COURSE MANUAL

Statistical Inference III

STA 321

University of Ibadan Distance Learning Centre
Open and Distance Learning Course Series Development

ISBN 978-021-275-2

General Editor: Prof. Bayo Okunade

University of Ibadan Distance Learning Centre
University of Ibadan,
Nigeria

Telex: 31128NG

Tel: +234 (80775935727)
E-mail: ssu@dlc.ui.edu.ng
Website: www.dlc.ui.edu.ng

Vice-Chancellor's Message

The Distance Learning Centre is building on a solid tradition of over two decades of service in the provision of External Studies Programme and now Distance Learning Education in Nigeria and beyond. The Distance Learning mode to which we are committed is providing access to many deserving Nigerians in having access to higher education especially those who by the nature of their engagement do not have the luxury of full time education. Recently, it is contributing in no small measure to providing places for teeming Nigerian youths who for one reason or the other could not get admission into the conventional universities.

These course materials have been written by writers specially trained in ODL course delivery. The writers have made great efforts to provide up to date information, knowledge and skills in the different disciplines and ensure that the materials are user-friendly.

In addition to provision of course materials in print and e-format, a lot of Information Technology input has also gone into the deployment of course materials. Most of them can be downloaded from the DLC website and are available in audio format which you can also download into your mobile phones, IPod, MP3 among other devices to allow you listen to the audio study sessions. Some of the study session materials have been scripted and are being broadcast on the university's Diamond Radio FM 101.1, while others have been delivered and captured in audio-visual format in a classroom environment for use by our students. Detailed information on availability and access is available on the website. We will continue in our efforts to provide and review course materials for our courses.

However, for you to take advantage of these formats, you will need to improve on your I.T. skills and develop requisite distance learning Culture. It is well known that, for efficient and effective provision of Distance learning education, availability of appropriate and relevant course materials is a *sine qua non*. So also, is the availability of multiple plat form for the convenience of our students. It is in fulfilment of this, that series of course materials are being written to enable our students study at their own pace and convenience.

It is our hope that you will put these course materials to the best use.

Prof. Abel Idowu Olayinka
Vice-Chancellor

Foreword

As part of its vision of providing education for "Liberty and Development" for Nigerians and the International Community, the University of Ibadan, Distance Learning Centre has recently embarked on a vigorous repositioning agenda which aimed at embracing a holistic and all encompassing approach to the delivery of its Open Distance Learning (ODL) programmes. Thus we are committed to global best practices in distance learning provision. Apart from providing an efficient administrative and academic support for our students, we are committed to providing educational resource materials for the use of our students. We are convinced that, without an up-to-date, learner-friendly and distance learning compliant course materials, there cannot be any basis to lay claim to being a provider of distance learning education. Indeed, availability of appropriate course materials in multiple formats is the hub of any distance learning provision worldwide.

In view of the above, we are vigorously pursuing as a matter of priority, the provision of credible, learner-friendly and interactive course materials for all our courses. We commissioned the authoring of, and review of course materials to teams of experts and their outputs were subjected to rigorous peer review to ensure standard. The approach not only emphasizes cognitive knowledge, but also skills and humane values which are at the core of education, even in an ICT age.

The development of the materials which is on-going also had input from experienced editors and illustrators who have ensured that they are accurate, current and learner-friendly. They are specially written with distance learners in mind. This is very important because, distance learning involves non-residential students who can often feel isolated from the community of learners.

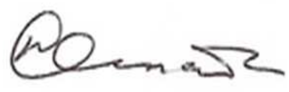
It is important to note that, for a distance learner to excel there is the need to source and read relevant materials apart from this course material. Therefore, adequate supplementary reading materials as well as other information sources are suggested in the course materials.

Apart from the responsibility for you to read this course material with others, you are also advised to seek assistance from your course facilitators especially academic advisors during your study even before the interactive session which is by design for revision. Your academic advisors will assist you using convenient technology including Google Hang Out, You Tube, Talk Fusion, etc. but you have to take advantage of these. It is also going to be of immense advantage if you complete assignments as at when due so as to have necessary feedbacks as a guide.

The implication of the above is that, a distance learner has a responsibility to develop requisite distance learning culture which includes diligent and disciplined self-study, seeking available administrative and academic support and acquisition of basic information technology skills. This is why you are encouraged to develop your computer skills by availing yourself the opportunity of training that the Centre's provide and put these into use.

In conclusion, it is envisaged that the course materials would also be useful for the regular students of tertiary institutions in Nigeria who are faced with a dearth of high quality textbooks. We are therefore, delighted to present these titles to both our distance learning students and the university's regular students. We are confident that the materials will be an invaluable resource to all.
We would like to thank all our authors, reviewers and production staff for the high quality of work.

Best wishes.



Professor Bayo Okunade
Director

Course Development Team

Content Authoring	O.I. Shittu
Content Editor	Prof. Remi Raji-Oyelade
Production Editor	Ogundele Olumuyiwa Caleb
Learning Design/Assessment Authoring	SkulPortal Technology
Managing Editor	Ogunmefun Oladele Abiodun
General Editor	Prof. Bayo Okunade

Table of Contents

Study Session 1: Statistical Inference

Introduction

In most scientific research, data are collected from a sample drawn from a given population and a number of summary statistics are obtained. But, no good scientist is content with a mere summary of his findings. More often than not, the interest is on drawing conclusions about the population from the sample data. Scientist may also want to ascertain with some degree of confidence that the effect of an independent variable is due to chance or really due to the variable on question.

In this study session, introduces you to the foundation of statistical inference; it explains the meaning of the basic terms and concepts. The various branches of statistical inference are introduced. They include the techniques of estimation and hypothesis tests.

Learning Outcomes for Study Session 1

At the end of this study session, you should be able to:

1.1 Explain the concept of Statistical Inference

1.2 Discuss Sampling Distribution.

1.1 Statistical Inference

Statistical inference is the act of generalising from the sample to a population with calculated degree of certainty probability.

In real life and in most statistical investigations, our goal is usually to measure some characteristics of a specified population. However, constraints such as time cost and the nature of the unit of inquiry may prevent us from such elaborate (census) exercise.

The power of statistical inference is that you can actually estimate the characteristics of a population from the characteristics of a random sample drawn from it. It also provides us with the methods and techniques for testing the validity (truth or otherwise) of a statistical (probability) statement. Thus, the essentials of statistical inference are estimation, hypothesis testing and confidence intervals.

Before we go further in our discussion, let us familiarise ourselves with some terms and concepts that would assist in understanding the content of this chapter and subsequent chapters of this book.

Population

This is the target group under investigation about which we want to gather information; for example, the population of children of school age.

Sample

A sample is a representative part of a population drawn according to probability rule.

Sample size

This is the number of units in the sample usually denoted by (small) n.

Parameter

This is some unknown but fixed quantity, usually characteristic of a population. It is usually denoted by θ or k(θ) in the parameter space Ω from a known probability density function (p.d.f0 $f(x;\theta)$. Examples of parameter are population mean (μ) and the population variance (σ^2).

Statistic

This refers to a quantity that depends on data usually computed from the sample. A statistic T(x) is a function of the sample X_1, ,,, X_n. Examples of statistic are the sample mean $(\bar{X})$, sample variance (S^2), the smallest order statistic $X_{(1)}$ and the largest order statistic $X_{(n)}$.

Estimator

Since the parameters are unknown, the onus is to use one of the statistics to approximate its value. Such statistic is known as an estimator. The numerical value of the estimator is called the **estimate**. E.g. the sample mean $\bar{X}$ is an estimator of the population mean μ.

Invariant Estimator

An estimator could be location invariant or scale invariant

a. **Location Invariant**

Given a(x_1, x_2……x_n) is a r.s. of size n, an estimator t(x_1, x_2……x_n) is said to be location invariant if

$$t(x_1 + C)\, x_2 + C, \ldots\ldots x_n + C\ ) = t(x_1, x_2 \ldots\ldots\ldots\ldots x_n) + C$$

b. **Scale Invariant**

An estimator t (x_1, x_2 ……x_n) is scale invariant if

$$t(Cx_1, Cx_2 \ldots\ldots\ldots\ldots Cx_n) = C\, t\,(x_1\ x_2 \ldots\ldots\ldots\ldots x_n)$$

In general, an estimator $\hat{\theta}$ of θ is invariant if the estimator of a function $\tau(\hat{\theta})$ is $\tau(\theta)$ where τ is a one valued function.

In-Text Question

What is Sample?

In-Text Answer

Sample is a descriptive part of a population drawn according to probability rule.

Example:

Is standard deviation a location invariant?

Solution

$$\sigma(x_1, x_2 \ldots\ldots\ldots\ldots x_n) = \sigma$$

$$\sigma(x_1 + C, x_2 + C \ldots\ldots\ldots\ldots x_n +) \neq \sigma + C$$

$$\text{but} = C\sigma$$

So standard deviation is not a location invariant but scale invariant.

1.2 Sampling Distribution

This refers to the probability distribution of the sample statistics assuming that the sample is drawn at random from the population. For example, if $E(X) = \mu$, $Var(X) = \sigma^2$, then we obtain $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$ meaning that among all possible sample of size n, the sample mean $\bar{X}$ will vary in such a way that its mean is μ and its variance is $\frac{\sigma^2}{n}$. [You can get more detailed explanation of this concept in our book on statistical inference III

Sampling error

This is the measure of the difference between a sample statistic and the real population value. For example, the difference between the sample mean $\bar{X}$ and the population mean μ is measured by the standard error of means $S.E_{(\text{mean})} = \sigma_{(\bar{x})} = \frac{\sigma}{\sqrt{n}}$

(a function of the sample size).

Estimation

This is the act of providing the most likely location of a population parameter with calculated degree of error.

Let X_1, ,,, X_n be identically and independently distributed sample from a population with the distribution $f(x/\theta)$, where θ is unknown. Estimation involves the construction of a good estimator for θ from the sample values.

In-Text Question

What is Sampling Error?

In-Text Answer

This is the measure of the difference between a sample statistic and the real population value.

There are two types of estimation:

a. point estimation, and
b. Interval estimation.

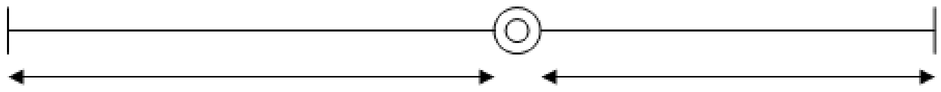
a. **Point Estimation:**

This involves the construction (or computation) of a single value statistic from observations realised from a population that helps to locate or can be approximated under certain assumptions to the population parameter.

