1) = \frac{\theta^2}{n}$$

We now calculate the variance of estimator T_2 as

$$\text{Var}(T_2) = \text{Var}\left[\frac{1}{n+1}\sum_{i=1}^{n} X_i\right] = \frac{1}{(n+1)^2}\left[\text{Var}(X_1 + X_2 + \ldots + X_n)\right]$$

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

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

$$\text{Var}(T_2) = \frac{n\theta^2}{(n+1)}$$

We now calculate the mean square errors as

$$\text{MSE}(T_1) = \text{Var}(T_1) + (\text{bias})^2 = \text{Var}(T_1) + \left[E(T_1) - \theta\right]^2$$

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

$$\text{MSE}(T_2) = \text{Var}(T_2) + (\text{bias})^2 = \text{Var}(T_2) + \left[E(T_2) - \theta\right]^2$$

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

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

Since, $\text{MSE}(T_2) < \text{MSE}(T_1)$, therefore, we conclude that the estimator T_2 is a more efficient estimator than the estimator T_1 for estimating the magnitude of the earthquake in the region.

Terminal Questions (TQs)

1. Refer to Sections 8.2 and 8.5.

2. Refer to Sections 8.3 and 8.6.

UNIT 9

SUFFICIENCY AND MINIMAL SUFFICIENCY

Structure

9.1 Introduction

Expected Learning Outcomes

9.2 Joint Probability Density (Mass) Function

9.3 Concept of Sufficiency

9.4 Fisher-Neyman Factorization Theorem

9.5 Properties of Sufficient Statistic

9.6 Minimal Sufficient Statistic

9.7 Summary

9.8 Terminal Questions

9.9 Solutions /Answers

9.1 INTRODUCTION

In Units 6, 7 and 8, you have studied the characteristics of a good estimator namely: unbiasedness, consistency and efficiency. Let us have a look at them.

- An estimator is said to be unbiased for a parameter θ if and only if the average/mean of the sampling distribution of the estimator is equal to the true value of the parameter. In other words, 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**, that is, $E(T) = \theta$
- An estimator T_n is said to be a consistent estimator of θ if T_n converges to θ in probability.
- If T_1 and T_2 are two unbiased estimators of a parameter θ with the same size, then the estimator T_1 is said to be more efficient than the estimator T_2 if $\mathrm{Var}(T_1) < \mathrm{Var}(T_2)$ for all n.

In the continuation of finding the best estimator, we introduce the concept of sufficiency in this unit.

This unit is divided into nine sections. Section 9.1 is introductory in nature. The joint probability density (mass) function which is used to find a sufficient statistic is defined in Section 9.2. Section 9.3 is devoted to explaining the concept of sufficient statistic. Section 9.4 explores the Fisher-Nayman factorization theorem. The properties of the sufficient statistic are described in Section 9.5. The concept of minimal sufficient statistic is described in Section

Tools You Will Need

The following terms are considered essential background material for this Unit. If you doubt your knowledge of any of these terms, you should review the appropriate Unit or section before proceeding:

- Sampling distributions (Units 2,3, 4 and 5).
- Basic terms of estimation (Unit 6).
- Unbiased, consistency and efficiency (Units 6, 7 and 8).
- Probability distributions (MST-012).

Unit Writer- Dr. Prabhat Kumar Sangal, School of Sciences, IGNOU, New Delhi

9.6. The unit ends by providing a summary of what we have discussed in this unit in Section 9.7. The terminal questions and the solution of the SAQs/TQs are given in Sections 9.8 and 9.9, respectively.

Expected Learning Outcomes

After studying this unit, you should be able to:

- define the joint probability density (mass) function and how to compute it;
- explain the concept of sufficient statistic and how to find sufficient statistic for a parameter;
- describe the Fisher-Nayman Factorization theorem and how to use it to find the sufficient statistic;
- explain the properties of sufficient statistic; and
- describe the concept of minimal sufficient statistic.

9.2 JOINT PROBABILITY DENSITY (MASS) FUNCTION

Before discussing sufficiency, we discuss the joint probability density (mass) function. This term is very useful in understanding the concept of sufficiency.

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 the sample values is denoted by $f(x_1, x_1, \ldots, x_n, \theta)$ and it is a function of the sample observations. We can define it as

For discrete case:

The joint probability mass function is defined as

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

If A and B are two independent events, then

P(A∩B) =P(A)P(B)

Since sample observations $X_1, X_2, \ldots, X_n$ are independent, therefore, we can write the above expression as

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

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

For continuous case:

The joint probability density function is defined as

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

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

Let us understand the process of finding the joint probability density (mass) function by taking some examples.

Example 1: Suppose the number of weekly accidents occurring on a mile stretch of a particular road follows a Poisson distribution with a parameter λ whose pdf is given by

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

To estimate the number of road accidents, a transport officer randomly selected a sample of the number of road accidents, say, $X_1, X_2, \ldots, X_n$, then find the joint probability mass function of $X_1, X_2, \ldots, X_n$.

