U.P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105
Advance Statistical Inference

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj
Prof. Anup Chaturvedi **Member**
Department of Statistics, University of Allahabad, Prayagraj
Prof. S. Lalitha **Member**
Department of Statistics, University of Allahabad, Prayagraj
Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur
Dr. Shruti **Member-Secretary**
School of Sciences, U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Prof. S. K. Pandey **Writer** (Block 1 & 4)
Department of Statistics, Lucknow University, Lucknow
Prof. V. K. Sehgal **Writer** (Block 1)
Department of Mathematical Sciences and Computer Applications, Bundelkhand University, Jhansi
Prof. Ram Ji Tiwari **Writer** (Block 1)
Department of Statistics, University of Jammu, Jammu, J&K
Dr. Shruti **Writer** (Block 1)
School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj
Prof. Umesh Singh **Writer** (Block 2)
Department of Statistics, Banaras Hindu University, Varanasi
Prof. A. H. Khan **Writer** (Block 3 & 4)
Department of Statistics, Aligarh Muslim University, Aligarh
Dr. Padmakar Singh **Reviewer** (Block 1)
Department of Statistics, U. P. Autonomous P. G. College, Varanasi
Prof. B. P. Singh **Reviewer** (Block 2)
Department of Statistics, Banaras Hindu University, Varanasi
Prof. Umesh Singh **Reviewer** (Block 3 & 4)
Department of Statistics, Banaras Hindu University, Varanasi
Prof. S. K. Pandey **Reviewer** (Block 3)
Department of Statistics, Lucknow University, Lucknow
Prof. V. P. Ojha **Reviewer** (Block 4)
Department of Statistics and Mathematics, D. D. U., Gorakhpur University, Gorakhpur
Prof. Umesh Singh **Editor** (Block 1)
Department of Statistics, Banaras Hindu University, Varanasi
Dr. Sanjay Singh **Editor** (Block 2)
Department of Statistics, Banaras Hindu University, Varanasi
Prof. K. K. Singh **Editor** (Block 3)
Department of Statistics, Banaras Hindu University, Varanasi
Prof. B. P. Singh **Editor** (Block 4)
Department of Statistics, Banaras Hindu University
Dr. Shruti **Course/ SLM Coordinator**
School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105 ADVANCE STATISTICAL INFERENCES

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)
Second Edition: ***July 2021***
ISBN : 978-93-94487-38-3

Printed By: K.C. Printing & Allied Works, Panchwati, Mathura - 281003 .

.

Blocks & Units Introduction

The present SLM on ***Advance Statistical Inference*** consists of fifteen units with four blocks.

The ***Block - 1 – Point Estimation,*** is the first block, which is divided into four units, and deals with theory of statistics in advance statistical inference.

Unit – 1- Introduction to Statistical Inference, is introductory and gives an idea about parameters, statistics and likelihood function of a random sample.

Unit – 2 – Point Estimation and Cramer Rao Inequality, describes point estimation along with desirable properties of a point estimator, Carmer Rao inequality and amount of information.

Unit – 3 – Sufficiency and Factorization Theorem, discuss property of sufficiency along with factorization theorem.

Unit – 4 – Complete Sufficient Statistics and Rao-Blackwell Theorem, Present the concept of complete sufficient statistics along with Rao Blackwell Theorem and its applications.

The ***Block - 2 – MVU Estimation*** is the second block with two units and deals with the problem of estimation particularly the procedures and concepts related to the minimum variance unbiased estimation.

Unit – 5 – MVU Estimators, provides the basic concepts related to minimum variance unbiased estimators.

Unit – 6 – Complete Sufficient Statistics, describes the basic concepts of complete sufficient statistics.

The ***Block - 3 – Testing of Hypothesis - I,*** deals with testing of hypothesis and consists of two units.

Unit – 7 – Preliminary Concepts in Testing, describes the concepts of critical regions, test function, two kinds of errors, size and power function of the test.

Unit – 8 – MP and UMP Tests discusses the concepts of most powerful and uniformly most powerful test is a class of size α tests with simple illustration.

The ***Block - 4 – Testing of Hypothesis - II*** based on testing of hypothesis and interval estimation consists of four units.

Unit – 9 – Neyman – Pearson Lemma, Likelihood Ratio Test and Their Uses, describes Neyman Pearson Lemma and likelihood ratio tests along with their uses in determination of test.

Unit – 10 – Testing of Means of Normal Population, discuss tests for significance of mean from a normal population and testing the equality of means from two independent normal populations

Unit – 11 – Interval Estimation, defines interval estimation for single unknown parameter of univariate population. Confidence intervals have been given for parameters of univariate normal population and one parameter exponential family.

Unit – 12 – Shortest and Shortest Unbiased Confidence Intervals, provides the concept of shortest and shortest unbiased confidence intervals.

At the end of every block/unit the summary, self assessment questions and further readings are given.

U.P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105
Advance Statistical Inference

Block: 1 Point Estimation

Course Preparation Committee

Prof. S. K. Pandey **Writer**
Department of Statistics, Lucknow University, Lucknow

Prof. V. K. Sehgal **Writer**
Department of Mathematical Sciences and Computer Applications, Bundelkhand University, Jhansi

Prof. Ram Ji Tiwari **Writer**
Department of Statistics, University of Jammu, Jammu, J&K

Dr. Padmakar Singh **Reviewer**
Department of Statistics, U. P. Autonomous P. G. College, Varanasi

Prof. Umesh Singh **Editor**
Department of Statistics, Banaras Hindu University, Varanasi

Dr. Shruti **Course/ SLM Coordinator**
School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105 ADVANCE STATISTICAL INFERENCES

First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)
Second Edition: ***July 2021***
ISBN : 978-93-94487-38-3

Printed By: K.C. Printing & Allied Works, Panchwati, Mathura - 281003.

Block & Units Introduction

The ***Block - 1 – Point Estimation,*** is the first block, which is divided into four units, and deals with theory of statistics in advance statistical inference.

Unit – 1- Introduction to Statistical Inference, is introductory and gives an idea about parameters, statistics and likelihood function of a random sample.

Unit – 2 – Point Estimation and Cramer Rao Inequality, describes point estimation along with desirable properties of a point estimator, Carmer Rao inequality and amount of information.

Unit – 3 – Sufficiency and Factorization Theorem, discuss property of sufficiency along with factorization theorem.

Unit – 4 – Complete Sufficient Statistics and Rao-Blackwell Theorem, Present the concept of complete sufficient statistics along with Rao Blackwell Theorem and its applications.

At the end of each unit the summary, self assessment questions and further readings are given.

Unit-1 Introduction to Statistical Inference

Structure

1.1 Introduction

In carrying out any statistical investigation, one starts by taking a suitable probability model for the phenomenon (X) that one seek to describe. According to the probability model, the distribution function (denoted by F) is supposed to be (unspecified) member of a more or less general class of distribution functions. Here one's goal may be the task of specifying F more completely than that is done by the model. The task is achieved by taking a random sample X_1, X_2,…..X_n from the parent population. These observations are the raw materials of the investigation and are used to make a guess about the distribution function F which is partly unknown. Thus statistical inference is the science of drawing the conclusions about the population on the basis of a random sample drawn from the parent population. So, we can term it as the calibration zone of statistics. Now onwards we will learn more about statistical inference.

1.2 Objectives

After reading this unit you should be able to :

- Define statistical inference
- Explain the parameters and statistics
- Understand parametric and non-parametric methods
- Know likelihood function of sample values
- Understand sampling distributions
- Discuss the problems of Statistical Inference

1.3 Parameter and Statistic

When we use sample observations to get an overview about population values it is called estimation e.g. We want to study average income of an industry workers in a metro city. For this first we will chalk out population of industry workers in that metro city. Then since the number of industry workers is large, we will find an appropriate sample of workers. Then a possible justified method of estimating average income of the workers is to obtain average income of the workers from the sample. This sample average income of the worker from the sample. This sample average may be an estimate of the population average. Let us define two more terms i.e. the parameter and the statistic.

Parameter: A parameter is defined as a constant of the population. In other words it is a measure which describe a population value i.e. a parameter provides information about population e.g. population mean, population variance etc.

Statistic: A statistic is defined as a function of sample observations. It is independent of unknown parameters. Sample mean, Sample median, i^{th} observation of a sample etc. are some examples of statistics. The purpose of estimation is to find that statistic which is a good representative of a parameter. This statistic is called an estimate of population parameter.

The Estimation, thus, is that branch of statistics where we learn about finding an estimate of population parameter through statistic.

Suppose the population under investigation is having the density function f(x; $\theta_1, \theta_2, \theta_3, \dots \dots . \theta_n$), where X is the variate and $\theta_1, \theta_2, \dots \dots . \theta_m$ are m parameters of the distribution. For example, in the case of normal distribution, the density function can be written as N (x; μ, σ^2). Suppose X_i (i = 1, 2,…., n) are n observations of a random sample. In estimation problems, we define estimators, for one or more of the parameters in terms of the sample values and these estimators, naturally will be function of the sample values.

1.4 Parametric and Non-Parametric Methods

In the development of Statistical methods the techniques of inference that were first to appear were those which involved may assumptions about the distribution of sample values X_1, X_2,…..X_n. In most of the cases, it is assumed that these are i.i.d. normal variables. In any case, it would be assumed that the join distribution has a particular parametric form like normal or exponential, only some or all of the parameters may be unknown. Statistical inference in these cases would relate solely to the value or values of some or all of the unknown parameters. This is called ***Parametric Inference***.

Comparatively, a large number of methods of inference have been developed in Statistics which do not make too many assumptions about the distribution of $X_1, X_2, \ldots X_n$. It may simply be assumed that these are i.i.d. random variables having a common continuous distribution but no parametric form of the common distribution may be assumed. Statistical inference under such a set up is called ***Non Parametric Inference*.**

1.5 Likelihood Function of Sample Values

Let $X_1, X_2, \ldots X_n$ be a random sample of size n taken from the population whose p.d.f. or p.m.f. is $f(x, \theta)$ θ is the parameter. θ may be single or vector valued. Then likelihood function of sample values denoted by L or $L(X_1, X_2, \ldots X_n, \theta)$ is defined as

$$L = L(X_1, X_2, \ldots X_n, \theta)$$

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

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

Actually, likelihood function of sample values gives the probability of getting a specific sample of size n from the population.

1.6 Some Illustrations of Likelihood Function

1. Let $X_1, X_2, \ldots X_n$ be a random sample from $N\ (\mu, \sigma^2)$. Then

$$f\ (x) = \frac{1}{\sigma\sqrt{2\pi}} exp - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right); -\infty < x < \infty$$

$$-\infty < \mu < \infty$$

$$-\infty < \sigma < \infty$$

and then the likelihood function of sample values is

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

2. If a random sample of size n has been taken from a Poisson population with p.m.f.

$$f\ (x) = \frac{e^{-\lambda^*}}{x!}, \qquad where\ 0 < \lambda < \infty, x = 0,1,2, \ldots\ldots, \infty$$

then likelihood function of the sample values is

$$L = \left(e^{-n\lambda} \lambda^{\Sigma x}\right) / \ x_1!\, x_2! \ldots x_{n!}$$

3. For a random sample of size one is drawn from a binomial population with parameters n and p, having p.m.f.

$$f(x) = {}^{n}C_{x}.\ p^{x}\ (1-p)^{\ n-x}$$

the likelihood function is

$$L = {}^{n}C_{x}.\ p^{x}\ (1-p)^{\ n-x}$$

It is important to note that the likelihood function in the case of binomial distribution for sample size one is same as its p.m.f.

4. For a random sample of size n from uniform population with p.d.f.

$$f\ (x_i, \theta) = \ 1/\theta; \theta < x < \theta$$

$$= 0 \ other\ wise$$

The likelihood function is

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

$$= \begin{cases} \left(\frac{1}{\theta}\right)^{n} ; 0 \leq x_1, x_{2,} \ldots \ldots x_n \ \leq \ \theta. \\ 0 \qquad\qquad other\ wise \end{cases}$$

1.7 Sampling Distribution

Statistical inference helps us to estimate the unknown parameter using statistics. We first obtain the statistic and on that basis we estimate the parameter. As we are aware a number of different samples can be obtained from the population. The values of the statistic computed from these different samples may not be equal. In statistical terms we can say that a statistic is a variable quantity whose values changes with each sample. Since each sample is obtained through some specified procedure and a probability of drawing each sample already exists, certain probability is also associated with each value of statistic. So we may say that a statistic is a random variable which takes on certain values with some probability law.

The probability distribution of a statistic is called its sampling distribution.

Thus the probability distribution of sample mean is called the sampling distribution of sample mean, and probability distribution of sample variances is called the sampling distribution of sample variance. In the same way we can have sampling distribution of sample proportion, sample median or of any other statistic we want to use.

Further it is also very important to note that the sampling distribution of a statistic is dependent on the population, the size of the sample and on the method by which the units are selected in the sample.

Some Examples:

1. If a sample of size n is taken form a normal distribution N (μ, σ^2) with known variance of the population then the sample mean μ is found to be normal distributed with mean and variance σ^2/n i.e.

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

2. If the sample is taken from a normal distribution with unknown variance then

$$\frac{(\bar{x}-\mu)\sqrt{n}}{s} \sim t_{n-1}$$

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

1.8 Standard Error of Statistic

The standard deviation of a statistic is called its standard error and the variance of this statistic is called its sampling variance. e.g. standard error of sample mean form a normal population with known variance is $\frac{\sigma}{\sqrt{n}}$

1.9 Problem of Statistical Inference

The problem of Statistical Inference can be divided into two parts

1. Estimation of parameters
2. Testing of hypothesis

1. Estimation of Parameters

On some occasions our interest will be in such feature as the central tendency or dispersion of the distribution of X_1, X_2,……X_n. In order to make conjecture about this feature we may use some statistic T, i.e. some measurable function of X_1, X_2, X_n. To be precise X_1, X_2, X_n if be the available set of observations then we put forward the corresponding value of T, say

$$T = T(X_1, X_2, X_n)$$

as the likely value of the parameter of the distribution. This t is then our estimate of the parameter an is also called the point estimate. The problem of inference is this case takes the form of point estimation i.e. estimation of the parameter by a single value.

In some cases one may give, instead of a single value as the likely estimate of the parameter as set of values this set being determined in terms of the observation, such that the actual value of the parameter may be considered likely to belong to that set. Estimation of the parameter is now achieved by means of a confidence set. Usually the set is taken to be an interval and then the statistical procedure is called interval estimation of the parameter of the distribution.

To summarize we can say that when a single number is used to estimate an unknown parameter, this is called point estimate and this method is termed as **Point Estimation.**

But sometimes we find that a point estimate is not sufficient as it may be either correct or incorrect. Thus we are not sure about its reliability. Also a point estimate is of no use if it is not accompanied by an estimate of the error that might be involved. Then we estimate the parameter by method of interval estimation where instead of a point value an interval is provided i.e. parameter is generally estimated to be within a range of the values rather than as a single number. So when an interval of values is used to estimate a population parameter it is called interval estimation and this estimate is called the interval estimate.

Thus if we say that the average height of men whose ages are between 25 to 30 years is 168 cm. on the basis of sample then it is a point estimate and when we say that this height is expected to lie between 165 cm. to 171. It is called an interval estimate.

2. Testing of Hypothesis

In some situations we start with tentative notion about the feature of the distribution that we are interested in. This idea may be suggested to us by some authority (e.g. a manufacture placing a new product in the market or a leading scientist propounding some new scientific theory) or by the results of the previous investigations conducted in the same field or in a similar field. We may then like to know tenable or valid the idea is in the light of the observations (X_1, X_2,……X_n). The inference problem is now one of testing a hypothesis about the unknown feature of the distribution. Note that the model used and the hypothesis being tested are both assumptions regarding the probability distribution of X_1, X_2,……X_n. However, the hypothesis is an assumption the validity of which is questioned, but is taken for granted.

In much simpler words any assumption that we make about a population parameter is called a hypothesis and the statistical procedures that are used to test the hypothesis of the basis of sampled observation are covered under the topic testing of hypothesis.

For example a doctor may set up a hypothesis that smoking increase the risk of throat cancer in human beings. To ascertain this he will collect some primary or secondary data and then after some statistical analysis he might approve or disapprove it. This is the problem of testing of hypothesis. Additionally the assumptions that we wish to test is called a null hypothesis and the assumptions that we accept in case the null hypothesis is rejected is called alternative hypothesis.

1.10 Key Words

1. Parametric Inference.
2. Non Parametric Inference.
3. Likelihood Function
4. Sampling Distribution
5. Standard Error
6. Estimation Of Parameters
7. Testing Of Hypothesis
8. Point Estimation
9. Interval Estimation
10. Hypothesis Testing

1.11 Exercises

1. What is statistical inference? Explain
2. Write an essay on the theory of estimation.
3. What are two types of problem of inference
4. Write a short note on likelihood function. Obtain the likelihood function of the sample if it belongs to a Bernoulli density function.

1.12 Summary

To summarize we can say that statistical inference is the process of arriving at conclusion about the population under study on the basis of data obtained from a sample. Probability theory forms the basis of statistical inference as it uses various probability methods for decision making.

The quantity or measure which describe a population value is called a **parameter**.

The quantity or measure which describe a sample value is called a **statistic.**

If X_1, X_2,……X_n are independently and identically distributed random variables we say that they constitute a **random sample** from the infinite population given by their common distribution.

The probability distribution of a statistic is called its sampling distribution.

The standard deviation of the distribution of a statistic is known as **standard error** of the statistic.

The **likelihood function** of sample values gives the probability of getting a specific sample of size n from the population.

The problem of statistical inference can be divided into two parts

1. Estimation of parameters
2. Testing of hypothesis.

When a single number is used to estimate an unknown parameter, this estimate is called **point estimate** and this method is termed as **point estimation**.

When an interval of values is used to estimate a population parameter it is called **interval estimation** and this estimate is called the **interval estimate**.

Any assumption that we make about a population parameter is called a **hypothesis** and the statistical procedure that is used to test the hypothesis on the basis of sampled observation is called **testing of hypothesis**.

1.13 Further Readings

- Cramer, H. *Mathematical Methods of Statistics*, Princeton Univ. Press 1946.
- Kendall, M.G. *A Course in Multivariate Analysis*, Charles Griffin, 1957.
- Lehmann, E.L. (1986). *Testing statistical hypothesis*. John Wiley, 1959
- Goon, A.M., Gupta, M.K., and Dasgupta, B. (2000). *An outline of statistical Theory*, world Press

Unit-2 Point Estimation and Cramer Rao Inequality

Structure

2.1 Introduction

Every one of us make estimates in our lives. For example while going away for vacation we estimate the possible expenditure. Similarly a student estimates the time he requires for doing revisions before examination. A sportsperson judges himself on the basis of practice sessions and so on. Business organizations, shopkeepers, institutions, governing bodies all estimate one thing or another with the hope that the estimates bear a reasonable resemblance to the outcome. The question here is what estimation is in statistical term? A one line answer to this query is that the estimation is that statistical method of obtaining the value of the parameter from a possible set of alternatives. In the ongoing text we will take a deeper look in topic.

2.2 Objectives

After reading this unit you will be able to understand -

- The concept of Point Estimation
- Learn about the properties of a good estimator which are unbiasedness, consistency and efficiency
- Understand Minimum Variance Unbiased Estimate,
- Understand Cramer-Rao inequality

- Learn about amount of information

2.3 Point Estimation

If we use the value of a statistic to estimate a population parameter, this value is a point estimate of the parameter. For example, in the case of binomial (n,p) population, if we use simple proportion to estimate the parameter θ, this estimate is called point estimate because it is single number, or point on the real axis. The statistic, whose value is used as the point estimate of a parameter, is called an estimator. Therefore, the statistic $\bar{X}$ is an estimator of μ and its value x is the point estimate of μ.

Since estimator are random variables we need to study their sampling distributions. For instance, when we estimate the variance of a population on the basis of a random sample, we can hardly expected the value of S^2 which one gets form the sample, to be actually equal to σ^2but it will be certainly reassuring if the value is close to σ^2. Also, when there are more than one statistics available to estimate the parameter of a population, (for example the mean and the median of the sample to estimate the population mean in N (μ, σ^2), it is important to know, among other things whether the sample mean or sample median is more likely to yield a value which is actually close to parameter.

Theory of Point Estimation

Let $x_1, x_2, \ldots\ldots x_n$ be a random of size n drawn from the population whose p.d.f. is f(x,θ), θ is the parameter of the population. We denote the sample observations $x_1, x_2, \ldots\ldots x_n$ by x, i.e.

$$x- (x_1, x_2, \ldots\ldots x_n)$$

Suppose we are interested to determine (or estimate) the true value of θ. It may be assued known that it lies in a certain set Ω, known as the parametric space (or parameter space).

For the purpose of estimation, we make use of some statistic T, a measurable function of sample values. The value of T at x is assumed to be t= T(x). One may purpose to estimate θ by this value t, knonw as estimate of θ corresponding to the given random sample x.

Since random sample x will differ one case to another, thus leading to different estimates in different one can't expect that the estimate in each case will be good in the sense of having only small deviation from true value of θ. Hence to judge the desirability (or otherwise) of any estimation procedure, one should really judge the properties of the estimation T. Obviously T may be regarded as a good estimator if it gives in general values of T that deviate from θ from θ only by a small amount, that is if the probability distribution to T has a high degree of concentration around true value of θ in Ω. The value of T for a specific x is also known as point estimate of θ. The problem of inference is in this case is known is 'Point Estimation' i.e. estimation by a point or a single value (on the basis of x drawn from the parent population).

Now the question is how to know about the estimator for the estimation of θ? The answer is provided in the form of describing different methods of estimation. Some of the various available are

(1) Method of Moments

(2) Methods of Maximum Likelihood

(3) Method of Minimum Variance

(4) Method of Least Squares

(5) Method of Minimum Chi-Square

These methods give different estimators for the block estimation of the same parameter. (These methods have been discussed in other units in detail). Now the question arises which estimator one should choose from and why?

The answer has been given by describing various desirable properties of a good estimator.

2.4 Properties of a Good Estimator

A very important decision, which an experimenter has to take is to decide which estimator one should choose among a number of possible estimators. Various statistical properties of the estimators like unbiasedness, minimum variance, consistency, efficiency and sufficiency, can be used to decide which estimator is most appropriate to a given situation. Following are termed as the desirable properties of a good estimator-

(i) Unbiasedness

(ii) Consistency

(iii) Efficiency

(iv) Sufficiency

2.5 Unbiasedness

Let $x_1, x_2, \ldots\ldots x_n$ be a random sample of size n taken from the population where p.d.f. or p.m.f. is $f(x, \theta)$, θ is the unknown parameter and $T = T(x_1, x_2, \ldots\ldots x_n)$ be an estimator of θ.

Then T is said to be an unbiased estimator of θ if

$$E(T) = \theta$$

If $E(T) \neq \theta$, T is known as a biased estimator of θ and bias in T is define as bias (T) = bias in T = E (T) – θ

If E (T) > θ, T is called positively biased estimator of θ and if E(T) < θ, t is called negatively biased estimator of θ

Some times it is noted that

$$E(T) \rightarrow \theta \text{ as } \rightarrow \infty$$

In this case, T is known as asymptotically unbiased estimator of θ. A very important are not unique. That is there may exist more than one unbiased estimator for a parameter. it is also to be noted that biased estimator does not always exists.

Some worked example

Example 1: If X has a binomial distribution then x/n, the observed proportion of success, in an unbiased estimator of the parameter θ.

Proof: Since E(x) = n θ, it follows that

$$E\left(\frac{x}{n}\right) = \frac{1}{n}, E(x) = \frac{1}{n}, n\theta = \theta.$$

Hence, x/n is an unbiased estimate of θ

Example 2: If $s^2 = \frac{1}{n-1}\sum_{n-1}^{n}(x_i - \bar{x})^2$ is the variance of a random sample from, N (μ, σ^2), μ, σ^2 both unknown then the s^2 in an unbiased estimator of σ^2.

Proof:

$$E\delta(s)^2 = \frac{1}{n-1} E\left\{\sum_{n-1}^{n}(x_i - \bar{x})^2\right\}$$

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

$$= \frac{\sigma^2}{n-1} . E\ (Y)$$

Where

$$Y = \sum_{i=1}^{n}\left(\frac{x-\bar{x}}{\sigma}\right)^2$$

Then Y follows a χ^2 distribution with (n-1) degrees of freedom as sample observations have been drawn N (μ, σ^2). Then obviously E(Y) = (n - 1).

Hence

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

$$= \sigma^2$$

2.6 Consistency

The statistics (T) is said to be a consistent estimator θ, if T converges to θ, in probability i.e.,

$$\Pr(|T-\theta| \leq \epsilon) \to 1$$

$$or\ Pr(|T-\theta| \leq \epsilon) \to 0$$

Consistency is an asymptotic property, namely a limiting property of an estimator i.e. when n is sufficiently large we can be certain that the error made with a consistent estimator will be less than any small preassigned constant.

A Sufficient Condition for Consistency:

T is a consistent estimator of θ if

(i) E(T) $\to \theta$

(ii) Var (T) $\to 0$.

Proof:

For a r.v. X having finite mean and variance, we have from Chebychev's Inequality.

$$\Pr[|x-E(x)| \leq \epsilon] \geq \left[1-\frac{var\ (x)}{\epsilon^2}\right]$$

Applying it to 'T' we get

$$\Pr[|T-E(T)| \leq \epsilon] \geq \left[1-\frac{var\ (T)}{\epsilon^2}\right]$$

Making $n \to \infty$ and applying (i) & (ii), we may write

$$\Pr(|T-\theta| \leq \epsilon) \geq 1 \; for \; n \to \infty$$

But probability can never exceed unity, therefore we write

$$\Pr(|T-\theta| \leq \epsilon) \to 1 \; for \; n \to \infty$$

Showing T to be a consistent estimator of θ under (i) & (ii).

Proved.

Mean Square Error (or M.S.E.)

Before defining the concept of efficiency let us define the concept of mean square error of an estimator (or statistic).

The mean square error of an estimator T is defined as

M.S.E. (T) $= E(T-\theta)^2$

$= E\,[T-E\,(T)+E(T)-\theta]^2$

$= E[T-E\,(T)]2 + [E(T)-\theta]^2$

(Cross term vanishes)

$= Var\,(T) + (\text{bias in } T)^2$

If T is an unbiased estimator of θ then bias in T is zero and in this case,

M.S.E. (T) = Var (T) or i.e. if E(T) = Q

Thus for an unbiased estimator T of θ its mean square error coincides with its variance.

2.7 Efficiency

Among the class of all possible estimators for estimating θ, one which has the minimum m.s.e. is called most efficient estimator of θ.

If T_1 and T_2 are two estimator for estimating θ, the T_1 is said to be more efficient then T_2 for estimation of θ if

M.S.E. $(T_1) <$ M.S.E. (T_2)

The efficiency of T_1 w.r.t., T_2, denoted by E (or e), is defined as

$$E\ (or\ e) = \frac{m.s.e.(T_2)}{m.s.e.(T_1)}$$

$$= \frac{m.s.e.(T_2)}{m.s.e.(T_1)} \times 100\%$$

However, if we are given the class of unbiased estimators for estimating θ, we may replace m.s.e. by variance for the concept of efficiency. It has already been stated that in case of unbiasedness m.s.e. coincides with the variance.

We have already indicated that when there are two unbiased estimators for a parameter, the estimator with less variance is more reliable. If T_1 and T_2 are two unbiased estimators of parameter θ and the variance of T_1 is less than the variance T_2 then T_1 is said to be relatively more efficient. The most efficient estimator, among a class of consistent estimator, is one whose sampling variance is less than that of any other estimator. Whenever such an estimator exists, it provides a criterion for measurement of efficiency of the other estimators.

If T_1 is the most efficient estimator with variance σ_1^2 and T_2 is any others estimate with variance σ_2^2, then the efficiency E of T_2 is defined as

$$E = \frac{\sigma 1^2}{\sigma 2^2} \qquad \text{(This is always < 1)}$$

For example, the efficiency of the sample median of normal population can be determined in relation to the most efficient estimator, ˜x (,mean of the sample). The efficiency of the median of the sample is (for large n)

$$E = \frac{Var\ X}{Var\ (median)} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}}$$

$$= \frac{2}{\pi} = 0.637$$

The minimize of M.S.E. for all $\theta \varepsilon \Omega$ it self is found to be a difficult task. One may resolve this problem if insistence is given to unbiased that is if confines to the class of unbiased estimators Minimization of M.S.E. will amount to the minimization of the variance.

Criterion of unbiasedness has no great merit. It only enables to find the processor of choosing estimators within a mathematically tractable framework.

The criterion of unbiasedness may be deemed defective in cases where biased estimators have smaller M.S.E. than unbiased ones. The question, then, arises why to learn them out of consideration? In some investigation it becomes necessary to pool the evidence collected from several sources. The evidence may be in nature of an estimate, perhaps with a standard error attached to it. if the estimates are unbiased then a combined estimate may be formed with reduced standard error and with the accumulation of more evidence the true value may be approached. On the hand, if biased estimates are combined without any indication regarding the magnitude of the bias, then nothing definite can be said about such combined estimates. The bias may actually exceed the standard error at some stage and combined estimate may never approach the true value.

2.8 Minimum Variance Unbiased Estimator (MVUE)

T is known as best estimator of θ if it is unbiased for θ and has the minimum variance among the class of all possible unbiased estimators for estimating θ. In this case, T is also known as (uniformly). Minimum Variance Unbiased Estimator of θ In other words the statistic T is known as UMVUE of θ if it is unbiased and has smallest variance (for each θ) among all possible unbiased estimator of θi.e. if

(i) E(T) $= \theta$ for every $\theta \varepsilon \Omega$

(ii) Var (T) < Var (T\`) for every $\theta \varepsilon \Omega$

Where T\` is any other estimator of θ satisfying (i).

We know that

$$\text{M.S.E. (T)} = (\text{Bias in T})^2 + \text{Var (T)}$$

One can observe that MVU estimator makes the first contrast (i.e. bias in T) of MSE a minimum (i.e. zero) and then also make the second contrast i.e. Var (T), a minimum for all θ.This, of course does not mean that T will have the minimum mean square for all θ. However it is evident that by minimizing the two contrasts separately, T will on the whole (i.e. through out the parametric space) keep the MSE at a low level.

MVUE and CR Inequality

While an estimator may be directly examined for unbiasedness it is not immediately apparent how to satisfy one self that an estimator has the smallest variance among the class of all possible unbiased estimators.

Some methods in literature are available to solve this problem. One method is based on the use of Cramer-Rao (or Rao-Camer or CR) inequality.