Definition: A point estimator of θ is $T(x) = \hat{\theta} = t(x_1, x_2, ..., x_n)$ that can approximate a single value population parameter. Any statistic ($\hat{X}, S^2, S$) is a point estimator.

b. **Interval estimation:**

Point estimation provides no information about the precision of an estimate. Thus, interval estimation provides two limits within which the population parameter is expected to lie with a given level of confidence. It is usually constructed by surrounding the point estimate with a margin of error so that there is an upper confidence limit and lower confidence limit.



The above interval is called a confidence interval.

(We shall discuss this concept fully in chapter-----)

Hypothesis

This is a widely held belief about a phenomenon (e.g. mean age of goats, average output and so on) which has not been validated by statistical test.

Hypothesis test

This is a procedure for measuring the strength of the evidence provided by data against a widely held belief (i.e. a hypothesis). In a scientific inquiry, the hypothesis of interest is formulated as a hypothesis concerning the values of an unknown parameter in a specified probability model.

A test statistic is then developed to obtain a value to be compared with a theoretical value on the basis of which a decision is taken about whether or not to reject the true hypothesis. Details of this procedure are given in chapter ………….

In-Text Question

Define Point Estimator?

In-Text Answer

A point estimator of θ is $T(x) = \hat{\theta} = t(x_1, x_2, ..., x_n)$ that can approximate a single value population parameter.

Summary for Study Session 1

In this study session, you have learnt that;

1. You have discussed the definition of basic terms and concepts necessary for better understanding of this lecture and other modules.
2. You also gave the basic logic of statistical inference, defining its nature and scope.

Self-Assessment Questions (SAQs) for Study Session 1

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Is standard deviation a location invariant?
2. What is Parameter?

Note SAQ (Solution)

1. $\sigma(x_1, \mathrm{x}_2 \ldots\ldots\ldots\ldots x_n) = \sigma$

$$\sigma(x_1 + \mathrm{C}, \mathrm{x}_2 + C \ldots\ldots\ldots\ldots x_n +) \neq \sigma + C$$

but = Cσ

So standard deviation is not a location invariant but scale invariant.

2. This is some unknown but fixed quantity, usually characteristic of a population. It is usually denoted by θ or k(θ) in the parameter space Ω from a known probability density function (p.d.f0 $f(x;\theta)$.

Notes on SAQ

1. Statistical inference is that you can actually estimate the characteristics of a population from the characteristics of a random sample drawn from it. It also provides us with the methods and techniques for testing the validity (truth or otherwise) of a statistical (probability) statement.
2. Estimation is the act of providing the most likely location of a population parameter with calculated degree of error.

   Let X_1, ,,, X_n be identically and independently distributed sample from a population with the distribution $f(x/\theta)$, where θ is unknown. While Hypothesis Test is a procedure for measuring the strength of the evidence provided by data against a widely held belief (i.e. a hypothesis).

3. The difference between the sample mean $\overline{X}$ and the population mean μ is measured by the standard error of means $S.E_{(\text{ mean})} = \sigma_{(\bar{x})} = \dfrac{\sigma}{\sqrt{n}}$

Sampling Distribution **is** probability distribution of the sample statistics assuming that the sample is drawn at random from the population

References

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics,* 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference*.

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Freund, J.E (1992): *"Mathematical Statistics" 5th edition,* Prentice Hall, New Jersey.

Bhat, B.R and Shah, D.K (1997): "Statistical Inference" Unpublished lecture note in the Department of Statistics, University of Botswana, Gaborone, Botswana.

Study Session 2: Properties of a Good Estimator

Introduction

The aim of this study session is to introduce you to the criteria to be adopted in comparing different estimators. You shall be discussed with examples the important properties of a good estimator such as unbiasedness, efficiency, consistency, sufficiency and completeness.

Learning Outcomes for Study Session 2

At the end of this study session, you should be able to:

2.1 Discuss the properties of a good Estimator.

2.1 Good Estimator

An estimator is considered good or better than the other, if it is the most appropriate in a given situation, if it will expose us to the smallest risk (variance) and will give us the most information at the lowest cost.

Above all, a good estimator should possess the following properties:

1. Unbiasedness,
2. Minimum variance,
3. Efficiency,
4. Consistency,
5. Sufficiency
6. Robustness.

You shall discuss each of these properties one after the other, illustrating with examples.

2.1.1 Unbiasedness

An estimator (T) is said to be unbiased if its expectation is equal to the population value (parameter).

Definition

For a random sample, $X_1, X_2, \ldots X_n$ from the distribution $f(x,\theta)$, a statistic
$T = t(x_1, x_2 \ldots.., x_n)$ is an unbiased estimator at θ if $E(T)=\theta$ i.e. $E(T)-\theta=0$
In other words, an estimator T is biased if $E(T)\neq\theta$

Example
Let $x_1, x_2, \ldots x_n$ be a random sample from a normal distribution $N(\mu,\sigma^2)$, show the $\overline{X}$ is an unbiased estimator of μ .

Solution

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

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

$$\frac{1}{\text{n}}.n\,\mu$$

$$=\mu \quad (\text{i.e. } \overline{\text{X}} \text{ is unbiased for } \mu)$$

Example 2

Let $x_1, x_2, \ldots x_n$ be a random ample from the binomial distribution B(n, P), show that the sample proportion, $\frac{X}{n}$ is an unbiased estimator of P.

Solution

Since $E(X)=nP,$ it follows that

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

$$=\frac{1}{\text{n}}.nP$$

$$=\text{P}$$

Hence $\frac{X}{n}$ is an unbiased estimate of P.

Example 3

Let $\{X_i\}; \text{i} = 1, 2 \ldots \text{n}$ be a random sample from a normal distribution $N(\mu, \sigma^2)$, derive the expectation of variance $\hat{S}^2 = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)2/n$ and show that is it a biased estimator of σ^2.

In-Text Question

What is an Unbiasedness?

In-Text Answer

An estimator (T) is said to be unbiased if its expectation is equal to the population value (parameter).

Solution

Let

$$S^2 = \hat{\sigma}^2 = \frac{1}{n}\sum\left(X_i - \overline{X}\right)^2$$

$$n\hat{\sigma}^2 = \sum (X_i - \mu) - \left(\overline{X} - \mu\right)^2$$

$$nE\left(\hat{\sigma}^2\right) = \sum\left[\sum (X_i - \mu)^2 - \sum\left(\overline{X} - \mu\right)^2\right] \text{ Since } \sum (X_i - \mu) = 0$$

$$nE\left(\hat{\sigma}^2\right) = n\sigma^2 - n\frac{\sigma^2}{n}$$

$$= (\text{n -1})\sigma^2$$

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

hence $\hat{\sigma}^2$ is a biased estimator of σ^2, why? To calculate σ^2, we have to extract the mean, assuming 1 degree of freedom. So we do not have n independent estimates of dispersion about the means; we have (n–1) degrees of freedom.

2.1.2 Minimum Variance

This is a useful criterion for assessing the quality of an estimator $\hat{\theta}$. A good estimator must have minimum variance in the class of estimators.

This criterion is usually that of an unbiased estimator. But sometimes no unbiased estimator exists. Therefore we define another concept called the Means Square Error.

Definition 1

Let $\hat{\theta}$ be a biased estimate of θ, then the Mean Square Error (MSE) of θ is the expected squared deviation of the estimator from the true value, that is

$$MSE = E(T - \theta)^2$$

Definition 2

For a random sample $x_1, x_2, \ldots x_n$ from $f(x, \theta)$, and a statistic T = t ($x_1, x_2, \ldots x_n$) which is an estimator of θ, the mean square error (MSE) is defined as $MSE = E(T - \theta)^2$.

The above can be expressed alternatively:

$$MSE = E(T - \theta)^2 = E[T - E(T) + E(T) - \theta]^2$$
$$= E[(\mathrm{T} - \mathrm{E}(\mathrm{T}))]^2 + [E(T) - \theta]^2$$
$$\text{Since } E(\mathrm{T} - E(\mathrm{T})) = 0$$
$$\mathrm{Var}(\mathrm{T}) + b_T^2(\theta) \quad \text{for } b_T^2\theta > 0$$

Note that: $b_T^2(\theta)$ is called the bias of the estimator T.

While the mean square error is a reasonable way to assess the quality of an estimator, it does not lead to a useful criterion.

Thus, the estimator that minimizes the MSE is simply a good estimator.

$$MSE_{(T)} = Var(T) + b_T^2(\theta)$$

If T is unbiased, then $b_T^2(\theta) = 0$;

That is MSE is minimum when $MSE = Var(T)$

An estimator that satisfies the above condition is called the Minimum Variance Unbiased Estimator.

Example 4:

Consider the problem of estimation of σ^2 based on a random sample of size n from $N(\mu,\sigma)$ (see Post-test question)

$$
\begin{aligned}
E(S^2) &= E\left[S^2 - E(S^2)\right]^2 \\
&= E\left[\frac{1}{\text{n -1}}\sum\left(X-\overline{X}\right)^2 - \sigma^2\right]^2 \\
&= \frac{1}{(\text{n -1})^2}\sum\left[(n-1)S^2 - \sigma^2\right]^2 \\
&= \frac{1}{(\text{n -1})^2}\left[\sum (n-1)^2 \text{S}^4 + (n-1)^2\sigma^4 - 2(n-1)n\text{S}^2\sigma^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{(\text{n -1})^2}\left[(n-1)^2 S^4 + (n-1)^2\sigma^4 - 2(n-1)n\sigma^4\right] \\
&= \frac{1}{(\text{n -1})^2}\left[2(n-1)^2\sigma^4 - 2n(n-1)\sigma^4\right] \\
&= \frac{1}{(\text{n -1})^2}\ 2(\text{n -1})\sigma^4(n-1-n) \\
&= \frac{2\sigma^4}{n-1}
\end{aligned}
$$

The variance of the sample variance S^2 .

Recall from example 3 that the expectation of the population variance is: $E\left(\hat{\sigma}^2\right) = \frac{n-1}{n}\sigma^2$.

The variance of the population variance is derived as:

$$
\begin{aligned}
Var\left(\hat{\sigma}^2\right) = \frac{(n-1)^2}{n^2}Var\left(S^2\right) &= \frac{(n-1)^2}{n^2}\left(\frac{2\sigma^4}{n-1}\right) \\
&= \frac{2(\text{n -1})}{n^2}\sigma^4
\end{aligned}
$$

Comparing the variances, we can see from the above that the

$$MSE\left(\hat{\sigma}^2\right) = Var\left(\hat{\sigma}^S\right) + bias\left(\hat{\sigma}^2\right)$$

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

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

We can see from the above that

$$MSE(S^2) > MSE(\sigma^2)$$
$$MSE(\hat{S}^2) > MSE(\sigma^2)$$
$$i.e.\ \frac{2\sigma^4}{\text{n - 1}} > \left(\frac{2n-1}{n^2}\right)\sigma^4$$

Therefore, the variance of the population, $\hat{\sigma}^2$ is minimum even though its estimator is biased.