Solution: The probability mass function of the Poisson distribution is given by

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

Since the Poisson distribution is a discrete distribution, therefore, by the definition of the joint probability mass function of the sample observations $X_1, X_2, \ldots, X_n$, we have

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

Since the number of weekly accidents (sample observations) follows the Poisson distribution with parameter λ, therefore, we can obtain the joint probability mass function by putting X as $x_1, x_2, \ldots, x_n$ in the probability mass function as mentioned above. Therefore,

$$f(x_1, x_1, \ldots, x_n, \lambda) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \ldots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

Collecting like terms, we get

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

Now, simplifying, by adding up all λs in the exponents as well as x_i's in the power of λ, we get the required joint pdf of the sample observations as

$$f(x_1, x_1, \ldots, x_n, \lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \quad \left[\prod_{i=1}^{n} x_i! - \text{repersents the product of } x_i!\right]$$

Example 2: The magnitude of the earthquakes recorded in a region modelled as an exponential distribution with an unknown parameter θ whose pdf is given by

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

If a seismologist measured the magnitude of the n random earthquakes in that region and denoted by $X_1, X_2, \ldots, X_n$ then find the joint probability density of the sample observations.

Solution: Since the exponential distribution is a continuous distribution, therefore, by the definition of the joint probability density function of the sample observations $X_1, X_2, \ldots, X_n$, we have

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

Since the magnitude of the earthquakes follows the exponential distribution with parameter θ, therefore, we can obtain the joint probability density function by putting X as $x_1, x_2, \ldots, x_n$ in the probability density function of the exponential distribution as

$$f(x_1, x_1, \ldots, x_n, \theta) = \theta e^{-\theta x_1} . \theta e^{-\theta x_2} \ \ldots \ \theta e^{-\theta x_n}$$

Collecting like terms, we get

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

Now, simplifying, by adding up all θs as well as x_i's in the exponents, we get the required joint pdf of the sample observations as

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

Let us check your understanding of the above by answering the following Self Assessment Question.

SAQ 1

To test the effectiveness of a new drug in controlling systolic blood pressure, a medicine scientist applied the drug to 10 systolic blood pressure patients. If the numbers of the patients who were cured the disease follow binomial distribution with parameters 10 and p, then find the joint probability mass function.

9.3 CONCEPT OF SUFFICIENCY

Any quantity calculated from sample values and does not contain any unknown parameter is known as a statistic.

In statistical estimation, the aim of the statistician is to estimate the value of the unknown parameter (θ) on the basis of an estimator/statistic. As the sample mean is used to estimate the population mean; the sample variance is used to estimate the population variation. Also, an estimator or a statistic T (function of sample values) is said to be the best estimator of a population parameter θ if T is close to the parameter θ (unbiased and efficient). In this way, a statistic which is used in estimation or drawing inferences about the population parameter is a function of sample observations, rather than the full data set. Obviously, there are lots of functions of the sample observations, and so lots of statistics. For example, suppose you are interested in knowing the average age of Facebook users on the basis of a sample $X_1, X_2, \ldots, X_n$ of size n, then some functions of the sample observations through which you can estimate/obtain the average age of Facebook users are given as follows:

- $T_1(X_1, X_2 \ldots, X_n)$ = Sample mean $\bar{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$
- $T_2(X_1, X_2 \ldots, X_n)$ = Sample median $\tilde{X} = \text{median}(X_1, X_2 \ldots, X_n)$
- $T_3(X_1, X_2 \ldots, X_n)$ = Sample mode $X_0 = \text{mode}(X_1, X_2 \ldots, X_n)$
- $T_4(X_1, X_2 \ldots, X_n) = \dfrac{\max(X_1, X_2, \ldots, X_n) + \min(X_1, X_2, \ldots, X_n)}{2}$

All estimators are statistics because they do not depend on the unknown population parameter. Obviously, there are lots of functions of $X_1, X_2, \ldots, X_n$ and so lots of statistics.

To understand when a statistic is said to be a sufficient statistic, we consider an example. Suppose we have three coins, and if we toss all three coins simultaneously, then the possible outcomes are:

(T, T, T), (T, T, H), (T, H, T), (H, T, T), (T, H, H), (H, T, H), (H, H, T), (H, H, H)

If we represent head (H) by 1 and tail (T) by 0, we can represent the outcomes as

(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)

Suppose we are interested in estimating the number of heads on the basis of samples, so we have to worry about all outcomes as mentioned above. We can also estimate the number of heads using a statistic, $T = X_1 + X_2 + X_3$ then it takes the values 0, 1, 2, and 3 as shown in the following table:

Outcomes	(0, 0, 0)	(0, 0, 1), (0, 1, 0), (1, 0, 0)	(0, 1, 1), (1, 0, 1), (1, 1, 0)	(1, 1, 1)
Statistic (T)	0	1	2	3

It means that using the statistic T instead of the sample, we condense the data into four subgroups instead of eight. And if we use the statistic T instead of the sample, we have to worry about 4 subgroups instead of 8. Now some questions may arise:

(i) Is some information lost by using a statistic instead of full data?

(ii) Does the random sample contain any more information about the population than this?

To get the answers to such types of questions, we have to study the concept of sufficient statistic.