2.9 CR Inequality

Let θ be a single parameter varying over the parametric space Ω and that $x_1, x_2, \ldots x_n$ be a random sample of size n taken from a continuous population having p.d.f. (x, θ). The likelihood function of sample values is given by

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

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

For sake of notational simplicity the multiple integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \ldots \int_{-\infty}^{\infty} f\ (x_1, x_2, \ldots x_n; \theta) dx_1, dx_2, \ldots dx_n$$

Will be denoted by $\int_{\underline{x}} \ L d\underline{x}$

Let us make the following assumptions known as regularity conditions of CR inequality:

(i) Ω is a non-degenerate open interval on the real line,

(ii) For almost all $\underline{x} = (x_1, x_2, \ldots x_n)$ and all $\theta \varepsilon \Omega, \left(\frac{\partial L}{\partial \theta}\right)$ exists, the exceptional set if any being independent of θ

(iii) The differentiation is possible at least one under the sign of integral that is $\frac{\partial}{\partial \theta} \int_{\underline{x}} \ L d\underline{x} = \int_{\underline{x}} \ \frac{\partial L}{\partial \theta} d\underline{x}$

(iv) T be an unbiased estimator of $\psi(\theta)$

(v) $\frac{\partial}{\partial \theta} \int_{\underline{x}} \ T.L d\underline{x} = \int_{\underline{x}} \ T.\frac{\partial L}{\partial \theta} d\underline{x}$

(vi) $E\left(\partial \ log \frac{L}{\partial \theta}\right)^2$ exists and is positive for each $\theta \varepsilon \Omega$

Under these assumptions

$$Var\ (T) = \sigma^2 \geq \left\{\frac{\left(\Psi'(\theta)\right)^2}{E(\partial \log L/\partial \theta)^2}\right\}$$

Where $\Psi'^{(\theta)} = \frac{\partial \Psi}{\partial \theta}$ which is finite and exists.

We may denote

$$E\left(\frac{(\partial \ logL)}{\partial\theta}\right)^2 \ by \ I(\theta)$$

Which is called by Fisher the amount of information about θ supplied by the sample and is reciprocal of the information limit to the Variance of T.

Proof:

We have $\int_{\bar{x}} \ Ld\underline{x} = 1$

Differentiating it w.r.t. 'θ' and using assumption (iii), we have

$$\int_{\underline{x}} \frac{\partial L}{\partial\theta} dx = 0$$

$$or \int \frac{1}{L}\frac{\partial L}{\partial\theta} dx = 0 \ or \int \frac{\partial \log L}{\partial\theta} Ld\underline{x} = 0$$

$$or \ \ E\left(\frac{(\partial \ logL)}{\partial\theta}\right) = 0 \ or \ E(Q) = 0 \qquad (2.1)$$

Where $Q = \left(\frac{\partial \ logL}{\partial\theta}\right)$

Again E(T) = $\Psi(\theta)$

$$or \int_{\underline{x}} TLD\underline{x} = \Psi(\theta)$$

Differentiating both side partially w.r.t. θ and a[pplying (v)

$$\int_{\underline{x}} T\frac{\partial L}{\partial\theta} d\underline{x} = \Psi^{'(\theta)} = \ \frac{\partial\Psi(\theta)}{\partial\theta}$$

Or

$$\int_{\underline{x}} T\left(\frac{1}{L}\frac{\partial L}{\partial\theta}\right) Ld\underline{x} = \Psi^{'(\theta)}$$

Or E(TQ) = $\Psi'(\theta)$ ------------------------- (2.2)

$$Var\ (Q^2) = E(Q^2) - [E(Q)]^2$$

$$= E(Q^2) as\ E(Q) = 0$$

$$= E\left(\frac{\partial\ logL}{\partial\theta}\right)^2 \qquad (2.3)$$

Also, Cov (TQ) = E(TQ) – E(T). E(Q)

= E(TQ) as E(Q) = 0

= $\Psi'(\theta)$ [using 2.8.1] (2.4)

We may write Cov (TQ) = p_{TQ}. $\left(\sqrt{Var(T)Var(Q)}\right)$

[Where QTQ is the correlation coefficient between T & Q].

So {Cov (TQ)}2 < Var (Q). Var (Q)

Or $$Var\ (T) \geq \frac{\{Cov\ (TQ)\}^2}{Var\ (Q)}$$

$$\geq \frac{\{\Psi'\ (\theta)\}^2}{E\left(\frac{\partial\ logL}{\partial\theta}\right)^2} \qquad [using\ (2.3) and\ (2.4)]$$

Proved.

The CR inequality remains valid even when r.v. x_1, x_2, ,….x_n (a random sample of size n drawn from the parent population) are all discrete. The proof remains the same. Only multiple integrals are replaced by appropriate multiple signs.

An unbiased estimator T of θ, which attains the lower bound of Cramer Rao inequality, is known as Minimum Variance Bound estimator (MVB estimator). One should keep in mind that MVBE and UMVUE may be different at times. The unbiased estimator which attains the lower bound of CR inequality is necessarily UMVUE.

Sometimes there may exists a class of unbiased estimators whose minimum variance may be more than the lower bound of CR inequality. Thus, though the variance of this estimator may not attain the lower bound of CR inequality it may not attain the lower bound of CR inequality, it may or may not be UMVUE.

It may also be noted that in case the regularity conditions underlying CR inequality do not hold, the least variance may be less than CR lower bound.

2.10 Remarks

Generally, two sets of criteria of a good point estimator viz (1) unbiased ness and minimum variance and (2) consistency and efficiency are considered. The criteria of having minimum variance and (asymptotic) efficiency are similar and in a way are necessary accompaniments of the basic criteria of unbiased ness and consistency respectively.

The criterion of unbiased is better in the sense that it is application irrespective of the number of random variable under consideration. The criterion of consistency and efficiency (particularly in case of asymptotic efficiency) relates to the asymptotic behavior of the statistic. In other words, a consistent estimator may be expect dot give a close estimate in case sample size is sufficient large but may leave completely in dark regarding its performance when sample size is small.

However, consistency may be a better criterion than unbiased ness in the sense that the central tendency of the distribution of the estimator may be towards θ or its parametric function as the case may be, for large n, without confirming to any particular measure of central tendency. Unbiased ness on the other hand only ensures that the mean of the estimator will be θ. Without bothering about the appropriateness of the mean as a measure of central tendency in the particular situation some times in a given situation, the mean of an estimator may not even exits. Even if it does, the criterion of unbiased ness may lead to undersirable estimators. Neither unbiased ness nor consistency leads to unique estimators but the scope of arbitrariness is much greater in the case of consistency than unbiased ness. Thus suppose T_n is a consistent estimator of θ. Then we may think of infinitely may others eg. $T_n + 1/\theta_{(n)}$ or $T_n (1+A/\theta_{(n)})$ where A is a constant independent of n and $\theta_{(n)}$ is an increasing function of n, are also consistent estimators of θ. This sort of arbitrariness does not arise in case of unbiased ness.

There is one point that consistency in its favour. Common sense requires that if T is considered a good estimator of, than and $\psi(\theta)$ be a function of θ then $\psi(T)$ should be deemed an equally good estimator of $\psi(\theta)$. From this point of view, unbiased ness may not be considered as a good criteria because ψ (T) will not be unbiased for ψ (θ) unless it is a linear function, even if T is unbiased for θ. The criterion of consistency may be supposed to meet this requirement because in a large class of problems consistent estimators have this desirable property of invariance.

2.11 Worked Examples

Example 1: Show that in sampling from a normal population with mean μ and variance σ^2 the sample mean is consistent estimator of μ.

Solution: In sampling from a normal population the sample mean x is also normally distributed with mean μ and variance $\frac{\sigma^2}{n}$.

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

$$As \qquad n \to \infty, E(\bar{x}) = \mu \text{ and } V(\bar{x}) = \frac{\sigma^2}{n}$$

x, thus, confirms to the conditions for consistency of the estimator i.e. sample mean $\bar{x}$ is a consistent estimator for population mean μ.

Example 2: If x_1, x_2 and x_3 form a random sample a normal population with mean μ and the variance σ^2, what is the efficiency of the estimator

$$t = \frac{x_1 + 2x_2 + x_3}{3} \text{ relative to } \bar{x}\,?$$

Solution: Here we have

$$\bar{x} = \frac{x_1 + 2x_2 + x_3}{3}$$

$$Since\ Var\ (x_i) = \sigma^2, i = 1,2,3\ var\ (\bar{x}) = \frac{1}{9}[var\ (x_1) + var\ (x_2) + var\ (x_3)]$$

$$var\ \bar{x} = \frac{\sigma^2}{3}(Variance\ of\ the\ sampling\ distribution\ of\ Means).$$

$$Var\ (t) = Var\frac{x_1 + 2x_2 + x_3}{3} = \frac{1}{16}[var\ (x_1) + var\ (x_2) + var\ (x_3)]$$

$$= \frac{\sigma^2}{16} + \frac{4}{16}\sigma^2 + \frac{\sigma^2}{16} = \frac{6\sigma^2}{16}$$

$$Efficiency\ of\ t\ relative\ to\ \bar{x} = \frac{Var\ (t)}{Var\ (\bar{x})} = \frac{\frac{\sigma^2}{3}}{\frac{6\sigma^2}{16}} = \frac{8}{9}$$

Example 3: If x_1 is the mean of random sample of size n form a normal population with the mean μ and the variance σ_1^2 and x_2 is the mean of a random sample of size n from a normal population with the mean μ and the variance σ_2^2 show that.

(a) $wx_1 + (1-\overline{w})\, x_2$ value $0 \leq w \leq 1$ is an unbiased estimator of μ.

(b) The variance of this estimator is minimum when $w = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$

Solution

(a) Let T = wx_1 + (1-w) x_2

E(T) = E (x_1) + 1-w) E (x_2)

=w μ +(1-w) $\mu = \mu$.

Hence T is an unbiased estimator of μ.

(b) $Var\ (T) = \ w^2 var\ (x_1) + (1-w)^2\ Var\ (x_2)$

$$= \ w^2 \frac{\sigma_1^2}{n} + (1-w)^2 \frac{\sigma_2^2}{n}$$

$$if\ var\ (T) is\ minimum, then \frac{d}{dw}\ \left(Var(T)\right) = \ 0$$

$$and\ \frac{d^2}{dw^2}\{Var(T)\} must\ be + ve.$$

$$\frac{d}{dw}\left(Var(T)\right) = \ 0\ gives$$

$$2w\frac{\sigma_1^2}{n} - 2(1-w)\frac{\sigma_2^2}{n} = 0\ \ i.e. w\ (\sigma_1^2 + \sigma_2^2) + \sigma_2^2$$

$$i.e. w = \ \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

For this value of w = $\frac{d^2\{Var\ \ (T)\}}{dw^2}$ is positive . Hence Var(T) is minimum when

$$w = \ \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Example 4: x_1, x_2 and x_3 is a random sample of size 3 from a population with mean μ and variance σ^2. T_1, T_2, and T_3 are the estimators used to estimate mean value μ where

$$T_1 = x_1 + x_2 - x_3, T_2 = 2x_1 + 3x_3 - 4x_2$$

$$and\ T_3 = \frac{\lambda x_1 + x_2 + x_3}{3}$$

I. Are T_1 and T_2 unbiased estimation?

II. Find value of λ such that T_3 is unbiased estimator of μ.

III. Which is the best estimator?

Solution Since x_1, x_2, x_3 is a random sample from a population with mean μ and variance σ^2

E (x_1) = μ, Var (x_1) = $\sigma^2$and Cov (x_1, x_2) = 0 i≠j= 1,2,….n.

I. E (T_1) = E (T_1)+ E (T_2)- E (T_3)

$= \mu + \mu - \mu = \mu$

i.e. T_1 is an unbiased estimator of μ.

E (T_2) = E ($2x_1$)+ E ($3X_3$)- E(X_4)

$= 2\mu + 3\mu - 4\mu = \mu$

i.e. T_2 is an unbiased estimator of μ.

II. i.e. T_3 is an unbiased estimator ⇒ E(T_3) = μ

1/3 { λE(X_1)+ E(X_2)+ E(X_3}= μ

i.e. 1/3 (λ+2) $\mu = \mu$

i.e. λ = 1

III. We have

Var (T_1) = Var (X_1)+ Var (X_2)+ Var (X_3) = $3\sigma^2$

Var (T_2) = 4Var (X_1)+ 9Var (X_3)+ 16Var (X_2) = $29\sigma^2$

$$Var\ (T_3) = \frac{1}{9}\{Var\ (X_1) + Var\ (X_2) + \ Var(X_3)\}$$

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

Since Var (T_3) is minimum, T_3 is the best estimator.

Example 5: If $x = \frac{1}{2}(x_1 + x_2)$ where x_1 and x_2 are most efficient estimators with variance S^2 then show that Var (x) = $\frac{1+p}{2} = \ S^2$ where Q is the correlation coefficient between X_1 and X_2.

Solution: Since both x_1 and x_2 are most efficient estimators

$$V(x) = \ V\left\{\frac{1}{2}(x_1 + x_2)\right\} = \frac{1}{4}V\ (x_1 + x_2)$$

$$= \frac{V(x_1) + V(x_2) + \ 2Cov\ (x_1, x_2)}{4}$$

$$= \frac{S^2 + S^2 + 2pS^2}{4} = \frac{2S^2 + 2S^2 + p}{4}$$

$$= (1 + p)\frac{S^2}{2}$$

2.12 Exercises

1. Show that if T is an unbiased estimator of θ the T^2 is not necessarily an unbiased estimator of θ^2. But if T is a consistent estimator of θ then T^2 is consistent estimator of θ^2 .

2. Find an unbiased estimator of (i) θ and (ii) θ^2 in case of a binominal distribution with parameters n and θ having p.m.f.

$$F(x, \theta) = {}^nC_x\, \theta^x\, (1- \theta)^{n-x};\ x= 0,1,2,\ldots,n,\ \ 0< \theta <1.$$

based on a random sample of size n.

3. Prove that the sample median is a consistent estimator for the mean of a normal population. Also, show that for a normal population, sample mean is efficient than the sample median.

4. A random sample $(x_1, x_2, x_3, x_4, x_5)$ of size 5 is drawn from a normal population with unknown mean μ. Considering the following estimators to estimate μ.

(i) $$T_1 = \frac{(x_1 + x_2 + x_3 + x_4 + x_5)}{5}$$

(ii) $$T_2 = \frac{(x_1 + x_2)}{5} + x_3$$

(iii) $$T_3 = \frac{(2x_1 + x_2 + \lambda x_3)}{3}$$

Where λ is such that T_3 is an unbiased estimator of μ. Find λ .Are T_1 and T_2 unbiased for μ. ? State giving reasons the estimator,which is best among T_1, T_2, and T_3.

2.13 Key Words

(a) Unbiased: An estimator "T" is unbiased for a parameter if E(T) = θ.

(b) Consistent: "T_n" is consistent estimator of θ if it converges to θ with probability 1 i.e. $n \to \infty$ $P[T_n \to \theta] = 1$.

(c) Efficient: In a class of consistent and unbiased estimators, one with the smallest variance is known as most efficient estimator.

2.14 Summary

In this unit we study about the theory of point estimation. An estimation is a statistic that is used to estimate a population parameter, while an estimate is a specific observed value of the estimator. A single number that is used to estimate an unknown parameter is called a point estimate. A good estimator is one that is (a) unbiased (b) consistent (c) efficient and (d) sufficient.

The C-R inequality provides the lower bound for the variance of an unbiased estimator of ψ(θ) and states that

$$V(t) = \frac{(\psi(\theta))^2}{E\left(\frac{\partial logL}{\partial\theta}\right)^2}$$

The denominator of this inequality is called the information on θ, supplied by the sample. This nomenclature is due to R.A. Fisher.

2.15 Further Readings

- Modd, A.M. Graybill, F.A.m Boes, D.C. (1974). *Introduction to the Theory of Statistics,* McGraw Hill international edition.
- Rao,C.R., Linear statistical inference and its applications, John Wiley and Sons, Inc.
- Wilks, mathematical statics, Jon Wiley and Sons.
- Kendall, Vol. 1,2,3. Hafner Publishing Company, New York.

Unit-3 Sufficiency and Factorization Theorem

Structure

3.1 Introduction

In the previous unit, we read about the properties of a good estimator. Sufficiency is another desirable property of an estimator. An estimator is sufficient if it makes so much use of the information in the sample that no other estimator could extract additional information from the sample about the population parameter being estimated.

According to R.A. Fisher, " A sufficient statistic summarizes the whole of the relevant information supplied by the sample".

Now we will study the concept of sufficiency in detail.

3.2 Objectives

After going through this unit you will be able to -

- Know what is sufficiency
- Understand the Neyman-Fisher factorization criterion of sufficiency
- Understand and Koopman's form of the distribution
- Have a concept of invariance property of sufficient statistics.

3.3 Sufficiency

The only information that guides the investigator in making a decision is supplied in the form of a random sample of size n drawn from the parent population. In most of the cases, it would be too numerous and too complicated set of observation to be directly dealt with. Therefore, a simplification or reduction to be desirable. Naturally one should use for such reduction of data, some statistics that loose as little of the information contained in the sample that is relevant to parameter θ.

It is this objective that leads to the concept of sufficient statistics. The principle of sufficiency is a principle for reducing or condensing the original random sample to a few statistics which may than be used for the purpose of drawing inference about the parent population characterized by θ. Loosely speaking, sufficiency amounts to replacing the sample observations x_1, x_2, …..x_n by few statistics T_1, T_2,……..T_k and thus discarding information, which is not relevant to and retaining every thing that is essential.

T is said to be a sufficient statistic of θ if conditional distribution of sample values given (T=t) is independent of θ. This definition is not very satisfactory because conditional distribution may not always be defined.

However where the random variables have purely discrete or purely continuous distribution, this definition is alright. Since these two are the cases, which we are concerned with at this level the above definition may be taken as adequate for our purpose.

A sufficient statistics T is said to be minimal sufficient if it is a function of every other sufficient statistic.

Note: The term ‘function’ is used here in a wide sense to include vector valued functions.

Example: Let (x_1, x_2, …..x_n) be a random sample from Bernoulli population with parameter ‘p’, $0 < p < 1$. i.e.

$$x_i = \begin{cases} 1 \; with \; probabiltiy \; p \\ 0 \; with \; probability \; (i-p) \end{cases}$$

Let $T = \sum X_i$

Hence $P(T=K) = ({}^nC_k)\, p^k\, (i-p)^{n-k}$

The conditional distribution of (x_1, x_2, …..x_n) given T, is

$$P(x_1 \cap x_2 \cap x_3 \cap \ldots . \cap x_n \cap T = K) = \frac{P(x_1 \cap x_2 \cap x_3 \cap \ldots . \cap x_n \cap T = K)}{P(T = K)}$$

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

Since the conditional distribution is independent of the parameter p, T=$\sum_{i-1}^{n} x_i$ is sufficient estimator for p.

It is tedious to check whether a statistic is sufficient for a given parameter using the concept of conditional distribution, one uses factorization theorem, to ascertain whether a statistic is a sufficient statistic or not.

3.4 Neyman – Fisher Factorization Theorem

The statistic T is a sufficient estimator of the parameter θ if and only if the likelihood function of sample values can be written as a product of two functions, one being the function of T and θ only, while other is the function of sample values independent of θ.

Mathematically, T is a sufficient statics for θ iff

$$L = G(T,\theta), H\ (x_1, x_2, \ldots \ldots x_n)$$

Where L stands for the Likelihood function of sample values, i.e.

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

(Here $x_1, x_2, \ldots \ldots x_n$ is the random sample of size n drawn form the population whose p.d.f. or p.m.f. is f(x, θ)).

G(T, θ) stands for the function of T and θ only and H $(x_1, x_2, \ldots \ldots x_n)$ denotes the function of sample values independent of θ.

Example: The statistic is a sufficient estimator of the mean of a normal population with t mean μ and variance $\sigma^2 (\mu\ unknown, \sigma^2\ known)$. Here the likelihood function of sample values is

$$L = \left\{\frac{1}{\sigma\sqrt{2\pi}}\right\}^n exp - \left[\frac{1}{2}\sum\left(\frac{x_i - \mu}{\sigma}\right)^2\right]$$

We may write-

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

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

$$\left(Because\ \sum_{i=1}^{n}(x_i - \bar{x})\,(\bar{x} - \mu) = 0\right)$$

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

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

Here the first factor on the right hand side only on the estimate x and the population mean μ and the second factor does not involve μ. Therefore according to the factorization theorem, x is a sufficient estimator of the mean μ of a normal population with the known variance σ^2.

3.5 Some Important Remarks About Sufficiency

1. The original sample $X_1, X_2, \dots \dots X_n$ is always a sufficient statistic.

2. A sufficient estimator is always a consistent estimator.

3. A sufficient estimator may or may not be an unbiased one.

4. A sufficient estimator is the most efficient one if a sufficient estimator exists.

3.6 The Koopman's From of the Distribution

The most general form of the distributions admitting sufficient statistic is Koopman's form given by

$$L = g(x).h(\theta).\ Exp\{a(\theta)\psi(x)\}$$

Where h(θ) and a(θ) are the functions of the parameter only and g(x) and {ψ(x)} are function of the sample observations only.

Binomial, Poisson, Normal, etc. are some distribution of this kind.

3.7 The Invariance Property of A Sufficient Estimator

If T is a sufficient statistic of parameter θ and ψ(T) is one to one function of T than ψ(T) is sufficient for ψ(θ).

3.8 Key Words

Sufficient Estimator: A statistic 'T' is said to be sufficient for estimating a parameter θ if it contains all the relevant information available in the sample about θ.

3.9 Exercises

1. Let X be a binomial variate with parameters 1 and p. if x_1, x_2, ……x_n constitutes a random sample from the distribution than show that $T_n = x_1 + x_2 + \ldots\ldots + x_n$ is a sufficient statistic for p.

2. Show that in estimating the parameter m in the Poisson distribution sample mean is a sufficient estimator of m.

3.10 Summary

In this unit we read about the sufficiency of an estimator. A statistic 'T' is said to be sufficient for estimating a parameter θ if it contains all the relevant information available in the sample about θ.

Neyman Fisher Factorization Theorem states that the static T is a sufficient statistic of parameter θ if and only if the likelihood function of sample values can be written as a product of two functions, one being the function of T and θ while other is the function of sample values independent of unknown parameter t.

3.11 Further Readings

- Modd, A.M. Graybill, F.A. Boes, D.C. (1974). *Introduction to the Theory of Statistics,* McGraw Hill International Edition.
- Rao,C.R., Linear statistical inference and its Applications, John Wiley and Sons, Inc.
- Wilks, Mathematical Statistics, Jon Wiley and Sons.
- Kendall, Vol. 1,2,3. Hafner Publishing Company, New York.

Unit-4 Complete Sufficient Statistics and RAO-Blackwell Theorem

Structure

4.1 Introduction

The concept of sufficient has already been discussed in unit 3. The principle of sufficient plays a very important role in various model of statistical inference. Here in this unit another important concept that of complete family of distribution has been discussed.

4.2 Objectives

After going through this unit you should be able to -

- Develop a perspective of the idea of complete family of distributions.
- Understand what is complete sufficient statistics.
- Know what is Rao-Blackwell theorem.

4.3 Complete Family of Distributions

Consider the statistic T based on a random sample of size n say x_1, x_2, ………x_n with joint distribution depending upon $\theta \varepsilon$.. The distribution of T itself will in general depend upon θ. Let = $\{f_\theta(t)\}$ be the family of distribution related to T.

The statistics T or more precisely the family of distributions $\{f_\theta(t\}; \theta \varepsilon \Theta$ is called complete, if for any measurable function $\emptyset$(T).

$$E\ \emptyset(T) = 0 \Rightarrow\ \emptyset(T) =\ 0 \text{ almost every where (for all } \theta \varepsilon \Theta)$$

Here $E\big(\emptyset(T)\big)$ denotes the expected value of $\emptyset(T)$.

If in addition to above property $\emptyset(T)$ is such that $\emptyset(T) < M,$ for some finite M then T is said to be boundedly complete.

Some Illustrations:

1. We have seen that if $X_1, X_2, \ldots\ldots X_n$ are a random sample from the binomial distribution with parameter $\theta(0 < \theta < 1)$ whose p.m.f.

$$\begin{cases} \theta^x(1-\theta)^{1-x} & if\ x = 0,1 \\ 0, & otherwise \end{cases}$$

Then the statistic $T = \sum_i X_i$ is sufficient for θ.

Now T has a binomial p.m.f.

$$g_\theta(t) = \{C_t^n \theta^t (1-\theta)^{n-t}\}\ if\ t = 0,1,2, \ldots \ldots \ldots . n$$

= 0 otherwise

Hence for any other function $\psi(T)$,

$$E_\theta \psi(T) \quad \sum_{t=0}^{n} (t). \theta^t . (1-\theta)^{n-t}$$

Where (t) = ${}^nC_t$. $\psi(T)$

Hence

$$E_\theta \psi(T) = o\ for\ all\ \theta\ such\ that\ 0 < \theta < 1$$

$$\Rightarrow a(0)(1-\theta)^n + a(1)\theta(1-\theta)^{n-1} + \ldots\ldots\ldots\ldots + a(n)\theta^n = 0$$

$$for\ all\ \theta\ such\ that\ 0 < \lambda = \theta(1-\theta).$$

The left hand side in this identity is a polynomial in λ, all the coefficient of which must be zero hence

$\Psi(t) = 0$ for t= 0,1,2,…..,n i.e. for all the values of T with non zero probabilities (for all 0< θ <1).

Hence T is a complete statistic. In other words, the binomial family of distribution of T is complete.

2. Let $X_1, X_2, \ldots\ldots X_n$ are a random sample from the Poisson distribution, whose p.m.f. may be written as

$$f_\theta(x) = \left\{\frac{\exp(-\theta)\theta^x}{x!}\right\}\ if\ t = 0,1,2, \ldots \ldots, \infty$$
$$= 0\ otherwise$$

$$where\ the\ parameter\ \theta \in (0, \infty).$$

We have already seen that T= $\sum_i X_i$is a sufficient statistic for.

Again T is also distributed in the Poisson from with parameter n θ i.e. with p.m.f.

$$f_\theta(t) = \left\{\frac{\exp(-\theta)\theta^t}{t!}\right\}\ if\ t = 0,1,2, \dots \dots, \infty$$
$$= 0\ otherwise$$

$$E_\theta \psi(T) = \sum_{t=0}^{\infty} \psi(T). \frac{\exp(-\theta)\theta^t}{t!}$$

$$= \exp(-n\theta) \sum_{t=0}^{\infty} a(t). \theta^t, say,$$

Where a(t) = $\frac{\psi(t).n^t}{t!}$

Consequently, $E_\theta \psi(T) = 0\ for\ all\ \theta\ such\ that\ 0 < \theta < \infty$

However, it is known form algebra that a convergent power series which is identically zero must have all the coefficient equal to zero. As such, a(t) = 0 for t = 0,1,2…..

i.e. Ψ(t) = 0 for t= 0,1,2,……

But for every $\theta \in (0, \infty)$ theses are the values of T with positive probabilities. Hence T is a complete statistic, or in other words, the Poisson family of distributions is complete.

3. Let X_1, X_2,……..X_n be a random sample from some normal distribution with unknown mean but known variance, say, from N (θ, σ^2), where $-\infty < \theta < \infty$.

We known that T= $\bar{X}$ is sufficient for θ and that it has the p.d.f.

$$g_\theta(t) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}. \exp\left(-\frac{n(t-\theta)^2}{2\theta^2}\right)$$

$$if E_\theta \psi(T) = 0 \forall \theta \in \Theta, then$$

$$\int_{-\infty}^{\infty} \psi(t) \exp\left(\frac{-nt^2}{2\sigma^2} + \frac{n\theta t}{\sigma^2}\right) dt = 0$$

For $-\infty < \theta < \infty$. However the left hand side it the bilateral Laplace transform of the function $E_\theta \psi(T)\left(\frac{-nt^2}{2\sigma^2}\right)$. From the unicity theorem of this type of transform it follows that.

$$\psi(T)exp\left(\frac{-nt^2}{2\sigma^2}\right) = 0(all\theta \in \Theta)$$

$$\psi(T) = 0 \; for \; all\theta \in \Theta$$

Hence the family of distribution$f_\theta(t), -\infty < \theta < \infty$ is complete.

4.4 Complete Sufficient Statistic

A statistic which is complete as well as sufficient is known as complete sufficient statistic.

Example 1: In case of a Poisson distribution with parameter λ i.e. when

$$f(x) = \frac{e^{-\lambda} d^x}{x!}, 0 < \lambda < \infty, x = 0,1,2, \ldots ..$$

$\bar{x}$ = sample mean is a complete sufficient statistic.

Example 2: For a binomial distribution B (n, p) i.e.

$f(x) = C_x^n . p^x . (1-p)^x, \bar{x}$ i.e. sample mean is complete sufficient statistic for np.

4.5 Rao -Blackwell Theorem

The Cramer-Rao inequality gives us a tool of judging whether or not a given unbiased estimator is also an M.V.U.F., Moreover the application of Cramer-Rao theorem is too restrictive because of regularity conditions under which it is valid.

Rao- Blackwell theorem enable us to obtain an M.V.U.E. form any unbiased estimator by using a sufficient, say T of parameter θ. The only condition that must be fulfilled is that T must also be complete.

Statement of the Theorem:

Let U be any unbiased estimator of r(θ). where r(θ) is an unknown function of . Let T be a sufficient of θ. Define

$$\emptyset(T) = E\left[\frac{U}{T}\right] \text{ which is independent of } \theta.$$

(It is guaranteed because of sufficiency of T for θ)

Then ∅(T) is itself and unbiased estimator of r(θ) and ∅Var (T) < Var (U)

Proof*:* We have,

E(U) = r(θ) (as U is an unbiased estimator of r(θ))

E(U) may be written as

E(U) = E [E(U/T)} (By the theory of conditional expectation)

=E(∅(T)) (As ∅(T)= E(U/T))

Hence, we have

E(U) = E (∅ (T)) = r(θ)

Which shows that ∅(T) is an unbiased estimator of r(θ)

Further, we may write Var(U) as

Var (U) = Var {E(U/T)}+ {V(U/T)} (By the theory of conditional Expectation)

Variance is a non-negative quantity and expectation of a non-negative quality is always nonnegative.

Hence E{V(U/T)}≥ 0

So that we have Var(U) ≥ Var E{V(U/T)}

≥ Var (∅(T)) as ∅(T) = E[V(U/T)}

Hence proved.

The implication of this result is that if one is given an unbiased estimator U of r(θ), then one may improve upon U by forming the new estimator $\emptyset(T)$ for r(θ), based on U and sufficient statistics T. This estimator $\emptyset(T)$ is unbiased for r(θ) and has smaller variance (or mean squared error) than U. This process of finding a new improved estimator is called "Black wellisation" after D. Blackwell.

The estimator ∅(T) will not be better estimator than U in sense of smaller variance but best in the sense of smallest variance provided T is also complete.

It T is a complete sufficient statistic of θ and one may find a function $\emptyset(T)$ of T such that E[$\emptyset(T)$]= r(θ) Then, $\emptyset(T)$ is necessarily an UMVUE of r(θ).