2.1.4 Consistency

This is another desirable property, which is asymptotic property. Asymptotic means the truth. An estimator is consistent if, a n gets large, the probability that the estimator T_n lies arbitrarily close to the parameter θ, being estimated becomes arbitrarily close to 1.

Definition

$T_n = t(x_1, x_2, \ldots x_n)$ is a consistent estimator of θ if

$$\lim_{n\to\infty} P[|T_n - \theta| \geq \varepsilon] = 0 \quad \text{for any } \varepsilon > 0$$

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

This is often referred as convergence in probability of T_n to θ.

Example 5

Show that the sample variance $S^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$ is a consistent estimate of σ^2.

Solution

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

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

i.e. $\lim_{n\to\infty}\left(\frac{2\sigma^4}{n-1}\right) \to 0$

Thus, we have shown that S^2 is a consistent estimator of the variance of a normal population.

Example 6

If T_n is an unbiased estimator for θ and $\sigma_n^2 = Var(T_n)$, show that T_n is a consistent estimate for θ.

Solution

By Chebyshev's inequality

$$P[|T_n - \theta| < \delta\sigma_n] > 1 - \frac{1}{\delta^2};\ \delta > 0$$

Choosing $\delta\sigma_n = \in$ as being fixed, then

$$P(|T_n - \theta| > \varepsilon) > 1 - \frac{\sigma_n^2}{\varepsilon^2} \to 1$$

$$\text{as} \quad \sigma_n^2 \to 0$$

2.1.5 Efficiency

This is also an important criterion for choosing one of several unbiased estimators of a given parameter, especially where estimators of relative efficiency are used.

Relative efficiency is defined as the ratio of the variances of two estimators.

In-Text Question

What do you understand Consistency?

In-Text Answer

An estimator is consistent if, a n gets large, the probability that the estimator T_n lies arbitrarily close to the parameter θ, being estimated becomes arbitrarily close to 1

2.1.6 Relative Efficiency

If T_1 and T_2 denote two unbiased estimators of θ, then relative efficiency is given as:

$$RE = \frac{Var(T_1)}{Var(T_2)}$$

An estimator T_2 is more efficient than T_1 if $RE > 1$, $[\text{i.e. Var}(T_2) < Var(T_1)]$

Note that it is only reasonable to compare estimators on the basis of variances if they are both unbiased.

For cases where one or both of the estimators are not unbiased, you define relative efficiency as:

$$RE = \frac{MSE(T_1)}{MSE(T_2)} = \frac{E(T_1 - \theta)^2}{E(T_2 - \theta)^2}$$

Example 7

Using our results in Examples 3 and 4, compare the efficiency of S^2 and $\hat{\sigma}^2$.

Solution

You compare the sample variances of S^2 and the population variance $\hat{\sigma}^2$.

The relative efficiency is defined as:

$$RE = \frac{S^2}{\hat{\sigma}_2} = \frac{2\sigma^4 / n-1}{(2n-1)\sigma^4 / n^2}$$

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

$$= \frac{2n}{2n-1} \cdot \frac{n}{n-1}$$

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

Thus $\quad \text{MSE}(\hat{\sigma}_2) < MSE(S^2)$

Example 8

Consider a random sample of size 2n+1 from a normal population. If the variance of the median is given as $Var(\tilde{X}) = \frac{\pi\sigma^2}{4n}$; (i) obtain the variance of the sample mean and (ii) determine the asymptotic relative efficiency of the mean and median.

Solution

i. The variance of the $\overline{X}$ can be obtained as $Var(\overline{X}) = \frac{\sigma^2}{2n+1}$

ii. The relative efficiency is given as

$$RE = \frac{Var(\overline{X})}{Var(\tilde{X})} = \frac{\sigma^2 / 2n+1}{\pi\sigma^2 / 4n} = \frac{4n}{\pi(2n-1)}$$

The asymptotic relative efficiency of the median with respect to the mean is

$$\lim_{n\to\infty} \frac{4\pi}{\pi(2n+1)} = \frac{2}{\pi} = 0.64 < 1$$

This implies that the mean $(\overline{X})$ is more efficient than the median $\tilde{X}$ as a measure of location.

This can be interpreted as follows: for large samples, the mean requires only 64 percent as many observation as the median to estimate μ with the same reliability.

Uniformly Minimum Variance Unbiased Estimator (UMVUE)

An estimator $\hat{\theta}$ is said to be UMVUE of θ, if there exists a random sample of size n from the distribution $f(x,\theta)$ such that:

i. $\hat{\theta}$ is unbiased for θ, and

ii. for any other unbiased estimator $\tilde{\theta}$, the variance of $\hat{\theta}$ is less than the variance of all other estimators for all $\tilde{\theta} \in \Omega$

i.e $V(\hat{\theta}) \le V(\tilde{\theta}) \quad \forall \quad \theta \in \Omega$

Summary for Study Session 2

In this study session, you have learnt that;

1. The various criteria with which you can judge an estimator as being better than the other(s).
2. This includes unbiasedness, sufficiency, efficiency, consistency and uniformly minimum variance unbiased estimator.

3. The definitions were given with some illustrations. Worked examples were also given in some cases.

4. $\hat{\theta}$ is unbiased for θ

5. for any other unbiased estimator $\tilde{\theta}$, the variance of $\hat{\theta}$ is less than the variance of all other estimators for all $\tilde{\theta} \in \Omega$

   i.e $V(\hat{\theta}) \leq V(\tilde{\theta}) \quad \forall \quad \theta \in \Omega$

Self-Assessment Questions (SAQs) for Study Session 2

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Let $x_{1,}$ x_2,….x_n be a random sample from a normal distribution $N(\mu, \sigma^2)$., show the $\overline{X}$ is an unbiased estimator of μ .

2. Let x_1, x_2, …x_n be a random ample from the binomial distribution B(n, P), show that the sample proportion, $\frac{X}{n}$ is an unbiased estimator of P.

3. Compare the efficiency of S^2 and $\hat{\sigma}^2$.

4. Let x_1, x_1 ,,, x_n be identically and independently distributed sample from a population with the distribution $X \sim (\mu, \sigma^2)$. .Estimate the parameter μ and σ^2 by the ,method of moments.

Note on SAQ

1.

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

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

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

$$= \mu \quad (\text{i.e. } \overline{X} \text{ is unbiased for } \mu)$$

2. Since $E(X) = nP,$ it follows that

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

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

$$= P$$

Hence $\dfrac{X}{n}$ is an unbiased estimate of P.

3. You compare the sample variances of S^2 and the population variance $\hat{\sigma}^2$.

The relative efficiency is defined as:

$$RE = \frac{S^2}{\hat{\sigma}_2} = \frac{2\sigma^4 / n-1}{(2n-1)\sigma^4 / n^2}$$

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

$$= \frac{2n}{2n-1}.\frac{n}{n-1}$$

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

Thus $\quad \text{MSE}(\hat{\sigma}_2) < MSE(S^2)$

4. Recall that $\quad \mu_{(1)}^{1} = m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} = \mu$

and $\quad \mu_2 = \mu_2' - (\mu_1')^2 = m_2' - (m_1')^2$

$$m_2' - (m_1')^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = S^2$$

Where $S^2 = \frac{1}{n-1}\sum\left(X - \overline{X}\right)^2$ (the sample variance)

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) "Statistical Inference" Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference.* Vol.2, Second Edition, Springer-Verlag,New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 3: Point Estimation Techniques I

Introduction

So far, you have discussed some desirable properties of a good estimator. In this study session, you shall discuss the methods of finding the point estimates with relevant examples. Some of the important estimation methods include Method of Moment; Method of Least Squares, Method of Maximum Likelihood and the Bayesian Method of Estimation.

Learning Outcomes for Study Session 3

At the end of this study session, you should be able to:

3.1 Discus the point estimates

3.2 Explain the Least Squares Method

3.1 Methods of Moment

In some cases, reasonable estimators can be found on the basis of intuition; however, some methods have been developed for deriving estimators given the probability distribution function.

3.1.1 Method of Moments

This is the oldest method of estimation given by Karl Pearson. The method consists of equating population moments with the corresponding sample moments, starting with the first order until enough equations are available to get a unique solution for the parameters being estimated. The solution of these equations is called moment estimators.

Consider a population with p.d.f. $f(x,\theta)$ depending on one or more parameter $\theta_1, \theta_2 ....\theta_k$. If $x_1, x_2,.......x_n$, is a random sample from $f(x,\theta)$ the first k sample and population moments are given by:

Sample Moment **Population Moment**

$$m_1' = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \mu_1' = E(X)$$

$$m_2' = \frac{\sum_{i=1}^{n} X^2{}_i}{n} \qquad \mu_2' = E(X^2)$$

....

$$m_k' = \frac{\sum_{i=1}^{n} X^k{}_i}{n} \qquad \mu_k' = E(X^k)$$

Example 1

Let x_1, x_1 ,,, x_n be identically and independently distributed sample from a population with the distribution $X \sim (\mu, \sigma^2)$. .Estimate the parameter μ and σ^2 by the ,method of moments.

Solution

Recall that $\mu^1_{(1)} = \mathrm{m}_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} = \mu$

and $\mu_2 = \mu_2' - (\mu_1')^2 = m_2' - (m_1')^2$

$$m_2' - (m_1')^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = S^2$$

Where $S^2 = \frac{1}{n-1}\sum (X - \overline{X})^2$ (the sample variance)

Example 2

Given a random sample of size n from a Gamma distribution, use the method of moment to obtain formulas' for estimating the parameters α and β.

Solution

The systems of equations are $m_1' = \mu_1'$ and $\mathrm{m}_2' = \mu_2'$

where $\mu_1' = \alpha\beta = \overline{X}$, $\mu_2' = \alpha(\alpha - 1)\beta^2$ and

$$\mu_2 = \mu_2' - (\mu_1')^2 = \alpha\beta^2 = S^2$$

solving for α and β we have:

$$\hat{\alpha} = \frac{\overline{X}^2}{S^2} \quad \text{or} \quad \hat{\alpha} = \frac{n\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \quad \text{and}$$

$$\hat{\beta} = \frac{S^2}{\overline{X}} \quad \text{or} \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n\overline{X}}$$

in term of the original observations.

In-Text Question

Write the 1st sample of moment?

In-Text Answer

$$m_1' = \frac{\sum_{i=1}^{n} X_i}{n}$$

Example 3

Suppose a random sample of size 8 from a Gamma distribution gives the following sample observations: 10.4, 13.2, 11.2, 10.8, 15.2, 9.8, 15.8 and 8.9. Obtain the estimates of α and β

.