Suppose there are two researchers, say, Rajesh and Prabhat. Rajesh knows a particular outcome (sample) of all possible cases (X_1, X_2, X_3), say (1,0,1), however, Prabhat only knows the value of a statistic $T = X_1 + X_2 + X_3$ that is 2. Since Prabhat knows that 2 heads come when three coins are tossed simultaneously, therefore, he can generate a random sample (X_1', X_2', X_3') such as (1, 1, 0) or (0, 1, 1) or (1, 0, 1) and he can use his random sample (X_1', X_2', X_3') to compute whatever Rajesh computes using his random sample (X_1, X_2, X_3). Therefore, we can say that there is no information lost using the statistic $T = X_1 + X_2 + X_3$. In other words, we can say that Prabhat who knows the value of T can do just as good a job of estimating the unknown parameter θ as Rajesh who knows the entire random sample. Thus, we can say that statistic T is a sufficient statistic. So, a statistic is sufficient if it is just as informative as the full data. We can define the sufficient statistic as

A sufficient statistic is a particular kind of statistic which condenses the data in such a way that no information about the parameter is lost.

The concept of sufficiency was introduced by Ronald. A. Fisher in the 1920s, and refined by Jerzy Neyman in the 1930s. We now formally define the sufficient estimator/statistic as

"A statistic T = T(X) is said to be a sufficient statistic for a parameter θ if it contains "all of the information" about θ that is available in the sample."

In other words, we do not lose any information about θ by reducing the sample X to the statistic T. This property of an estimator is called sufficiency.

According to statistician Ronald A. Fisher,

"…no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter."

It means a sufficient statistic contains all information about θ, that is, contained in the sample and if we know the value of the sufficient statistic, then the sample values themselves are not needed and can nothing tell you more about θ.

Now the question may arise:

How do we check whether a statistic contains all the information about θ that is contained in the sample?

As you know the actual data has a certain probability distribution such as Normal, Bernoulli, Binomial, etc., which in general also, involve the parameter(s). A statistic $T = t(X_1, X_2, \cdot\cdot\cdot, X_n)$ is a sufficient statistic for a parameter θ if, for each t, the conditional distribution of $X_1, X_2, \ldots, X_n$ given $T = t$ does not depend on θ.

Mathematically,

For discrete distribution

$$P\left[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T = t\right] = \frac{P\left[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, T = t\right]}{P\left[T = t\right]}$$

$$= g\left(x_1, x_2, \ldots, x_n\right)$$

For continuous distribution

$$f\left(x_1, x_2, \ldots, x_n \mid t\right) = \frac{f\left(x_1, x_2, \ldots, x_n, t\right)}{f\left(t\right)} = g\left(x_1, x_2, \ldots, x_n\right)$$

where the numerator is the joint probability density (mass) function of the sample values and the function $g\left(x_1, x_2, \ldots, x_n\right)$ does not depend on the parameter θ.

Let us take an example to understand the above definition.

Example 3: Consider the example of tossing three coins simultaneously and assume that the probability of getting head is p. Check whether the statistic $T = X_1 + X_2 + X_3$ is sufficient or not.

Solution: To check whether the statistic $T = X_1 + X_2 + X_3$ is sufficient or not, we have to find the conditional distribution of the sample observations given T. For that, we have to use some concepts of probability distributions which you have studied in Unit 9 of the course MST-012: Probability and Probability Distributions.

The probability of getting a head in a single coin is p and not getting the same is 1 – p. Since we perform a random experiment (tossing a coin independently) and the outcome of each has two categories: head and tail, then the probability distribution of a random variable which takes the value 1 if the outcome is a head (success) and 0 if the outcome is a tail (failure) is known as Bernoulli distribution. Therefore, we can write the probability mass function as

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

Since T = $X_1 + X_2 + X_3$, is the sum of the Bernoulli distributed random variables, therefore, T follows a binomial distribution with parameters n = 3 and p whose probability mass function is given as

$$P[T = t] = {}^nC_t p^t (1-p)^{n-t}; \quad t = 0,1,2,3 \ \& \ n = 3$$

We are now ready to find the conditional distribution of the sample observations given T and check whether it is independent of the parameter p or not. Therefore, we consider

$$P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n \mid T = t] = \frac{P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n, T = t]}{P[T = t]}$$

Suppose we observed a random sample of size n = 3 in which X_1 = 1, X_2 = 0 and X_3 = 1. In this case:

$$P[X_1 = 1, X_2 = 0, X_3 = 1, T = 0] = 0$$

$$P[X_1 = 1, X_2 = 0, X_3 = 1, T = 1] = 0$$

$$P[X_1 = 1, X_2 = 0, X_3 = 1, T = 3] = 0$$

Because for X_1 = 1, X_2 = 0, X_3 = 1, $T = \sum_{i=1}^{3} X_i = 1 + 0 + 1 = 2$ is possible whereas T =0, 1 and 3 are not possible. So the above expression becomes impossible and their probability becomes 0 whereas only event (X_1 = 1, X_2 = 0, X_3 = 1, T = 2) is possible. In this case, we have, by independence:

$$P[X_1 = 1, X_2 = 0, X_3 = 1, T = 2] = P[X_1 = 1]P[X_2 = 0]P[X_3 = 1]$$

$$= p(1-p)p = p^2 (1-p)^{3-1}$$

So, in general

$$P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n, T = t] = P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n] \text{ if } \sum_{i=1}^{n} x_i = t$$

and

$$P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n, T = t] = 0 \text{ if } \sum_{i=1}^{n} x_i \neq t$$