Example 1: Let X_1, X_2,……..X_n be a random sample from N(μ, σ^2), μ known and σ^2 unknown. We wish to find out MVUE of σ^2. We know that T= $\sum_{i=1}^{n}(x-\mu)^2$ sufficient statistic of σ^2. Moreover $\sum_{i=1}^{n}\left(\frac{x-\mu}{\sigma}\right)^2$ following a chi square distribution with n degree of freedom.

$$Hence \qquad \left\{\sum_{i=1}^{n}\left(\frac{x-\mu}{\sigma}\right)\right\} n$$

$$or \qquad E\left[\frac{1}{n}\sum_{i=1}^{n}(x-\mu)\right] = \sigma^2$$

$$or \qquad E\left(\frac{T}{n}\right) = \sigma^2$$

$$or \qquad E(S_o) = \sigma^2 \qquad where\ S_0 = \frac{T}{n}$$

This shows that S_0 = is MVUE of σ^2.

Example 2: Let X_1, X_2,……..X_n be a random sample of size n taken from a Poisson distribution with parameter θ i.e. its p.m.f. is

$$p(x,\theta) = \frac{e^{-\theta}.\theta^{x}}{x!}; x = 0,1,2, \dots \dots .\infty, \theta > 0.$$

Let θ is unknown. We wish to find out MVUE of

$$r(\theta) = P_r(X = m) \ (when\ m\ is\ known)$$

$$= \frac{e^{-\theta}.\theta^{xm}}{m!}$$

Let us define a r.v. U such that

U= 1 if X_i = m

= 0 otherwise

Then,

$$E(U) = \frac{e^{-\theta}.\theta^x}{x!} = 1 \times P_r(X = m) + 0 \times P_r(X \neq m)$$

$$= r(0)$$

Implying U is an unbiased estimator of r(θ).

We know that T = $\sum_{i=1}^{n} x_i$ sample total is a complete sufficient statistics of θ and its distributions is P(n θ) i.e.

$$p(t) = P_r(T = t) = \frac{e^{-n\theta}.(n\theta)^1}{t!}, \qquad t = 0,1,2, \ldots ..$$

Let us consider

$$(t) = E\left(\frac{U}{T} = t\right)$$

$$= \frac{\Pr[X_i = m, \sum_{i=2}^{n} X_i = t - m]}{Pr[T = \sum_{i=1}^{n} X_i = t]}$$

$$= \frac{\frac{e^{-\theta}\theta^m}{m} . \frac{e^{-(n-1)\theta}[(n-1)\theta]^{t-m}}{(t-m)!}}{\frac{e^{-n\theta}.(n\theta)^1}{t!}}$$

$$= C_m^t . \frac{(n-1)^{t-m}}{n!}$$

$$Thus = \emptyset(t) = C_m^t . \frac{(n-1)^{t-m}}{n^T} \text{ is an unbiased estimator of } r(\theta) = \frac{e^{-\theta}.\theta^m}{m!}$$

But T is also complete sufficient statistic of θ. Hence $\emptyset$(T) is a MVUE of $r(\theta) = \frac{e^{-\theta}.\theta^m}{m!}$

Remark: here $\emptyset$(T) has been defined as per norms of Rao-Blackwell theorem. It is unbiased estimator estimator of r(θ) with a variance that is at least as small as the variance of U. In this way one may start from any unbiased estimator for r(θ) and get a new one from it by using the conditional expectation of this estimator for given T. However all these estimators are equal because T is complete and therefore $\emptyset$(T) is MVUE of r(θ).

Example 3: Let X_1, X_2,……..X_n be a random sample of size n taken from U(0,θ) i.e.

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}, & 0 < x <;\ \theta > 0 \\ 0 & otherwise \end{cases}$$

Our problem is to find out MVUE of θ.

Let T = $X_{(n)}$ be the nth order statistics of θ.

Then T is a sufficient and compete statistics of θ and its p.d.f is given by

$$f(t) = \frac{nt^n}{\theta^n}, 0 \le t \le \theta, \qquad \theta > 0$$

$$Hence \qquad E(T) = \int_0^\theta t.\frac{nt^n}{\theta^n}dt$$

$$= \frac{n}{\theta^n}\int_0^\theta t^n.dt = \frac{n}{\theta^n}\left(\frac{\theta^{n+1}}{n+1}\right) = \frac{n}{n+1}\theta$$

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

$$or \qquad \frac{n+1}{n}E\,[T] = \theta$$

$$or \qquad E\left[\frac{n+1}{n}T\right] = \theta$$

$$where \qquad \emptyset(T) = \frac{n+1}{n}T$$

As T is complete and sufficient statistic of θ and E $[\emptyset(T)] = \theta$ therefore $\emptyset(T) = \frac{n+1}{n}T$ of θ

4.7 Summary

A statistics which is complete as well as sufficient is known as compete sufficient statistic.

If U is an unbiased estimator of r (θ) then one may improve upon U by forming the new estimator $\emptyset$(T) for r(θ), based on U and sufficient statistics T. This estimator $\emptyset$(T) is unbiased for r(θ) and has smaller variance (or mean squared error) than U. This process of finding a new improved estimator in the sense of smaller variance, starting from an unbiased estimator is called 'Blackwellsation.'

If T is a complete sufficient statistic of and one may find a function $\emptyset(T)$ of T such that E[$\emptyset(T)$] = r(θ) then, r(θ) is necessarily an UMVUE or r(θ).

4.8 Key Words

Complete Statistic: The statistic T or more precisely the family of distributions$\{f_\theta(t); \theta \in \Theta\}$ is called complete, if for any measurable function $\emptyset$(T)

$$E(\emptyset(T)) = 0 \Rightarrow \emptyset(T) = 0 \text{ almost every where (for all } \theta \in \Theta).$$

Complete Sufficient Statistic: A statistic which is complete as well as sufficient is known as complete sufficient statistic.

4.9 Exercises

1. Find out UMVUE of in case of θ^2 Poisson distribution unit parameter θ.

2. In case of binomial distribution with parameters n and find out UMVUE of θ (1- θ).

4.10 Further Readings

- Outlines of Statistics by Goon Gupta and Dasgupta, Volumes II
- Kendall, Vol. 1,2,3. Hafzer Publishing Company, New York.

U.P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105
Advance Statistical Inference

Block: 2 MVU Estimation

Course Design Committee

Course Preparation Committee

Prof. Umesh Singh **Writer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. B. P. Singh **Reviewer**
Department of Statistics, Banaras Hindu University, Varanasi

Dr. Sanjay Singh **Editor**
Department of Statistics, Banaras Hindu University, Varanasi

Dr. Shruti **Course/ SLM Coordinator**
School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105 ADVANCE STATISTICAL INFERENCES
©UPRTOU
First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)
Second Edition: ***July 2021***
ISBN : 978-93-94487-38-3

Printed By: K.C. Printing & Allied Works, Panchwati, Mathura - 281003.

Block & Units Introduction

The ***Block - 2 – MVU Estimation*** is the second block with two units and deals with the problem of estimation particularly the procedures and concepts related to the minimum variance unbiased estimation.

Unit – 5 – MVU Estimators, provides the basic concepts related to minimum variance unbiased estimators.

Unit – 6 – Complete Sufficient Statistics, describes the basic concepts of complete sufficient statistics.

At the end of each unit the summary, self assessment questions and further readings are given.

Unit-5 MVU Estimators UMVUE

Structure

5.1 Introduction

In this unit, we will introduce you to the concept of minimum variance unbiased estimation procedure. In Block-I you have studied the general point estimators problem along with the various properties of the estimators.

5.2 Objectives

After reading this unit you should be able to :

- Obtain the MVUE
- Result related to MVU Estimation.

5.3 Minimum Variance Unbiased Estimation

You have already studied the general point estimation problem. You may recall that in point estimation problem we have a sample of specified size, say n ($\geq$1), and assume that the sample has been drawn from a population having a distribution function F $(X|\theta)$ where the functional form of F is assumed to be known however the arbitrary constant (s) θ involve therein is (are) not known except that its values lies in a given set of values called parameter space denoted as . On the basis of the information supplied by the sample we wish estimate the value of the constant(s) θ or a function of θ, say g(θ). For generally we will consider the estimation of g (θ).

Let us denote by $X_1, X_2, \ldots\ldots\ldots, X_n$ a random sample of, size n. The given sample is therefore a random observation on it and may be denoted as $x_1, x_2, \ldots\ldots\ldots x_n$. since the estimation of g (θ) is to be done on the basis of the observed value of $X_1, X_2, \ldots\ldots\ldots, X_n$. We should search a function of sample observation T = T (x_1, x_2, ……….x_n) such that its value t = T (X_1, X_2,………,X_n) for given sample may be taken as the estimate of g(θ). T is called estimator of g(θ) and t is called the estimate of g (θ). It is worthwhile to mention here that a number of functions of X_1, X_2,………,X_n (estimators) may be define for the estimation of g(θ). You may further note here that from the given population if we draw samples of specified size n again and again then the estimates may differ form sample to sample even for the given function T. No doubt, the variation in the estimates can be studied from the distribution of T. However, one would expect in this case that the estimator T should be chosen such that it always gives the estimate to be equal to g(θ) for every $\theta \in \Theta$ or in other words T should be chosen such that P[T = g(θ)] = 1. This can only be achieved if T is a constant (degenerate random variable) equal to g(θ). But g(θ) is unknown, thus it can never be met.

In such a situation one would like to choose an estimator T which gives the estimate close to the true value of g(θ). In other words the estimator should be such the estimates are densely clustered around g (θ) i.e. the central tendency of the distribution of T should be g(θ) and the dispersion should be as small as possible.

If the central tendency of the distribution of the estimator is equal to g(θ), the estimator is called unbiased. Depending on the various measures of the central tendency various types of unbiasedness can be defined which are below:

Mean Unbiased: An estimator T is said to be mean unbiased (often called unbiased only) for g(θ), if mean of T is g (θ), if mean of T is g (θ) for every $\theta \in \Theta$.

Median Unbiased: An estimator T is said to be median unbiased for g(θ), if median of T is g(θ) for every $\theta \in \Theta$.

Modal Unbiased: An estimator T is said to be modal for g(θ), if mode of T is g(θ) for every $\theta \in \Theta$.

Remark: The mean unbisedness is the most popular unbiased criterion which was proposed by Prof. R. A. Fisher. We will see in the following paragraph the reason for its popularity.

Consider the dispersion of T. How the values of T are dispersed around g(θ) can be measured by noting average of the differences between T and g (θ). The differences should be measured regardless of there sign. Boscovitch and Laplace suggested the use of |T-g(θ)| and Gauss and Legendre suggested the use of [T-g(θ)]. For sake of simplicity and mathematical tractability consider Gauss and Legendre suggestion. Since [T-g(θ)] is random variable we consider the average of $[T-g(\theta)]^2$ overall possible values of t as measure of the dispersion. In other words E $[T-g(\theta)]^2$ = (Mean Square Error) MSE (T) can be taken as the measure of dispersion. Thus if T_1 and T_2 are the two estimators of g(θ) we should prefer T_1 over t2 if MSE (T_1) ≤ MSE (T_1) for all possible values of $\theta \in \Theta$.

You may recall that by Tchebychev's inequality for any estimator T

$$P\ [|\ T\text{-}g(\theta)\ | < \varepsilon\ ]\ 1\text{-}\ MSE\ (T)/\ \varepsilon^2$$

Thus, we may conclude that we prefer an estimator with greater concentration around g(θ) which corresponds to the smaller MSE. Now the question arises that does MSE criterion lead to unique choice of estimator. The answer is no. Consider the following example:

Example 1: Let X_1, X_2,………,X_n be a random sample form a normal population with unknown mean θ and variance 1. We wish to estimate θ. Consider the estimator $T_1 = \bar{X}$, the sample mean $T_2 = \theta_0$. You know that the distribution of sample mean $\bar{X}$ is normal with mean θ and variance 1/n. thus MSE (T_1) = 1/n. Similarly it is easy to check that MSE (T_2) = $(\theta - \theta_0)^2$. Now for the values of $\theta \in \left(\theta_0 - \frac{1}{n}, \theta_0 + \frac{1}{n}\right)$, MSE ($T_2$) < MSE ($T_1$) and for other values of θ MSE (T_2) > MSE (T_1) and hence we cannot choose T_1 over T_2 or T_2 over T_1 for all the values of the parameter θi.e. whole parameter space. It may also be noted here that T_1 is minimal sufficient statistics where as T_2 does not depend on the observation at all.

From the above example it is clear that minimization of the mean square error itself will be a difficult task. In fact it can be readily seen that there exist no estimator for which the mean square error is minimum for all values of $\in \Theta$. On the other hand minimization of MSE provides a greater concentration of estimated values around g(θ). Note that mean square error of the estimator T is defined as

$$MSE\ (T) = \ E\ [T - g(\theta)]^2 = E[\ T - E\ (T) + \ E(T) - g(\theta)]^2$$

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

$$= V\ (T) + \{B(T)\}^2$$

Where B(T) = E (T)- g(θ) defines the bias of the estimator T. One way of minimization of the MSE would be to restrict our self to those estimators only for which the B(T) is zero i.e. the estimators are mean unbiased (Now on wards we will call mean unbiased estimators as unbiased estimators only). Thus for searching the best estimator we should confine our self to unbiased estimators only and choose among them that estimator for which the variance is least. Now we can define the minimum variance unbiased estimators.

Definition: A statistic T is said to be minimum variance unbiased estimators of g (θ) if.

1. E(T) g (θ) for all $\theta \in \Theta$
2. V(T) ≤ V(T) for all and for all $\theta \in \Theta$ the estimators T satisfying the condition 1.

5.4 Some Theorems of MVUE

Theorem 1:

A MVUE estimator is unique in the sense that if T_1 and T_2 both are MVUE of g(θ) the T_1 = T_2 almost every where.

Proof:

It is given that T_1 and T_2 both are unbiased for g(θ), therefore,

$$E(T_1) = E\ (T_2)\ g\ (\theta) \text{ for all } \theta \in \Theta \qquad (3.1)$$

And both are minimum variance estimators,

$$V(T_1) \leq V(T_2) = V \text{ (say) for all } \theta \in \Theta \qquad (3.2)$$

Consider now the new estimator

$$T = (T_1 + T_2)/2 \qquad (3.3)$$

$$= g\ (\theta) \qquad \text{form (3.1)}$$

Hence T is also unbiased estimator of g(θ).

Further,

$$V(T) = V\left[\frac{\{(T_1) + (T_2)\}}{2}\right]$$

$$= \frac{1}{4}[V(T_1) + V(T_2)\ 2Cov\ (T_1 T_2)]$$

$$= \frac{1}{4} V(T_1) + V(T_2) + 2\rho\sqrt{V(T_1)V(T_2)}$$

$$= \frac{V(1+\rho)}{2} \text{ (From 3.2)} \qquad (3.4)$$

Where ρ is the Pearson's correlation coefficient between T_1 and T_2 Since T_1 is MVUE

$$V(T) \geq V(T_1)$$

$$\Rightarrow \frac{V(1+\rho)}{2} \geq V$$

$$\Rightarrow \frac{1+\rho}{2} \geq 1\ i.e. \geq 1$$

Since Pearson's correlation coefficient can not be greater than 1, we must have ρ =1. Therefore T_1 and T_2 must have a linear relationship of the form

$$T_1 = a + bT \qquad (3.5)$$

Where a and b are constants independent of sample observation but does not depend on θ.

Taking expecting of both the sides of (3.5) and using (3.1) we have

$$g(\theta) = a + b\, g(\theta) \qquad (3.6)$$

Further

$$V(T_1) = V(a + bT_2)$$

$$= b^2 V(T_2)$$

$$1 = b^2$$

$$\Rightarrow b = \pm 1 \qquad \text{(from 3.4)}$$

But since ρ $(T_1, T_2) = +1$, The coefficient of regression of T_1 and T_2 must be positive.

$b = 1 \Rightarrow a = 0$ substitution in (3.5), we get $T_1 = T_2$ as desired.

Theorem 2 :

If T1 is MVU for θ and T_2 be any other unbiased estimator of θ then no linear combination of T_1 and T_2 is a MVU estimator.

Proof:

Let us consider a linear combination

$$T = k_1T_1 + k_2T_2 \qquad (3.7)$$

It will be unbiased estimator of θ if

$$E(T) = k_1E(T_1) + k_2E(T_2) = \theta$$

$$\Rightarrow k_1 + k_2 = 1 \qquad (3.8)$$

We have

$$e = \frac{Var\ (T_1)}{Var\ (T_2)} \Rightarrow Var\ T_2 = \frac{Var\ T_1}{e} \qquad (3.9)$$

Now Var(T) = Va ($k_1T_1 + k_2T_2$)

$$= k_1^2 Var\ (T_1) + k_2^2 Var\ (T_2) + 2k_1k_2\ cov\ (T_1, T_2)$$

$$= k_1^2 Var\ (T_1) + k_2^2 Var\ (T_2) + 2k_1k_2\ \rho\ (var\ (T_1) var\ (T_2))1/2$$

$$= \ Var\ (T_1)\left(k_1^2 + \frac{k_2^2}{e} + 2k_1k_2 \frac{\rho}{\sqrt{e}}\right)$$

$$= \ Var\ (T_1)\left(k_1^2 + 2k_1k_2 + \frac{k_2^2}{e}\right) \qquad (\rho = \sqrt{e})$$

$$= \ Var\ (T_1)(k_1^2 + 2k_1k_2 + k_2^2)$$

$$= \ Var\ (T_1)(k_1+k_2)^2$$

$$= \ Var\ (T_1)$$

$\Rightarrow$ T can not be MVU estimator.

Example 1: If T_1 and T_2 be two unbiased estimate of parameters θ with variance σ_1^2, σ_2^2 and correlation ρ what is the best unbiased linear combination of T_1 and T_2 and what is the variance of such combination?

Let T_1 and T_2 be two unbiased estimate of parameters θ.

$$\therefore \ E(T_1) = E(T_2) = \theta \qquad (3.10)$$

Let T be a linear combination of T_1 and T_2 given by

$$T = l_1T_1 + l_2T_2$$

Where l_1, l_2 are arbitrary constants.

$$E(T) = \ l_1E(T_1) + l_2E(T_2) = (l_1 + l_2)\theta$$

$\therefore$ T is also an unbiased estimate of θ if and only if

$$l_1 + l_2 = 1 \qquad (3.11)$$

$$V(T) = \ (l_1T_1 + l_2T_2)$$

$$= l_1^2 V\ (T_1) + l_2^2 V\ (T_2) + 2 l_1 l_2\ cov\ (T_1, T_2)$$

$$= l_1^2 \sigma_1^2 + l_2^2 \sigma_2^2 + 2 l_1 l_2 \rho \sigma_1 \sigma_2 \qquad (3.12)$$

We want the minimum value of (3.12) for variations in $l_1 and\ l_2$ subject to the condition (3.11).

$$\frac{\partial}{\partial \mathrm{l}_1} V(T) = 0 = l_1 \sigma_1^2 + l_2 \rho \sigma_1 \sigma_2$$

$$\frac{\partial}{\partial \mathrm{l}_1} V(T) = 0 = l_2 \sigma_2^2 + l_2 \rho \sigma_1 \sigma_2$$

Subtracting, we get

$$l_1(\sigma_1^2 - l_2 \rho \sigma_1 \sigma_2) = l_2(\sigma_2^2 - l_2 \rho \sigma_1 \sigma_2)$$

$$\Rightarrow \frac{l_1}{\sigma_1^2 - l_2 \rho \sigma_1 \sigma_2} = \frac{l_2}{\sigma_2^2 - l_2 \rho \sigma_1 \sigma_2} = \frac{l_1 + l_2}{\sigma_1^2 + \sigma_2^2 + 2\rho \sigma_1 \sigma_2} = \frac{1}{\sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2}$$

$$\therefore\ l_1 = \frac{\sigma_2^2 - \rho \sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2}\ and\ \ l_2 = \frac{\sigma_1^2 - \rho \sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2}$$

With these values of given (*) is an unbiased combination of T_1 and T_2 and its variance is given by (3.12).

Example 2: Suppose T_1 in the above example is an unbiased minimum variance estimate and T_2 is any other estimate with variance $\frac{\sigma^2}{e}$ .Then prove that the correlation between T_1 and T_2 is $\sqrt{e}$.

Sol. The coefficients of the best linear combination of T_1 and T_2, given by

$$l_1 = \frac{\sigma_2^2 - \rho \sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2}\ and\ \ l_2 = \frac{\sigma_1^2 - \rho \sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2} \qquad (3.13)$$

We are given that $\sigma_1^2 = V(T1) = \sigma^2\ and$

$$e = \frac{V(T_1)}{V(T_2)} = \frac{\sigma^2}{V(T_2)} \Rightarrow V(T_2) = \sigma_2^2 = \sigma^2/e$$

Substituting in (3.13) we get,

$$l_1 = \frac{1 - \rho\sqrt{e}}{D}$$

$$l_2 = \frac{e - \rho\sqrt{e}}{D}$$

Here D= 1+e-2 $\rho\sqrt{e}$ (3.14)

Hence from T= l_1T_1 + l_2T_2 the unbiased statistic is

$$T = \frac{[1 - \rho\sqrt{e}]T_1 + (e - \rho\sqrt{e})T_2}{D}$$

And From V(T)= V (l_1T_1 + l_2T_2)

$$= l_1^2 V\ (T_1) + l_2^2 V\ (T_2) + 2l_1 l_2\ cov\ (T_1, T_2)$$

$$= l_1^2\sigma_1^2 + l_2^2\sigma_2^2 + 2l_1 l_2 \rho\sigma_1\sigma_2$$

$$V(T) = \frac{1}{D^2}\left[\left(1 - \rho\sqrt{e}\right)^2\sigma^2 + \left(e - \rho\sqrt{e}\right)^2\frac{\sigma^2}{e} + 2\left(1 - \rho\sqrt{e}\right)\left(e - \rho\sqrt{e}\right)\rho.\sigma.\sigma/\sqrt{e}\right]$$

$$= \frac{\sigma^2}{D^2}\left[\left(1 + \rho^2 e - 2\rho\sqrt{e}\right)\ \frac{1}{e} + \left(e^2 + \rho^2 e - 2\rho e\sqrt{e}\right) + 2\left(1 - \rho\sqrt{e}\right)\left(\sqrt{e} - \rho\right)\rho\right]$$

$$= \frac{\sigma^2}{D^2}\left[\left(1 - \rho^2 e - \rho^2 - 2\rho\sqrt{e} + 2\rho^3\sqrt{e}\right)\ \right]$$

$$= \frac{\sigma^2(1 - \rho^2)\left(1 + e - 2\rho\sqrt{e}\right)}{\left(1 + e - 2\rho\sqrt{e}\right)^2} = \frac{\sigma^2(1 - \rho^2)}{\left(1 + e - 2\rho\sqrt{e}\right)^2}$$

$$= \frac{\sigma^2(1 - \rho^2)}{(1 - \rho^2) + \left(\sqrt{e} - \rho\right)^2}$$

$$\therefore \frac{V(T)}{\sigma^2} = \frac{1 - \rho^2}{(1 - \rho^2) + \left(\sqrt{e} - \rho\right)^2} \leq 1 \qquad (3.15)$$

Since T_1 is the most efficient statistics,

$$V(T_1)\sigma^2 \Rightarrow \frac{V(T)}{\sigma^2} \geq 1 \qquad (3.16)$$

From (3.15) and (3.16) we get

$$\frac{V(T)}{\sigma^2} = 1, i.e. \frac{1-\rho^2}{(1-\rho^2) + \left(\sqrt{e} - \rho\right)^2} = 1$$

$$\left(\sqrt{e} - \rho\right)^2 = 0 \ \Rightarrow \rho = \sqrt{e}$$

5.5 Summary

Unbiasedness is an important concept of an estimator and unbiased estimator with minimum variance among all unbiased estimator is called minimum variance unbiased estimator. A minimum unbiased estimator is unique in the sense that T_1 and T_2 both are MVUE of g (θ) the $T_1 = T_2$ almost every where. The correlation coefficient between MVUE and other estimator is ratio of the variance of MVU and variance of other estimator.

5.6 Further Readings

- Rohatgi V.K. (1984): Statistical Inference John Wiley & Sons, New York.
- Lehman E. L. (1986) Testing Statistical Hypothesis John Wiley & sons, New York.
- Goon A.M., Gupta M.K. & Das Gupta B. (1977) An Outline of Statistical Theory. Vol. I The World Press Pvt. Ltd., Calcutta.
- Goon A.M., Gupta M.K. & Das Gupta B (1983) Fundamentals of Statistics Vol. I The World Press Pvt. Ltd., Calcutta.
- Hogg R.V. Craig A. (2003). Introduction to Mathematical Statistics

Unit-6 Use of Complete Sufficient Statistics in the Construction of UMVUE

Structure

6.1 Introduction

In statistics, we often represent our observations as X= (X_1,….X_n), a random sample of size n form some population. The model can be written in the from $\{f\theta(x): \theta \in \Omega\}$ where Ω is the parameter space of or set of permissible values of the parameter and $f\theta(x)$ is the probability density (mass) function. A statistic T(X) is a function of the observation, which does not depend on the unknown parameter. Although a statistic T(X) is not a function of θ, its distribution can depend on θ. An estimator is a statistic considered for the purpose of estimating a given parameter. One of our objectives is to find a 'good' estimator of the parameter θ, in some sense of the wood 'good'. How do we ensure that a statistic T(X) is estimating the correct parameter and not consistently too large or too small and that as much variability as possible has been removed? The problem of estimating the correct parameter is often dealt with by requiring that the estimator be unbiased.

We will denote an expected value under the assumed parameter value θ by E_θ (.). Thus in the continuous case

$$E_\theta[h(X)] = \int_{-\infty}^{\infty} h(x) f_\theta(x) dx,$$

and in the discrete case

$$E_\theta[h(X)] = \sum_{all\ x} h(x) f_\theta(x)$$

Provided in integral/sum exists and is finite.

Definition: A statistic T(X) is an unbiased estimator of θ if E_θ(T(X)) for all $\in \Omega$. Also if there exists an unbiased estimator of , θ is called an estimable parameter.

For example suppose that x_i; i=1,….,n; is an independent Poisson variate with parameter θ, so $E_\theta(X_i)$. We may then have

$$E_\theta(\bar{X}) = \frac{1}{n}\sum_i^n E_\theta(X_i) = \frac{1}{n}\sum_i^n \theta = \theta.$$

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

In general, in fact there can be so many unbiased estimators of the parameter and attempt is always made to find one which outperforms all others at all values of the parameter. In order to achieve an optimal estimator, it is necessary to restrict ourselves to the class of unbiased estimators and select the best within that class.

Definition: An estimator T(X) is said to be a uniformly minimum variance unbiased estimator (UMVUE) of the parameter θ if

(i) It is an unbiased estimator of θ and
(ii) Among all unbiased estimators of θ it has the smallest variance.

6.2 Sufficient Statistic and Completeness

The derivation of UMVUE is relatively easy if we have notion of sufficiency and the completeness for the unknown parameter θ or for the family of distribution under consideration.

A sufficient statistic is one that, from a certain perspective, contains all the necessary information for making inference about the unknown parameter(s) in a given model. It is important to remember that a statistic is sufficient for inference on a specific parameter in specific model.

Suppose the data is in vector X and T = T(X) is a sufficient statistic for θ. The intuitive basis for sufficiency is that the conditional distribution of X given T(X) does not depend on θ. Thus X provides no additional value in addition to T for estimating θ. The assumptions is that random variables carry information on a statistical parameter θ only in so far as their distributions (or conditional distributions) change with the value of the parameter and that since given T(X) we can randomly generate at random values for the X without knowledge of the parameter and with the correct distribution, these randomly generated values cannot carry additional information. All of this, of course, assumes that the model is correct and is the only unknown. The distribution of X, given a sufficient statistic T, will often have value for other purpose such as measuring the variability of the estimator or testing the validity of the model.

Definition: A statistic T(X) is sufficient for the parameter if the conditional distribution of data (X_1,….X_n), given T(X) = t is independent of the unknown parameter θ.

Sufficient statistics are not unique. For example if the sample mean $\bar{X}$ is a sufficient statistic, then any other statistic, that allows us to obtain $\bar{X}$ is also sufficient. This will include all one-to-one functions of $\bar{X}$ (these are essentially equivalent) like $\bar{X}^3$ and all statistics T(X) for which we can write $\bar{X}$ = g(T) for some, possibly many –to-one function g(.).

One result which is normally used to verify whether a given statistic is sufficient is the Factorization Criterion for Sufficiency. The criterion is both necessary and sufficient condition which states that the likelihood function can be factorized in to non-negative functions g(T, θ) and h(X) where h(.) is independent of θ.

Definition: A sufficient statistic T (.) or, more precisely, a family of distribution of T(.) is said to be complete if there exists no non-trivial unbiased estimate of zero. That is T(X) is complete if for any function h(.) of T(.), $E_\theta[h(T(X))] = 0 \; for \; all \; \theta \in \Omega$ implies h (T(X)) = 0 almost everywhere.

6.2.1 Lehmann-Scheffe' Theorem

This theorem is an immediate consequence of Rao-Blackwell theorem in the sense that it adds uniqueness feather for the existence of UMVUE. The statement of the theorem is as under.

Let T be a sufficient statistic for θ and suppose further that T is complete. Then every estimable function g(θ) possesses an unbiased estimate with uniformly minimum variance and this estimate is unique unbiased estimate of g(θ) and is function of T.

Proof: It follows from Rao-Blackwell theorem that there is at most one unbiased estimator of g(θ), which is a function of sufficient statistic. If possible, let there be two functions of T, i.e. φ_1(T) and φ_2 (T), both unbiased estimators for g(θ).

Then φ(T) = φ_1(T) - φ_2 (T) is an unbiased estimator of zero. So E_θ [φ(T)]=0. Since T is complete, this implies that φ(T) =0 almost every where. It in turn implies that for almost all T. So, there cannot be two unbiased estimators of g(θ).

6.2.2 Simple Examples on the Construction of UMVUE

The construction of UMVUE can be easily done if we have a complete sufficient statistic. At first we need to obtain the complete sufficient statistic T and its distribution. We then try to assess a function φ(T) such that E_θ [φ(T)] is related to the parameter of interest, say g(θ). if need arises we may require to manipulate the function to some other function say φ'(T) such that E_θ [φ'(T)]= g(θ) .

Example 1: Let $X_1, \ldots X_n$ be independent and identically distributed according to the Poission distribution with parameter θ (> 0). Find the UMVE of θ, θ^2 and e^{θ}.

Solution: It can be shown that T = $\sum_i^n X_i$ is sufficient for θ and the Poisson family is complete. It can be further shown that T is a Poisson variate with parameter n θ. Thus, E_θ (T) = n θ which ultimately suggests that T/n is UMVE of θ.

For getting the UMVUE of θ^2 let us write.

V(T) = n θ which implies that $E_\theta[T^2] - \{E_\theta[T]\}^2 = n\theta$. This last relationship can be simplified to get $E_\theta[(T^2 - T)/n^2] = \theta^2$. Thus, $[(T^2 - T)/n^2]$ is UMVUE of θ^2.