Solution

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{95.3}{8} = 11.91; \qquad \mathrm{S}^2 = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{X_i} - \overline{X})^2}{\mathrm{n}} = 5.643$$

to obtain

$$\hat{\alpha} = \frac{\overline{X}^2}{S^2} = 25.14 \quad \text{and} \quad \hat{\beta} = \frac{S^2}{\overline{X}} = 0.473$$

Example 4: Let X: (x_1, x_2 ,,, x_n) be a random sample of size n from $U(\theta_1, \theta_2)$. Find the estimators of θ_1 and θ_2.

Solution

$$\mu_1' = E(X) = \frac{\theta_1 + \theta_2}{2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ i.}$$

$$\mu_2' = E(X^2) = \frac{\theta_2^2 + \theta_1\theta_2 + \theta_1^2}{3} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ ii.}$$

$$\mu_2 = \mu_2' - (\mu_2')^2 = \frac{(\theta_2 - \theta_1)^2}{12} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \text{ iii.}$$

On equating $m_1' = \mu_1' = \overline{\mathrm{X}}$ and $\mathrm{m}_2 = \mu_2 = S^2$ the sample variance, we obtain the following equations

from i. $\quad \theta_1 + \theta_2 = 2\overline{X}$

from iii. $\quad \theta_2 - \theta_1 = 2\sqrt{3}S$

Solving for θ_1 and θ_2, we obtain

$\theta_2 = \overline{X} + \sqrt{3}S$ and

$\theta_1 = \overline{X} - \sqrt{3}S$

Alternatively

For a uniform distribution $E(x) = \frac{\theta}{2} = \mu$ equate the sample mean to population mean i.e.

$$\overline{X} = \frac{\theta}{2} \Rightarrow \hat{\theta} = 2\overline{X}$$

To show that $\hat{\theta}$ is an unbiased estimate for unbiasedness $\in (\hat{\theta}) = \theta$

$$E(\hat{\theta}) = E(2\overline{X})$$
$$= 2E(\overline{X})$$
$$= 2\mu$$

The method of moment is simple and easy to derive even without the knowledge of the functional form of the population. However, in some cases, moments do not exist and in those cases this method can not be used. Method of moment sometimes leads to a negative estimate of an essentially positive parameter which is unacceptable.

Example 5

Consider a left truncated exponential distribution $f(x;\theta,\mu) = \left\{\frac{1}{\theta}\exp\left[-\frac{1}{\theta}(x-\mu)\right]\right.;\quad 0 < \mu < X < \infty;\quad \theta > 0$ with $E(X) = \mu + \theta$ and $Var(X) = \theta^2$. The moment estimator for μ and θ are given by

$$\mu + \theta = \overline{X} \qquad \text{and}$$

$$\theta = S \qquad \text{(the sample standard deviation)}$$

If the sample mean is less than the sample standard deviation, the moment estimator of μ will be negative which is unacceptable because $\mu > 0$.

3.1.2 Remarks on Method of Moments Estimators

Moment estimators:

1. are very easy to compute;
2. always give an estimator to start with;
3. are generally consistent (since sample moments are consistent for population ones);
4. are not necessarily the best or most efficient estimators.

3.2 Least Squares Method

The Method of Least Squares is quite important in certain specific type of model. That is linear models $Y_i = \beta_0 + \beta x_1 + \beta x_2 + ... + \beta x_p + \varepsilon_i$ where Y_i is the dependent variable, x_i is the explanatory variable α and β are unknown parameters to be estimated and ε_i is the random error term. Here, it is assumed that the mean of a random variable Y is a linear function of unknown parameter ($(\beta_0, \beta_1, ...\beta_p)$ and p factors $X = \{x_1, ...x_p\}$ that can be fixed or measured without error.

Assumption

To find the least squares estimator of $\beta_i \quad \mathrm{i} = 1,2..., \mathrm{p}$ we make the following assumptions:

1. $E(\mathrm{Y_i}) = \sum \beta_i x_i$ Linear function of the parameters;
2. $\mathrm{V}(\mathrm{Y}) = \sigma^2$ where σ^2 is unknown but does not depend on $\beta_0, \beta_1, \beta_2, ...\beta_p$;
3. $Cov(X_i, X_i')$, No serial correlation in the explanatory variables
4. The observation y_i, x_i are uncorrelated $E(\varepsilon_i) = 0$ and $\mathrm{V}(\varepsilon_\mathrm{i}) = \sigma^2$

3.2.1 Simple Linear Model

Consider a simple linear model $Y_i = \alpha + \beta x_i + \varepsilon_i$ from n pairs of data (y_1, x_1), ……; (y_n, x_n).

The ideal situation would be for pairs (y_i, x_i) to all fall on a straight line $y = \hat{\alpha} + \hat{\beta} x$ with all the $\varepsilon_i = 0$, so that we could predict Y without error.

But when the points do not all fall on the straight line, we choose a line $\hat{y} = \hat{\alpha} + \hat{\beta} x$ that minimises some functions of the $\varepsilon_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$.

This is obtained by minimising the sum of squares

$$Q = \sum_{i=1}^{n} \varepsilon_i^2 \quad = \sum_{i=1}^{n} [y_i - (\alpha + \beta x_i)]^2$$

$$\frac{dQ}{d\alpha} = 2\sum_i [y_i - (\alpha + \beta\, x_i)](-1) = 0$$

$$\frac{dQ}{d\beta} = 2\sum_i [y_i - (\alpha + \beta\, x_i)](-\mathrm{x_{i)}} = 0$$

Solving simultaneously gives

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(\mathrm{y_i} - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{\mathrm{x}}$$

$$SSE = \sum \varepsilon_i^2 = \sum [y_i - (\hat{\alpha} + \beta x_i)]^2$$

$$= \sum (\mathrm{y_i} - \bar{y}_i)^2$$

Such that $\hat{\sigma}_2 = \dfrac{SSE}{n-2}$

The expectations and variances and the parameters are

$E(\hat{\beta}) = \beta$ $\quad$ $\mathrm{Var}(\hat{\beta}) = \sigma^2 \Big/ \sum (x_i - \bar{x})^2$

$E(\hat{\alpha}) = \alpha$ $\quad$ $\mathrm{Var}(\hat{\alpha}) = \dfrac{\sigma^2 \sum x_i}{n \sum (x_i - \bar{x})^2}$

3.2.2 Properties of Least Squares Estimator (LSE)

1. $\hat{\alpha}$ and $\hat{\beta}$ are linear function of $y_1 .... y_k$
2. $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estuimator s
3. *Among* all linear unbiased estimates the L.S.E of α and β has the smallest v ariance i.e. BLUE
4. If $\varepsilon_i \sim N\left(0, \sigma^2\right)$ i.i.d the L.S.C. $\equiv$ M.L.E

Note that using different methods to obtain an estimator could yield different results.

Summary for Study Session 2

In this study session, you have learnt that;

1. You have described two point estimation techniques; Method of Moment and the Least Squares method. We also illustrated their applications with some examples.
2. The situations under which each method may not be applicable were also given

Self-Assessment Questions (SAQs) for Study Session 3

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Let x_1, x_1 ,,, x_n be identically and independently distributed sample from a population with the distribution $X \sim \left(\mu, \sigma^2\right)$. .Estimate the parameter μ and σ^2 by the ,method of moments.
2. Given a random sample of size n from a Gamma distribution, use the method of moment to obtain formulas' for estimating the parameters α and β .

3. Suppose a random sample of size 8 from a Gamma distribution gives the following sample observations: 10.4, 13.2, 11.2, 10.8, 15.2, 9.8, 15.8 and 8.9. Obtain the estimates of α and β.

NOTE ON SAQ

1. Recall that $\mu^{1}_{(1)} = \mathrm{m}'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = \mu$

and $\mu_2 = \mu'_2 - (\mu'_1)^2 = m'_2 - (m'_1)^2$

$$m'_2 - (m'_1)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = S^2$$

Where $S^2 = \frac{1}{n-1}\sum\left(X - \bar{X}\right)^2$ (the sample variance)

2. The systems of equations are $m'_1 = \mu'_1$ and $\mathrm{m}'_2 = \mu'_2$

where $\mu'_1 = \alpha\beta = \bar{X}$, $\mu'_2 = \alpha(\alpha-1)\beta^2$ and

$$\mu_2 = \mu'_2 - (\mu'_1)^2 = \alpha\beta^2 = S^2$$

solving for α and β we have:

$$\hat{\alpha} = \frac{\bar{X}^2}{S^2} \quad \text{or} \quad \hat{\alpha} = \frac{n\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{and}$$

$$\hat{\beta} = \frac{S^2}{\bar{X}} \quad \text{or} \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n\bar{X}}$$

in term of the original observations.

3. $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{95.3}{8} = 11.91$; $\mathrm{S}^2 = \frac{\sum_{i=1}^{n}(\mathrm{X_i} - \bar{X})^2}{\mathrm{n}} = 5.643$

to obtain

$$\hat{\alpha} = \frac{\overline{X}^2}{S^2} = 25.14 \quad \text{and} \quad \hat{\beta} = \frac{S^2}{\overline{X}} = 0.473$$

References

Bain, L.J. and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) "Statistical Inference" Unpublished lecture note in the Department of Statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992): *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 4: Method of Maximum Likelihood (M.L.E)

Introduction

Here, we shall continue our discussions on important methods of obtaining point estimates. In this study session, we shall describe the Maximum Likelihood Method of estimating parameters of a probability distribution and estimation based on frequency tables. Examples shall be given to illustrate their applicability.

Learning Outcomes for Study Session 4

At the end of this study session, you should be able to:

1.1 Discuss the likelihood function of a given probability function;

4.1 Likelihood Function

The **Likelihood Function** of a random sample x_1, x_2,…x_n from a population X with p.d.f. $f(x,\theta)$ is defined as the joint density function denoted by

$$L(\theta; x) = f_{(x_1,\theta)}\ f_{(x_2,\theta)} \ldots\ldots\ldots\ldots f(x_n,\theta)$$
$$= \prod_{i=1}^{n} \mathrm{f}(\underline{\mathrm{x}}, \theta)$$

This may be a function of θ only defined on the parameter space Ω.

Let $\mathrm{L}(\theta) = f(x_1, \mathrm{x}_2 \ldots\ldots x_n, \theta)$, $\quad \theta \in \Omega$ be the joint probability function (p.d.f) of x_1, x_2, ..x_n. For a given set of observation (x_1, x_2……..x_n) a value $\hat{\theta}$ in Ω at which $\mathrm{L}(\theta, x)$ is a maximum is called a *maximum likelihood estimate* (MLE) of θ.

If $L(\theta, x)$ is differentiable with respect to θ and $\hat{\theta}$ is an interior point of Ω, the MLE is a solution to equation $\frac{d}{d\theta}\mathrm{L}(\theta, x) = 0$ at which $\frac{d^2}{d\theta^2}\mathrm{L}(\theta, x) > 0$.

If one or more solutions exist, it should be verified which one maximizes $L(\theta)$. But what maximizes $L(\theta)$ also maximizes its log likelihood in $L(\theta)$, which is commonly used for computational convenience. $i.e. \frac{d}{d\theta} \text{In L}(\theta) = 0$.