Therefore,

$$P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n \mid T = t] = \frac{P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n, T = t]}{P[T = t]}$$

If each trial of a random experiment is termed in one of the two possible categories traditionally known as a success or a failure then such a trial is known as Bernoulli trial. If we perform a random experiment and the realisation of a trial has only two categories success or failure, then probability distribution of a random variable which takes value 1 if outcome is a success and 0 if outcome is a failure is known as Bernoulli distribution. The probability mass function of Bernoulli random variable X is given by

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

$$= \frac{p^2(1-p)^{3-2}}{{}^3C_2 p^2 (1-p)^{3-2}} = \frac{1}{{}^3C_2}$$

We have just shown that the conditional distribution of the sample $X_1, X_2, \ldots, X_n$ given T = t does not depend on the parameter p. Therefore, T is indeed sufficient for p. That is, once the value of T is known neither the sample nor other function of $X_1, X_2, \ldots, X_n$ will provide any additional information about the possible value of p.

Now, it is time for you to check your understanding of the condition distribution by solving the following Self Assessment Question.

SAQ 2

Suppose the time between customers who enter a certain shop follows exponential distribution with parameter θ whose pdf is given as follows:

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

To estimate the parameter θ, the market researcher proposed the statistic/estimator $T = \bar{X}$ based on sample $X_1, X_2, \ldots, X_n$ then show that T is a sufficient statistic for the parameter θ using condition distribution.

9.4 FISHER-NEYMAN FACTORIZATION THEOREM

In the previous section, you studied the definition of sufficient statistic for a parameter θ and a statistic $T = t(X_1, X_2, \cdot\cdot\cdot, X_n)$ is said to be a sufficient statistic if, for each t, the conditional distribution of $X_1, X_2, \ldots, X_n$ given T = t does not depend on θ.

While this definition of sufficient statistic is fairly simple, but finding the conditional distribution is the tough part. In fact, most statisticians consider it **extremely difficult**. One, slightly easier, way to find the conditional distribution is to use the Factorization Theorem. The Factorization theorem was given by Fisher and Nayman so it is called the Fisher-Nayman factorization theorem or simply called the factorization theorem. Suppose we would like to get information about the parameter value θ from our sample. The concept of factorization theorem allows us to separate information contained in the sample into two parts. One part contains all valuable information as long as we are concerned with parameter θ, while the other part contains pure noise in the sense that this part has no valuable information. Thus, we can ignore the latter part.

Let us learn what the factorization theorem states.

Statement of Fisher-Nayman Factorization Theorem: Let $X_1, X_2, \ldots, X_n$ 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 sufficient for a parameter θ if and only if the joint probability density (mass) function of sample observations $X_1, X_2, \ldots, X_n$ can be factored as

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

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

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

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

Note 1: The factorization theorem should not be used to show that a given statistic or estimator T is not sufficient.

Let us learn how to apply the factorization theorem to obtain a sufficient statistic with the help of some examples.

Example 4: Suppose the number of visitors visiting a website per hour follows Poisson distribution with parameter λ. Find a sufficient statistic for λ.

Solution: To find the sufficient statistic for λ, we can use the factorization theorem. For that, we have to find the joint probability density function of the sample values. We know that the probability mass function of Poisson distribution with parameter λ is

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

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

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

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

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

$$f(x_1, x_2, ..., x_n, \lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

We now try to factor the above joint probability mass function as the product of two functions, one of which is a function of the parameter (λ) and another is independent of the parameter (λ). We can factor the joint probability mass function as

$$f(x_1, x_2, ..., x_n, \lambda) = \left(e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}\right) \times \left(\frac{1}{\prod_{i=1}^{n} x_i!}\right)$$

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

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

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

But, wait a second! We can also write the joint probability mass function as:

$$f\left(x_1, x_2, ..., x_n, \lambda\right)=\frac{e^{-n\lambda}\lambda^{n\bar{x}}}{\prod_{i=1}^{n}x_i!} \qquad \left(\text{since } \bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\right)$$

And factor as

$$f\left(x_1, x_2, ..., x_n, \lambda\right)=\left(e^{-n\lambda}\lambda^{n\bar{x}}\right)\times\left(\frac{1}{\prod_{i=1}^{n}x_i!}\right)$$
$$=g\left[t(x),\lambda\right].h\left(x_1, x_2, ..., x_n\right)$$

Therefore, the Factorization Theorem of sufficiency tells us that $\bar{X}$ is a sufficient statistic for λ.