Finally, let f(t) be UMVUE of e^{θ} . We then have

$$E_\theta[f(t)] = \sum f(t) \frac{e^{n\theta}(n\theta)^t}{t!} = exp(-\theta)$$

$$\Rightarrow \sum f(t) \frac{(n\theta)^t}{t!} = \exp[(n-1)\theta].$$

Since this series is absolutely convergent, we may equate the coefficients of θ^t from both the sides. Thus

$$f(t)\frac{n^t}{t!} = \frac{(n-1)^t}{t!}$$

$$\Rightarrow f(t) = \left(\frac{n-1}{n}\right)^t$$

Which the required UMVUE of exp (-θ).

Example 2: Let $x_1, \ldots x_n$ be i.i.d. from the uniform distribution on (o, θ), θ > 0. Find the UMVUE of θ.

Solution: It can be shown that T= X_n, the largest order observation is sufficient and the family of distribution is complete. It can be further shown that T has the distribution $\frac{nt^{(n-1)}}{\theta^n}$ Thus

$$E_\theta(T) = \int_0^\theta \frac{nt^n}{\theta^n} dt$$

Which on simplification gives

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

and, therefore, [(n+1)/n] T is UMVUE of θ.

6.2.3 Some Important Distribution and Corresponding Complete Sufficient Statistics

The table below gives standard forms of some important distributions and corresponding complete sufficient statics. Verify the result in each case.

Distributions	**Complete Sufficient Statistic**
$Possion\ (\theta)$	$\sum_{i=1}^{n} X_i$
$Binomial\ (n, \theta)$	$\sum_{i=1}^{n} X_i$
$Negativ\ e\ Binomial\ (k, \theta)$	$\sum_{i=1}^{n} X_i$
$Geometric\ (\ \theta)$	$\sum_{i=1}^{n} X_i$
$Normal\ (\mu, \sigma^2)$	$if\ \sigma^2\ known\ \sum_{i=1}^{n} X_i$
$Normal\ (\mu, \sigma^2)$	$\mu\ known\ \sum_{i=1}^{n} (X_i - \mu)^2$
$Normal\ (\mu, \sigma^2)$	$\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$
$Exponential\ (\theta)$	$\sum_{i=1}^{n} X_i$

6.3 Unsolved Exercises

E- 1. Obtain UMVUE of (i) p and (ii) pq in f(x|p) = pxq1 when x = 0, and 1 and 0 <p = (1-q) <1.

E-2. Show that the sample mean $\bar{X}$ is UMVUM for μ in Normal (μ, σ^2) when is known.

E-3. Obtain the UMVUE of σ^2in Normal (μ, σ^2) where μ is known.

E-4. Consider the distribution $(x|\theta) = \theta \exp(-\theta x), x > 0.$. obtain the UMVUE of $\theta \ and 1/\theta.$

E-5. Let X_1, ….X_n be a random sample from N(θ, θ^2). Show that T =$(\sum x_i, \sum x_i^2)$ is sufficient for θ although it is not complete.

6.4 Summary

Unbiasedeness is an important concept of an estimator but the problem that we may several unbiased estimators of a parameter and we may require one which is optimal in some sense. We, therefore come across yet another concept of sufficiency. A simple definition of sufficiency suggests that a statistics is said to be sufficient if it contains all the information about the parameter contained in the sample. But this is not a working definition that helps us to obtain the sufficient statistics unless one goes for factorization criterion. Rao-Blackwell theorem consider sufficiency and obtains the minimum variance unbiased estimator of a parameter as a function of sufficient statistics. Thus minimum variance unbiased estimator is in some sense optimal that restricts to the class of unbiased estimators and recommends the one that has minimum variance.

Completeness says that there is no unbiased estimate of zero. Other than zero itself. So if there is a family which offers both complete and sufficient statics, Lehmann-Scheffe theorem further adds into the Rao-Blackwell theorem and suggests that the function of sufficient statistic, which is unbiased as well is unique. Thus we finally obtain the notion of uniformly minimum variance unbiased estimators that is easily obtainable for the families where we are in a position to asses both complete sufficient statistic. UMVUE is the optimal estimator in the class of unbiased estimators.

For obtaining UMVUE, we need to ascertain that the family does provide compete-sufficient statistics, say T, and then assess a function of T which is unbiased for the parameter under consideration. This unbiased estimator if UMVUE.

6.5 Further Readings

- Rohatgi V.K. (1984): Statistical Inference John Wiley & Sons, New York.
- Lehman E. L. (1986) Testing Statistical Hypothesis John Wiley & sons, New York.
- Goon A.M., Gupta M.K. & Das Gupta B. (1977) An Outline of Statistical Theory. Vol. I The World Press Pvt. Ltd., Calcutta.

- Goon A.M., Gupta M.K. & Das Gupta B (1983) Fundamentals of Statistics Vol. I The World Press Pvt. Ltd., Calcutta.
- Hogg R.V. Craig A. (2003). Introduction to Mathematical Statistics

U.P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105 Advance Statistical Inference

Block: 3 Testing of Hypothesis - I

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj
Prof. Anup Chaturvedi **Member**
Department of Statistics, University of Allahabad, Prayagraj
Prof. S. Lalitha **Member**
Department of Statistics, University of Allahabad, Prayagraj
Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur
Dr. Shruti **Member-Secretary**
School of Sciences, U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Prof. A. H. Khan **Writer**
Department of Statistics, Aligarh Muslim University, Aligarh
Prof. Umesh Singh **Reviewer**
Department of Statistics, Banaras Hindu University, Varanasi
Prof. S. K. Pandey **Reviewer**
Department of Statistics, Lucknow University, Lucknow
Prof. K. K. Singh **Editor**
Department of Statistics, Banaras Hindu University, Varanasi
Dr. Shruti **Course/ SLM Coordinator**
School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105 ADVANCE STATISTICAL INFERENCES
©UPRTOU
First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)
Second Edition: ***July 2021***
ISBN : 978-93-94487-38-3

Printed By: K.C. Printing & Allied Works, Panchwati, Mathura - 281003.

Block & Units Introduction

The ***Block - 3 – Testing of Hypothesis - I,*** deals with testing of hypothesis and consists of two units.

Unit – 7 – Preliminary Concepts in Testing, describes the concepts of critical regions, test function, two kinds of errors, size and power function of the test.

Unit – 8 – MP and UMP Tests discusses the concepts of most powerful and uniformly most powerful test is a class of size α tests with simple illustration.

At the end of every unit the summary, self assessment questions and further readings are given.

Unit-7 Preliminary Concepts in Testing

Structure

7.1 Introduction

Let X be a random variable (rv) whose form of the distribution f (x,θ) may be known except perhaps for parameter(s).

For example:

(i) Let a random variable X follow Poisson distribution with probability mass function (pmf)

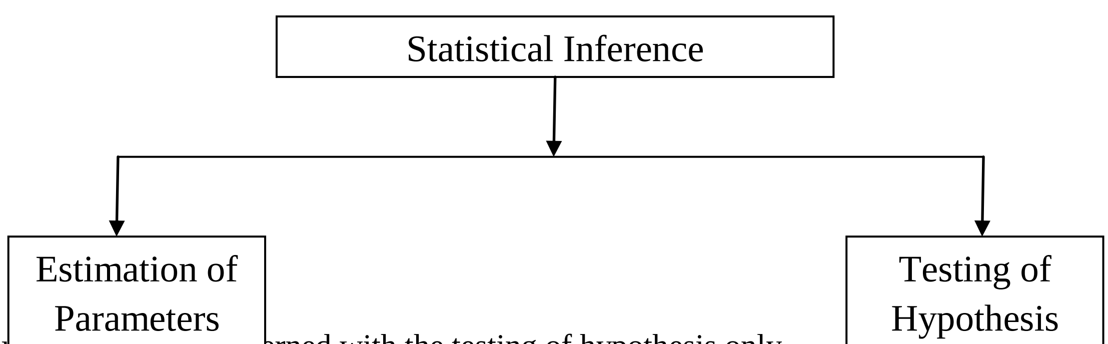
$$f(x,\theta) = \frac{e^{\theta}\theta^{x}}{x!}, \quad x = 0,1,2, \dots \dots$$

Here $f(x,\theta)$ is the pmf of Poisson distribution with θ unknown.

(ii) Let a random variable X follow normal n (θ_1, θ_2) distribution with θ_1, θ_2 unknown.

In statistical inference we are concerned with the estimation/ testing of unknown parameter(s).

Therefore statistical inference may be broadly divided into two groups.



In this unit, we will be concerned with the testing of hypothesis only.

We may have information about some thing but we want to insure that the available information is correct. This is testing problem. For example, if an Electric Company is producing light bulbs and claim made by the Company we will randomly choose n bulbs from their produce and their life length. On the basis of this random sample, we may reach to a conclusion that the claim is tenable or not.

To develop the theory of testing, we first introduce the concepts needed.

7.2 Objectives

After reading this unit you should be able to understand :

- Testing of Statistical Hypothesis
- Null and Alternative Hypothesis
- Two Types of Errors
- Critical Region
- Size and Power of the Test

7.3 Basic Concepts

Statistical Hypothesis: It is an assertion/assumption about probability density function/ probability mass function or its parameters. We have two types of hypothesis.

(i) Null Hypothesis (ii) Composite Hypothesis

Null Hypothesis: A hypothesis under test is called a Null Hypothesis and is denoted by Ho. It is also called hypothesis of no difference in some context (if H_o: $\theta_1 = \theta_2$ then it means there is no difference between $\theta_1 \, and \, \theta_2$)

***Alternative Hypothesis*:** A hypothesis under consideration is called an Alternative Hypothesis and is denoted by H_1.

This (H_1) is needed to specify our problem. For example, if

Let a random variable X follow Poisson distribution P(θ), then we may have

(i) H_o: $\theta = \theta_0$	(ii) H_o: $\theta = \theta_0$	(iii) H_o: $\theta = \theta_0$
H_1: $\theta > \theta_0$	H_1: $\theta < \theta_0$	H_1: $\theta \neq \theta_0$

Here H_0 is to be tested against H_1.

(i) and (ii) are one tailed test where as (iii) is a two tailed test.

Simple and Composite Hypothesis: If all the parameters of a distribution are completely specified then it is simple otherwise composite hypothesis.

Examples: Let X follow normal distribution N (H_1: θ,1) δ.

then H_0: $\theta = 1$; is known as simple hypothesis, whereas, H_0: $\theta > 10$ is composite hypothesis.

(i) Let X follow normal distribution $N\,(\theta_1, \theta_2)\,\delta$.

Then $H_0: \theta_1 = 10, \theta_2 = 1$; is a simple hypothesis whereas

$H_0: \theta_1 = 10; H_0: \theta_1 = 10, \theta_2 > 1$

$H_0: \theta_1 = 10;, \theta_2 > 1 \; and \; H_0: \theta_2 = 1 \; composite$

Two Kinds of Errors: While taking decisions we may commit two types of errors:

Type I Error: We may reject H_0 When H_o is true.

Type II Error: We may accept H_o when H_1 is true.

Types of Error

Decision \ Truth	H_o	H_1
Accept H_o	True decision	Type II error
Reject H_1	Type I error	True decision

$$\alpha = P\,(type\ I\ error) = P\,(Reject\ H_0\ when\ H_0\ is\ true) = P\,(RejH_0 \mid H_0)$$
$$\beta = P\,(Type\ II\ error) = P\,(Acc\ H_0 \mid H_1)$$

Type I Error is consider to be more serious error. It is like “Hang a person who is innocent” This is more serious error than ‘Let a criminal be set free’

In Statistical inference, our concern is to minimize both α and β. But minimization of α results in maximization of β and vice versa. That is it is not possible to minimize both α and β simultaneously. Therefore, we fix α and minimize β.

The maximum value of α which we fix is called Level of Significance of size of the test.

Critical Region (CR): Critical region is that part of the sample space which corresponds to the rejection of null hypothesis.

Let w be critical region or rejection region, then $\overline{w}$ is acceptance region.

That is if we take a random sample $\underline{X} = (X_1, X_2, \ldots\ldots X_n)'\, from\, f(x, \theta), then$

$Reject\ H_0\ if\ x\ \in\ \ w$

$Accept\ H_0\ if\ x\ \in\ \ w$

Power Function ($P_\theta(w)$):

$$P_\theta(w) = \ P\ (Reject\ H_0\ if\ \theta\ is\ the\ true\ value)$$
$$= P\ (Rej\ H_0|\theta)$$
$$= P(\underline{X}\ \in w\mid\theta)$$
$$=\ \alpha\ if\ \theta\ \in\ H_0$$
$$= 1 -\ \beta\ if\ \theta\ \in\ H_1$$

α is the size of the test or size of the critical region and $(1 -\ \beta)$ is power of the test. A curve between $\theta\ and\ P_0(w)$ is called power curve.

7.4 Some Examples

Example 1: Let the $pdf\ of\ a\ rv\ X\ be$

$$f(x,\theta) = \begin{cases} \dfrac{1}{\theta}, & o < x < \theta \\ 0, & elsewhere \end{cases}$$

[This is called Uniform or Rectangular Distribution and is dented by U(0, θ) or R (0, θ)].

Let the hypothesis to be tested be

$H_0 : \theta = 2 \quad vs\ \ H_1 : \theta = 3$

On the basis of a single observation obtain

(i) Power function $P_\theta(w)$ of the test.

(ii) α, the size of the test and

(iii) $1 -\ \beta\ the\ power\ of\ the\ test$

If critical region be w = {x : x >1}

Solution: We have

Power of the test

$$P_\theta(w) = \ P\ (Rej\ H_0|\theta)$$
$$= P(x\ \in w\mid\theta)$$

$$= \int_{w}^{\theta} f\ (x, \theta) dx$$

$$= \int_{1}^{\theta} \frac{1}{\theta} dx = \frac{\theta - 1}{\theta}$$

$$\alpha = P(X\ \in w \mid \theta \in\ H_0) = \frac{2-1}{2} = \frac{1}{2} = 0.5 \quad \textit{is size of the test}$$

$$1 - \beta =\ P(X\ \in w \mid \theta \in\ H_1) = \frac{3-1}{3} = \frac{2}{3} \quad \textit{is power of the test}$$

$$\beta = \frac{1}{3}$$

Example 2: Let for the Example 1 considered above, we have to test.

$$H_0 : \theta = 2 \quad vs\ \ H_1 : \theta = 3$$

If the size of the test be α = 0.25 find k is the critical region be w ={x : x >k}.

Solution: We have,

$$P(X\ \in w \mid \theta) = \int_{k}^{\theta} \frac{1}{\theta} dx = \frac{\theta - k}{\theta}$$

$$\alpha = P(X\ \in w \mid\ H_0) = \frac{2-k}{2}$$

$$0.25 = \frac{2-k}{2}$$

$$k = \frac{3}{2}$$

Thus critical region is w = $\left\{x > \frac{3}{2}\right\}$.

Example 3: Let for an exponential distribution

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & o < x < \theta \\ 0, & elsewhere \end{cases}$$

The hypothesis to be tested by

(a) $H_0 : \theta = 1 \quad vs\ \ H_1 : \theta = 2\ with\ CR\ w = \{x : x > 1\}$.

(b) $H_0 : \theta = 2 \quad vs\ \ H_1 : \theta = 1\ with\ CR\ w = \{x : x < 5\}$.

Obtain α, β and power function of the test on the basis of a single observation.

Solution:

(a) Power function $P_\theta(w) = P\,(Rej\; H_0|\theta)$

$$= P(x \in w \mid \theta)$$

$$= \int_1^\infty \frac{1}{\theta} e^{-\left(\frac{x}{\theta}\right)} dx = e^{-\frac{1}{\theta}}$$

$\alpha = P(X \in w \mid H_0) = e^{-1}$

$1 - \beta = P(X \in w \mid H_1) = e^{-1/2}$

$\beta = 1 - e^{-1/2}$

$(\boldsymbol{b})$ $P_\theta(w) = P(x \in w \mid \theta)$

$$= \int_0^5 \frac{1}{\theta} e^{-x/\theta}\, dx = 1 - e^{-5/\theta}$$

$\alpha = P(X \in w \mid H_0) = 1 - e^{-5/2}$

$1 - \beta = P(X \in w \mid H_1) = 1 - e^{-5}$

$\beta = e^{-5}$

Example 4: For an exponential distribution

$$f(x,\theta) = \begin{cases} \theta\, e^{-\theta x}, & x,\theta \geq 0 \\ 0, & elsewhere \end{cases}$$

Let the hypothesis to be tested be

(a) $H_0 : \theta = 1$ vs $H_1 : \theta = 2$ $with\; CR\; w = \{x : x < 10\}$.

(b) $H_0 : \theta = 2$ vs $H_1 : \theta = 1$ $with\; CR\; w = \{x : x > 1\}$.

Obtain α, β and power function of the test on the basis of a single observation.

Solution:

(a) Power function $P_\theta(w) = P\,(Rej\; H_0|\theta) = P(x \in w \mid \theta) = \int_0^{10} \theta e^{-\theta x} dx = 1 - e^{-10\theta}$

$$\alpha = P(X \in w \mid H_0) = 1 - e^{-10}$$

$$1-\beta = P(X \in w \mid H_1) = e^{-20}$$

$$\beta = 1 - e^{-20}$$

$(\boldsymbol{b})$ $$P_\theta(w) = P(x \in w \mid \theta) = \int_0^\infty \theta e^{-x\theta}\, dx = e^{-\theta}$$

$$\alpha = P(X \in w \mid H_0) = e^{-2}$$

$$1-\beta = P(X \in w \mid H_1) = e^{-1}$$

$$\beta = 1 - e^{-1}$$

Example 5: Let p be the probability of getting head in a coin tossing experiment. Suppose that the hypothesis

$H_0 : p = 0.3 \quad vs \;\; H_1 : p = 0.5$

Is rejected if 10 trials result in 6 or more heads. Calculate the probability of Type I error and Type II error.

Solution: Let X be the number of heads obtained in ten trials.

$\alpha = P[\,Type\ I\ error]$

$= P[X \geq 6 | p = 0.3]$

$$= \sum_{x=6}^{10} \binom{10}{x} (0.3)^x (0.7)^{10-x} = \sum_{x=0}^{5} \binom{10}{x} (0.3)^x (0.7)^{10-x}$$

$= 1 - 0.9527 = 0.0473.$

$\beta = P[Type\ II\ error]$

$= P[X < 6 | p = 0.5]$

$$= \sum_{x=0}^{5} \binom{10}{x} (0.5)^{10} = 0.6230$$

Example 6: An urn contains θ whine and (6- θ) black marbles. In order to test the hypothesis

$$H_0 : \theta = 3 \quad vs \quad H_1 : \theta = 4$$

Two marbles are drawn (without replacement) and H_0 is rejected if both the marbles are white. Calculate $and\ \beta$.

Solution: Let X be the number of white marbles drawn

$$P(X = x) = \frac{\binom{\theta}{x}\binom{6-\theta}{2-x}}{\binom{6}{2}}$$

Therefore,

$$\alpha = P[\ Rej\ H_0|\ H_0]$$

$$P(X = 2|\theta = 3) = \frac{\binom{3}{2}\binom{3}{0}}{\binom{6}{2}} = \frac{1}{5} = 0.2.$$

$$1 - \beta = P[\ Rej\ H_0|\ H_0] = \frac{\binom{4}{2}\binom{2}{0}}{\binom{6}{2}} = \frac{2}{5}$$

$$\beta = \frac{3}{5}$$

Example 7: Let p the probability that a given die shows even number. To test

$$H_0 : p = \frac{1}{2} \quad vs \quad H_1 : p = \frac{1}{3}$$

a die is thrown twice and H0 is accepted if even number is obtained on both the throws. Calculate probability of Type I error and Type II error.

Solution:

$$\alpha = P[\ Rej\ H_0|\ H_0] = 1 - P[\ Acc\ H_0|\ H_0] = 1 - \left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) = \tfrac{3}{4}$$

$$\beta = P[\ Acc\ H_0|\ H_1] = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}.$$

Example 8: The hypothesis $H_0: \mu = 50$ is rejected if mean of the sample of size 25 is either greater than 70 or less than 30. Assuming the distribution to be normal with standard deviation (σ) = 50, obtain level of significance.

Solution: Level of significance is $\alpha = P[\ Rej\ H_0 |\ H_0]$

Here under H_0, X is distributed as N (50, 2500) and $\bar{X}$ is distributed as N (50, 100)

That is, $Z = \frac{\bar{X}-50}{10}$ is distributed as N (0,1)

Therefore, $\alpha = P[\ \bar{X} \geq 70\ or\ \bar{X} \leq 30 |\ H_0]$

$$= 1 - P[\ 30 < \bar{X} < 70 |\ H_0]$$

$$= 1 - P\left[\frac{30-50}{10} < Z < \frac{70-50}{10}\right]$$

$$= 1 - P[-2 < Z < 2]$$

$$= 1 - 2 \times 0.4772 = 0.0456.$$

Example 9: Let X_1, ……X_9 be a random sample from normal N (θ,25) distribution. If for testing.

$$H_0 : \theta = 20 \quad vs \quad H_1 : \theta = 26$$

The critical region be w = {x | $\bar{x}$ >24}, find the size and power of the test.

Solution:

$$P_\theta(w) = [\ Rej\ H_0 |\ H_0] = P\ [\bar{X} > |\theta]$$

$$\alpha = P\ [\bar{X} > 24 | \theta = 20] = P\left[Z > \frac{(24-20)3}{5}\right] = P[Z > 2.4] = 0.0082$$

$$1 - \beta = P\ [\bar{X} > 24 | \theta = 26] = P\left[Z > \frac{(24-26)3}{5}\right] = P[Z > -1.2] = 0.8849$$

$$\beta = 0.1151.$$

7.5 Problems and Exercises

1. Explain the terms: null hypothesis, alternative hypothesis, simple and composite hypothesis by giving examples.

2. Define the terms: type I error, type II error, power function, size of the test, level of significance and power of the test.

3. Let p be the probability of coming up head in a coin tossing experiment. To test the hypothesis

$$H_0 : \frac{3}{4} \quad vs \quad H_1 : \frac{1}{2}$$

a coin is tossed 5 times and H_0 is rejected if more than 3 tails are obtained. Calculate size and power of the test.

Hint: H_0 is rejected if X 3, where X is number of 'Heads'.

$$\alpha = P[X \leq 3|H_0] = \sum_{x=0}^{3} \binom{5}{x} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{5-x}$$

$$1 - \beta = P[X \leq 3|H_1] = \sum_{x=0}^{3} \binom{5}{x} \left(\frac{1}{2}\right)^5$$

4. Let for a rv X with pdf

$$f\ (x;\theta) = \ \theta\, x^0, 0 < x < 1.$$

The hypothesis to be tested be

$$H_0 : \theta = 2$$

$$H_1 : \theta > 2.$$

If the CR based on a single sample be x > 9/19. Obtain size of the test and also power of the test if under $H_1 : \theta = 2$ (Ans. $\alpha = 0.19, \beta = 0.729$).

5. It is given that scores in a test follow normal N (μ,144) distribution. To test

$$H_0 : \mu = 500, \ \ H_1 : \mu \neq 500$$

a random sample of size 36 is taken and H_0 is rejected if $|\bar{x} - 500|$ > 4. Evaluate α.

Hint:

$$\alpha = P[|\bar{X} - 500| > 4/\ \mu = 500] = P\left[\frac{|\bar{X} - 500|}{2} > \frac{5}{2}/\mu = 500\right] = P[|Z| > 2] = 0.0456.$$

6. To test $H_0 : \mu = 200, H_1 : \mu > 200$ a random sample of size 20 is drawn from a normal population N (μ, σ^2) with σ^2 = 80. Ho is accepted if x ≤ 204.

i) Evaluate α and ii) if μ = 207 under H_1.

(Ans: $\alpha = 0.0228, \quad \beta = 0.0668$).

7. A random sample of size 80 is drawn from a normal N (μ, σ^2) distribution with σ^2 = 720. In testing

$$H_0 : \mu = 1000 \quad vs \quad H_1 : \mu < 1000,$$

H_o is accepted if $\bar{x}$ > 994. Obtain α *and* β if μ = 991 under H_1.

(Ans. $\alpha = 0.0228, \beta = 0.1587$).

8. To test the mean $H_0 : \mu = 100 \quad vs \quad H_1 : \mu \neq 100$ of a normal N $(\mu, 200)$ distribution a random sample of size 50 is drawn. If the null hypothesis is rejected when $|\bar{x}$ -100| > k at = 0.05, obtain k. (Ans. k = 3.92).

9. A random sample is taken from binomial B (5,p) distribution, to test $H_0 : p = 0.5 \ against \ H_1 : p = 0.7$ if H_0 is rejected if X ≥ 3, find α *and* β.

(Ans. $\alpha = 0.5, \beta = 0.1631$).

10. A random sample of size 5 is taken from a Poisson distribution P(λ), to test H_0 : λ =1 against H_1 : λ = 2. If it is decided to reject H_0 if find $\sum_{i=1}^{5} X_i \geq 8, \alpha$ *and* β.

(Ans. $\alpha = 1334, \ \beta = 0.2202$).

7.6 Summary

Statistical Hypothesis is an assertion about probability density function/ probability mass function or its parameters. We have two types of hypothesis.

(i) Null Hypothesis (ii) Composite Hypothesis

A hypothesis under test is called a Null Hypothesis and is denoted by H_O. A hypothesis under consideration is called an Alternative Hypothesis and is denoted by H_1.

If all the parameters of a distribution are completely specified then it is simple otherwise composite hypothesis.

Another interesting concept is two types of error. These arise while taking decisions. Type I error is probability of rejecting H_0 when H_0 is true, whereas, Type II error is probability of accepting H_0 when H_1 is true. Type I error is denoted by α and Type II error is denoted by β. The quantity 1- β is called the power of the test.

Another important concept is Critical region. It is that part of the sample space which corresponds to the rejection of null hypothesis.

7.7 Further Readings

- Goon, A.M., Gupta, M.K., and Dasgupta, B. (2000). *An outline of statistical Theory*, Vol. 2, The world Press Private Limited.
- Hogg, R.V. and Craig, A. (2005). *Introduction to Mathematical Statitics*, 6th edition, Prentice Hall.
- Modd, A.M. Graybill, F.A.m Boes, D.C. (1974). *Introduction to the Theory of Statistics*, McGraw Hill
- Mohr, L.B. (1994). *Understanding Significance Testing*, Sage Publications, USA.

Unit-8 MP and UMP Tests

Structure

8.1 Introduction

The discussion in the previous unit was aimed to provided some preliminary concepts in testing of statistical hypothesis. In the same very spirit, we have discussed the notion of two kinds of hypothesis, level of significance, size power, two kinds of errors, etc. We shall now provide two very important concepts, i.e., most powerful and uniformly most powerful tests.

Let X_1, X_2,……..X_n be random sample from the probability density (mass) function f(x,) or f(x,). We can say that X_1, X_2,……,X_n is a random sample from one or other member of the parametric family $\{f\ (x,\theta): \theta = \theta_0\ or\ \theta = \theta_1\}$. Thus , $\Omega = \{\theta_0, \theta_1\}$ is a parameter space that consists of only two points $\theta_0\ and\ \theta_1$. Suppose the values $\theta_0\ and\ \theta_1$ are known. We want to test $H_0: \theta = \theta_0$versus $H_1: \theta = \theta_1$. We know that corresponding to any test say ɼ of H_0 versus H_1, there exists a power function, say $P_\theta(\omega)$where ω is the critical region. A good test is a test for which $P_{\theta_0}(\omega)$ = Type I error = P (reject H_0 when H_0 is true) is small (ideally zero) and $P_{\theta_1}(\omega)$ = Power = P (reject H_0 when H_1 is true) is large (ideally unity). One might reasonably use the two values $P_{\theta_0}(\omega)$ and $P_{\theta_1}(\omega)$ to set up criteria for defining a best test. It is to be noted that $P_{\theta_0}(\omega)$ is size of Type I error whereas 1- $P_{\theta_1}(\omega)$ is the size of Type II error; and, therefore, our goodness criterion might desire making power of the two errors as small as possible and consequently making the power of the test as large as possible. For example, one might define as best test the one which makes sum of two errors sizes as small as possible. There may be several ways of defining best test. We shall provide some such criteria in the following sections. For simplicity, we shall assume that the indexing parameter $\theta$is single valued. Say for instance, Poisson P(λ) distribution has a single parameter λ. Normal N(μ, σ^2) distribution has two parameters μ and σ^2. However if one of the parameters is known, it reduces to single parameter case.

The justification for fixing the size of the Type I error (usually small and often taken as 0.05 or 0.01) seems to arise from those testing situations where the two hypotheses are formulated in such a way that one type of error is more serious than the other. The hypotheses are stated so that the Type I error is more serious and hence one wants to make sure that it is kept small.

8.2 Most Powerful Test

Let the hypothesis to be tested be

$$H_0: \theta = \theta_0$$

$$H_1: \theta = \theta_1$$

Where both H_0 and H_1 are simple hypothesis. Since defining a test is same as defining a critical region, we start with the following simple statement.

Suppose given a sample X_1, X_2, X_3 a test for H_o is define as Reject H_0 if $x_1^2 + x_2^2 + x_3^2 \leq 5$. This statement is equivalent to say that given a sample X_1, X_2, X_3, the test is reject H_o if X* = (X_1, X_2, X_3) $\in \omega$ where $\omega = \{(x_1, x_2, x_3): x_1^2 + x_2^2 + x_3^2 \leq 5\}$ where small font is as usual denotes the observed value.

For further illustration, suppose that Kamal an Uni both have tests of size α. This means by definition, that they each have a set ωk and ωu (K= Kamal, U= Uni) such that

$$P\big((X_1, X_2, \ldots . X_n) \in \ \omega k | H_0\big) = \alpha$$

$$P\big((X_1, X_2, \ldots . X_n) \in \ \omega u | H_0\big) = \alpha$$

To find out the better test between these two if we look at the probabilities that the sample $X_1, X_2, \ldots . X_n$ is not in the respective sets we won't get an answer since this will be (1- α) in either case. The only way to differentiate between the tests will be in the case that the alternative hypothesis is true. So we will consider the probabilities.

$$P\big((X_1, X_2, \ldots . X_n) \in \ \omega k | H_1\big)$$
$$and$$
$$P\big((X_1, X_2, \ldots . X_n) \in \ \omega u | H_1\big)$$

Further suppose that

$$P\big((X_1, X_2, \ldots . X_n) \in \ \omega k | H_1\big) > \ P\big((X_1, X_2, \ldots . X_n) \in \ \omega u | H_1\big)$$

Thus the question is to find out a better test we need to know what exactly we are doing. In fact, we are looking at probabilities that the sample falls in a critical region when the alternative hypothesis H_1 is true. These probabilities are nothing but the power of the tests associated with the two critical regions. Since from the above relation we find that Kamal has a higher probability of rejecting H_0 and H_1 is true than the same of Uni. That is the power of the test considered by Kamal is better than that of Uni and therefore we can say that Kamal's test is better. If it found that the above inequality holds with Kamal's ωk in the left hand side and any other critical region of size α in the right hand side then Kamal's test will be the best and it will be referred to as the MP test. Following the above discussion, one can formulate the MP test in the following manner.