If $\theta = (\theta_1, \theta_2, ..., \theta_n)$ is a vector of k parameters, the equation will be

$\frac{d}{d\theta} \text{L}(\theta_j, x) = 0, \quad j = 1,2,...,k$ and $\frac{d}{d\theta_j} \text{In L}(\theta_j, x) = 0, \quad j = 1,2,...,k$

The solution vector is called the MLE of θ_j.

If L (θ, x) is monotonic increasing or decreasing function of θ and $\Omega \subset (-\infty, \infty)$, the maximum likelihood of θ would be the upper end of Ω.

When L (θ, x) is monotonically increasing and it would be the lower end of Ω , when L(θ, x) is monotonically decreasing.

However, if the sample values give a value $\hat{\theta}$ that is not in Ω, the MLE of θ cannot be obtained.

We shall illustrate this technique with some example as follows:

Example 1

Let X: (x_1, $x_{2...}$ x_n) be a random sample from a Bernoulli distribution with p.d.f.

$$f(x, \theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & x = 0, 1, 2.....; \quad 0 < \theta < 1. \\ 0 & \text{elsewhere} \end{cases}$$

Obtain the M.L.E. for θ

Solution

The likelihood function is given by

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

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

The log likelihood is also given as:

$$In\, \mathrm{L}(\theta) = \sum x_i \;\mathrm{In}\, \theta + \left(\mathrm{n} - \sum \mathrm{x}_{\mathrm{i}}\right) \mathrm{In}\,(1 - \theta)$$

Differentiating with respect to θ, we have

$$\frac{d}{d\theta} In\, \mathrm{L}(\theta) = \frac{\sum x_i}{\theta} + \frac{n - \sum x_i}{1-\theta}$$

Equating $\frac{d}{d\theta} In\, \mathrm{L}(\theta)$ to zero

$\frac{d}{d\theta} In\, \mathrm{L}(\theta) = 0$, we have

$$\frac{\sum x_i}{\theta} + \frac{\left(n - \sum x_i\right)}{1-\theta} = 0$$

$$\Rightarrow \qquad \hat{\theta} = \frac{\sum x_i}{n}$$

The sample mean is the MLE of θ.

In-Text Question

State the likelihood function?

In-Text Answer

$$L(\theta; x) = f_{(x_1, \theta)}\; f_{(x_2, \theta)} \ldots\ldots\ldots\ldots\ldots f(x_n, \theta)$$
$$= \prod_{i=1}^{n} \mathrm{f}\,(\underline{\mathrm{x}}, \theta)$$

Example 2

Consider a random sample of size n from the normal distribution with the p.d.f given by

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

Obtain the M.L.E. for μ, and σ^2

Solution

The likelihood function is

$$L(\mu, \sigma) = \prod_{i=1}^{n} f\left(x_i, \mu, \sigma^2\right)$$

$$= (2\pi\sigma)^{-n/2} \exp\left\{ \frac{1}{2\sigma^2} \sum (X_i - \mu)^2 \right\}$$

and the log likelihood function is

$$InL(\mu, \sigma) = \frac{-n}{2} \text{In}(2\pi\sigma) - \frac{1}{2\sigma^2} \sum (X_i - \mu)^2$$

Differentiating In L(μ, σ^2) w.r.t μ gives

$$\frac{d \text{ In } L\left(\mu, \sigma^2\right)}{d\mu} = -\frac{2}{2\sigma^2} \sum (X_i - \mu) = 0$$

$$\text{i.e.} \sum X_i = n\mu$$

$$\Rightarrow \mu = \frac{\sum X_i}{n} = \bar{X}$$

Again differentiating w.r.t σ^2, we have

$$\text{In } L\left(\mu, \sigma^2\right) = \frac{-n}{2}(2\pi) - \frac{n}{2} \text{In}\,\sigma^2 - \frac{1}{2\sigma^2} \sum (X_i - \mu)^2$$

$$\frac{d \text{ In } L\left(\mu, \sigma^2\right)}{d\sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2} \sum (X_i - \mu) = 0$$

$$\Rightarrow \quad n\,\hat{\sigma}^2 = \sum \left(X_i - \bar{X}\right)^2$$

$$\Rightarrow \quad \hat{\sigma}^2 = \frac{\sum \left(X_i - \bar{X}\right)^2}{n}$$

the population variance.

Example 3 *Consider* $X_i \sim \text{POI}(\theta)$ with p.d.f $f(x,\theta) = \dfrac{\ell^{-\theta}\theta^{x}}{x!}$.

Find the MLE of θ.

Solution

The likelihood function is

$$L(\theta) = \frac{\ell^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

while the log likelihood function is also given as

$$In\ \text{L}(\theta) = -n\,\theta + \sum_{i=1}^{n} \text{X}_i\ \text{In}\,\theta\ \text{- In}\left(\prod_{i=1}^{n} \text{X}_i\ \ \middle|.\right)$$

Differentiating w.r. t θ and equate to zero to obtain the sample mean.

$$\frac{d}{d\theta}\ \text{In L}(\theta) = \frac{-n}{\theta} + \frac{\sum X_i}{\theta} = 0$$

$$\text{n}\,\theta = \sum_{\text{i=1}}^{\text{n}} \text{X}_\text{i}$$

$$\hat{\theta} = \frac{\sum_{\text{i=1}}^{\text{n}} \text{X}_\text{i}}{\text{n}}$$

To verify that this is the maximum

$$\frac{d^2}{d\theta^2}\ \text{In L}(\theta) = \sum \frac{X_i}{\theta^2}$$

which is negative when evaluated at? $\bar{x}$, i.e $\dfrac{\text{- n}}{\bar{\text{x}}} < 0$

Example 4

Consider the task of obtaining the M.L.E for estimating the parameters of a gamma distribution based on a random sample of size n.

Solution

The p.d.f. is given by

$$f(x,\alpha,\theta)=\frac{1}{\theta^{\alpha}\Gamma_{(\varepsilon)}}x_i^{\alpha-1}\ell^{-x/\theta}$$

and the likelihood function

$$L(\theta,\alpha)=\frac{1}{\theta^{n\alpha}\left[\Gamma_{(\alpha)}\right]^n}\left(\prod_{i=1}^{n}X_i\right)^{\alpha-1}\exp\left[-\sum X_i/\theta\right]$$

and with the log likelihood

$$In\ \mathrm{L}(\theta,\alpha)=-n\alpha\ \mathrm{In}\ \theta\ \text{-}\ \mathrm{n\ In}\ \Gamma_{(\alpha)}+(\alpha-1)\,In\prod_{\mathrm{i=1}}^{\mathrm{n}}\mathrm{X_i}-\sum X_i/\theta$$

The Maximum likelihood equations are

$$\frac{d\ \mathrm{In\ L}(\theta,\alpha)}{\mathrm{d}\theta}=\frac{-n\alpha}{\theta}+\sum X_i/\theta$$

$$\frac{d\ \mathrm{In\ L}(\theta,\alpha)}{\mathrm{d}\alpha}=-In\,\theta\ \text{-}\ \mathrm{n}\Gamma^{1}_{(\alpha)}\,/\,\Gamma_{(\alpha)}+In\prod_{\mathrm{i=1}}^{\mathrm{n}}\mathrm{X_i}$$

Letting $\tilde{x} = (\pi \, x_i)^{\frac{1}{n}}$ denote the geometric mean of the sample and $\varphi(\alpha) = \Gamma^1_{(\alpha)} / \Gamma_{(\alpha)}$ denote the psi function, then setting the derivatives equal to zero gives the equation.

$$\hat{\theta} = \tilde{x} / \hat{\alpha}$$

$$In\,(\alpha) - V(\hat{\alpha}) - In\left(\tilde{X} / \overline{X} \right) = 0$$

which can not be solved in closed form, although numerical solution can be obtained from the last equation?

From the theory by method of moment as before, it can be shown that for a gamma random variable that

$\mu = E(X) = \alpha\beta;$ and $V(x) = \alpha\beta^2 = \sigma^2$

Using the method of moment (as before)

$$\overline{X} = \alpha\,\beta; \quad S^2 = \alpha\,\beta^2 \Rightarrow \quad \hat{\alpha} = \frac{S^2}{\beta^2}$$

$$\Rightarrow \quad \hat{\alpha} = \frac{\overline{X}^2}{S^2}$$

The same with Weibull Distribution

$$f(x, \theta, \beta) = \left(\beta / \theta \right) \left(x / \theta \right)^{\beta - 1} \ell^{-\left(x/\theta \right)^{\beta}}$$

$$x > 0$$
$$\theta > 0$$
$$\beta > 0$$

Properties of Maximum Likelihood Estimates

1. If $\hat{\theta}$ exists and is unique
2. $\hat{\theta}_n$ is a consietnt estimator of θ
3. $\hat{\theta}_n$ is asymptotically normal with asymptotic mean θ and variance

$$1/n \in \left(\frac{\partial}{\partial \theta} \, \text{In f}(x, \theta) \right)^2 \text{ and}$$

4. $\hat{\theta}_n$ is asy, mptotically effciient

Thus for large n $\hat{\theta}_n \sim N(\theta, \text{CRLB})$

Minimum Likelihood based on Frequency Tables

Sometimes, it may not be possible/easy to obtain a function that best describes the evolution of an experiment such that it can be handled by methods described above.

Outcomes of n independent repetition of an experiment can be summarised in a frequency table like this:

Event/Class	k_1	k_2	k_3…….. k_j	Total
Observed freq.	f_1	f_2	f_3 fj	n
Expected freq.	np_1	np_2	np_3 np_j	

$$\textit{Such that} \quad S = K_1 \cup K_2 \cup .....\cup k_j$$

$$\textit{and} \quad \sum_{i=1}^{j} f_i = n \quad \text{and} \quad \sum_{i=1}^{j} P_i = 1$$

It is desired to obtain the estimator of p_i in order to obtain the expected frequency np_i's for comparison with observed frequencies f_i's. The model that generates the experiment involves an unknown parameter θ, then p_i's will generally be function of θ.

The probability of observing a particular frequency table is given by the multinomial distribution.

$$P(E,\theta) = \begin{pmatrix} n \\ f_1, f_2, ... f_j \end{pmatrix} p_2^{f_2} \; p_2^{f_2} .... p_j^{f_j}$$

where E is the experiment, and θ is the parameter involved in the model. e.g. Binominal distribution, Normal distribution, Poisson distribution, and so on.

The likelihood function of the multinomial distribution is

$$L(\theta) = C \;\; p_1^{f_1} \;\; p_2^{f_2} .... p_j^{f_j}$$

where the multinomial coefficient has been absorbed into the constant C.