If you think about it, it makes sense that $\bar{X}$ and $\sum_{i=1}^{n}X_i$ are both sufficient statistics, because if we know $\bar{X}$, we can easily find $\sum_{i=1}^{n}X_i$. And, if we know $\sum_{i=1}^{n}X_i$, we can easily find $\bar{X}$. Also, both condense the data the same.

Note 2: Since throughout the course we are using a capital letter for statistic or estimator, therefore, in the last line of the above example, we use $\sum_{i=1}^{n}X_i$ in place of $\sum_{i=1}^{n}x_i$. Thus, in all the examples and exercises relating to sufficient statistic, we are using a similar approach.

Note 3: It is easy to see that if f(t) is a one-to-one function and T is a sufficient statistic for the parameter θ, then f(T) is also a sufficient statistic for θ. In particular, we can multiply/divide a sufficient statistic by a non-zero constant and get another sufficient statistic.

Example 5: If the life of the lithium batteries used in cars follows a normal distribution with mean μ months and standard deviation 5 months, then show that the sample mean $\bar{X}$ is a sufficient statistic for μ. Are $\bar{X}^2$ and $\bar{X}^3$ sufficient statistics for μ.

Solution: To find the sufficient statistic for the parameter μ, we use the factorization theorem. We know that the probability density function of the normal distribution with mean μ and standard deviation σ as

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

In our case, σ is known as 5, therefore,

$$f\left(x, \mu\right) = \frac{1}{\sqrt{2\pi \times 25}} e^{-\frac{1}{2\times 25}(x-\mu)^2} = \frac{1}{\sqrt{50\pi}} e^{-\frac{1}{50}(x-\mu)^2}$$

We can obtain the joint probability density function of the sample observations $X_1, X_2, ..., X_n$ as

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

$$= \frac{1}{\sqrt{50\pi}} e^{-\frac{1}{50}(x_1-\mu)^2} . \frac{1}{\sqrt{50\pi}} e^{-\frac{1}{50}(x_2-\mu)^2} ... \frac{1}{\sqrt{50\pi}} e^{-\frac{1}{50}(x_n-\mu)^2}$$

$$f\left(x_1, x_2, ...x_n, \mu\right) = \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\sum_{i=1}^{n}(x_i-\mu)^2}$$

We now try to factor the joint probability density function as the product of two functions, one of which is a function of parameter μ and statistic and another is independent of the parameter μ. But there is no separate term which is a function of x_i's in $\sum_{i=1}^{n}(x_i-\mu)^2$. Therefore, we use a trick to factor of the joint pdf.

An easier task is to add and subtract $\overline{x}$ to the quantity in parentheses in the summation. That is

$$f\left(x_1, x_2, ...x_n, \mu\right) = \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-\mu)^2}$$

Now, squaring the quantity in parentheses, we get

$$f\left(x_1, x_2, ...x_n, \mu\right) = \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\sum_{i=1}^{n}\left[(x_i-\overline{x})^2+(\overline{x}-\mu)^2-2(x_i-\overline{x})(\overline{x}-\mu)\right]}$$

And then distributing the summation, we get

$$f\left(x_1, x_2, ...x_n, \mu\right) = \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2+\sum_{i=1}^{n}(\overline{x}-\mu)^2-2(\overline{x}-\mu)\sum_{i=1}^{n}(x_i-\overline{x})\right]}$$

But the last term in the exponent $\sum_{i=1}^{n}\left(x_i-\overline{x}\right)=0$ (by property of mean), and the second term can be added up n times because it does not depend on the index i, therefore, we get

$$f\left(x_1, x_2, ...x_n, \mu\right) = \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\mu)^2-0\right]}$$

$$= \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\mu)^2\right]}$$

$$f(x_1, x_2, \ldots x_n, \mu) = \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{n}{50}(\bar{x}-\mu)^2}$$

Now, we can easily factor the joint probability density function as

$$f(x_1, x_2, \ldots x_n, \mu) = e^{-\frac{n}{50}(\bar{x}-\mu)^2} \times \left[\left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$$

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

Where $g\left[t(x), \mu\right] = e^{-\frac{n}{50}(\bar{x}-\mu)^2}$ is a function of parameter μ and sample values $x_1, x_2, \ldots, x_n$ only through $t(x) = \bar{x}$, whereas $h(x_1, x_2, \ldots, x_n) = \left(\frac{1}{\sqrt{50\pi}}\right)^n e^{-\frac{1}{50}\sum_{i=1}^{n}(x_i-\bar{x})^2}$ is independent of μ. Hence, by the factorization theorem of sufficiency, the statistic sample mean $\bar{X}$ is a sufficient statistic for μ when σ^2 is known.

We now check whether $\bar{X}^2$ and $\bar{X}^3$ are sufficient statistics for μ.

Since $y = \bar{X}^2$ is not a one-to-one function of $\bar{X}$ because we get two possible values, namely $-\bar{X}$ and $\bar{X}$. Therefore, $\bar{X}^2$ is not a sufficient statistic for μ.

If we are given the value of $y = \bar{X}^3$, we can easily get the single value of $\bar{X}$ through the one-to-one function $y^{1/3}$. Therefore, $\bar{X}^3$ is also sufficient for μ.