Take a random sample $X_1, X_2, \ldots . X_n$ of size n from $f\ (x, \theta)$. Let w be a critical region of size α, w is called most powerful critical region (MPCR) of size α if the power corresponding to CR w is greater than the power corresponding to w*, where w* is any other CR of size α.

That is,

(i) $$P[X \in w|H_0] = \alpha = P[X \in\ w^*|H_o]$$

(ii) $$P[X \in w|H_1] \geq P[X \in\ w^*|H_1]$$

Or

(i) $$P_{\theta_0}(w) = \alpha = P_{\theta_0}(w^*)$$

(ii) $$P_{\theta_1}(w) \geq P_{\theta_1}(w^*)$$

A test based on MPCR is called MP test. From the above discussion you may have learnt that for testing a simple hypothesis against a simple alternative hypothesis, it is easy to choose the MP test among the given tests. You may be wondering at this stage that how to be get a particular test (namely the MP test) such that exists no other test which has greater power than this. There are procedures to get the MP tests which will be discussed in other unit.

8.3 Uniformly Most Powerful Test

Let us now consider a situation where a simple hypothesis $H_0: \theta = \theta_0$ is to be tested against a composite alternative $H_1: \theta > \theta_0$ Before commenting on such a testing scenario, we instead consider two points $\theta_1\ and\ \theta_2$ where both these points fall in the region specified by H_1 and decide to begin with $H_0: \theta = \theta_0$ versus $H_1^{'}: \theta = \theta_1$ and $H_0: \theta = \theta_1$ versus $H_1: \theta = \theta_2$. Thus we have two testing problems, each concerned with testing of simple null versus simple alternative. We can therefore, proceed to obtain MP test in the way described in section 1.2 for each of the two pairs of hypotheses.

Consider now $H_0: \theta = \theta_0$ versus $H_1^{'}: \theta = \theta_1$ and suppose ω is a critical region of size α such that ω is MPCR and the test based on ω is MP test. Further suppose that same happens to offer MPCR for $H_0: \theta = \theta_0$versus $H_1^{'}: \theta = \theta_1$. If this is true we may use the same ω and obtain a single MP test for testing $H_0: \theta = \theta_0$ against either of two alternatives. Now extend this discussion for a series of points in the region specified by H_1 and suppose every time our null hypotheses is $H_0: \theta = \theta_0$. If the same ω offer MPCR for a series of alternative hypothesis against the same null hypothesis, we may call the MPCR as uniformly most powerful critical region (UMPCR) and the test based on it will be uniformly most powerful (UMP) test. The above discussion can be extended similarly if the two hypotheses are $H_0: \theta = \theta_0\ versus\ H_1: \partial < \theta_0\ or\ H_0: \theta = \theta_0\ versus\ H_1: \theta \neq \theta_0$ although the two sided case requires some additional arguments.

We shall formalize our definition of UMPCR and UMP test. Let the hypotheses be $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$. Then a critical region *w* is UMPCR of size α if

$$(i) \qquad P[X \in w|H_0] = \alpha = P[X \in\ w^*|H_o]$$

$$(ii) \qquad P[X \in w|H_1] \geq P[X \in\ w^*|H_1]$$

Or

$$(i) \qquad P_{\theta_0}(w) = \alpha = P_{\theta_0}(w^*)$$

$$(ii) \qquad P_{\theta_1}(w) \geq P_{\theta_1}(w^*)$$

Here under H_1, θ may have many values and conditions (ii) for all $\theta \in H_1$ should be satisfied with strict inequality for at least one θ.

A test on *UMPCR* is called *UMP* test. The major difference between MP and UMP test is in *MP* test H_0 and H_1 both are simple whereas, in *UMP* test H_0 is simple but H_1 is composite.

Remark: For every MP or UMPCR the power of the test is always greater than its size. The test for which the power is greater than size is known as an unbiased test. This way every MP or UMP critical region is necessarily unbiased.

8.4 Summary

A critical region for testing a simple hypothesis against a simple alternative is said to be most powerful if it is of size and the power corresponding to this critical region is greater than the power corresponding to any other critical region of the same size. The test based on most powerful critical region is called most powerful test.

`The simple vs. simple case is not of much practical relevance from statistical point of view. Thus we define another important concept.

A critical region which is most powerful for a series of alternative hypothesis when tested against the same null hypothesis is called uniformly most powerful critical region. A test based on uniformly most powerful critical region is called uniformly most powerful test. Such tests are usually meant for composite alternative hypothesis when tested against the simple null hypotheses.

Uniformly most powerful tests do not exist, in general except for a few restrictive scenarios. These restrictive families are called monotone likelihood ratio families for one-sided problems. For two sided problems, the situation becomes even worse and a particular family can be one parameter exponential where existence of uniformly most powerful tests can be ascertained.

8.5 Further Readings

- Goon, A.M., Gupta, M.K., and Dasgupta, B. (2000). *An Outline of Statistical Theory*, Vol. 2, The world Press Private Limited.
- Hogg, R.V. and Craig, A. (2005). *Introduction to Mathematical Statistics,* 6th edition, Prentice Hall.
- Mood, A.M. Graybill, F.A. Boes, D.C. (1974). *Introduction to the Theory of Statistics,* McGraw Hill
- Mohr, L.B. (1994). *Understanding Significance Testing,* Sage Publications, USA.
- Lehmann, E.L. (1986). *Testing Statistical Hypothesis*. Springer-Verlag, New York, Inc. 2nd edition.

U.P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105
Advance Statistical Inference

Block: 4 Testing of Hypothesis - II

Unit – 9 : Neyman – Pearson Lemma, Likelihood Ratio Test and Their Uses

Unit – 10 : Testing of Means of Normal Population

Unit – 11 : Interval Estimation

Unit – 12 : Shortest and Shortest Unbiased Confidence Intervals

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Department of Statistics, University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Department of Statistics, University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur

Dr. Shruti **Member-Secretary**
School of Sciences, U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Prof. S. K. Pandey **Writer**
Department of Statistics, Lucknow University, Lucknow

Prof. A. H. Khan **Writer**
Department of Statistics, Aligarh Muslim University, Aligarh

Dr. Shruti **Writer**
School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Umesh Singh **Reviewer**
Department of Statistics, Banaras Hindu University, Varanasi

Prof. V. P. Ojha **Reviewer**
Department of Statistics and Mathematics, D. D. U., Gorakhpur University, Gorakhpur

Prof. B. P. Singh **Editor**
Department of Statistics, Banaras Hindu University

Dr. Shruti **Course/ SLM Coordinator**
School of Sciences, U. P. Rajarshi Tandon Open University, Prayagraj

DECSTAT – 105 ADVANCE STATISTICAL INFERENCES
©UPRTOU
First Edition: *March 2008* (Published with the support of the Distance Education Council, New Delhi)
Second Edition: ***July 2021***
ISBN : 978-93-94487-38-3

Printed By: K.C. Printing & Allied Works, Panchwati, Mathura - 281003.

Blocks & Units Introduction

The ***Block - 4 – Testing of Hypothesis - II*** based on testing of hypothesis and interval estimation consists of four units.

Unit – 9 – Neyman – Pearson Lemma, Likelihood Ratio Test and Their Uses, describes Neyman Pearson Lemma and likelihood ratio tests along with their uses in determination of test.

Unit – 10 – Testing of Means of Normal Population, discuss tests for significance of mean from a normal population and testing the equality of means from two independent normal populations

Unit – 11 – Interval Estimation, defines interval estimation for single unknown parameter of univariate population. Confidence intervals have been given for parameters of univariate normal population and one parameter exponential family.

Unit – 12 – Shortest and Shortest Unbiased Confidence Intervals, provides the concept of shortest and shortest unbiased confidence intervals.

At the end of every unit the summary, self assessment questions and further readings are given.

Unit-9 Neyman- Pearson Lemma, Likelihood Ratio Test and Their Uses

Structure

9.1 Introduction

In previous Block 3, we introduced the concept of most powerful and uniformly most powerful tests. In this section, we shall develop methods to obtain these tests. It is important to note that Neyman-Pearson lemma provides a procedure for getting a Most Powerful tests and thus, in its turn, it can be used for developing UMP test provided it exists. In addition to this, the use of likelihood ratio test will also be discussed for obtaining UMP tests.

9.2 Objectives

After reading this unit you should be able to:

- Understand Neyman-Pearson fundamental lemma.
- Derive most-powerful test for simple Vs. simple hypothesis
- Understand likelihood Ratio test and its uses in testing of hypothesis.

9.3 Neyman- Pearson Lemma

Neyman and Pearson gave a simple rule, known as their Lemma to obtain MP test:

Let X has pdf/pmf f (x, θ) where θ is a single parameter. Take a random sample X_1, X_2, …..X_n of size n from f(x,θ). Define likelihood function as

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

Lemma: Let for f(x, θ) the simple hypothesis to be tested be

$$H_0: \theta = \theta_0$$

$$H_1: \theta = \theta_1$$

Then w, a CR of size α is MPCR of size α if

$$Inside\ w: \ L_1 \geq k\ L_0$$

$$and\ Outside\ w: \ L_1 < k\ L_0$$

Where L_0 and L_1 are likelihood functions of the sample observations under H_0 and H_1 respectively and k is such that size of the CR w is α.

Proof: Since w is CR of size α,

$$\alpha = P\ (Type\ I\ error)$$

$$= P\ (Reject\ H_0|H_0)$$

$$= P\ (X\ \in w|H_0)$$

$$= \int_w L_0 dx \qquad (1)$$

And

$$1 - \beta = P\ (Reject\ H_0|H_1)$$

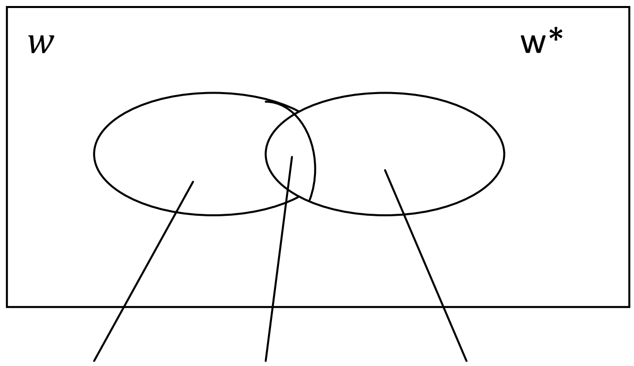
$$= P\ (X\ \in w|H_1)$$

$$= \int_w L_1 dx \qquad (2)$$

Let w* be another CR of size α, then

$$\alpha = \int_{w^*} L_0 dx \qquad (3)$$

If we can show that power corresponding to w is greater than that of w* then w is MPCR. That is, we have to prove that



$$\int_{w} L_1 dx \geq \int_{w^*} L_0 dx \qquad (4)$$

To prove (4), we note that

$w = w\,\bar{w}^* \cup ww^*$

$and\ w^* = \bar{w}w^* \cup ww^*$

$where\ w\,\bar{w}^* \subset w\ and\ \bar{w}w^* \not\subset w$

$where\ w\,\bar{w}^* and\ ww^*\ \ \bar{w}w^*$

Therefore, from (1) and (3)

$$\int_{w} L_0 dx = \int_{w^*} L_0 dx$$

Or

$$\int_{w\bar{w}^* \cup ww^*} L_0 dx = \int_{\bar{w}w^* \cup ww^*} L_0 dx$$

Or,

$$\int_{w\bar{w}^*} L_0 dx + \int_{ww^*} L_0 dx = \int_{\bar{w}w^*} L_0 dx + \int_{ww^*} L_0 dx$$

Or,

$$\int_{w\bar{w}^*} L_0 dx = \int_{\bar{w}w^*} L_0 dx \qquad \text{.... (5)}$$

Similarly,

$$\int_{w} L_1 dx - \int_{w^*} L_1 dx = \int_{w\bar{w}^*} L_1 dx - \int_{\bar{w}w^*} L_1 dx \qquad \text{....... (6)}$$

But inside $w: L_1 \geq kL_0$

and outside $w: L_1 < kL_0$

Therefore in $ww^* \subset w,$ we have

$$\int_{w\bar{w}^*} L_1 dx \geq k \int_{\bar{w}w^*} L_0 dx \qquad \text{...... (7)}$$

and in $ww^* \not\subset w,$ we have

$$\int_{w\bar{w}^*} L_1 dx < k \int_{\bar{w}w^*} L_0 dx \qquad \text{........(8)}$$

Hence in view of (7) and (8)

$$\int_{w} L_1 dx - \int_{w^*} L_1 dx \geq k \int_{w\bar{w}^*} L_0 dx - k \int_{\bar{w}w^*} L_0 dx$$

$$\geq k \left(\int_{w\bar{w}^*} L_0 dx - k \int_{\bar{w}w^*} L_0 dx \right) \geq 0 \; from \; (5)$$

That is

$$\int_{w} L_1 dx \geq \int_{w^*} L_1 dx$$

This completes proof of the Lemma.

This Lemma gives a very simple and powerful method to obtain MP test. It also helps in finding UMP test in some cases.

For example, if $X \sim f(x, \theta)$

$$I: \quad H_0: \theta = \theta_0 \qquad Simple$$

$$H_1: \theta > \theta_0 \qquad Composite$$

Then we convert δ both to the simple hypothesis.

$$II: \quad H_0: \theta = \theta_0$$

$$H_1: \theta = \theta_1 (\theta_1 > \theta_0)$$

Now we use Neyman Pearson Lemma for II to find MPCR. If this CR does not depend on θ_1 the MPCR of II will be UMPCR for I. This statement will be elaborated by examples in sequel.

9.4 Examples on Neyman-Pearson Lemma

Example 1: Let $X_1, X_2, \ldots\ldots X_n$ be Bernoulli variates with *pmf*

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

Find MP test of size α for

(a) $\quad H_0: \theta = \theta_0$

$\quad H_1: \theta = \theta_1 (> \theta_0)$

(b) $\quad H_1: \theta = \theta_1$

$\quad H_1: \theta = \theta_1 (< \theta_0)$

Solution:

(a) We are given

$$P\,(X = x) = f(x, \theta) = \theta^x (1-\theta)^{1-x}$$

Therefore,

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

From Neyman-Pearson (NP) Lemma,

Inside $\quad w: L_1 \geq kL_0$

$$\frac{\theta_1 \sum_{i=1}^{n} x_i \,(1-\theta_1)^{n-\sum_{i=1}^{n} x_i}}{\theta_0 \sum_{i=1}^{n} x_i \,(1-\theta_0)^{n-\sum_{i=1}^{n} x_i}} \geq k$$

$$\left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right]^{\sum_{i=1}^n x_i} \geq k \qquad (*)$$

$$or, \left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right]^{\sum_{i=1}^n x_i} \geq k_1$$

Here in (a), $\theta_1 > \theta_0$

Therefore, $\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} > 1$

And $log\ \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} > 0$

Taking log of both sides in (*), we have

$$\sum_{i=1}^{n} x_i\ log\ \left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right] \geq k_2$$

$$or, \sum_{i=1}^{n} x_i \geq\ k_3\ as\ log\ \left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right] \geq 0.$$

$$Here\ k_1 = \frac{k}{\left(\frac{1-\theta_1}{1-\theta_0}\right)^n}, k_2 = \log k_1\ and\ \ k_3 = \frac{k_2}{log\,\frac{(1-\theta_0)\theta_1}{(1-\theta_1)\theta_0}}$$

But it is not needed to keep these records. It is enough to note that $y = \sum_{i=1}^n x_i \geq k_3$ and thus MPCR of size α is

$$w = [y \geq k_3],$$

Where k_3 is so chosen that

$$\alpha = P[y \geq k_3 | H_0].$$

It may be noted that if x_i are independent Bernoulli variates B(1, θ) then

$Y = \sum_{i=1}^n x_i$ is distributed as Binomial B(n, θ) and therefore,

$$\alpha = \sum_{k_3}^{n} \binom{n}{y} \theta_0^y (1-\theta_0)^{n-y} \qquad (**)$$

$$or, 1-\alpha = \sum_{0}^{k_3-1} \binom{n}{y} \theta_0^y (1-\theta_0)^{n-y}.$$

Since this is a discrete distribution, we may or may not get k_3 corresponding to which sum is α in (**). In that case we choose nearest k_3 for which

$$\sum_{0}^{k_3-1} \binom{n}{y} \theta_0^y (1-\theta_0)^{n-y} \geq 1-\alpha.$$

Numerical Example: From the table n= 10, θ_0 = 0.30 and α = 0.05, we have k_3 = 6

CR is {y ≥ 6} i.e. Reject H_0 if $\sum X_i \geq 6$ otherwise accept H_0.

(b) Here $H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (< \theta_0)$

$Since\ \theta_1 < \theta_1, therefore,$

$$log\ \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} < 0$$

Proceeding as above in (a), we get MPCR from (*),

$$\mathrm{w} = [\mathrm{y} \leq \mathrm{k}_3], \mathrm{y} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x}_{\mathrm{i}}$$

where k_3 is obtained as

$$\alpha = P[\mathrm{y} \leq \mathrm{k}_3 | \mathrm{H}_0]$$

$$= \sum_{0}^{k_3-1} \binom{n}{y} \theta_0^y (1-\theta_0)^{n-y}$$

Here also, as explained earlier we may choose k_3 such that

$$\sum_{0}^{k_3-1} \binom{n}{y} \theta_0^y (1-\theta_0)^{n-y} \leq \alpha.$$

Numerical Example: From the table n= 10, θ_0 = 0.30 and α = 0.05, we have k_3 = 3

CR is $\{y \leq 3\}$.

Remark: In the example, we have obtained MPCR $w = \{\sum_{i=1}^{n} x_i \geq 3\}$ of size α for testing

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (> \theta_0)$

Since this CR does not depend on θ_1, hence $w = \{\sum_{i=1}^{n} x_i \geq 3\}$ is also UMPCR for testing

$H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

similarly

$$w = \left\{\sum_{i=1}^{n} x_i \geq 3\right\} \text{ is UMPCR is size } \alpha \text{for testing}$$

$H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$

Since the UMPCR obtained for testing

$H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

and for

$H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$

are different, hence there does not exist UMPCR for tesing

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

Example 2: Let X has Poisson distribution $P(\theta)$ with pmf

$$f(x, \theta) = \frac{e^{-\theta} \theta^x}{x!}, x = 0,1,2, \dots \dots$$

$Find\ MPCR\ of\ size\ \alpha\ for\ testing$

$(\boldsymbol{a})\ \ H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (> \theta_0)$

and

$(\boldsymbol{b}) H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (< \theta_0)$

$based\ on\ a\ random\ sample\ X_1, \ldots\ldots X_n\ of\ size\ n\ from\ P(\theta)$

Solution:

(a) We are given

$$f(x,\theta) = \frac{e^{-\theta}\theta^x}{x!}$$

$$L = \prod_{i=1}^{n} f(x_i,\theta) = \frac{e^{-n\theta}\theta^{\sum_1^n x_i}}{\prod_{i=1}^n x!}$$

Therefore MPCR *w* by *NP* Lemma is:

$$Inside\ w: \frac{L_1}{L_0} \geq k$$

That is,

$$\frac{e^{-n\theta_1}(\theta_1)^{\sum_1^n x_i}}{\prod_{i=1}^n x!} \ \frac{\prod_{i=1}^n x!}{e^{-n\theta_0}(\theta_1)^{\sum_1^n x_i}} \geq k$$

$$or\ \ e^{-n(\theta_1-\theta_0)}\left(\frac{\theta_1}{\theta_0}\right)^{\sum x_i} \geq k \qquad (1)$$

$$Thus \qquad \left(\frac{\theta_1}{\theta_0}\right)^{\sum x_i} \geq k_1$$

$$or, \quad \sum_1^n x_i log\left(\frac{\theta_1}{\theta_0}\right) \geq k_2$$

$Since \qquad \theta_1 > \theta_2, \qquad \log\frac{\theta_1}{\theta_0} \geq 0,$

and

$$\sum_{1}^{n} x_1 \geq k_3$$

Hence MPCR of size α is

$$w = \left\{\sum_{i=1}^{n} x_i \geq k_3\right\}$$

$$\text{where } \alpha = P\left\{\sum_{i=1}^{n} x_i \geq k_3 | H_0\right\}.$$

It may be noted that if Xi, i= 1,…….,n are independently distributed as Poisson P(θ), then y = $\sum_{i=1}^{n} x_i$ is distributed as Poisson P(n θ).

Therefore,

$$\alpha = \sum_{y=k_3}^{\infty} \frac{e^{-n\theta_0}(n\theta_0)^y}{y!}$$

and thus for given θ_0 and α we can obtain k_3 from the table. If we do not get k_3 for which the sum is α then we take that value for which the sum is close to α but ≤ α.

$$\text{i. e.} \qquad \sum_{y=k_3}^{\infty} \frac{e^{-n\theta_0}(n\theta_0)^y}{y!} \leq \alpha \text{ or } \sum_{y=0}^{k_3-1} \frac{e^{-n\theta_0}(n\theta_0)^y}{y!} \geq 1-\alpha.$$

Numerical Example: For n=10, θ_0=1, α =0.5, it can be obtained by N-P Lemma that CR is $\{\sum_1^{10} x_i \geq 16\}$, i.e. Reject H_0 if $\sum_1^{10} x_i \geq 16$ and accept otherwise.

(b) For $\theta_1 < \theta_0$,

$$\log\frac{\theta_1}{\theta_0} < 0,$$

and consequently,

$$\sum_{i=1}^{n} x_i \le k_3 \;\; as \; \sum_{i=1}^{n} x_i \; \log\frac{\theta_1}{\theta_0} \ge k_2$$

Where k_3 is obtained from

$$\alpha = P\left\{\sum_{i=1}^{n} x_i \le k_3 | H_0\right\}.$$

$$= \sum_{y=0}^{k_3} \frac{e^{-n\theta_0}(n\theta_0)^y}{y!}$$

Numerical Example: For n=6, θ_0=2, α =0.05, CR is $\{\sum_1^6 x_i \le 6\}$, i.e. Reject H_0 if $\sum_1^6 x_i \le 6$ and accept otherwise.

Remark: In this example (Poisson distribution), we have seen that in both the cases, i.e.

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (> \theta_0)$

and

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (< \theta_0)$

MPCR of size α does not depend on θ_1 , hence MPCR of size α for testing $H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (> \theta_0)$

is also UMPCR of size α for testing

$H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

and MPCR of size α for testing
$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (< \theta_0)$

is also UMPCR of size α for testing

$H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$

Since these two UMPCR are different hence we conclude that UMPCR for testing

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

does not exist.

Example 3: Let X be distribution as normal *N(θ,1)*. Find MPCR of size α for testing

$(\boldsymbol{a})$ $H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (> \theta_0)$

and

$(\boldsymbol{b})$ $H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (< \theta_0)$

$based\ on\ a\ random\ sample\ X_1, \ldots\ldots X_n\ of\ size\ n\ from\ N(\theta, 1)$

Solution: The *pdf* of normal $N(\theta, 1)$distribution is

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

The likelihood function based on a random sample is

$$L = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_1^n (x-\theta)^2}$$

The MPCR *w* by *NP* Lemma is

$$Inside\ w: \qquad \frac{L_1}{L_0} = \frac{e^{-\frac{1}{2}\sum_1^n (x-\theta_1)^2}}{e^{-\frac{1}{2}\sum_1^n (x-\theta_0)^2}} \geq k$$

Or,

$$e^{-\frac{n}{2}\left(\theta_1^2-\theta_0^2\right)} \quad e^{(\theta_1-\theta_0)\sum x_i} \geq k$$

Or,

$$e^{(\theta_1-\theta_0)\sum x_i} \geq k_1$$

Or,

$$(\theta_1-\theta_0)\sum x_i \geq \log k_1$$

$(\boldsymbol{a})$ $\quad For(\theta_1 > \theta_0),$

$$\sum x_i \geq k_2 \quad or \quad \bar{x} \geq k_3,$$

Where k_3 is obtained by

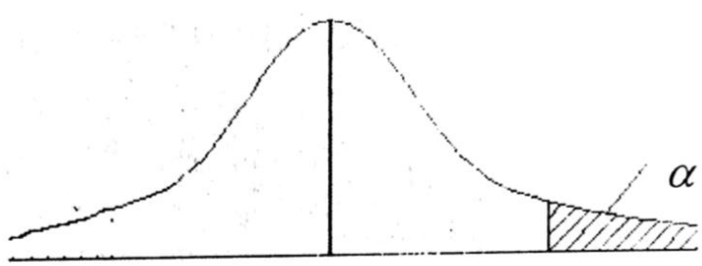
$$\alpha = P(\bar{x} \geq k_3|H_0).$$

$$Note\ that \qquad if\ X \sim N\left(\theta, \frac{1}{n}\right), then\ \bar{X} \sim N\left(\theta, \frac{1}{n}\right), and\ Z = \sqrt{n}(X-\theta) \sim N(0,1).$$

Hence

$$\alpha = P\left[Z \geq \sqrt{n}(k_3-\theta_0)\right] = P(Z \geq z_\alpha)$$

and thus for given α we can find z_α from normal table and MPC R is $\left\{x \geq \theta_0 + \frac{z_\alpha}{\sqrt{n}}\right\}$



This is also UMPCR for testing

$H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

Numerical Example: If n = 100, θ_0 = 50 and α = .05 then z_α = 1.65 and k_3 = 50+ $\frac{1.65}{10}$ = 50.165

i.e. Reject H_0 if $\bar{x} \geq 50.165$ and accept otherwise.

(b) For $H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 (< \theta_0)$

$We\ can\ show\ that\ MPCR\ of\ size\ \alpha\ is$

$w = \{\bar{x} \leq k_4\},$

$where\ k_4 = \theta_0 - \frac{z_\alpha}{\sqrt{n}}$

We conclude that $w = \{\bar{x} \leq k_4\}$ is UMPCR for testing

$H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$

Numerical Example: Let n = 100, θ_0 = 50 and α = .05 then z_α = 1.65 and k_4 = 49.835

i.e. Reject H_0 if $\bar{x} \leq 49.835$ and accept otherwise.

As noted above UMPCR of size α for testing

$H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

$is\ \left\{x \leq \theta_0 - \frac{z_\alpha}{\sqrt{n}}\right\}.$

And for

$H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$

UMPCR of size α is $\left\{\bar{x} \leq \theta_0 - \frac{z_\alpha}{\sqrt{n}}\right\}$

and since these CR are different, we conclude that UMPCR does not exist for

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0.$

Example 4: Let X be distributed as normal N (μ, σ^2) with μ known. Find MPCR of size α for testing.

$(\boldsymbol{a})$ $\quad H_0: \sigma = \sigma_0$

$H_1: \sigma = \sigma_1 (> \sigma_0).$

and

$(\boldsymbol{b})$ $\quad H_0: \sigma = \sigma_0$

$H_1: \sigma = \sigma_1 (< \sigma_0).$

based on a random $X_1, \ldots, X_n$ of size n from N (μ, σ^2) .

Solution: The pdf of Normal N(μ, σ^2) is

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

$$L = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_1^n (x-\mu)^2}$$

Therefore, MPCR *w* by *NP* Lemma is

$$Inside\ w: \qquad \frac{L_1}{L_0} = \frac{e^{-\frac{1}{2\sigma_1^2}\sum_1^n (x_i-\mu)^2}}{e^{-\frac{1}{2\sigma_0^2}\sum_1^n (x_i-\mu)^2}} \left(\frac{\sigma_0}{\sigma_1}\right)^n \geq k$$

Or,

$$e^{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_1^n (x_i-\mu)^2} \geq k_1$$

Or,

$$-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_1^n (x_i - \mu)^2 \geq \log k_1$$

(a) Since $\sigma_1 > \sigma_0$ therefore $-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) \geq 0$

Thus MPCR of size α is

$$w = \left\{ \sum_{1}^{n} (x_i - \mu)^2 \geq k_2 \right\}$$

To obtain k_2 note that $U = \sum_1^n \left(\frac{x_i - \mu}{\sigma}\right)^2$ has χ^2 distribution with n degree of freedom (df).

Thus,

$$\alpha = P[U \geq \chi_n^2, \alpha]$$

Where χ_n^2, α is obtained from χ^2 table.

$\chi^2_{n,\alpha}$

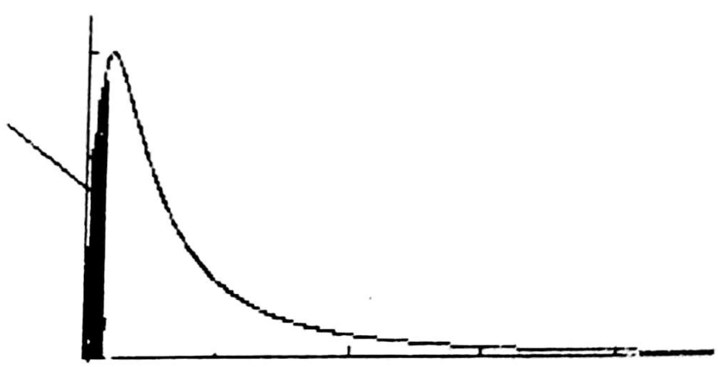
(b) Similarly for

$H_0: \sigma = \sigma_0$

$H_1: \sigma = \sigma_1 (< \sigma_0).$

$MPCR\ of\ size\ \alpha\ is$

$$w = [\mu \leq \chi^2_{n,1-\alpha}]\ where, u = \sum_{1}^{n} \left(\frac{x_i - \mu}{\sigma_0}\right)^2$$



$\chi^2_{n,1-\alpha}$

Since both these MPCR does not depend on σ_1 hence UMPCR for testing

$$H_0: \sigma = \sigma_0$$

$$H_1: \sigma > \sigma_0$$

$$is \qquad \mu \leq \chi^2_n, \alpha.$$

and for

$$H_0: \sigma = \sigma_0$$

$$H_1: \sigma < \sigma_0$$

$$is \qquad \mu \leq \chi^2_n, 1 - \alpha.$$

As obtained above. Further as explained earlier UMPCR for testing

$$H_0: \sigma = \sigma_0$$

$$H_1: \sigma \neq \sigma_0$$

Does not exist.

Numerical Example: Let X has normal distribution *N (50,*σ^2*).* Based on a random sample of size n = 15, obtain MPCR of size α = .05 for

$$(a) \qquad H_0: \sigma = 10$$

$$H_1: \sigma = 15$$

$$Here \qquad \chi^2_{n,\ \alpha} = 24.996$$

$$Reject\ H_0\ if \sum_1^n \left(\frac{x_i - 50}{15}\right)^2 \geq 24.996$$

and

$$(b) \qquad H_0: \sigma = 15$$

$$H_1: \sigma = 10$$

$$Here \qquad \chi^2_{n,\ \alpha} = 7.261$$

$$Reject\ H_0\ if \qquad \sum_1^n \left(\frac{x_i - 50}{15}\right)^2 \leq 7.261.$$

9.5 Likelihood Ratio Test

In the Neman-Pearson (NP) Lemma, we used likelihood ratio to obtain MPCR if both the hypotheses are simple. But NP Lemma fails if one or both the hypotheses are composite. To overcome this problem another method, known as likelihood Ratio Test (LRT) is used, which is also ratio of two likelihood's. For this define a concept parameter space.