The maximum likelihood estimate of $\hat{\theta}$ is the value of θ which maximizes the likelihood function. Using $\hat{\theta}$, we can compute expected frequencies $n\hat{p}_i$ with a view to comparing with the observed frequency f_i.'s.

Example 5

On each of the 220 consecutive working days, a Quality Control Manager draws a random sample of 10 items from the production line for non-conformity with specification. The result of the inspection is given in the following frequency table.

No of items not conforming	0	1	2	3	≥ 4	Total
Frequency observed	140	62	14	4	0	220

If the sample inspector is thought to follow the binominal distribution, find the MLE of θ, the probability that one item is non-conforming, and compute the estimate of the expected frequencies under the binominal distribution model.

Solution

The probability of observing i detective item out of 10 is

$$p_i = \binom{10}{i}\theta^i(1-\theta)^{n-i}; \quad \text{i} = 0, 1, 2, ..., 10$$

The probability of observing θ in a particular frequency table, according to binominal distribution, is:

$$p(\in, \theta) = C\, \text{P}_0^{140}\ \text{P}_1^{62}\ \text{P}_2^{14}\ \text{P}_3^{4}\ \text{P}_{4+}^{0} \quad \text{for } 0 \le \theta \le 1$$

Then,

$$\text{P}_0 = (1-\theta)^{10}; \text{P}_1 = \theta(1-\theta)^9;\ \text{P}_2 = \theta^2(1-\theta)^8;\ \text{P}_3 = \theta^3(1-\theta)^7$$

$$where\ \text{C} = \frac{1}{\binom{10}{0}^{140}\binom{10}{1}^{62}\binom{10}{2}^{14}\binom{10}{3}^{4}}$$

The Likelihood function is obtained as:

$$L(\theta) = \left[(1-\theta)^{10}\right]^{140}\left[\theta(1-\theta)^{9}\right]^{62}\left[\theta^{2}(1-\theta)^{8}\right]^{14}\left[\theta^{3}(1-\theta)^{7}\right]^{4}$$

$$= \theta^{102}(1-\theta)^{2098}$$

$$InL(\theta) = 102\, Log\theta - 2098\, Log(1-\theta)$$

$$\frac{dInL(\theta)}{d\theta} = \frac{102}{\theta} - \frac{2098}{1-\theta}$$

$$\frac{dInL\theta}{d\theta} = 0 \Rightarrow \hat{\theta} = \frac{102}{2200} = 0.04636$$

The estimated probability for class $i = 0$ is

$$\hat{P}_0 = \binom{10}{0}\hat{\theta}^{0}\left(1-\hat{\theta}\right)^{10} = (1-0.04636)^{10} = 0.622$$

The estimated expected frequency for this class is: $n\hat{P}_0 = 220(0.6221) = 136.86$

Similarly, we compute $n\hat{P}_i$ for $\mathrm{i} = 1, 2, 3$ and $4+$

The estimated expected frequency is given the table below:

No of non-conforming	0	1	2	3	≥ 4	Total
Observed frequency	140	62	14	4	0	220
Expected frequency	136.86	66.53	14.55	1.88	0.18	220

It can be observed that the agreement between the observed and the expected frequencies is good. The observed differences are due to chance variation, since the items are drawn at random from the production line.

Summary for Study Session 4

In this study session, you continued our discussion on methods of obtaining the estimate of parameters using the Maximum Likelihood Method. The methods were illustrated with examples both for the probability models and the frequency table.

Self-Assessment Questions (SAQs) for Study Session 4

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. On each of the 220 consecutive working days, a Quality Control Manager draws a random sample of 10 items from the production line for non-conformity with specification. The result of the inspection is given in the following frequency table.

No of items not conforming	0	1	2	3	$\geq$ 4	Total
Frequency observed	140	62	14	4	0	220

If the sample inspector is thought to follow the binominal distribution, find the MLE of θ, the probability that one item is non-conforming, and compute the estimate of the expected frequencies under the binominal distribution model.

NOTE ON SAQ

1. The probability of observing i detective item out of 10 is

$$p_i = \binom{10}{i}\theta^i (1-\theta)^{n-i}; \quad \mathrm{i} = 0, 1, 2, ..., 10$$

The probability of observing θ in a particular frequency table, according to binominal distribution, is:

$$p(\in, \theta) = C\ \mathrm{P}_0^{140}\ \mathrm{P}_1^{62}\ \mathrm{P}_2^{14}\ \mathrm{P}_3^{4}\ \mathrm{P}_{4+}^{0} \quad \text{for } 0 \leq \theta \leq 1$$

Then,

$$P_0 = (1-\theta)^{10};\ P_1 = \theta(1-\theta)^9;\ P_2 = \theta^2(1-\theta)^8;\ P_3 = \theta^3(1-\theta)^7$$

$$where\ \ C = \frac{1}{\binom{10}{0}^{140}\binom{10}{1}^{62}\binom{10}{2}^{14}\binom{10}{3}^{4}}$$

The Likelihood function is obtained as:

$$L(\theta) = \left[(1-\theta)^{10}\right]^{140}\left[\theta(1-\theta)^9\right]^{62}\left[\theta^2(1-\theta)^8\right]^{14}\left[\theta^3(1-\theta)^7\right]^4$$

$$= \theta^{102}(1-\theta)^{2098}$$

$$InL(\theta) = 102\,Log\theta - 2098\,Log(1-\theta)$$

$$\frac{dInL(\theta)}{d\theta} = \frac{102}{\theta} - \frac{2098}{1-\theta}$$

$$\frac{dInL\theta}{d\theta} = 0 \Rightarrow \hat{\theta} = \frac{102}{2200} = 0.04636$$

The estimated probability for class $i = 0$ is

$$\hat{P}_0 = \binom{10}{0}\hat{\theta}^0\left(1-\hat{\theta}\right)^{10} = (1-0.04636)^{10} = 0.622$$

The estimated expected frequency for this class is: $n\hat{P}_0 = 220(0.6221) = 136.86$

Similarly, we compute $n\hat{P}_i$ for $i = 1, 2, 3$ and $4+$

The estimated expected frequency is given the table below:

No of non-conforming	0	1	2	3	≥ 4	Total
Observed frequency	140	62	14	4	0	220
Expected frequency	136.86	66.53	14.55	1.88	0.18	220

It can be observed that the agreement between the observed and the expected frequencies is good. The observed differences are due to chance variation, since the items are drawn at random from the production line.

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997): "Statistical Inference" Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 5: Bayesian Estimation

Introduction

In the last two estimation techniques, you have assured that the parameter you are trying to estimate are unknown constant. The freqentist view was adopted. In this study session, you shall take a Bayesian view of the estimation process. You shall discuss the use of prior knowledge of a parameter and the likelihood information from data to obtain the posterior distribution of the parameter.

Learning Outcomes for Study Session 5

At the end of this study session, you should be able to:

5.1 Discus the Prior and Posterior Distribution Function.

5.1 Prior and Posterior Distribution

In the last two estimation techniques, the frequestist view was adopted by interpreting probability as the relative frequency of success in a conceptually infinite sequence of independent trials. The frequestist claims that all the information about the parameter is contained in the sample as summarised by the likelihood function. $L\left(X/\theta\right)$.

A Bayesian, however, interprets probability as a person rational degree of belief in a proposition. Under this framework, one does not have any difficulty in revealing the parameter θ as a random variable having distribution of its own and the data X, once drawn is fixed.

Thus in classical or frequency approach, the parameter is fixed and one is suspicious about the sample, but in Bayesian context, the sample is fixed or given and one is fixed or given and one is suspicious about the parameters.

In Bayesian approach, θ is considered as a random quantity whose variation can be described by a probability distribution called *prior distribution*.

In-Text Question

State the likelihood function?

In-Text Answer

$L\left(X/\theta\right)$.

Prior and Posterior Distribution

Prior distribution reflects the strength of one's belief about the possible value of θ prior to the experiment. Such knowledge about θ is summarised by the p.d.f g(θ) , i.e. $g(\theta) = f(X/\theta)$

After specifying the pre-sample or prior information about θ, we then proceed to collect data X and the likelihood function $L(X/\theta)$ gives additional information about θ. The Baye's theorem is then used to combine the prior information and the information contained in these sample to obtain the revised degree at belief about θ given by the posterior distribution $\pi(\theta/X)$.

Let $f(X/\theta)$ be the sampling distribution of X given θ, $g(\theta)$ in the prior distribution function, $f(\theta, X)$ be the joint distribution of θ and X at x, and h (x) be the marginal distribution of X at x, then the posterior distribution is defined as

$$\pi(\theta/X) = \frac{f(\theta, X)}{h(x)} = \frac{g(\theta)f(X/\theta)}{h(x)} \quad \propto \quad g(\theta)f(X/\theta)$$

Given the data set X, we may now write

$$\pi(\theta/X) = kg(\theta).L(\theta/X)$$

$$[Posterior \propto \text{ prior X likelihood}]$$

where k is the normalising constant such that over the parameter space Ω

$$\int_{\Omega} \pi(\theta / X) d\theta = 1$$

$$i.e. \int_{\Omega} g(\theta) f(X / \theta) d\theta = K$$

Choice of Prior Distribution

Since prior information is subjective, one may then ask "what should be the right prior distribution to start with?" According to Kalbfleich (1985), various attempts have been made to formulate prior distribution which represents a state of total ignorance about the parameter. Sometimes, they are derived scientifically from arguments of mathematical symmetry and invariance.

In the absence of any scientific choice of prior distribution, it is only reasonable to result to subjective prior distribution. This may just be a constant, the choice of which depends on individual prior perception (belief) about θ. Such prior is called the *subjective prior distribution* as prior is determined by introspection, and it is a measure of personal opinion concerning what the value of θ. is likely to be.

In-Text Question

Define the Posterior Distribution?

In-Text Answer

$$\pi(\theta / X) = \frac{f(\theta, X)}{h(x)} = \frac{g(\theta) f(X / \theta)}{h(x)} \quad \propto \quad g(\theta) f(X / \theta)$$

The Baye's theorem is then used to modify such prior opinion on the basis of experimental data. It has been argued however that subjective Bayesian approach may prove to be valuable in personal decision for problems of scientific inference.

Example 1

Let $(X_1, X_2, \ldots X_n)$ be a binominal random variable with parameter θ and given the prior distribution is a beta distribution with parameter α and β. Show that the posterior distribution of $\theta.$ given X is also a beta distribution with the parameter $x+\alpha$ and $n-x+\beta$.