Example 6: Suppose the time between medication doses follows a gamma distribution whose pdf is given as follows:

$$f(x, a, b) == \frac{b^a}{\overline{|a}} e^{-bx} x^{a-1}; \quad x > 0, \quad a, b > 0$$

Obtain sufficient statistic for

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

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

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

Solution: To find the sufficient statistic, first, we have to find the joint probability density function of the sample values of the gamma distribution. Let $X_1, X_2, \ldots, X_n$ be a random sample taken from the gamma distribution with parameters 'a' and 'b'.

We can obtain the joint probability density function as

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

$$= \frac{b^a}{\overline{|a}} e^{-bx_1} x_1^{a-1} . \frac{b^a}{\overline{|a}} e^{-bx_2} x_2^{a-1} \ldots \frac{b^a}{\overline{|a}} e^{-bx_n} x_n^{a-1}$$

After simplification, we get

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

Since there are two parameters (a, b) so we consider the following cases:

Case I: When 'b' is known then we treat 'b' as a constant and find the sufficient statistic for 'a'.

We can factor the joint probability density function as

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

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

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

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

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

Case II: When 'a' is known then we treat 'a' as a constant and find the sufficient statistic for 'b'.

We now factor the joint probability density function as

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

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

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

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

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

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

Since we cannot separate any term of the joint probability density function which is independent to both 'a' and 'b', therefore, we can factor the joint probability density function as

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

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

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

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

Example 7: If $X_1, X_2, ..., X_n$ is a random sample taken from a uniform distribution U(α, β), find the sufficient statistics for α and β.

Solution: The probability density function of $U(\alpha, \beta)$ is given by

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

We can obtain the joint probability density function as

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

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

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

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

Since the range of variables depends upon the parameters so we consider ordered statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ instead of the sample observations $X_1, X_2, \ldots, X_n$. Therefore, we can write the joint probability density function as

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

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

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, $h(x_1, x_2, ... x_n) = 1$ and independent of parameters 'α' and 'β'.

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

Now, you will get more clearly about how to factor the joint probability density (mass) function and obtain the sufficient statistic, when you try the following Self Assessment Question.

SAQ 3

(i) If the time between two customers arriving in a bank follows an exponential distribution with parameter θ, then find the sufficient statistic for θ.

(ii) Consider Example 5 of the life of the lithium batteries used in cars. If the life of the lithium battery follows a normal distribution with a mean 95 months and variance σ^2. Find the sufficient statistic for σ^2.

(iii) The time interval between two metro trains follows the uniform distribution [0, θ]. Find the sufficient statistic for θ.

Let us discuss the properties of the sufficient statistic in the next section.

9.5 PROPERTIES OF SUFFICIENT STATISTIC

After understanding the concept of sufficient statistic, we now describe some important properties of the sufficient statistic as follows:

1. A sufficient estimator/statistic is always a consistent estimator.

2. A sufficient estimator/statistic 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 statistics because both contain all information about a parameter of the population.

5. If T is a sufficient statistic for the parameter θ and f(T) is a one-to-one function of T then f(T) is also sufficient for θ. For example, if $T = \sum X_i$ is sufficient statistic for the 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.

After understanding the sufficiency, we now discuss the concept of minimal sufficient statistic in the next section.

9.6 MINIMAL SUFFICIENT STATISTIC

In Sections 9.3 and 9.4, you studied the concept of sufficient statistic and how to use the Fisher-Nayman factorization theorem to find a sufficient statistic/ estimator, respectively. In section 9.5, you studied the properties of the sufficient statistic and one property of the sufficient statistic is that every one-to-one function of a sufficient estimator is also sufficient statistic for the same parameter and also the whole sample and order statistics are sufficient for a parameter. In Example 2, you have seen that the sample mean $\bar{X}$ and $\bar{X}^3$ are sufficient statistics for the population mean when the sample is taken from normal distribution. Also, the whole sample $X_1, X_2, \ldots, X_n$ and order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are also sufficient statistics for the population mean μ. Now the question may arise:

(i) Are they all as "good" as one another?

(ii) Is there some reason to prefer one over another?

It means that to give the answer of the same, we require sometimes more. The minimal sufficient statistic does the same job for us. Actually, we use the concept of sufficient statistic to condense the data $X_1, X_2, \ldots, X_n$ using a statistic in such a way that no information will be lost. In general, if a statistic condenses the data more than the other then we prefer that statistic. A sufficient statistic is said to be a minimal sufficient statistic if no other sufficient statistic condenses the data more.

We can formally define minimal sufficient statistic as

A sufficient statistic is defined to be a minimal sufficient statistic if and only if it is a function of every other sufficient statistic.

Like the definition of the consistent estimator and sufficient statistic, the definition of the minimal sufficient statistic has little use in finding a minimal sufficient statistic. It is noted that if the joint probability density function is properly factored, then the Fisher-Nayman factorization criterion will give a minimal sufficient statistic. So, we can say that the statistic found in Examples 3 to 7 are minimal.