Parameter space Θ is set of all possible values of the parameters in a distribution.

Example: $X \sim P(\theta), \Theta = \{\theta: \theta > 0\}$

$X \sim B(n,p), \Theta = \{(n,p): 0 \leq p \leq 1,\ \ n = 1,2, \dots\dots\}$

$X - N(\mu, \sigma^2), \Theta = \{(\mu, \sigma^2): -\infty < \mu < \infty, \sigma > 0\}$.

We will denote parameter space under H_0 as Θ_0 and under H_1 as Θ_1. The Hypothesis under H_0 and H_1 may or not be simple.

Definition: Let $\sim f(x, \theta), \theta \in \Theta$. Suppose on the basis of a random sample X_1, ….X_n of size n, we have to test.

$H_0: \theta \in \Theta_0$

$H_1: \theta \in \Theta_1$

$where\ \Theta_0 \cup \Theta_1 = \Theta\ and\ \Theta_0 \cap \Theta_1 = \emptyset$

Define likelihood ratio as

$$\lambda = \frac{\underset{\Theta_0}{Sup}\, L(\theta)}{\underset{\Theta}{Sup}\, L(\theta)} = \frac{N}{D},$$

Where $L = \prod_{i=1}^{n} f(x_i, \theta)$ is the likelihood function and $\underset{\Theta_0}{Sup}\, L(\theta)$ is the maximum of the likelihood function $L(\theta)$ under H_0. If some of the parameters are unspecified then are replace with their maximum likelihood estimates (MLE). Similarly $\underset{\Theta}{Sup}\, L(\theta)$ will denote the maximum of the likelihood function $L(\theta)$ when $\theta \in \Theta$.

The LRT says,

$Reject\ H_0\ \ if \quad \lambda \leq \lambda_0, \qquad where \qquad \alpha = P[\lambda \leq \lambda_0 | H_0]$.

Here it may be seen that $0 \leq \lambda \leq 1\ as\ \Theta_0 \subset \Theta, and\ therefore\ N \leq D..$

For finding CR of size α the distribution of λ should be known. However, if the distribution of λ is not known, we may use asymptotic result:

$-2log\lambda - \chi^2_{(r)}$.

9.6 Examples based on Likelihood Ratio Test

Example 5: Let $X \sim N(\theta, \sigma^2)$. Using LRT, obtain CR of size α on the basis of a random sample $X_1, \ldots .X_n$ of size *n* for testing the hypothesis.

(a) $H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

(b) $H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$.

(I) $H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$.

(i) If σ is known

(ii) if σ is unknown

Solution;

a (i) if σ is known let $\sigma = \sigma_0$ and then

$$\Theta_0 = \{(\theta, \sigma^2): \theta = \theta_0,\ \ \sigma = \sigma_0\}$$

$$\Theta = \{(\theta, \sigma^2): \theta \geq \theta_0,\ \ \sigma = \sigma_0\}$$

Now

$$f(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2}$$

$$L = \prod_{i=1}^{n} f(x_i, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_1^n \left(\frac{x_i - \theta}{\sigma}\right)^2}$$

$$N = \underset{\Theta_0}{Sup\ L(\theta)} = \frac{1}{\sigma_0^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_1^n \left(\frac{x_i - \theta_0}{\sigma_0}\right)^2}$$

Note that maximum likelihood (MLE) for θ in Θ is

$$\hat{\theta} = \begin{cases} \bar{x}, & if \quad \bar{x} > \theta_0 \\ \theta_0, & if \quad \bar{x} \leq \theta_0 \end{cases}$$

Therefore,

$$D = \begin{cases} \underset{\Theta_0}{Sup\ L(\theta)} = \frac{1}{\sigma_0^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_1^n \left(\frac{x_i - \theta_0}{\sigma_0}\right)^2}, & if \quad \bar{x} > \theta_0 \\ N, & if \quad \bar{x} \leq \theta_0 \end{cases}$$

And

$$\lambda = \begin{cases} \dfrac{\underset{\Theta_0}{Sup\ L(\theta)}}{\underset{\Theta}{Sup\ L(\theta)}} = \dfrac{e^{-\frac{1}{2}\sum_1^n \left(\frac{x_i - \theta_0}{\sigma_0}\right)^2}}{e^{-\frac{1}{2}\sum_1^n \left(\frac{x_i - \bar{x}}{\sigma_0}\right)^2}}, & if \quad \bar{x} > \theta_0 \\ 1, & if \quad \bar{x} \leq \theta_0 \end{cases}$$

$$= \begin{cases} e^{-\frac{1}{2\sigma_0^2}\sum_1^n \left[(x_i - \theta_0)^2 - (x_i - \bar{x})^2\right]}, & if \quad \bar{x} > \theta_0 \\ 1, & if \quad \bar{x} \leq \theta_0 \end{cases}$$

$$= \begin{cases} e^{-\frac{1}{2\sigma_0^2}\sum_1^n \left[(\bar{x} - \theta_0)^2\right]} & \bar{x} > \theta_0 \\ 1, & \bar{x} \leq \theta_0 \end{cases}$$

$$= \begin{cases} e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta_0)^2} & \bar{x} > \theta_0 \\ 1, & \bar{x} \leq \theta_0 \end{cases}$$

$For\ \bar{x} > \theta_0, \quad reject\ H_0 \quad if \quad e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta_0)^2} \leq \lambda_0$

$$or, \qquad \frac{n}{2}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)^2 \leq \log \lambda_0$$

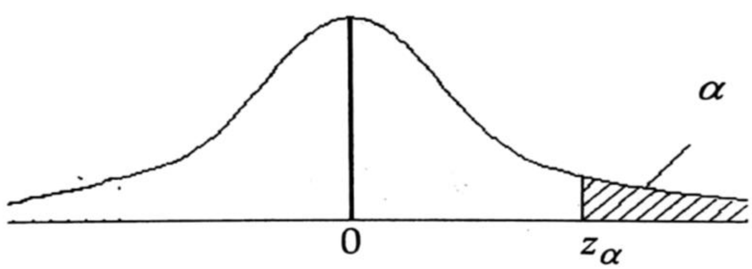
$$or, \qquad \frac{n}{2}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)^2 \geq \lambda_0$$

$$or, \qquad \sqrt{n}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)^2 \geq \lambda_0\ as\ \bar{x} > \theta_0.$$

Note that if $X \sim N\ (\theta_0, \sigma_0^2), then$

$$Z = \sqrt{n}\left(\frac{\bar{X}-\theta_0}{\sigma_0}\right) \sim N(0,1)$$

Therefore reject H_0 if $z \geq \lambda_2$, where λ_2 is such that $\alpha = P(Z \geq \lambda_2 | H_0)$. Hence $\lambda_2 = Z_\alpha$ and CR is $\left\{\bar{x} \geq \theta_0 + \sigma_0 \frac{Z_\alpha}{\sqrt{n}}\right\}$



as obtained by NP Lemma because it was simple vs simple hypothesis. This is UMPCR

a (ii) σ is unknown.

$$Here\ \Theta_0 = \{(\theta, \sigma^2) : \theta = \theta_0, \sigma > 0\}$$

$$\Theta = \{(\theta, \sigma^2) : \theta = \theta_0, \sigma > 0\}$$

and maximum likelihood estimate of σ^2 in Θ_0 is $\frac{1}{n}\sum_1^n (x_i - \theta_0)$ and mle of $\theta, \sigma^2 in\ \Theta\ is$

$$\hat{\theta} = \begin{cases} \bar{x}, & if \quad \bar{x} > \theta_0 \\ \theta_0, & if \ \bar{x} \leq \theta_0 \end{cases}$$

$$\hat{\sigma}^2 = \begin{cases} \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})^2, & \bar{x} > \theta_0 \\ \\ \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})^2, & \bar{x} \leq \theta_0 \end{cases}$$

Hence,

$$\lambda = \begin{cases} \dfrac{\underset{\Theta_0}{Sup\ L(\theta)}}{\underset{\Theta}{Sup\ L(\theta)}} = \dfrac{\left(\frac{n}{\sum(x_i - \theta_0)^2}\right)^{n/2} e^{-\frac{n}{2}}}{\left(\frac{n}{\sum(x_i - \theta_0)^2}\right)^{n/2} e^{-\frac{n}{2}}}, & \bar{x} > \theta_0 \\ \\ 1, & \bar{x} \leq \theta_0 \end{cases}$$

Now

$$\sum(x_i - \theta_0)^2 = \sum[(x_i - \bar{x}) + (x_i - \theta_0)]^2$$

$$= \sum(x_i - \bar{x})^2 + n(x_i - \theta_0)^2$$

Therefore for $\bar{x} > \theta_0$ reject H_0 if

$$\lambda = \left(\frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \theta_0)^2}\right)^{n/2} = \left(\frac{1}{\frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \theta_0)^2}}\right)^{n/2}$$

$$= \left(\frac{1}{1 + \frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \theta_0)^2}}\right)^{n/2} \leq \lambda_0$$

Or

$$1 + \frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \theta_0)^2} \geq \lambda_1$$

Or,

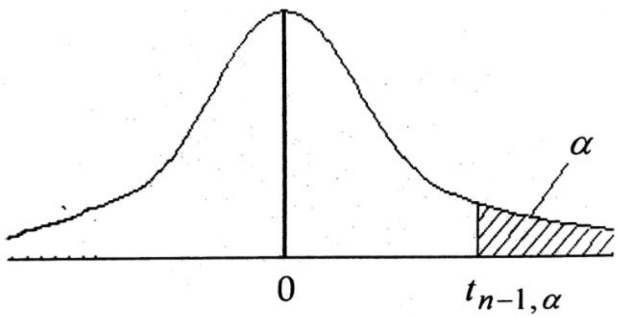
$$\frac{n(\bar{x} - \theta_0)^2}{\sum(x_i - \bar{x})^2} \geq \lambda_2$$

Or,

$$\frac{\sqrt{n}(\bar{x} - \theta_0)^2}{s} \geq \lambda_3, \quad as \quad \bar{x} > \theta_0, \quad where \quad (n-1)s^2 = \sum(x_i - \bar{x})^2.$$

But t $= \frac{\sqrt{n}\ (\bar{x}-\theta_0)}{s}$ has t distribution with (n-1) df.

Hence reject $H_0\ if\ t \geq t_{n-1,\alpha}$



$Thus\ CR\ is \quad \left\{x: \bar{x} \geq \theta_0 + t_{n-1,\alpha}\frac{s}{\sqrt{n}}\right\}.$

b (i) $H_0: \theta = \theta_0$

$H_1: \theta < \theta_0,$

Where $\sigma = \sigma_0$ is known

$\Theta_0 = \{(\theta, \sigma^2): \theta = \theta_0, \sigma > 0\}$

$\Theta = \{(\theta, \sigma^2): \theta = \theta_0, \sigma > 0\}$

Here in denominator θ will be replace by its MLE

$$\hat{\theta} = \begin{cases} \bar{x}, & if \quad \bar{x} > \theta_0 \\ \theta_0, & if\ \bar{x} \leq \theta_0 \end{cases}$$

Thus,

$$\lambda = \begin{cases} \dfrac{\underset{\Theta_0}{Sup}\, L(\theta)}{\underset{\Theta}{Sup}\, L(\theta)} = \dfrac{e^{-\frac{1}{2}\sum_1^n\left(\frac{x_i-\theta_0}{\sigma_0}\right)^2}}{e^{-\frac{1}{2}\sum_1^n\left(\frac{x_i-\bar{x}}{\sigma_0}\right)^2}}, & if \quad \bar{x} > \theta_0 \\ \\ 1, & if \quad \bar{x} \le \theta_0 \end{cases}$$

For $\bar{x} \le \theta_0$ reject H_0 if

$$\lambda = e^{-\frac{n}{2}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)} \le \lambda_0, \qquad as\ shown\ in\ (a).$$

Or,

$$\frac{n}{2}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right) \ge \lambda_1$$

Or,

$$z = \sqrt{n}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right) \ge \lambda_2 \qquad as\ \bar{x} < \theta_0,$$

Where λ_2 is such that $\alpha = P(Z \ge \lambda_2 | H_0)$.

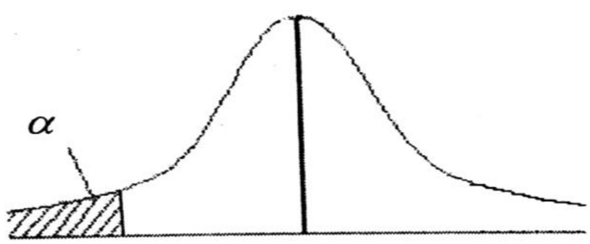
Therefore, reject H_0 if $\frac{\sqrt{n}(\bar{x}-\theta_0)}{\sigma_0} \le -z_\alpha$

Or,

$$\bar{x} < \theta_0 - z_\alpha \frac{\sigma_0}{\sqrt{n}}$$

Thus CR is $\left\{x: \bar{x} \le \theta_0 - z_\alpha \frac{\sigma_0}{\sqrt{n}}\right\}$.

This is UMPCR of size α as obtained through NP Lemma.



b (ii) if σ is known then

$\Theta_0 = \{(\theta, \sigma^2): \theta = \theta_0, \sigma > 0\}$

$\Theta = \{(\theta, \sigma^2): \theta \leq \theta_0, \sigma > 0\}$.

And MLE for σ^2 in Θ_0 is $\frac{1}{n}\sum_1^n (x_i - \theta_0)^2$

and for θ, σ^2 for in Θ are

$$\hat{\theta} = \begin{cases} \bar{x}, & if \quad \bar{x} > \theta_0 \\ \theta_0, & if \ \bar{x} \geq \theta_0 \end{cases}$$

and,

$$\hat{\sigma}^2 = \begin{cases} \frac{1}{n}\sum_1^n (x_i - \bar{x})^2, & if \quad \bar{x} > \theta_0 \\ \frac{1}{n}\sum_1^n (x_i - \bar{x})^2, & if \quad \bar{x} \geq \theta_0 \end{cases}$$

thus,

$$\lambda = \begin{cases} \dfrac{\left(\frac{n}{\sum (x_i - \theta_0)^2}\right)^{n/2} e^{-\frac{n}{2}}}{\left(\frac{n}{\sum (x_i - \theta_0)^2}\right)^{n/2} e^{-\frac{n}{2}}}, & if \quad \bar{x} > \theta_0 \\ 1, & if \quad \bar{x} \leq \theta_0 \end{cases}$$

Or,

$$\lambda = \begin{cases} \dfrac{(\sum (x_i - \theta_0)^2)^{n/2}}{(\sum (x_i - \theta_0)^2)^{n/2}} & if \quad \bar{x} > \theta_0 \\ 1, & if \quad \bar{x} \geq \theta_0 \end{cases}$$

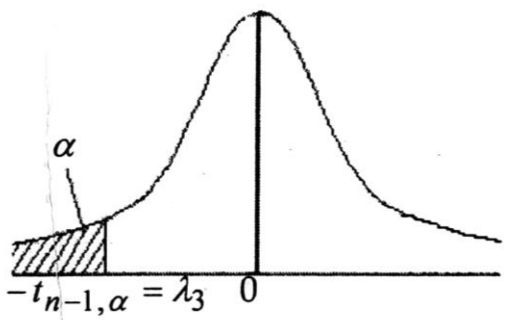
Proceeding as in (a) we get

$$\frac{n(\bar{x}-\theta_0)^2}{(n-1)s^2} \geq \lambda_2 \qquad if \quad \bar{x} \leq \theta_0$$

Implying that,

$$t = \frac{\sqrt{n}(\bar{x}-\theta_0)}{s} \leq \lambda_3,$$

$$where\ \alpha = P\left(\frac{\sqrt{n}(\bar{x}-\theta_0)}{s} \leq \lambda_3 | H_0\right)$$



i.e. $\frac{\sqrt{n}(\bar{x}-\theta_0)}{s} \leq -t_{n-1,}\alpha.$

Hence CR is $\left\{\underline{x}: \bar{x} \leq \theta_0 - t_{n-1,}\alpha \frac{s}{\sqrt{n}}\right\}.$

c (i) if $\sigma = \sigma_0$ is known then

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

$\Theta_0 = \{(\theta, \sigma^2): \theta = \theta_0, \sigma = \sigma_0\}$

$\Theta = \{(\theta, \sigma^2): \theta \leq \theta < \infty, \sigma = \sigma_0\}.$

$$N = \frac{1}{\sigma_0^n (2\pi)^{n/2}} e^{-\frac{1}{2}\Sigma_1^n \left(\frac{x_i-\theta_0}{\sigma_0}\right)^2}$$

$$D = \frac{1}{\sigma_0^n (2\pi)^{n/2}} e^{-\frac{1}{2}\Sigma_1^n \left(\frac{x_i - \bar{x}}{\sigma_0}\right)^2}$$

Therefore

$$\lambda = \frac{N}{D} = e^{-\frac{1}{2\sigma_0^2}\Sigma_1^n\left[(x_i-\theta_0)^2 (x_i-\bar{x})^2\right]} = e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta_0)^2}$$

Reject H_0 if $\quad e^{-\frac{n}{2}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)^2} \leq \lambda_0$

Or,

$$-\frac{n}{2}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)^2 \leq \ log\lambda_0$$

Or,

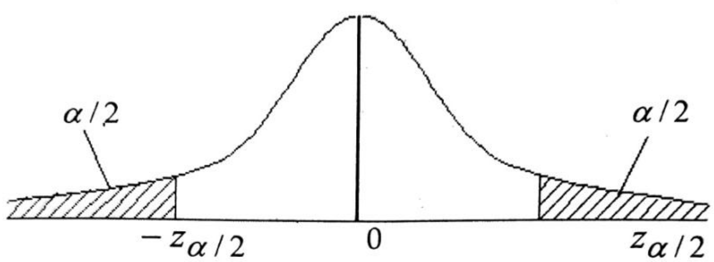
$$\frac{n}{2}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)^2 \leq \ \lambda_1$$

Or,

$$\left|\sqrt{n}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right)\right| \geq \lambda_2,$$

Where $\alpha = P(|Z| \geq \lambda_2 | H_0)$

$$z = \sqrt{n}\left(\frac{\bar{x}-\theta_0}{\sigma_0}\right) \sim N(0,1), and\ therefore\ \lambda_2 = z_{\alpha/2}$$



c (ii) If σ unknwon

$$\Theta_0 = \{(\theta, \sigma^2): \theta = \theta_0, \sigma > 0\}$$

$\Theta = \{(\theta, \sigma^2) : -\infty < \theta < \infty, \sigma > 0\}.$

MLE for σ^2 in Θ_0 is

$\hat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_i - \theta_0)^2$

and MLE for θ *and* σ^2 in Θ_0 is

$$\hat{\theta} = \bar{x}.$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2$$

Therefore,

$$\lambda = \frac{N}{D} = \frac{\left(\frac{n}{\sum(x_i - \theta_0)^2}\right)^{n/2}}{\left(\frac{n}{\sum(x_i - \bar{x})^2}\right)^{n/2}}$$

$$= \left(\frac{\sum(x_i - \bar{x})^2}{\sum(x_i - \theta_0)^2}\right)^{n/2}$$

Hence proceeding as in (a) (ii) for unknown

we reject H_0 if,

$$\frac{n(\bar{x} - \theta_0)^2}{(n-1)s^2} \geq \lambda_2$$

Or,

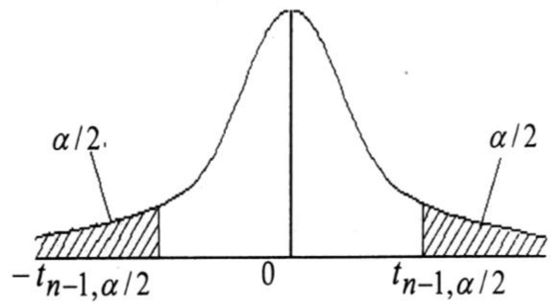
$$\left|\frac{\sqrt{n}(\bar{x} - \theta_0)}{s}\right| \leq \lambda_3$$

Now $t = \frac{\sqrt{n}(\bar{x} - \theta_0)}{s}$ has t- distribution with (n-1) df and therefore,

$$\alpha = P\left(|t| \geq t_{n-1,}\alpha/2\right).$$

Thus CR is

$$\left\{ \underline{x} : \ \left| \frac{\sqrt{n}(\bar{x} - \theta_0)}{s} \right| \geq t_{n-1,}\alpha/2 \right\}$$



Example 6: Let X be Bernoulli variate B(1, θ), Using LRT, obtain CR of size α for testing the hypothesis.

(a) $H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

(b) $H_0: \theta = \theta_0$

$H_1: \theta < \theta_0.$

(C) $H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0.$

On the basis of a random sample of size n.

Solution:

(a) Here

$$\Theta_0 = \{(n, \theta): \theta = \theta_0, n = 1,2, \dots..\}$$

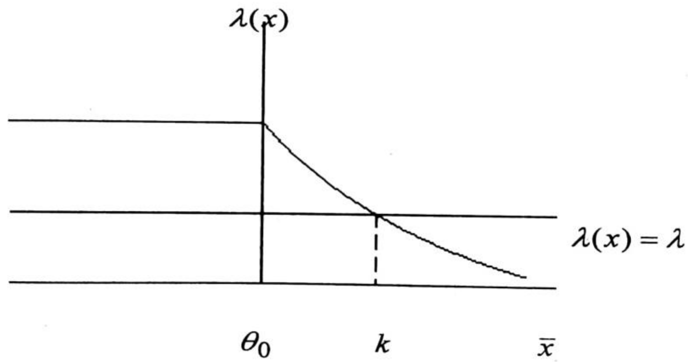
$$\Theta = \{(n, \theta): \theta < \theta_0, n = 1,2, \dots..\}.$$

and MLE of θ in Θ are

$$\hat{\theta} = \begin{cases} \bar{x}, & if \quad \bar{x} > \theta_0 \\ \theta_0, & if \quad \bar{x} \leq \theta_0 \end{cases}$$

Therefore,

$$\lambda(x) = \frac{\underset{\Theta_0}{Sup}\, L(\theta)}{\underset{\Theta}{Sup}\, L(\theta)} = \begin{cases} \dfrac{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}{\bar{x}^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}, & if \quad \bar{x} > \theta_0 \\ 1, & if \quad \bar{x} \le \theta_0 \end{cases}$$



Now for $\bar{x} > \theta_0$

$$\lambda(x) = \frac{\theta_0^{\sum x_i}(1-\theta_0)^{n-n\bar{x}}}{\bar{x}^{\sum x_i}(1-\theta_0)^{n-n\bar{x}}}$$

$$= \left[\left(\frac{\theta_0}{\bar{x}}\right)^{\bar{x}}\left(\frac{1-\theta_0}{1-\bar{x}}\right)^{1-\bar{x}}\right]^n$$

$$log\lambda(x) = n[\bar{x}log\theta_0 - \bar{x}log\bar{x} + (1-\bar{x})\log(1-\theta_0) - (1-\bar{x})\log(1-\bar{x})]$$

$$\frac{\partial}{\partial\bar{x}}log\lambda(x) = n[log\theta_0 - 1 - log\bar{x} - \log(1-\theta_0) + 1 + \log(1-\bar{x})]$$

$$= n\left[log\frac{\theta_0(1-\bar{x})}{\bar{x}(1-\theta_0)}\right] < 0$$

As $\bar{x} > \theta_0$ therefore $1 - \bar{x} < 1 - \theta_0$

and $\frac{\theta_0(1-\bar{x})}{\bar{x}(1-\theta_0)} < 1$

Therefore, the function $\lambda(x)$ is decreasing for $\bar{x} > \theta_0$

And as shown in curve $\lambda(x) < \lambda \Rightarrow \bar{x} > k$, That is reject H_0 if $\sum_{i=1}^{n} x_i \geq k_1$

Where $\sum_{i=1}^{n} X_i$ has Binomial B(n,θ) distribution and k_1 is such that

$$\alpha = P\left[\sum_i X_i \leq k_1 \mid H_0\right]$$

This is UMPCR as obtained by NP lemma.

(b) Here

$$\Theta_0 = \{(n, \theta): \theta = \theta_0, n = 1,2, \ldots\ldots\}$$

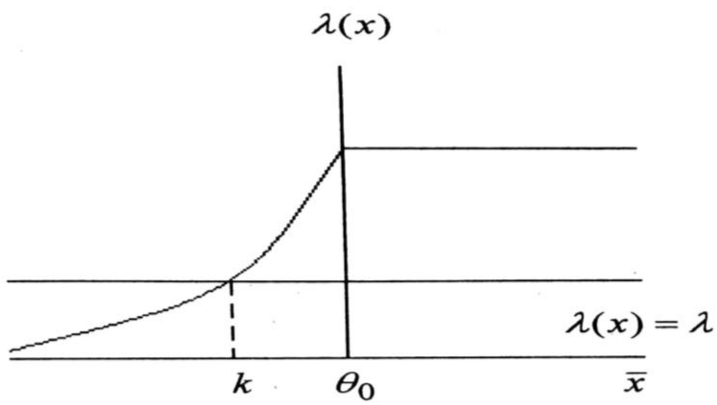
$$\Theta = \{(n, \theta): \theta < \theta \leq \theta_0, n = 1,2, \ldots\ldots\}.$$

and MLE of θ in Θ is

$$\hat{\theta} = \begin{cases} \bar{x}, & if \quad \bar{x} > \theta_0 \\ \theta_0, & if \quad \bar{x} \leq \theta_0 \end{cases}$$

Therefore,

$$\lambda(x) = \begin{cases} = \dfrac{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}{\bar{x}^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}, & if \quad \bar{x} < \theta_0 \\ 1, & if \quad \bar{x} \geq \theta_0 \end{cases}$$



Thus for $\bar{x} > \theta_0$, $\lambda(x)$ is an increasing function and therefore

$$\lambda(x) \le \lambda \Rightarrow \quad \bar{x} \le k$$

$$\Rightarrow \sum x_i \le k_1$$

and hence

reject H_0 if $\sum_{i=1}^{n} x_i \le k_1$

$$where \quad \sum_{i=1}^{n} x_i \sim N(n, \theta) \qquad and \quad k_1 \; is \; such \; that$$

$$\alpha = P\left[\sum_{i=1}^{n} x_i \le k_1 \,|H_0\right].$$

This CR is also UMPCR as obtained by NP lemma

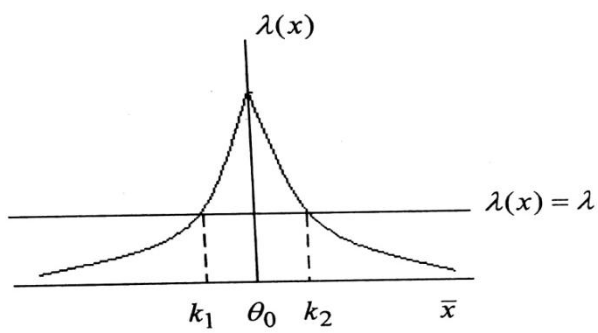
(c) Here

$$\Theta_0 = \{(n, \theta) \colon \theta = \theta_0, n = 1,2, \ldots\ldots\}$$

$$\Theta = \{(n, \theta) \colon \theta < \theta \le 1, n = 1,2, \ldots\ldots\}.$$

and MLE of θ in Θ is $\bar{x}$ therefore,

$$\lambda(x) = \frac{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}{\bar{x}^{\sum x_i}(1-\theta_0)^{n-\sum x_i}}$$



Here it may be seen that $\lambda(x)$ increase for $\bar{x} < \theta_0$ and decrease for $\bar{x} > \theta_0$.

Thus $\lambda(x) \leq \lambda$

$\Rightarrow \bar{x} \leq k_1 \quad or \quad \bar{x} \geq k_2$

$\Rightarrow \sum x_i \leq k_3 \quad or \quad \sum x_i \geq k_4$

$where \quad \alpha = P\left[\sum x_i \leq k_3 \,|H_0\right] + P\left[\sum x_i \geq k_4 \,|H_0\right]$

Because of symmetry, we have

$$\frac{\alpha}{2} = P\left[\sum_i X_i \leq k_3 \,|H_0\right] = P\left[\sum_i X_i \leq k_4 \,|H_0\right]$$

Example 7: Based on a random sample $X_1, X_2, \ldots\ldots X_n$ from Poisson P(θ) distribution test the hypothesis.

$(a) \quad H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

$(b) \quad H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$.

$(C) \quad H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$.

Solution: We have

(a) Here $\Theta_0 = \{\theta: \theta = \theta_0\}$

$\Theta = \{\theta: \theta \geq \theta_0\}$

$and\ MLE\ of\ \theta\ in\ \Theta\ is$

$$\hat{\theta} = \begin{cases} \bar{x}, & if \quad \bar{x} > \theta_0 \\ \\ \theta_0, & if \quad \bar{x} > \theta_0 \end{cases}$$

$Therefore$

$$\lambda(x) = \begin{cases} \dfrac{e^{-n\theta_0}\, \theta_0 \sum x_i}{e^{-n\bar{x}} \bar{x} \sum x_i}, & if \quad \bar{x} > \theta_0 \\ \\ 1, & if \quad \bar{x} \le \theta_0 \end{cases}$$

Now for $\bar{x} > \theta_0$

$$\lambda(x) = e^{-n(\theta_0 - \bar{x})} \left(\frac{\theta_0}{\bar{x}}\right)^{n\bar{x}}$$

$$log\lambda(x) = n[-(\theta_0 - \bar{x}) + \bar{x}\, \log \theta_0 - \bar{x}\, log\bar{x}]$$

$$\frac{\partial}{\partial \bar{x}} log\lambda(x) = n[1 + \log \theta_0 - 1\, log\bar{x}] = nlog\frac{\theta_0}{x} < 0$$

That is $\lambda(x)$ is decreasing function $\bar{x}$ of for $\bar{x} > \theta_0$.

Therefore,

$$\lambda(x) \le \lambda \Rightarrow \bar{x} \ge k$$

$$\Rightarrow \sum X_i \ge k_1,$$

Where $\sum_{i=1}^{n} X_i \sim P(n\theta)$

and k_1 is such that

$$\alpha = P\left[\sum_i X_i \le k_1 | H_0\right]$$

This is the UMPCR as obtained by NP lemma.

Proceeding as above, we can show that for **(b)**

Reject H_0 if $\sum_{i=1}^{n} X_i \le k_1,$

Where $\alpha = P[\sum_i X_i \le k_1 | H_0]$

And for (c) Reject H_0 if

$$\sum X_i \le k_3 \qquad or \sum X_i \ge k_4,$$

Where $\alpha/2 = P[\sum_i X_i \le k_3 | H_0]$

and Where $\alpha/2 = P[\sum_i X_i \le k_4 | H_0]$

9.7 Problems and Exercises

1. Find UMP test of size α =0.05 for testing $H_0: \mu = 0$ against $H_0: \mu > 0$ on the basis of a random sample of size 25 from normal $N\ (\mu, 16)$ distribution.