Solution

$$f(X/\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x};\ for \quad x=0,1,2,\ldots\ldots..n$$

$$g(\theta)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1};\ for \quad 0<0<1\end{cases}$$

and hence

$$f(\theta,x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}x\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

$$=\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

for $0<\theta<1$

Nothing that $\int_{\partial}^{1}x^{\alpha-1}(1-x)^{\beta-1}=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\chi)}$

The marginal distribution function $h_{(x)}$ is obtained as

$$h_{(x)}=\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)};\frac{\Gamma(\alpha+x).\Gamma(n-x+\beta)}{\Gamma(n+x+\beta)}$$

for x = 0, 1, 2, .. ,n, and hence

$$\pi(\theta/X)=\begin{cases}\dfrac{\mathrm{f}(\theta,\mathrm{x})}{\mathrm{h}_{(\mathrm{x})}}=\dfrac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x).\Gamma(n-x+\beta)} \quad \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1};\ for 0<\theta<1\\ 0, elsewhere\end{cases}$$

By inspection, the above is a beta density function with parameters $(x+\alpha)$ and $(\mathrm{n}-\mathrm{x}+\beta)$.

Example 2

Let $(x_1, x_2 \ldots x_n)$ be a sample from a Poisson population with parameter θ. Then $T = \sum X_i$ is sufficient for θ whose distribution is also poisson with parameter $n\theta$. given that the prior distribution of θ is exponential, obtain the posterior distribution function.

Solution

$$f(X/\theta) = \ell^{-n\lambda} (n\theta)^x \big/ x! \quad , \qquad x = 0, 1, 2 \ldots\ldots\ldots\ldots$$

$$\text{Prior distribution}: g(\theta) = \frac{1}{\alpha} \ell^{-\theta/\alpha}; \qquad \theta > 0$$

The joint distribution of θ is

$$f(\theta, x) = f(X/\theta) g(\theta)$$

$$f(\theta, x) = \frac{1}{\alpha} \ell^{-\theta/\alpha} \frac{\ell^{-n\theta} (n\theta)^x}{x!}; \qquad (\theta > 0, \quad x = 0, 1, 2, \ldots)$$

$$\textit{But} \quad h(x) = \frac{1}{x!} \frac{1}{\alpha} \int_0^\infty \ell^{-\theta\left(\frac{1}{\alpha}+n\right)} n^x \theta^x d\theta$$

$$= \frac{n^x}{\alpha\, x!} \int \ell^{-y} \left(\frac{y}{\frac{1}{\alpha} + n} \right)^x \frac{dy}{\frac{1}{\alpha} + n}$$

for $y = (\alpha^{-1} + n)\theta$ and using product rule and noting that $\Gamma(x+1) = \int_0^\infty \ell^{-y} y^x dy$

we have $h(x) = \dfrac{n^x}{\alpha x!} \displaystyle\int_0^\infty \frac{\Gamma(x+1)}{\left(\frac{1}{\alpha} + n \right)^{x+1}}$

but

$$\pi(\theta / X) = \frac{f(\theta, x)}{h(x)}$$

$$= \frac{\ell^{-\theta(\alpha-1+n)} n^x \theta^x}{\alpha x!} \, x \, \frac{\alpha(\alpha^{-1}+n)^{x+1}}{n^x \Gamma(x+1)}$$

$$= \ell^{-\theta(\alpha^{-1}+n)} (\alpha^{-1}+n)^{x+1} . \frac{\theta^2}{x! \Gamma(x+1)}$$

which is a gamma p.d.f. with parameters $(\alpha^{-1}+n)$ and $(x+1)$.

Bayes' Estimator

When we estimate a parameter θ by a statistic $T_{(x)}$, we incure certain loss certain loss represented by the loss function $L(\theta, T_{(x)})$. *Bayes' Estimator* $\hat{\theta}$ is one that minimizes the expected loss *i.e.* $E[L(\theta; T(x))]$ where the expectation is with respect to the posterior distribution $\pi(\theta / X)$.. Even though, there are different types of loss function, we will use the *squared error loss function* defined by

$$L(\theta; T(x)) = [T(x) - \theta]^2$$

Let $T(x)$ be an estimator of θ. Then $C(\theta - T(x))^2$ where C is a constant, is a measure of error (or loss) incurred when $T(x)$ is used for θ. The "best" estimator is the one that minimizes this error on the average.

The average error is

$$\int_\theta \int_x C(\theta - T_{(x)})^2 \, f(\theta; x) \, d\theta \, dx \quad = \int_x \int_\theta C(\theta - T_{(x)})^2 \, h(x) \pi(\theta / X) d\theta \, dx$$

$$= \int_x h(x) \left[\int_\theta C(\theta - T(x)\ ^2 \pi(\theta / x) d\theta \right] dx$$

The above is minimised if we choose $T(x)$ such $\wedge$ that the integral within the brackets is a minimum for all x-values for which $h(x)$, the p.d.f. If X is positive, such an estimator is called the Bayes' estimator for the squared error loss function.

We already know that for any random variable X, $E(X-a)^2$ is minimum if $a = E(X)$. Therefore, $\in \left[\left(\theta - T(x)^2\right)/x\right]$ is minimized when $T(x) = E(\theta/x)$.. Thus the Bayes' estimator is the mean of the posterior distribution.

Application 1

For the problem in example 1 above, find the mean of the posterior distribution (Bayes' estimator) of the 'true' probability of a success, if 45 successes are obtained in 130 binomial trials and the prior distribution of θ in a beta distribution with $\alpha = \beta = 43$

Solution

Since the posterior distribution of θ is a beta distribution with parameters $(x+\alpha)$ and $n-x+\beta$, it follows from our knowledge in distribution theory that

$$E(\theta/x) = \frac{x+\alpha}{\alpha+\beta+n}$$

$$\text{i.e. } E(\theta/x) = \int_0^1 \theta\pi(\theta/x)\, \mathrm{d}\theta$$

$$= \frac{\Gamma_{(\alpha+\beta+\mathrm{n})}}{\Gamma_{(\alpha+\mathrm{x})}\,\Gamma_{(\mathrm{n}+\beta-\mathrm{x})}} \int \theta^{\alpha+x}(1-\theta)^{n+\beta-x-1}\, d\theta$$

$$= \frac{\Gamma_{(\alpha+\beta+\mathrm{n})}}{\Gamma_{(\alpha+\mathrm{x})}\,\Gamma_{(\mathrm{n}+\beta-\mathrm{x})}} \quad \beta(\alpha+\mathrm{x}+1,\ \mathrm{n}+\beta-\mathrm{x})$$

$$= \frac{\Gamma_{(\alpha+\beta+\mathrm{n})}}{\Gamma_{(\alpha+\mathrm{x})}\,\Gamma_{(\mathrm{n}+\beta-\mathrm{x})}} \quad \frac{\Gamma_{(\alpha+x+1)}\Gamma_{(\mathrm{n}+\beta-\mathrm{x})}}{\Gamma_{(\alpha+\beta+n+1)}}$$

$$= \frac{\alpha+\mathrm{x}}{\alpha+\beta+\mathrm{n}}$$

Thus the Bayes' estimator for the Binominal parameter is obtained by substituting $x = 45$, $\mathrm{n} = 130$, $\alpha = 43 = \beta$ in

$$\theta^* = E(\theta/x) = \frac{\alpha+x}{\alpha+\beta+n} = \frac{43+45}{43+43+130} = 0.41$$

Note that without the prior information, the estimator of the proportion would be the sample proportion $\hat{\theta} = \frac{x}{n} = \frac{45}{130} = 0.35$

Application 2

For example 2 above, derive the Bayes' estimator and use the data in application 1 to evaluate your results.

Solution

For example 2 above, the mean of the posterior distribution (i.e. Bayes' Estimator) is

$$\theta^{*} = E(\theta / x) = \int_{0}^{\infty} \theta \, \pi \; (\theta / x) \, \mathrm{d}\theta$$

$$= \left(\mathrm{n} + \alpha^{-1}\right)^{x+1} \frac{1}{\Gamma(x+1)} \int_{0}^{\infty} \theta^{x+1} \ell^{-\theta\left(n+\alpha^{-1}\right)} d\theta$$

$$= \frac{\left(\mathrm{n} + \alpha^{-1}\right)^{x+1}}{\left(\mathrm{n} + \alpha^{-1}\right)^{x+2}} \; \frac{1}{\Gamma(x+1)} \int y^{t+1} \ell^{-y} dy$$

for $y = \theta\left(\alpha^{-1} + n\right)$

$$= \frac{\Gamma(x+2)}{\left(\mathrm{n} + \alpha^{-1}\right)\Gamma_{(x+1)}}$$

$$= \frac{\mathrm{x} + 1}{\left(\mathrm{n} + \alpha^{-1}\right)}$$

Thus $\theta^{*} = \frac{45 + 1}{130 + (43)^{-1}} = 0.354$

Again without the prior information, the estimator of the parameter would have been $\hat{\theta} = \frac{45}{130} = 0.346$

Summary for Study Session 5

In this study session, you have learnt that;

You have considered the Bayesian method of estimation. The procedure for obtaining the Bayes' estimate and Bayes' estimators were also given. The methods were illustrated with relevant examples.

Self-Assessment Questions (SAQs) for Study Session 6

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Let (X_1, X_2, …X_n) be a binominal random variable with parameter θ and given the prior distribution is a beta distribution with parameter α and β. Show that the posterior distribution of θ. given X is also a beta distribution with the parameter $x+\alpha$ and $n-x+\beta$.
2. Let (x_1, x_2… x_n) be a sample from a Poisson population with parameter θ. Then $T=\sum X_i$ is sufficient for θ whose distribution is also poison with parameter $n\theta$. given that the prior distribution of θ is exponential, obtain the posterior distribution function.

Notes On SAQ

1. $$f(X/\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x};\ for\quad x=0,1,2,........n$$

$$g(\theta)=\begin{cases}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1};\ for\quad 0<0<1\end{cases}$$

and hence

$$f(\theta, x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} x \binom{n}{x} \theta^{x}(1-\theta)^{n-x}$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

for $0 < \theta < 1$

Nothing that $\int_{\partial}^{1} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\chi)}$

The marginal distribution function $h_{(x)}$ is obtained as

$$h_{(x)} = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}; \frac{\Gamma(\alpha+x).\Gamma(n-x+\beta)}{\Gamma(n+x+\beta)}$$

for x = 0, 1, 2, .. ,n, and hence

$$\pi(\theta / X) = \begin{cases} \frac{\mathrm{f}(\theta, \mathrm{x})}{\mathrm{h}_{(\mathrm{x})}} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x).\Gamma(n-x+\beta)} \quad \theta^{\mathrm{x}+\alpha-1}(1-\theta)^{n-x+\beta-1}; \; for 0 < \theta < 1 \\ 0, elsewhere \end{cases}$$

By inspection, the above is a beta density function with parameters $(x+\alpha)$ and $(\mathrm{n} - \mathrm{x} + \beta)$.