9.7 SUMMARY

In this unit, we have covered the following points:

- The joint probability mass function for discrete distribution sample values is defined as

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

- The joint probability density function for continuous distribution sample values is defined as

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

- A statistic T = T(X) is said to be a sufficient statistic for a parameter θ if it contains "all of the information" about θ that is available in the sample. In other words, we do not lose any information about θ by reducing the sample X to the statistic T. This property of an estimator is called sufficiency.

- A statistic $T = t(X_1, X_2,\ldots, X_n)$ is a sufficient statistic if, for each t, the conditional distribution of $X_1, X_2,\ldots, X_n$ given T = t and θ does not depend on θ.
- A statistic or estimator T is said to be sufficient for a parameter θ if and only if the joint probability density (mass) function of $X_1, X_2,\ldots, X_n$ can be factored as

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

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

- Explain the properties of sufficient statistic.
- A sufficient statistic is defined to be a minimal sufficient statistic if and only if it is a function of every other sufficient statistic.

9.8 TERMINAL QUESTIONS

1. Consider Example 3 and find the sufficient statistic for the parameter θ using the Fisher-Nayman factorization theorem. Is it a minimal sufficient statistic.
2. Define sufficiency and minimal sufficient statistic.

9.9 SOLUTIONS / ANSWERS

Self Assessment Questions (SAQs)

1. Since the number of patients who were cured the disease follows a binomial distribution with parameters n = 10 and p, therefore, its probability mass function is given by

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, 2, \ldots n \;\&\; 0 \le p \le 1$$

$$= \binom{10}{x} p^x (1-p)^{10-x}; \quad x = 1, 2, \ldots, 10 \;\; (\text{since } n = 10)$$

If $X_1, X_2,\ldots, X_n$ denote the outcomes of the drug then by the definition of the joint probability mass function of the sample observations $X_1, X_2, \ldots, X_{10}$, we have

$$f(x_1, x_1, \ldots, x_n, p) = P[X_1 = x_1]P[X_2 = x_2]\ldots P[X_n = x_{10}]$$

We can obtain the joint probability mass function by putting X as $x_1, x_2, \ldots, x_{10}$ in the probability mass function as mentioned above. Therefore,

$$f(x_1, x_1, ..., x_{10}, p) = \binom{10}{x_1} p^{x_1} (1-p)^{10-x_1} . \binom{10}{x_2} p^{x_2} (1-p)^{10-x_2}$$
$$... \binom{10}{x_{10}} p^{x_{10}} (1-p)^{10-x_{10}}$$

Collecting like terms, we get

$$f(x_1, x_1, ..., x_{10}, p) = \prod_{i=1}^{10} \binom{10}{x_i} p^{x_1+x_2+...+x_{10}} (1-p)^{\underbrace{10+10+...+10}_{10-\text{times}}-(x_1+x_2+...+x_{10})}$$

On simplifying, we get the required joint probability mass function as

$$f(x_1, x_1, ..., x_{10}, p) = \prod_{i=1}^{10} \binom{10}{x_i} p^{\sum_{i=1}^{10} x_i} (1-p)^{100-\sum_{i=1}^{10} x_i}$$

2. To show that the statistics $T = \bar{X}$ is a sufficient statistic, we have to find the condition distribution of the sample observations given T. Therefore, we consider

$$f(x_1, x_2, ..., x_n \mid t) = \frac{f(x_1, x_2, ..., x_n, t)}{f(t)}$$

It means that we require $f(x_1, x_2, ..., x_n, t)$ and the distribution of the statistic $T = \bar{X}$. We first find $f(x_1, x_2, ..., x_n, t)$ as discussed in Example 3, that is,

$$f(x_1, x_2, ..., x_n, t) = f(x_1, x_2, ..., x_n, t)$$
$$= f(x_1, \theta) f(x_2, \theta) ... f(x_n, \theta)$$
$$= \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \frac{1}{\theta} e^{-\frac{x_2}{\theta}} ... \frac{1}{\theta} e^{-\frac{x_n}{\theta}}$$
$$= \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}$$

$$f(x_1, x_2, ..., x_n, t) = \frac{1}{\theta^n} e^{-\frac{n}{\theta}t} \quad \left(\because t = \frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

We now find the distribution of T. Since the sample comes from the exponential distribution, therefore, $X_1, X_2, \ldots, X_n$ follow the same exponential distribution with parameter θ. Also, we know that if $X_1, X_2, \ldots, X_n$ follow the exponential distribution then $\sum_{i=1}^{n} X_i$ will follow a gamma distribution with parameter $(n, 1/\theta)$ and the statistic $T = \bar{X}$ will follow the gamma distribution with parameters $(n, n/\theta)$. Therefore, the pdf of $T = \bar{X}$ is given as follows:

$$f(t) = \frac{\left(\frac{n}{\theta}\right)^n e^{-\frac{n}{\theta}t} t^{n-1}}{\Gamma n}; \quad t > 0$$

Therefore, we can find the condition distribution of the sample observations given T as

$$f(x_1, x_2, ..., x_n \mid t) = \frac{f(x_1, x_2, ..., x_n, t)}{f(t)} = \frac{\dfrac{1}{\theta^n} e^{-\frac{n}{\theta}t}}{\dfrac{\left(\dfrac{n}{\theta}\right)^n e^{-\frac{n}{\theta}t} t^{n-1}}{\overline{|n}}}$$

$$= \frac{\overline{|n}}{n^n t^{n-1}}$$

Since the conditional distribution of $X_1, X_2, \ldots, X_n$ given $T = \overline{X}$ does not depend on the parameter θ. Therefore, T is indeed sufficient for the parameter θ.