Hint: Let $H_1: \mu = \mu_1 (> 0), then$

$$\alpha = P[\bar{X} > k | H_0] = P\left[Z > \frac{5}{4}k\right] = 0.05$$

$$\Rightarrow k = \frac{4}{5} \times 1.645 = 1.316.$$

∴ MPCR for testing $H_0: \mu > 0$ against $H_0: \mu = \mu_1 > (0) is\ \{\bar{x} > 1.316\}$. Since this CR does not depend on the choice of μ_1 hence this is also UMP for $H_0: \mu = 0$ against $H_1: \mu > 0$.

2(a) Based on a random sample of size n from an exponential distribution with

$pdf\ \ f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, x, \theta > 0,$ obtain UMPCR of size for testing $H_0: \theta = \theta_0$ against $H_0: \theta > \theta_0$

Hint : Let $\frac{y}{2} = \frac{x}{\theta}$ then pdf of y is $\frac{1}{2} e^{-y/2}, x, y > 0$

Which is the pdf of χ_2^2 i.e. the distribution of

$$Y = \frac{2X}{\theta} \sim \chi_2^2$$

$$and\ U = \sum_{i=1}^{n} Y_i = \frac{2}{\theta} \sum_{i=1}^{n} X_i \sim \chi_{2n}^2$$

UMPCR for $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1\ (> \theta_0)$by NP lemma is

$$\left\{\sum_{i=1}^{n} X_i > k\right\} \; or \left\{U > \frac{2k}{\theta}\right\}, where\ \alpha = \left[U > \frac{2k}{\theta} \mid H_0\right].$$

From table.

$$\frac{2k}{\theta_0} = \chi^2_{2n,\alpha} \;\; or\ k = \frac{\theta_0}{2}\chi^2_{2n,\alpha}$$

This is also UMP for $H_1: \theta = \theta_0$

(b) Also obtain UMP test for $H_0: \theta = \theta_0\ against\ H_1: \theta < \theta_0$ and show that UMP test does not exit for $H_0: \theta = \theta_0\ vs\ H_1: \theta \neq \theta_0$.

3.(a) Based on a random sample of size n from an exponential distribution.

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

Obtain UMP test for testing$H_0: \theta = \theta_0\ against\ H_1: \theta < \theta_0$

Hint: It can be seen that $Y = 2X\theta$ has. Therefore χ^2_2

$$U = \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

Here for, $H_1: \theta = \theta_1\ (> \theta_0)$ MP test is $\{\sum_{i=1}^{n} X_i > k\}$, where

$$\alpha = \left\{\sum_{i=1}^{n} X_i < k | H_0\right\} = P[U \leq 2k\theta_0].$$

(b) Also obtain UMP test for $H_0: \theta = \theta_0\ against\ H_1: \theta < \theta_0$ and show that UMP test does not exit for $H_0: \theta = \theta_0\ vs\ H_1: \theta \neq \theta_0$.

4. Based on a random sample of size *n* from $N\ (\mu, \sigma^2)$ obtain CR for testing the hypotheses

(a) $\mathrm{H}_0: \sigma = \sigma_0$

$\mathrm{H}_1: \sigma > \sigma_0$

(b) $\mathrm{H}_0: \sigma = \sigma_0$

$\mathrm{H}_1: \sigma < \sigma_0$

(c) $\mathrm{H}_0: \sigma = \sigma_0$

$H_1 : \sigma \neq \sigma_0$

Using LRT method when (i) μ is known and (ii) μ is unknown.

9.8 Summary

Neyman-Pearson lemma provides a fundamental rule for obtaining MPCR and hence MP test for testing simple verse simple hypotheses. According to the lemma, a CR of size α is said to be MPCR if $L_1 \geq kL_0$ inside w and $L_1 < kL_0$ outside *w* where L_o and L_1 are likelihood functions of the sample observations under H_0 and H_1, respectively, and k is to be obtained such that of the *CR* w is α.

The likelihood ratio test plays the same important role in testing of statistical hypothesis as does the maximum likelihood method or estimation in point estimation of a parameter. It is important to mention that in Neman-Pearson (NP) lemma, we actually use likelihood ration to obtain MPCR if both the hypotheses are simple. NP Lemma however fails if one or both the hypotheses are composite. To override this problem the method known as likelihood Ratio Test (LRT) is used which is also the ratio of the two likelihood's.

9.9 Further Readings

- Goon A.N., Gupta M.K. & Das Gupta B (2000) *An Outline of Statistical Theory* Vol. 2 The World Press Private Limited.
- Hogg, R.V. and Craig, A. (2005). *Introduction to Mathematical Statistics* 6th edition, Prentice Hall.
- Mood, A.M. Graybill, F.A., Boes, D.C. (1974). *Introduction to the Theory of Statistics*, McGraw Hill.
- Mohr, L.B. (1994), *Understnading Significance Testing*, Sage Publications, USA.
- Lehmann, E.L. (1986). *Testing statistical hypothesis*. Springer-Verlag, New York, Inc. 2nd edition.

Unit-10 Testing of Means of Normal Population

Structure

10.1 Introduction

In Unit 9 we have develop methods- Neyman-Pearsn lemma and likelihood ration test to obtain most powerful critical regions for simple and composite hypotheses. These methods will now be employed here to test the equality of means in a one sample and two sample problems from distributions.

10.2 Objectives

After readings this unit you should be able to:

- Understand the testing procedure for testing the significance of mean of a normal population when variance is known or unknown.
- Understand the testing of equality of means of two normal populations based on two independent random samples.

10.3 One Sample Problem

Let a random sample X_1, ……X_n of size n be a drawn from normal N(θ, σ^2) distribution.

If $X \sim N(\theta, \sigma^2)$ then

$$\bar{X} \sim N\left(\theta, \frac{\sigma^2}{n}\right) \quad and \quad Z = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma} \sim N(0,1),$$

Where $\bar{X} = \frac{X_1 + X_2 + \cdots \ldots + X_n}{n}$ is sample mean Z is standard normal variate

A. For σ known

Suppose we have to test

1. $\mathrm{H_0}: \theta = \theta_0$

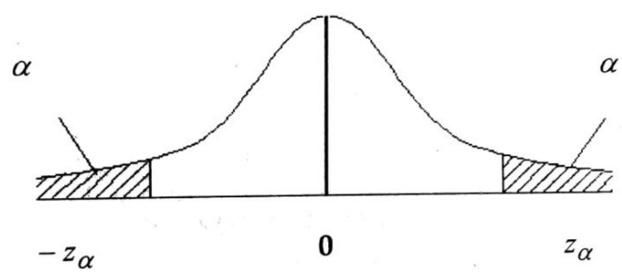
$\mathrm{H_1}: \theta > \theta_0$

$$Here\ we\ calculate \qquad Z = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma}$$

$$and\ reject\ H_0 \qquad if\ \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma} > z_\alpha$$

$$i.e. reject\ H_0\ if \qquad \bar{x} > \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}},$$

$$where \qquad P[Z > Z_\alpha] = \alpha$$



2. For $\mathrm{H_0}: \theta = \theta_0$

$\mathrm{H_1}: \theta > \theta_0$

$$\text{Reject } H_0\ if \qquad \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} < -Z_\alpha$$

$$i.e. reject\ H_0\ if \qquad \bar{x} < \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}},$$

Where P[Z<-z_α] = α.

This is because of symmetry that

$$P[Z < -z_\alpha] = \alpha = P[z > z_\alpha].$$

3. For the hypothesis
$\mathrm{H_0}: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

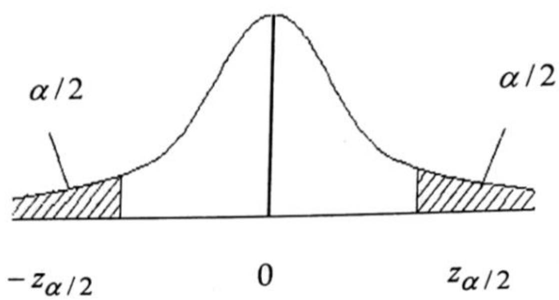
$$\text{Reject } H_0 \ if \qquad \left| \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} < Z_{\alpha/2} \right|$$

Or

$$\text{Reject } H_0 \ if \qquad \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} < -Z_{\frac{\alpha}{2}} \qquad or \qquad \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} > Z_{\alpha/2}.$$

$$or \ reject \ H_0 \ if \qquad \bar{x} < \theta_0 - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \qquad or, \qquad \bar{x} > \theta_0 + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},$$

$$where \qquad P\left[|Z| < -z_{\alpha/2}\right] = P\left[z < -z_{\alpha/2}\right] + P\left[Z > z_{\alpha/2}\right] = \alpha.$$



This CR was obtained earlier in the previous unit.

Example 1: Let a sample of size 25 is drawn from N (θ,4) distribution and let the sample be $\bar{x}$ = 18.9

Then to test

$H_0: \theta = 19.5$

$H_1: \theta < 19.5$

at $\alpha = 0.05$, we have $Z_\alpha = 1.645$

$$Z = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} = \frac{5(18.9 - 19.5)}{2} = \frac{5 \times 0.6}{2} = -1.5.$$

$$Since \qquad \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} > -Z_\alpha$$

We accept H_0 and conclude that

θ = 19.5

If further we have to test

$H_0: \theta = 19.5$

$H_1: \theta \neq 19.5$

at $\alpha = 0.05$, we have $Z_{\alpha/2} = 1.96$

Here $\left|\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma}\right| = |-1.5| = 1.5 < 1.96,$

and therefore, H_0 *is accepted*

B. For σ unknown

$$U = \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} \sim N(0,1)$$

$$V = \frac{\sum_1^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

U and V are independently distributed, therefore,

$$t = \frac{U}{\sqrt{\frac{V}{n-1}}} = \frac{\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}} = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}.$$

Thus if σ is not known then we have to calculate sample variance and have to use t-table instead of normal table.

1. $H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

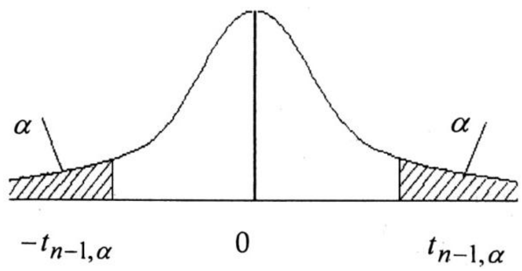
Calculate $\frac{\sqrt{n}(\bar{x} - \theta_0)}{S}$

Where $\bar{x} = \frac{x_1 + x_2 + \cdots . + x_n}{n}, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$Reject\ H_0\ if \qquad \frac{\sqrt{n}(\bar{x} - \theta_0)}{S} > t_{n-1,}\alpha$$

$$or\ reject\ H_0\ if \qquad \bar{x} > \theta_0 + t_{n-1}, \alpha\ \frac{s}{\sqrt{n}},$$

Where t_{n-1} is the value of t- distribution with (n-1) df at α level of significance



This CR was obtained earlier in the previous unit.

2. $H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$

$$Reject\ H_0\ if \qquad \frac{\sqrt{n}(\bar{x} - \theta_0)}{S} > -t_{n-1,}\alpha$$

$$or \qquad reject\ H_0\ if \qquad \bar{x} < \theta_0 - t_{n-1}, \alpha\ \frac{s}{\sqrt{n}},$$

as obtain in the previous unit.

3. $H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

$$Reject\ H_0\ if \qquad \left|\frac{\sqrt{n}(\bar{x} - \theta_0)}{S}\right| > t_{n-1,}\alpha/2$$

$$or \quad reject\ H_0\ if \qquad \bar{x} < \theta_0 - t_{n-1}, \frac{\alpha}{2}\frac{s}{\sqrt{n}}, \qquad or \qquad \bar{x} > \theta_0 - t_{n-1}, \frac{\alpha}{2}\frac{s}{\sqrt{n}}$$

as discussed in the previous unit.

Example 2: Consider a sample of 6 cylinder blocks whose cope hardness values are 70, 75, 60, 75, 65 and 80. Is there any evidence that the average cope hardness has changed from its specified value of 75?

Solution: Here n= 6

$$\bar{x} = \frac{70 + 75 + 60 + 75 + 65 + 80}{6} = 70.8$$

$$s = \sqrt{\frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x})^2} = 7.4$$

Hypothesis to be tested at $\alpha = 0.05$ is

$H_0: \theta = 75$

$H_1: \theta \neq 75$

$$t = \frac{\sqrt{n}(\bar{x} - \theta_0)}{s} = \frac{\sqrt{6}(70.8 - 75)}{7.4} = -1.39.$$

$at \quad \alpha = 0.05,\ t_{5,0.025} = 2.57$

Since calculate value is less than the tabulated value,

i.e $\quad |t| < t_{n-1}\alpha/2$

we accept H_0 and conclude that there is no evidence that average cope hardness has changed from its value 75.

10.4 Two Samples Problem

Let $\sim N(\mu_1, \sigma_1^2)$. Take a random sample of size n_1, X_1, X_2, ……, X_{n1} from this population.

Let $\sim N(\mu_2, \sigma_1^2)$. Take a random sample of size n_1, Y_1, Y_2, ……, Y_{n2} from this population.

Define

$$\bar{X} = \frac{X_1 + X_2 + \cdots . + X_{n_1}}{n_1}$$

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

$$\bar{Y} = \frac{Y_1 + Y_2 + \cdots . + Y_{n_2}}{n_2}$$

$$S_2^2 = \frac{1}{(n_2 - 1)} \sum_{i=1}^{n_1} (Y_i - \bar{Y})^2$$

then

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

and if two populaiton are independent then

$$\bar{X} \sim \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

and thus

$$Z = \frac{(\bar{X} - \bar{Y})(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Here the problem reduces to one sample problem and tests are carried out as explained in NP and LRT sections.

A. σ_1, σ_2 known

1. To test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

Calculate $Z = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, as under $H_0: \mu_1 = \mu_2$

and reject H_0 if $z = z_\alpha$.

2. To test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

reject H_0 if $Z = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, < -z_\alpha$

3. To test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

reject H_0 if $|Z| = \left|\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right|, > z_\alpha/2$

or reject H_0 if $z < -\frac{z_\alpha}{2}$ or $z < z_\alpha/2$.

However if population variances σ_1^2 *and* σ_2^2 are not known but n_1 *and* n_2 are large (>30), then we calculate sample variances.

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

$$S_2^2 = \frac{1}{(n_2 - 1)} \sum_{i=1}^{n_1} (Y_i - \bar{Y})^2$$

and consider s_1 *and* σ_1 *and* s_2 *as* σ_2.

Example 3: A simple random sample of heights of 6400 Englishmen has mean 67.85 inches and standard deviation (s.d.) inches, while a simple random sample of heights of 1600 Australian has mean 68.55 inches and s.d. 2.52. Do the data suggest that Australian are on the average, taller than Englishmen?

Solution: Let Englishmen have $N(\mu_1, \sigma_1^2)$ distribution.

Australian have $N(\mu_2, \sigma_2^2)$ distribution.

Then we have to test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

It is given that $n_1 = 6400$, $\bar{x} = 67.85$, $s_1 \cong \sigma_1 = 2.56$

$n_2 = 1600$, $\bar{y} = 68.55$, $s_2 \cong \sigma_2 = 2.52$

$$Z = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}} = \frac{-0.70}{0.07066} = -9.50$$

$at\ \alpha = .05,$ $z_\alpha = 1.645,$

Reject H_o as z< - z_α and conclude the data do suggest that Australians are taller than Englishmen.

B. If σ_1, σ_2 are unknown but equal.

That is $\sigma_1, = \sigma_2 = \sigma$ (unknown)

$then\ defines \quad S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}$

It can be seen that

$$V = \frac{(n_1 + n_2 - 2)S^2}{\sigma^2} = \frac{(n_1 - 1)S_1^2}{\sigma^2} = \frac{(n_2 - 1)S_2^2}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2}{\sigma^2} + \frac{\sum_{i=1}^{n_1}(Y_i - \bar{Y})^2}{\sigma^2}$$

is distributed as χ^2 with $(n_1 + n_2 - 2)$ df and is independent of Z.

Therefore,

$$t = \frac{Z}{\sqrt{\frac{V}{n_1+n_2-2}}} = \frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\frac{(n_1+n_2-2)S^2}{(n_1+n_2-2)\sigma^2}}}$$

$$= \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Thus, when $\sigma_1, = \sigma_2 = \sigma$ (unknown), we use t statistic rather than Z statistic to test hypothesis.

1. To test

$H_0: \mu_1 = \mu_2$

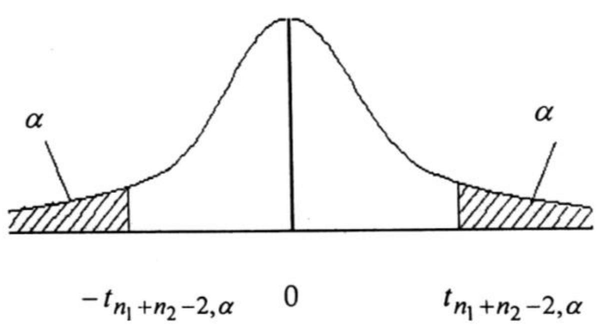
$H_1: \mu_1 > \mu_2$

we Calculate $t = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$,

and reject H_0 if t

$> t_{n_1+n_2-2}, \alpha$ *where* $t_{n_1+n_2-2}, \alpha$ *is the value obtained from* t

$-$ *table at* $(n_1+n_2-2)df$ *and* α *level of significance*



2. To test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

reject H_0 if $t_{n_1+n_2-2}, \alpha \;\; and$

3. To test

$\mathrm{H}_0: \mu_1 = \mu_2$

$\mathrm{H}_1: \mu_1 \neq \mu_2$

reject H_0 if $t_{n_1+n_2-2}, \frac{\alpha}{2} \quad i.e. t < -t_{n_1+n_2-2}, \frac{\alpha}{2} \quad or \;\; t > t_{n_1+n_2-2}, \frac{\alpha}{2}$

where $\quad \mathrm{t} = \dfrac{(\bar{X} - \bar{Y})}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad is\ as\ defined\ above.$

Example 4: Samples of two of electric light bulbs were tested for length of life (in hours) and following data were obtained

	Type-I	Type-II
Sample size	$n_1=8$	$n_2=7$
Sample means	$\bar{x} = 1234\ hrs$	$\bar{x} = 1036\ hrs$
Sample s.d.	$s_1=36$	$s_2=40$

Where $S_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$ and $S_2^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (y_i - \bar{y})^2$.

Test whether the difference in the means suggest that Type I is superior to Type II with regard to the length of life, assuming that population variance of two samples are same.

Solution: Here it is assumed the Type I sample of size n_1 is coming $N(\mu_1, \sigma_1^2)$ from and Type II of size n_2 is coming from $N(\mu_2, \sigma_2^2)$ and $\sigma_1, = \sigma_2 = \sigma$ (unknown).

$$Calculate\ s^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_1} (y_i - \bar{y})^2\right]$$

$$= \frac{1}{n_1 + n_2 - 2}[n_1 s_1^2 + n_2 s_2^2]$$

$$= \frac{1}{8 + 7 - 2}[8 \times 36^2 + 7 \times 40^2] = \frac{1}{13}[10368 + 11200] = 1659.08$$

Then

$$t = \frac{(\bar{X} - \bar{Y})}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1234 - 1036}{\sqrt{1659.08\left(\frac{1}{8} + \frac{1}{7}\right)}} = 9.39.$$

Now to test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

reject H_0 if $\quad t > t_{n_1+n_2-2,\alpha}$

where $t_{n_1+n_2-2,\alpha} = 1.77$ at 13 df and α = .05 from the t- table.

Since calculated t > tabulated t, we reject H_0 and conclude that the data suggest that the Type I bulbs are superior to Type II bulbs.

Example 5: Two horses A and B were tested according to time (in seconds) to run in a particular track with the following results:

Horse A:	28	30	32	33	33	29	34
Horse B:	29	30	30	24	27	29	

Test whether the two horses have the same running capacity.

Solution: It is assumed that the time taken by

Horse A has $N(\mu_1, \sigma_1^2)$ distribution

Horse B has $N(\mu_2, \sigma_2^2)$ distribution

with $\sigma_1, = \sigma_2 = \sigma$ (unknown)

Under this assumption, we have to test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Here $n_1 = 7$, $n_2 = 6$

$$\bar{x} = \frac{28+30+32+33+29+34}{7} = \frac{219}{7} = 31.2857$$

$$\bar{y} = \frac{29+30+30+24+27+29}{6} = \frac{169}{6} = 28.1667$$

$$\sum_{i=1}^{n_1}(x_i - \bar{x})^2 = \sum_{i=1}^{n_1} x_i^2 - n_1\bar{x}^2 = 6883 - 6851.5714 = 31.4285$$

$$\sum_{i=1}^{n_1}(y_i - \bar{y})^2 = \sum_{i=1}^{n_1} y_i^2 - n_1\bar{y}^2 = 4787 - 4760.1667 = 26.8333$$

$$s^2 = \frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{i=1}^{n_1}(y_i-\bar{y})^2\right]$$

$$= \frac{1}{11}[31.4285 + 26.8333] = \frac{58.2618}{11} = 5.2965$$

$$\mathrm{t} = \frac{(\bar{X}-\bar{Y})}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{31.2857 - 28.1667}{\sqrt{5.2965\left(\frac{1}{7}+\frac{1}{6}\right)}}$$

$$= \frac{3.119}{\sqrt{1.6392}} = \frac{3.119}{1.2803} = 2.4361.$$

Tabulated value of t at (7+6-2) =11 df and $\frac{\alpha}{2}$ = 0.025 is 2.20

Since calculated t > tabulated t, we reject H_0 and calculated that two horses have not the same running capacity.

C. If $\sigma_1 \neq \sigma_2$ (unknown) then to test sample mean is said to be **Fisher Behren's** problem. This has been resolved as:

As explained earlier

$$\bar{X} \sim \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$U = \frac{(\bar{X} - \bar{Y})(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad (i)$$

Let V be a random variable having χ_k^2 distribution, then

E(v) = k

Var (v) = 2k

Suppose

$$CV = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) \qquad \dots\dots\dots. \qquad (ii)$$

$Now\ that$

$$\frac{(n_1 - 1)S_1^2}{\sigma_1^2} = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2}{\sigma_1^2} \sim \chi_{n_1-1}^2$$

Therefore,

$$E\left(\frac{(n_1 - 1)S_1^2}{\sigma_1^2}\right) n_1 - 1 \qquad or \qquad E(S_1^2) = \sigma_1^2$$

and

$$V\left(\frac{(n_1 - 1)S_1^2}{\sigma_1^2}\right) = \frac{(n_1 - 1)}{\sigma_1^4} V(S_1^2) = 2(n_1 - 1)$$

Or

$$Var(S_1^2) = \frac{2\sigma_1^4}{(n_1 - 1)} \qquad similarly, \quad E(S_2^2) = \sigma_2^2$$

$$and \qquad Var\ (S_2^2) = \frac{2\sigma_2^4}{(n_2 - 1)}$$

Thus

$$E(CV) = Ck = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Var\ (CV) = \ 2C^2k = \frac{2\sigma_1^4}{n_1^2(n_1-1)} + \frac{2\sigma_2^4}{n_2^2(n_2-1)}$$

$$2CCK = 2\left[\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}\right]$$

$$C = \frac{\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$k = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}}$$

$$= \frac{\frac{\sigma_1^4}{n_2^2}\left(\frac{\sigma_1^2}{n_1}\frac{n_2}{\sigma_2^2} + 1\right)^2}{\frac{\sigma_2^4}{n_1^2}\left(\frac{\sigma_1^4}{n_1^2}\frac{n_2^2}{\sigma_2^4}\frac{1}{n_1-1} + \frac{1}{n_2-1}\right)}$$

$$= \frac{(1+R)^2}{\frac{R^2}{n_1-1} + \frac{1}{n_2-1}} \qquad where \qquad R = \frac{\sigma_1^2/n_1}{\sigma_2^2/n_2}$$

$$or \qquad \frac{1}{k} = \left(\frac{R}{1+R}\right)^2 \frac{1}{n_1-1} + \left(\frac{1}{1+R}\right)^2 \frac{1}{n_2-1}$$

Therefore,

$$\frac{U}{\sqrt{\frac{V}{k}}} = \frac{\frac{(\bar{X}-\bar{Y})(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}{ck}}} \sim t_k$$

$$or \qquad T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{k'}$$

Where k is given in (iii) with $\sigma_1 replace\ by\ s_1\ and\ \sigma_2\ by\ s_2$. Now to test

1. $H_0: \mu_1 = \mu_2$

   $H_1: \mu_1 > \mu_2$

Calculate $t = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$,

and reject H_0 if $t > t_{k'}\ \alpha$

2. $H_0: \mu_1 = \mu_2$

   $H_1: \mu_1 < \mu_2$

reject H_0 if $t < t_k, \alpha$

3. $H_0: \mu_1 = \mu_2$

   $H_1: \mu_1 \neq \mu_2$

   reject H_0 if $t_k, \frac{\alpha}{2}$

as explained in 2.2 B.

Example 6: For the data given in Example 4 test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

If it is given that $\sigma_1 = \sigma_2$ (unknown).

Solution: It is given that

$n_1 = 8, n_2 = 7, \bar{x} = 1234, \bar{y} = 1036, s_1 = 362\ and\ s_2 = 40,$

Where $S_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$ and $S_2^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(y_i - \bar{y})^2$.

$$n_1 s_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$$

$$\therefore s_1^{*2} = \frac{n_1}{n_1 - 1}s_1^2 = \frac{8}{7}\times(36)^2 = 1481.143$$

$$or\ \frac{s_1^{*2}}{n_1} = \frac{s_1^2}{n_1 - 1} = 185.1429$$

$$\frac{s_2^{*2}}{n_1} = \frac{s_2^2}{n_1 - 1} = 266.6667$$

$$\text{t} = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{s_1^{*2}}{n_1} + \frac{s_2^{*2}}{n_2}}} = \frac{1234 - 1036}{\sqrt{185.1429 + 266.6667}} = \frac{198}{21.2558} = 9.3150$$

$$R = \frac{s_1^2/n_1}{s_2^2/n_2} = 0.693$$

$$or\ \frac{1}{k} = \left(\frac{R}{1+R}\right)^2\frac{1}{n_1 - 1} + \left(\frac{1}{1+R}\right)^2\frac{1}{n_2 - 1}$$

$$k = 12.18 \cong 12$$

$$t_{12,\ 0.025} = 2.179 \qquad at \quad \alpha = 0.05.$$

Since calculated t > tabulated t, we therefore reject H_0 and conclude that life lengths of two types of electric bulbs are not same.

D. When the two normal samples are not independent, rather they are paired with $n_1 = n_2 = n$.

In this case we consider difference of two variates as

$$d_i = x_i - y_i$$

$$\bar{d} = \bar{x} - \bar{y}$$

$$s^2 = \frac{1}{n-1} = \sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2$$

Then under H_0: $\mu_1 = u_2$ it can be seen that

$$\frac{\sqrt{nd}}{S} \sim t_{n-1}$$

Example 7: In a certain experiment to compare two types of animal food A and B, the following results of increase in weights were observed in animals:

Animal Number		1	2	3	4	5	6	7	8	Total
Increase of wt. in lb	Food A	49	53	51	52	47	50	52	53	407
	Food B	52	55	52	53	50	54	54	53	423

i) Assuming that the two samples are independent can we conclude that food B is better than food A (Here you may assume that $\sigma_1 = \sigma_2 = \sigma$ (unknown))

ii) Also examine the case when same set of animals were used in both the cases.

Solution:

(i) Suppose food A has $N\ (\mu_1, \sigma_1^2)$ distribution

B has $N\ (\mu_2, \sigma_2^2)$distribution

We have to test

$\mathrm{H}_0: \mu_1 = \mu_2$

$\mathrm{H}_1: \mu_1 < \mu_2$

if $\sigma_1 = \sigma_2 = \sigma$ (unknown)

Now

$$\bar{x} = \frac{407}{8} = 50.875, \quad \bar{y} = \frac{423}{8} = 52.875$$

$$\sum_{i=1}^{n_1}(x_i-\bar{x})^2=\sum_{i=1}^{n_1}x_1^2-n_1\bar{x}^2=20737-20706.125=30.875$$

$$\sum_{i=1}^{n_2}\left(y_j-\bar{y}\right)^2=22383-22366.125=16.875$$

$$s^2=\frac{\sum_{i=1}^{n_1}(x_i-\bar{x})^2+\sum_{i=1}^{n_2}\left(y_j-\bar{y}\right)^2}{n_1+n_2-2}=\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}=3.41$$

$$t=\frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=-\frac{2}{0.9223}=-2.17$$

$t_{14,0.05}=1.761.$

Since calculated t < tabulated ($t_{14,0.05}$) we reject H_0 and conclude that food B is superior to A.

(i) Here d_i=-3,-2, -1, -1, -3, -4, -2, 0, $\sum d_i=-16, \bar{d}=-2$

$$\sum\left(d_i-\bar{d}\right)^2=\sum d_i^2-8\bar{d}=44-32=12$$

$$s^2=\frac{1}{n_1-1}\sum\left(d_i-\bar{d}\right)^2=\frac{12}{7}=1.7142$$

$$\sqrt{\frac{s^2}{n}}=\sqrt{\frac{1.7142}{8}}=0.4629$$

$$t=\frac{\bar{d}}{s/\sqrt{n}}=-\frac{2}{0.4629}=-4.32$$

$t_{7,0.05}=1.895$

Since -4.32< -1.895, we reject H_0 and conclude that food B is better than food A.

10.5 Problems and Exercises

1. Ten individuals are chosen at random from a normal population and their heights are found to be 63, 63, 66, 68, 69, 70, 70, 71, 71 inches. Test if the sample belongs to the population whose mean height is 66 inches. (Given $t_{0.05}$ = 2.62 for 9 df).

2. A random sample of 8 envelopes is taken from letter box of a post office and weights in grams are found to be 12.1, 11.9, 12.4, 12.3, 11.9, 12.1, 12.4, and 12.1.

Does this sample indicate at 1% level that the average weight of envelopes received at the post office is 12.35 gms?

3. Two independent groups of 10 children were tested to find how many digits they could repeat from memory after hearing them. The results are as follows:

Group A	8	6	5	7	6	8	7	4	5	6
Group B	10	6	7	8	6	9	7	6	7	7

Is the different between the mean scores of the two groups significant?

4. Eleven school boys were given a test in Statistics. They were given in month's tuition and second test was held at the end of it. Do the marks give evidence that the students have benefited by the extra coaching?

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks in 1st test	23	20	19	21	18	20	18	17	23	16	19
Marks in 2nd test	24	19	22	18	20	22	20	20	23	20	18

5. Let X_i, i= 1,2,……,m be a random sample from normal $N(\mu_1, \sigma_1^2)$ distribution and $Y_j, j = 1,2, \ldots\ldots, n$ be another random sample from normal $N(\mu_1, \sigma_2^2)$ distribution. Using LRT method, obtain best critical region for testing.

$H_0: \mu_1 = \mu_2$

$H_0: \mu_1 \neq \mu_2$

When (i) σ_1, σ_2 known and (ii) $\sigma_1 = \sigma_2 = \sigma$ (unknown).

Hint: The likelihood function

$$L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \left[\left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{1}{\sigma_1^m} e^{-\frac{1}{2\sigma_1^2} \sum_{i=1}^{m} (x_i - \mu_1)^2} \right]$$

$$\times \left[\left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma_1^n} e^{-\frac{1}{2\sigma_2^2}\sum_{i=1}^{n}(y_j-\mu_2)^2} \right]$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{m+2}{2}} \frac{1}{\sigma_1^m \sigma_1^n} e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(x_i-\mu_1)^2-\frac{1}{2\sigma_2^2}\sum_{i=1}^{n}(y_j-\mu_2)^2}$$

Here

$$\Theta = \{\mu_1 = \mu_2 = \mu, \sigma_1 > 0, \sigma_2 > 0\}$$

$$\Theta = \{(\mu_1, \mu_2): -\infty < \mu_1, \mu_2 < \infty, \sigma_1 > 0, \sigma_2 > 0\}$$

Under Θ_0,

$$\hat{\mu} = \frac{m\bar{x} + n\bar{y}}{m+n}$$

and under Θ

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = \bar{y}$$

$$\lambda(x) = \frac{\underset{\Theta_0}{Sup\ L}}{\underset{\Theta}{Sup\ L}} = \frac{e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(x_i-\hat{\mu})^2-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n}(y_j-\hat{\mu})^2}}{e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(x_i-\bar{x})^2-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n}(y_j-\bar{y})^2}}$$

Now

$$\sum_{i=1}^{m}(x_i-\hat{\mu})^2 = \sum_{j=1}^{m}(x_i-\bar{x})^2 + m(\bar{x}-\hat{\mu})^2$$

$$= \sum_{i=1}^{m}(x_i-\bar{x})^2 + m\left(\frac{m}{m+n}\right)^2 (\bar{x}-\bar{y})^2$$

and

$$\sum_{j=1}^{n}(y_i-\hat{\mu})^2 = \sum_{j=1}^{n}(y_i-\bar{y})^2 + n\left(\frac{m}{m+n}\right)^2 (\bar{x}-\bar{y})^2$$

$$\therefore \lambda(x) = e^{-\frac{1}{2\sigma_1^2}m\left(\frac{m}{m+n}\right)^2(\bar{x}-\bar{y})^2-\frac{1}{2\sigma_2^2}n\left(\frac{m}{m+n}\right)^2(\bar{x}-\bar{y})^2}$$

Reject H_0 if

$$\lambda(x) \leq \boldsymbol{\lambda}$$

$or\ (\bar{x}-\bar{y})^2 \geq \lambda_1$

$$\left(\frac{\bar{x}-\bar{y}}{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right)^2 \geq \lambda_2, \left|\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\right| \geq \lambda_3$$

$Since\ \bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)$

$and\ Z = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0,1), under\ H_{0,}$

Therefore reject H_0 if $|z| > z_{\alpha/2}$

For case (ii) when $\sigma_1 = \sigma_2 = \sigma$ (unknown) we have

$$\Theta_0 = \{(\mu_1,\mu_2,\sigma_1,\sigma_2): \mu_1 = \mu_2 = \mu, \sigma_1 = \sigma_2 = \sigma > 0\}$$

$$\Theta = \{(\mu_1,\mu_2,\sigma_1,\sigma_2): -\infty < \mu_1,\mu_2 < \infty, \sigma_1 = \sigma_2 = \sigma > 0\}$$

Under Θ_0,

$\hat{\mu} = \frac{m\bar{x}+n\bar{y}}{m+n}$

$\hat{\sigma}^2 = \frac{1}{(m+n)}\left[\sum_{i=1}^{m}(x_i-\hat{\mu})^2 + \sum_{i=1}^{m}(y_i-\hat{\mu})^2\right]$

and under Θ_0

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = \bar{y}$$

$\hat{\sigma}^{*^2} = \frac{1}{(m+n)}\left[\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{j=1}^{m}(y_i-\bar{y})^2\right] = \frac{(m+n-2)S^2}{m+n}$

$$where\ (m+n-2)S^2 = \sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{j=1}^{m}(y_i-\bar{y})^2$$

$$\therefore \boldsymbol{\lambda}(\boldsymbol{x}) = \left(\frac{\hat{\sigma}^{*2}}{\hat{\sigma}^2}\right)^{\frac{\boldsymbol{m+n}}{\boldsymbol{2}}}$$

$$\hat{\sigma}^2 = \frac{1}{m+n}\left[\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{j=1}^{m}(y_i-\bar{y})^2 + m\left(\frac{n}{m+n}\right)^2(\bar{x}-\bar{y})^2 + n\left(\frac{m}{m+n}\right)^2(\bar{x}-\bar{y})^2\right]$$

$$= \hat{\sigma}^{*2} + (\bar{x}-\bar{y})^2\frac{mn}{(m+n)^2}(n+m)$$

$$= \hat{\sigma}^{*2} + (\bar{x}-\bar{y})^2\left(\frac{1}{\frac{1}{m}+\frac{1}{n}}\right)\frac{1}{(m+n)}$$

and

$$\frac{\hat{\sigma}^2}{\hat{\sigma}^{*2}} = 1 + \left(\frac{\bar{x}-\bar{y}}{\hat{\sigma}^*\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}}\right)^2\frac{1}{(m+n)}$$

Reject H_0 if

$$\lambda(x) \le \boldsymbol{\lambda}$$

$$or\ \left(\frac{1}{1+\frac{\hat{\sigma}^2}{\hat{\sigma}^{*2}}}\right)^{\frac{m+n}{2}} \le \boldsymbol{\lambda}$$

$$\left(\frac{\hat{\sigma}^2}{\hat{\sigma}^{*2}}\right) \ge \lambda_1$$

$$\left(\frac{\bar{x}-\bar{y}}{s\left(\frac{1}{m}+\frac{1}{n}\right)}\right)^2 \ge \lambda_2$$

$$\left|\frac{\bar{x}-\bar{y}}{s\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}}\right| \geq \lambda_3$$

Where $\lambda_3 = t_{m+n-2,\frac{\alpha}{2}}$ the table value of t at df (m+n-2) for $\alpha/2$

That is reject H_0 if

$$\left|\frac{\bar{x}-\bar{y}}{s\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}}\right| \geq t_{m+n-2,\frac{\alpha}{2}}.$$

6. For Problem 5 above, use LRT method to obtain best critical region for testing

$(a)\ H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

$(b)\ H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

When (i) σ_1, σ_2 known and (ii) $\sigma_1 = \sigma_2 = \sigma$ (unknown).

10.6 Summary

This unit is exclusively the application on Neyman-Pearson lemma and likelihood ration criterion for the construction of tests in one-sample and two-sample problems. In on-sample problem, we have a single normal population from which the sample has come out and we wish to test for the mean of the population against one or two sided alternatives. The discussion has been done separately for the situations when the population variance is known or unknown. In two-sample problems, we have two normal populations from which the samples have come out. Here again we consider the problem of testing the equality of means against both one and two sided alternatives under various assumptions on population variance.

10.7 Further Readings

- Goon A.N., Gupta M.K. & Das Gupta B (2000) *An Outline of Statistical Theory* Vol. 2 The World Press Private Limited.

- Goon A.N., Gupta M.K. & Das Gupta B (2000) *Fundamentals of Statistics* Vol. I The World Press Pvt. Ltd., Kolkata.
- Hogg, R.V. and Craig, A. (2005). *Introduction to Mathematical Statistics* 6th edition, Prentice Hall.
- Mood, A.M. Graybill, F.A., Boes, D.C. (1974). *Introduction to the Theory of Statistics*, McGraw Hill.
- Mohr, L.B. (1994), *Understnading Significance Testing*, Sage Publications, USA.
- Lehmann, E.L. (1986). *Testing statistical hypothesis*. Springer-Verlag, New York, Inc. 2nd edition.

Unit-11 Interval Estimation

Structure

11.1 Introduction

Estimation of an unknown parameter or unknown parametric function by a single value calculated on the basis of a random sample of size n drawn from the parent population is referred to as point estimation. A single estimator, however good it may be is not expected to coincide with the true value of parameter and the distribution of errors is determined by the sampling distribution of the estimator. It is, therefore, desirable to give for any population parameter an interval in which the population parameter may be expected to lie, with a specific degree of confidence. The procedure of providing an interval within which the true value of parameter may be expected to lie with some degree of confidence in terms of probability, is called the interval estimation and the intervals for the parameters are called confidence intervals. The degree of confidence in terms of probability is known as confidence coefficient. For example, an electronic charge may be estimated to be $(4.770 \pm 0.005)\ 10^{-10}$ electronic unit with the idea that the first factor is very unlikely to be outside the range (4.765 to 4.775). A cost accountant for a publishing company in trying to allow for all factors which enter into cost of production of a certain book may estimate the cost to be Rs. 83 ± 40 paise per book with the implication that the correct cost very probably lies between (82.60, 83.40) per book. The Central Statistics Organization may estimate the number of unemployed in a certain area to be (12.4 ± 0.5) millions at a given time, feeling sure that the actual number is between (11.9 ± 12.9) million. This indicates in the form intervals.

Let θ be the parameter and T be a statistic based on a random sample of size n from the corresponding population. Let T be sufficient for θ. In many cases it may be possible to find a function say $\psi(T, \theta)$ whose distribution is independent of θ.

The statement

$$\psi_{1-\frac{\alpha}{2}} \leq \psi(T, \theta) \leq \psi_{\frac{\alpha}{2}}$$

can often be written in an equivalent form as, say

$$\theta_1(T) \leq \theta \leq \theta_2\ (T)$$

$$\Rightarrow P\{\theta_1(T) \leq \theta \leq \theta_2\ (T)\} = P\left\{\psi_{1-\frac{\alpha}{2}} \leq \psi(T,\theta) \leq \psi_{\frac{\alpha}{2}}\right\} = 1-\alpha$$

Whatever the true value of θ may be $\theta_1(T) and\ \theta_2\ (T)$ will then be called confidence limits to θ with confidence coefficient 1- α.

11.2 Objectives

This unit will enable you to understand the:

- Problem of interval estimation
- Some applications of interval estimation
- Concept of shortest confidence interval

11.3 Confidence Interval and Confidence Coefficient

Let T_1 and T_2 be two statistics based on the random sample of size n and let $T_1 < T_2$ such that

$$P\{T_1 > \psi(\theta)\} = \alpha_1$$

$$P\{T_2 > \psi(\theta)\} = \alpha_2$$

Where

$$0 < \alpha_i < 1\alpha_1 + \alpha_1\ and\ 0 < \alpha < 1$$

Where α_1 and α_2 are independent of θ. We may have the result that

$$P[T_1 \leq \psi(\theta) \leq T_2] = 1-\alpha\ for\ all\ \theta\varepsilon\Omega.$$

This probability statement does not mean $\psi(\theta)$ is a random variable and, in fact, it is constant. This statement implies that the probability is $(1-\alpha)$ that the random interval $[T_1, T_2]$ will cover, whatever the true value of θ may be. Let us take α to be. 05 so that $1-\alpha$ = 0.95. Let T_1 and T_2 takes the value $T_1(x) = t_1$ and $T_2(x) = t_2$ for a specific sample $\underline{x}$. Then in about 95% of the cases the interval $[t_1, t_2]$ will include $\psi(\theta)$and it will fail to do so in about 5% of the cases only. Hence higher the probability $(1-\alpha)$, the more confidence one has, for a given set of observation that the interval $[t_1, t_2]$ will actually cover the true value of $\psi(\theta)$.

Corresponding to the given set of observation x, the interval (t_1,t_2) is called a confidence interval for $\psi(\theta)$, t_1 is called the lower and t_2, the upper confidence limit to $\psi(\theta)$. The number $(1-\alpha)$ which may be attached to the confidence limit to $\psi(\theta)$ is not to be regarded as probability. It serves as a measure of the confidence with which one may asserts that $\psi(\theta)$ lies between t_1 and t_2. Hence $(1-\alpha)$ is called the confidence coefficient associated with the confidence interval [t_1,t_2].

Example: Suppose that a random sample (2.2, 4.4, 1.6, 5.6) of four observations is drawn form a normal population with an unknown mean μ and a known standard deviation 3. The maximum-likelihood estimate of μ is the mean of the sample observations: x = 3.7.

In general for samples of size 4 from the given distribution the quantity.

$$z = \frac{\overline{x}-\mu}{3/2}$$

will be normally distributed with mean zero and unit variance, $\overline{x}$ is the sample mean and 3/2 is $\frac{\sigma}{\sqrt{n}}$. Thus quantity z has a density

$$f_z(z) = \emptyset(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

which is independent of the true value of unknown parameter. We can compute the probability that z will be between any two arbitrarily chosen numbers.

$$P[-1.96 < z < 1.96] = \int_{1.96}^{1.96} \emptyset(z)dz = 0.95$$

In this relation the inequality

$$-1.96 < z \ \ or -1.96 < \frac{\overline{x}-\mu}{3/2}$$

is equivalent to the inequality

$$\mu < \bar{x} + \frac{3}{2}(1.96) = \bar{x} + 2.94$$

and the inequality z < 1.96 is equivalent to $\mu > \ \bar{x} - 2.94$ and this is equivalent to

$$P(\bar{x} - 2.94 < \mu < \bar{x} + 2.94) = 0.95$$

and subtracting 3.7 for $\bar{x}$ we obtain the interval (0.76, 6.64). It is observed that the random interval $(\bar{x} - 2.94, \bar{x} + 2.94)$ and the interval (+0.76, 6.64) are each called a confidence interval or more precisely a 95 percent confidence interval. The interval (+0.76, 6.64) is the value of the random interval $(\bar{x} - 2.94, \bar{x} + 2.94)$ when $\bar{x}$=3.7

Remark: The interpretation runs as follows. The probability that the random interval $(\bar{x} - 2.94, \bar{x} + 2.94)$ covers the unknown true mean μ is 0.95. That is, if samples of size 4 were repeatedly drawn from the normal population and if the random interval $(\bar{x} - 2.94, \bar{x} + 2.94)$ were computed for each sample, then the relative frequency of those intervals that contain the true unknown mean μ would approach 95 percent. We therefore have considerable confidence that the observed interval, here (+0.76, 6.64) covers the true mean. The measure of out confidence is 0.95 because before the sample was drawn 0.95 was the probability that the interval that we construct would cover the true mean 0.95 is called the confidence coefficient.

Intervals with any desired degree of confidence between 0 and 1 can be obtained.

$$P[-2.58 < z < 2.58] = 0.99$$

a 99 percent confidence interval for the true mean is obtained by converting the inequalities as before to get

$$p[\bar{x} - 3.87 < \mu < \bar{x} + 3.87] = 0.99$$

and substituting 3.7 for $\bar{x}$ to get the interval (-0.17, 7.50). It is to be observed that there in fact, many possible intervals with the same probability (with the same confidence coefficient)

$$P[-1.68 < z < 2.7] = 0.95$$

another 95 percent confidence interval for μ is given by the interval (-.35, 6.22). This interval is inferior to the one obtained before, because its length 6.57 is greater than the length 5.88 of the interval (+0.76, 6.64) it gives less precise information about the location of μ. Any two numbers a and b such that 95 percent of the area under $\emptyset$(z) lies between a and b will determine a 95 percent confidence interval.

Definition of Confidence Interval

Let us suppose that we have a random sample $x_1, x_2, \ldots x_n$ from a density $f(x;\theta)$ characterized by θ. Let $T_1 = t_1(x_1, x_2, \ldots x_n)$ and $T_2 = t_2(x_1, x_2, \ldots\ldots, x_n)$ be two statistic satisfying $T_1 < T_2$ for which

$$P[T_1 < \psi(\theta) < T_2] = 1 - \alpha$$

Where $(1-\alpha)$ does not depend on ; thus the random interval (T_1, T_2) is called a $(1-\alpha)$% confidence interval for $\psi(\theta)$, $(1-\alpha)$ is called confidence coefficient and T_1 and T_2 are called lower and upper confidence limits respectively for $\psi(\theta)$. A value (t_1,t_2) of the random interval (t_1,t_2) is also called $(1-\alpha)$% confidence interval for $\psi(\theta)$. We can observed that one or the other but not of the two statistics t_1 (x_1,……x_n) and t_2 (x_1,…..x_n) may be constant; that is one of two end points of the random interval (t_1,t_2) may be constant.

Let T_1 = t_1 (x_1, x_2,…..x_n) be a statistic for which p[T_1< $\psi(\theta)$] = $1-\alpha$, then t1 is called a one sided lower confidence limit for $\psi(\theta)$. Similarly, let T_2 =t_2 (x_1,….,x_n) be a statistic for which p[$\psi(\theta)$<T_2] = $1-\alpha$, then T_2 is called one sided upper confidence limits for $\psi(\theta)$.

If a confidence interval for θ has been determined, then, in fact, a family of confidence intervals has been determined. Precisely, for a given $(1-\alpha)$% confidence interval estimator of θ, a $(1-\alpha)$% confidence interval estimator of $\psi(\theta)$ can be obtained. Then $(\psi(T_1)<\psi(T_2))$ is $(1-\alpha)$% confidence interval for $\psi(\theta)$ since

$$P[\psi(T_1)<\psi(T_2)] = p[T_1<\theta<T_2] = 1-\alpha$$

11.4 Confidence Interval for the Sample Mean from a Normal Population

Let the random variables x_1, x_2,…..x_n denote respectively the outcomes to be obtained on these n repetitions of the experiment. Let the random sample be from N(μ,σ^2) σ^2known . Then it is known that x is distributed as N($\mu,\sigma^2/n$) and $\frac{(\bar{x}-\mu)}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ is N (0,1)

Thus

$$P\left(-1.96<\frac{(\bar{x}-\mu)}{\left(\frac{\sigma}{\sqrt{n}}\right)}<1.96\right)=0.95$$

The events:

$$-1.96<\frac{(\bar{x}-\mu)}{\left(\frac{\sigma}{\sqrt{n}}\right)}<1.96$$

$$-1.96\frac{\sigma}{\sqrt{n}}<\bar{x}-\mu<1.96\frac{\sigma}{\sqrt{n}}$$

and

$$\bar{x}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$$

are equivalent and these events have the same probability.

$$P\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

σ is known and each of the random variables

$(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}})$ and $(\bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$ is a statistics. The interval $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ is a random interval. This implies in repeated independent performances of the experiment, The probability is 0.95 that the random intervals $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ includes the unknown fixed point μ.

Let us look into the situation when σ^2 is not known.

It is known that $\frac{Ns^2}{\sigma^2}$, where $s^2$is the variacen of a random sample of size n from a distribution N(μ, σ^2) is $\chi^2_{(n-1)}$.

Implying $\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$to be N(0,1) $\frac{Ns^2}{\sigma^2}$ to be $\chi^2_{(n-1)}$ and the two are stochastically independent. We have

$$t = \frac{\left[\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma}\right]}{\left[\frac{\sqrt{Ns^2}}{\sigma^2(\bar{x} - \mu)}\right]} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}}$$

has a t- distribution with (n-1) degree of freedom. We can find two numbers t_1 and t_2 such as

$$P\left(t_1 < \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}} < t_2\right) = 0.95$$

Since t is a symmetrical distribution, we would take $t_1 = t_2$, $t_2 > 0$, with $t_1 = t_2$

$$P\left(\bar{x} - \frac{t_1 S}{\sqrt{n-1}} < \mu < \bar{x} + \frac{t_1 S}{\sqrt{n-1}}\right) = 0.95$$

Then the interval $\left[\bar{x} - \frac{t_1 S}{\sqrt{n-1}}, \bar{x} + \frac{t_1 S}{\sqrt{n-1}}\right]$is a random internal having probability 0.95 of including the unknown fixed point μ. if the experimental values of $x_1, x_2, \ldots . x_n$ are $x_1, x_2, \ldots . x_n$ with $S^2 = \sum \frac{(x_i - x)^2}{\sqrt{n}}, where\ \bar{x} = \sum \frac{X_i}{n}$, then the interval $\left[\bar{x} - \frac{t_1 S}{\sqrt{n-1}}, \bar{x} + \frac{t_1 S}{\sqrt{n-1}}\right]$is a 95 percent confidence interval for μ everyσ^2 > 0. This interval of μ is found by adding and subtracting a quantity, here $\frac{t_1 S}{\sqrt{n-1}}$ to the point estimate $\bar{x}$.

11.5 Confidence Intervals for Differences of Means from two Normal Population

Let x_{11}, x_{12},…..x_{1n1} and x_{21}, x_{22},…..x_{2n2} denotes independent samples from the two independent distributions N(μ_1, σ^2) and N(μ_2, σ^2). Let us denote the means of the samples by $\bar{x}_1$and $\bar{x}_2$and the variances by $s_1^2$and s_2^2respectively. $\bar{x}_1$and $\bar{x}_2$are normally and stochastically independently distributed with means μ_1 and μ_2 and variances $\frac{\sigma^2}{n_1}$ and $\frac{\sigma^2}{n_2}$ respectively and their differences ($\bar{x}_1$- $\bar{x}_2$) is normally distributed with mean $(\mu_1 - \mu_2)$ and variance $\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$.

Then the random variable

$$\frac{(\bar{x}_1 - \bar{x}_2)(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

is normally distributed with zero mean and unit variance $n_1 \frac{S_1^2}{\sigma^2}$ and $n_2 \frac{S_2^2}{\sigma^2}$ have stochastically independent chi-square distribution with (n_1-1) and (n_2-1) degrees on freedom so that their sum $\frac{\left(n_1 \frac{S_1^2}{\sigma^2} + n_2 \frac{S_2^2}{\sigma^2}\right)}{\sigma^2}$ has a chi-square distribution with (n_1+n_2-2) degrees of freedom. Then the random variable.

$$T = \frac{(\bar{x}_1 - \bar{x}_2)(\mu_1 - \mu_2)}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Has a t- distribution with $n_1 + n_2 - 2$degrees of freedom. We can find a positive number t_1 from t- table such that

[P-t_1< T <t_1)=0.95]

$$R = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and the probability may be written in the form

$$P[(\bar{x}_1 - \bar{x}_2) - t_1 R < \mu_1 - \mu_2 < (x_1 - x_2) + t_1 R] = 0.95$$

i.e. the random interval

$$\left[(\bar{x}_1 - \bar{x}_2) - t_1 \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right]$$

$$\left[(\bar{x}_1 - \bar{x}_2) - t_2 \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right]$$

has probability 0.95 of including the unknown fixed point $(\mu_1 - \mu_2)$. In practical situations, the experimental values of x_1, x_2, s ♠$, s \leftrightarrow namely\ x_1, x_2, s$♠$, s$ ↔will provide 95 percent confidence interval for $(\mu_1 - \mu_2)$ when the variance of the two independent normal distributions are unknown but equal.

11.6 Problems and Exercises

1. Explain interval estimation. In what respect, it is better than point estimate?
2. What is confidence interval? Explain the terms : confidence limits, confidence coefficient.
3. Obtain 95% confidence interval for the mean of a normal population when (i) variance is known (ii) variance is unknown.
4. Obtain confidence interval for difference of means of two normal populations when the variance are known.
5. A random sample of size 25 is drawn from a normal population N(μ,16). The mean of sample is calculated as 12.5, obtain 95% C.I. for μ.

11.7 Summary

An interval estimate describe range of values within which a population parameter is likely to lie. A point estimate does not give an idea that how far the estimates deviates from the true value of parameter being estimated. Confidence intervals provide the interval estimates in which the true value of parameter is expected to lie with pre-assigned confidence coefficient (probability). In practice, most commonly used confidence intervals are 95 percent and 99 percent.

11.8 Further Readings

- Goon, A.M.; Gupta, M.K., and Dasgupta, B.: An outline of Statistical Theory, Vol. 2. The World Press Private Limited, 2000.
- Goon, A.M.; Gupta, M.K., and Dasgupta, B.: Fundamental of Statistics, Vol. I. The World Press Private Limited, 2000.
- Mood, A.M., Graybill, F.A.; Boes, D.C.: Introduction to the theory of Statistics, McGraw Hill, 1974.

Unit-12 Shortest and Shortest Unbiased Confidence Interval

Structure

12.1 Introduction

We have identified the statistic T_1 and T_2 such that $\alpha_1 = \alpha_2 = \alpha/2$ in constructing various confidence intervals. It is clear that $\alpha_1 and\ \alpha_2$may be chosen in infinitely many ways, each satisfying the conditions $\alpha_1 \geq 0$ and $\alpha_1 + \alpha_2 = \alpha$. Infinitely many confidence intervals are available with the same confidence coefficient (1-α). Some criterion is clearly called for in order that we may make a choice among this infinite set of confidence intervals. An obvious choice may be to give weight to the width of the interval. Let us suppose $T_1^{`}$ and T' are such that.

$$p\left[T_1^{`} < \psi(\theta) < T'\right] = 1 - \alpha.$$

Then the confidence interval given by $T_1^{`}$ and $T_2^{`}$ will be said to be better than the interval given by, say T_1 and T_2 which also satisfies the above stated condition, if

$$(T_1^{`} - T_2^{`}) \leq (T_2 - T_1) \quad \text{for all } \theta \in \Omega\ , \qquad \text{the parametric space}$$

If this holds for every other pair of statistics T_1 and T_2 then the confidence interval given by $T_1^{`}$ and $T_2^{`}$ will be known as shortest confidence interval for $\psi(\theta)$ based on the statistic T.

Remark: The confidence intervals constructed for sample mean, σ is known, σ is not known, confidence interval for difference between mean, confidence interval for the proportion and difference between proportions all belongs to this category of shortest confidence interval.

4.2 Neyman's Principle of Shortest Confidence Interval

Neyman has introduced a slightly different aim (different from the aim of minimizing the length of a confidence interval for a given confidence coefficient) in seeking a 'best' confidence interval. Neyman adopted the following criterion of goodness of confidence interval.

For a fixed confidence coefficient the confidence interval which minimizes (under the assumptions that θ is true) the probability of covering any other $\theta' s(\neq \theta)$.

Neyman's criterion of goodness of a confidence interval as enunciated above aborts to minimizing the probability of a confidence interval covering false values and thus is equivalent to minimizing the second kind of error in testing problems.

As a against this Wilks defines the following criterion of goodness of a confidence interval.

"For a fixed confidence coefficient minimize the length of confidence interval". (This has already been discussed above).

In what follows now, we shall adopt only Neyman's criterion of shortest confidence interval.

Let X be a r.v. with d.f. f(x, θ), θ is a single values parameter and the form of F is known. We denote the sample point x_1, x_2,….x_n (a random sample of size an drawn from the given distribution) by E, i.e., E= (x_1, x_2,….x_n).

Definition: $\delta_0(E) = (\underline{\theta}(E), (\bar{\theta}, (E))$ is the shortest confidence interval of confidence coefficient (1-α) for θ if

(i) $\delta_0(E)$ is confidence interval of confidence coefficient (1-α)

(ii) $Pr\delta_0(E) \subset \theta/\theta' \leq Pr(\delta(E) \subset \theta/\theta')$

For all θ and θ' for any other $\delta(E)$satisfying (i)

Notation: $\subset$ means 'covers'

If a shortest confidence interval exists it is surely the one to be used always. Unfortunately, such shortest confidence intervals exits in very rare cases (just as UMP tests exists in very are cases) and so further principle must be adopted by which one confidence interval may be preferred to another even when a shortest confidence interval does not exists. Neyman has advanced such principles and it is to be noted that development of such principle follows very closely to the development of tests of hypothesis. These principles are precisely those which when used for seeking for a good test when no UMP test exists such as unbiasedness etc.

12.3 Unbiased Confidence Interval

$\delta(E) = (\underline{\theta}(E), (\bar{\theta}, (E))$ is an unbiased confidence interval of confidence coefficient (1- α) for θ if

(i) $Pr(\delta(E) \subset \theta/\theta') = 1 - \alpha$

(ii) For each fixed θ, the function $\psi_\theta\ (\theta') = Pr(\delta(E) \subset \theta/\theta')$ has a maximum at $\theta = \theta'$

Remark: It can be easily verified that if {w(θ)} is one parameter fairly of unbiased critical region which give rise to a confidence interval then confidence interval is unbiased.

12.4 Shortest Unbiased Confidence Interval

$\delta_0(E)$ is a shortest unbiased coefficient (1-α) for θ if

(i) $\delta_0(E)$ is an unbiased confidence interval of confidence coefficient (1-α) for θ

(ii) $Pr(\delta_0(E) \subset \theta/\theta') \leq Pr(\delta(E) \subset \theta/\theta')$

For all $\theta\ used\ \theta'$& for any $\delta(E)$satisfying (i).

Example: In case of N(θ, σ^2); σ^2 known the confidence interval $(\underline{\theta}, \theta)$=$\left(\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + 1.96\sigma/\sqrt{n}\right)$ is a shortest unbiased confidence interval with confidence coefficient 0.95 for θ.

$\bar{x}$ is the sample mean.

12.5 Case of Discrete Random Variables

The case of discrete random variables requires to be separately dealt with. In this case one cannot hope to get for each α (0< α <1) a confidence interval that will have confidence coefficient exactly equal to (1-α).

One way to avoid this problem is to consider the confidence coefficient to be at least (1-α). Then the statistics T_1 and T_2 will provide confidence limits to parametric function ψ (θ) if

$$Pr[T_1 \leq \psi(\theta) \leq T_2] \geq (1 - \alpha)\ for\ all\ \theta \in \Omega\ and\ 0 < \alpha < 1$$

The actual determination of the confidence intervals may be carried out by drawing confidence belts.

Remark: it may be noted that it is generally not possible to get confidence limit in discrete cases with confidence coefficient exactly equal to (1-α).

The procedure described above may be used will make the interval unnecessarily wide.

The difficulty is comparable to the difficulty in obtaining a test for a hypothesis is the discreet case with the level of significance exactly equal to α, if one confines one self to tests based on critical regions. One overcomes this difficulty in testing of hypothesis by considering randomized tests. Here to a similar procedure may be applied on order to get confidence intervals with confidence coefficient exactly equal to (1- α), which would naturally be shorter. The method is due to Stevens. And for detailed study is referred to

Stevens, W.L. “Fiducial limits of the parameter of discontinuous distributions”, Biometrika 37 (1950) pp 117-129.

12.6 Problems and Exercises

1. What is length of a confidence interval? Explain confidence interval with shortest length.
2. Describe Neyman’s principle of shortest confidence interval.
3. Obtain confidence interval with shortest length for mean of normal population with known variance.

12.7 Summary

We can construct a number of confidence interval for any parameter with same confidence coefficient. In fact, there are infinite many confidence interval for a parameter with same probability of inclusion. Then, it is desirable to choose interval with expected shortest length and it should also unbiased.

12.8 Further Readings

- Mood, A.M., Graybill, F.A.; Boes, D.C.: Introduction to the theory of Statistics, McGraw Hill, international edition, 1974.
- Rao,C.R., Linear statistical inference and its applications, John Wiley and Sons, Inc.
- Wilks, mathematical statics, Jon Wiley and Sons.

- Kendall, Vol. 1,2,3. Hafner Publishing Company, New York.