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

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

To find the sufficient statistic, first, we have to find the joint probability density function of the sample values of the exponential distribution. Let $X_1, X_2, \ldots, X_n$ be a random sample taken from the exponential distribution with parameter θ. We can obtain the joint probability density function of the exponential distribution as

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

$$= \theta e^{-\theta x_1}.\theta e^{-\theta x_2} \ldots \theta e^{-\theta x_n} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

$$f(x_1, x_2, ...x_n, \theta) = \theta^n e^{-\theta n \overline{x}} \quad \left(\because \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i\right)$$

Since we cannot separate any term of the joint probability density function which is independent to θ, therefore, we can factor the joint probability density function as

$$f(x_1, x_2, ...x_n, \theta) = \left(\theta^n e^{-\theta n \overline{x}}\right).1$$

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

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

3(ii) To find the sufficient statistic for the parameter σ^2, we use the factorization theorem. We know that the probability density function of the normal distribution with parameter μ and variance σ^2 as

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

In our case, μ is known as 95, therefore,

$$f(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-95)^2}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample taken from the above normal distribution. We can obtain the joint probability density function of the sample observations $X_1, X_2, \ldots, X_n$ as

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

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_1-95)^2} . \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_2-95)^2} \ldots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_n-95)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-95)^2}$$

We can factor the joint probability density function as

$$f\left(x_1, x_2, \ldots x_n, \sigma^2\right) = \left(\left(\frac{1}{\sqrt{\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-95)^2}\right)\left(\frac{1}{\sqrt{2\pi}}\right)^n$$

$$= g\left[t(x), \sigma^2\right].h\left(x_1, x_2, \ldots, x_n\right)$$

where $g\left[t(x), \sigma^2\right] = \left(\frac{1}{\sqrt{\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-95)^2}$ is a function of parameter σ^2 and sample values $x_1, x_2, \ldots, x_n$ only through $t(x) = \sum_{i=1}^{n}(x_i - 95)^2$, whereas $h\left(x_1, x_2, \ldots, x_n\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n$ is independent of σ^2. Hence, by the factorization theorem of sufficiency, $\sum_{i=1}^{n}(X_i - 95)^2$ is a sufficient estimator for σ^2 when μ is known as 95.

3(iii) The probability density function of uniform distribution U $[0,\theta]$ is given as follows:

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

To find the sufficient statistic, first, we have to find the joint probability density function of the sample values. Let $X_1, X_2, \ldots, X_n$ be a random sample taken from $U[0, \theta]$. Therefore, we can obtain the joint probability density function of the sample $X_1, X_2, \ldots, X_n$ as

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

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

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

Therefore, we can factor the joint probability density function as

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

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

where,

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

Therefore,

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

where $g\left[t(x), \theta\right] = \frac{1}{\theta^n} I\left(x_{(n)}, \theta\right)$ is a function of θ and sample values only through $t(x) = x_{(n)}$, whereas $h(x_1, x_2, \ldots, x_n) = 1$ is independent of θ.

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

Terminal Questions (TQs)

1. If the outcome of each coin is represented by random variable X then it follows the Bernoulli distribution with the probability of getting head p. Therefore, the probability mass function of X is given as follows:

$$P\left[X = x\right] = p^x (1-p)^{1-x}; \quad x = 0,1,\ 0 < p < 1$$

We can obtain the joint probability mass function of the three coins as

$$f(x_1, x_1, x_3, p) = P\left[X_1 = x_1\right]P\left[X_2 = x_2\right]P\left[X_3 = x_3\right]$$

We can obtain the joint probability mass function by putting X as x_1, x_2, x_3 in the probability mass function as mentioned above. Therefore,

$$f(x_1, x_1, x_3, p) = p^{x_1} (1-p)^{1-x_1} .p^{x_2} (1-p)^{1-x_2} .p^{x_3} (1-p)^{1-x_3}$$

Collecting like terms, we get

$$f(x_1, x_1, x_3, p) = p^{x_1+x_2+x_3} (1-p)^{1+1+1-(x_1+x_2+x_3)}$$

On simplifying, we get the required joint probability mass function as

$$f(x_1, x_2, x_3, p) = p^{\sum_{i=1}^{3} x_i} (1-p)^{3-\sum_{i=1}^{3} x_i}$$

We now try to write the joint mass function as the product of two functions, but we cannot separate any term of the joint probability mass function which is independent of p, therefore, we can factor the joint probability mass function as

$$f(x_1, x_2, x_3, p) = \left(p^{\sum_{i=1}^{3} x_i} (1-p)^{3-\sum_{i=1}^{3} x_i}\right).1$$

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