2.

$$f(X / \theta) = \ell^{-n\lambda} (n\theta)^{x} \big/ x! \quad , \qquad \mathrm{x} = 0, 1, 2 \, ... \, ..$$

Prior distributi on : $\mathrm{g}(\theta) = \frac{1}{\alpha} \ell^{-\theta/\alpha}; \qquad \theta > 0$

The joint distributi on of θ is

$$\mathrm{f} \; (\theta, \mathrm{x}) = f(X / \theta) g(\theta)$$

$$\mathrm{f} \; (\theta, \mathrm{x}) = \frac{1}{\alpha} \; \ell^{-\theta/\alpha} \; \frac{\ell^{-\mathrm{n}\theta} (n\theta)^{x}}{x!}; \qquad (\theta > 0, \quad \mathrm{x} = 0, 1, 2, ...)$$

But $\quad h(x) = \frac{1}{x!}\,\frac{1}{\alpha}\int_0^\infty \ell^{-\theta\left(\frac{1}{\alpha}+n\right)} n^x \theta^x d\theta$

$$= \frac{n^x}{\alpha\, x!}\int \ell^{-y}\left(\frac{y}{\frac{1}{\alpha}+n}\right)^x \frac{dy}{\frac{1}{\alpha}+n}$$

for $y = \left(\alpha^{-1}+n\right)\theta$ and using product rule and noting that $\Gamma(x+1) = \int_0^\infty \ell^{-y}\, y^x dy$

we have $h(x) = \frac{n^x}{\alpha x!}\int_0^\infty \frac{\Gamma(x+1)}{\left(\frac{1}{\alpha}+n\right)^{x+1}}$

but

$$\pi(\theta / X) = \frac{f(\theta, x)}{h(x)}$$

$$= \frac{\ell^{-\theta(\alpha-1+n)} n^x \theta^x}{\alpha x!}\, x\, \frac{\alpha\left(\alpha^{-1}+n\right)^{x+1}}{n^x \Gamma(x+1)}$$

$$= \ell^{-\theta\left(\alpha^{-1}+n\right)}\left(\alpha^{-1}+n\right)^{x+1} . \frac{\theta^2}{x!\Gamma(x+1)}$$

which is a gamma p.d.f. with parameters $\left(\alpha^{-1}+n\right)and\left(x+1\right)$.

References

Bain, L.J. and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997): “Statistical Inference” Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 6: Cramer-Rao (C –R) Lower Bound

Introduction

The concept of relative efficiency provides a criterion for choosing between two competing estimators. If there are more than two estimators, it does not give us any assurance that either of the estimators is better or has minimum variance. In this study session, you extend the concept of Minimum Variance Unbiased Estimator (MVUE), using the theorem provided by Rao and Cramer (1945 and 1946).

Learning Outcomes for Study Session 6

At the end of this study session, you should be able to:

6.1 Explain the theorem of Cramer – Rao lower bound for the variance of an estimator;

6.1 Cramer – Rao lower bound

When an estimator is unbiased and unique, it is also a MVUE. However, when there are more than one unbiased estimator, it is not easy to determine the estimator with minimum variance.

The Cramer Rao theorem provides a criterion that gives the lower bound for any estimator that may be considered to be the best estimator among all estimators $\hat{\theta} \in \Omega$ (estimators in the set Ω). Thus, any unbiased estimator whose variance is equal to the C – R lower bound is MVUE.

6.1.1 Cramer-Rao Inequality (Theorem)

Let $T = t(x)$ be an estimator from a random sample (X_1, X_2, …..X_n) from the distribution $f(x, \theta)$. Then the minimum variance $V(\hat{\theta})$ of estimate T of θ is one that satisfies the inequality:

$$V(T) \geq \frac{1}{n\left\{E\left[\frac{\partial \text{ In}}{-\partial\theta} f(x,\theta)\right]^2\right\}}$$

for the class of unbiased estimators, and

$$V(T) \geq \frac{1 + b_T'(\theta)}{n\left\{E\left[\frac{\partial \text{ In}}{-\partial\theta} f(x,\theta)\right]^2\right\}}$$

for the class of biased estimators.

Proof

The validity of the proof of the C-R lower bound depends on the validity of the following regularity conditions, which permit the interchange of the integration and differentiation operations, the existence integrability of the various partial derivatives. They are:

i. $\frac{\partial}{\partial\theta} In\, \mathrm{f}(\mathrm{x},\theta)$ exist for all x and all θ

ii. $\frac{\partial}{\partial\theta}\int f(x,\theta)dx = \int \frac{d}{d\theta} f(x,\theta)d$

iii. $\frac{\partial}{\partial\theta}\int \mathrm{t}(\mathrm{x})\, f(x,\theta)dx = \int t(x)\frac{\partial}{\partial\theta} f(x,\theta)dx = \tau(\theta)$

iv. $0 < E\left\{\left[\frac{\partial}{\partial\theta}\,\mathrm{In}\; f(x,\theta)\right]^2\right\} < \infty \;\forall\; \theta \in \Omega$

Note that (a) $f\left(\underset{-}{x},\theta\right)dx = 1$

a.. $\int \mathrm{t}\left(\underset{-}{\mathrm{x}}\right) f\left(\underset{-}{x},\theta\right)dx = \tau(\theta)$

b. $\frac{\partial\,\mathrm{In}\,\mathrm{f}(\mathrm{x},\theta)}{\partial\theta} = \frac{1}{f}\,\frac{\mathrm{d\,f}(\mathrm{x},\theta)}{\mathrm{d}\theta}\; f\partial x_i \Rightarrow \frac{\mathrm{d\,f}(\mathrm{x},\theta)}{\mathrm{d}\theta}\partial x_i = f\,\frac{\partial\,\mathrm{In}\,\mathrm{f}(\mathrm{x},\theta)}{\partial\theta}$

Let

$$\tau^1(\theta) = \frac{\partial}{\partial\theta}\tau(\theta) = \frac{\partial}{\partial\theta}\int t\left(\underset{-}{x}\right)\; f\left(\underset{-}{x},\theta\right)dx_i$$

$$= \int \mathrm{t}(\mathrm{x})\frac{\partial}{\partial\theta}\;\; f\left(\underset{-}{x},\theta\right)dx_i - \tau(\theta)\frac{\partial}{\partial\theta} f\left(\underset{-}{x},\theta\right)dx_i$$

$$= \int [\mathrm{t}(\mathrm{x}) - \tau(\theta)\;]\frac{\partial}{\partial\theta} f\left(\underset{-}{x},\theta\right)dx_i$$

$$= \frac{\partial}{\partial\theta}\int [\mathrm{t}(\mathrm{x}) - \tau(\theta)\;]\; f\left(\underset{-}{x},\theta\right)dx_i$$

$$\tau^1(\theta) = E\left[t\left(\underset{-}{x}\right) - \tau(\theta)\right]\frac{\partial}{\partial\theta}\,\mathrm{In}\,\mathrm{f}\left(\;\underset{-}{\mathrm{x}},\theta\right)$$

Using Cauchy-Schwarz inequality

$$\left[\tau^1(\theta)\right]^2 \le E\left[t(\underline{x})-\tau(\theta)\right]^2 \ \mathrm{E}\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\ \underline{x}\theta\right)\right]^2$$

i.e. $\left[\tau^1(\theta)\right]^2 \le \mathrm{V}\left(\hat{\theta}\right).\ \mathrm{E}\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underset{-}{\mathrm{x}}\ \theta\right)\right]^2$

$$\Rightarrow \qquad \mathrm{V}\left(\hat{\theta}\right) \ge \frac{\left[\tau^1(\theta)\right]^2}{E\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underline{x}\theta\right)\right]^2}$$

$\sin ce\ \ \underline{x}$ is a vector

$$and \quad E\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underset{-}{\mathrm{x}}\ \theta\right)\right]^2 = E\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underset{-}{\mathrm{x}_\mathrm{i}}\ \theta\right)\right] + \ldots + \left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underset{-}{\mathrm{x}_\mathrm{n}}\ \theta\right)\right]$$

$$\sum_\mathrm{i}\sum_j E\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underset{-}{\mathrm{x}_\mathrm{i}}\ \theta\right)\right]^2$$

$$\therefore \quad E\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underset{-}{\mathrm{x}}\ \theta\right)\right]^2 = \mathrm{n}E\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underset{-}{\mathrm{x}_\mathrm{i}}\ \theta\right)^2\right]$$

And

$$\mathrm{V}\left(\hat{\theta}\right) \ge \frac{\left[\tau^1(\theta)\right]^2}{\mathrm{n}\ E\left[-\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underline{x}\theta\right)\right]^2}$$

If T is unbiased, $\tau'(\theta) = 1$ then

$$\mathrm{V}\left(\hat{\theta}\right) \ge \frac{1}{\mathrm{n}\ E\left[-\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underline{x}\theta\right)\right]^2}$$

But if T is biased, $\frac{\partial}{\partial\theta}\left[Var(T) + b_T(\theta)\right] = \left[1 + b'_T(\theta)\right]$

$$\therefore \qquad \mathrm{V}\left(\hat{\theta}\right) \ge \frac{\left[1 + b'_T(\theta)\right]^2}{\mathrm{n}\ E\left[-\frac{\partial}{\partial\theta}\mathrm{In\ f}\left(\underline{x}\theta\right)\right]^2}$$

The denominator of the expression on the right hand side is referred to as the information in the sample denoted by $I_x(\theta)$.

and $\quad \mathrm{Var}(V) = E\left\{\left[\frac{\partial}{\partial\theta}\mathrm{In\ f}_{(\mathrm{x},\theta)}\right]^2\right\} = \mathrm{I}_{\mathrm{x}}(\theta)$

The above inequality becomes an equality whenever $\frac{\partial}{\partial\theta}\mathrm{In\ f}_{(\underline{\mathrm{x}},\theta)}$ is proportional to $t(\underline{x}) - \tau(\theta)$ or that there exist $k = k(\theta, n)$ such that

$$\frac{\partial}{\partial\theta}\mathrm{In\ f(x,}\theta) = k(\theta,\mathrm{n})[\mathrm{t(x)} - \tau(\theta)]$$

Example 1

Let $f(x,\theta) = f(x,\lambda) = \frac{\ell^{-\lambda}\lambda^x}{x!}; \quad \mathrm{x} = 0, 1, 2, \ldots$ **;** obtain the Cramer-Rao lower bound for the variance of λ

Solution

It can be shown that the regularity conditions are satisfied

$$\frac{\partial}{\partial\theta}\ \mathrm{In\ f}_{(\mathrm{x},\lambda)} = \frac{\partial}{\partial\theta}\mathrm{In}\frac{\ell^{-\lambda}\lambda^x}{x!}$$

$$= \frac{\partial}{\partial\theta}[-\lambda + x\,Log\,\lambda\ \text{- Log x!}]$$

$$= \text{-}1 + \frac{\mathrm{x}}{\lambda}$$

$$but \in \left[\frac{\partial}{\partial\theta}\mathrm{In\ f}_{(\mathrm{x},\lambda)}\right]^2 = E\left[\frac{x}{\lambda} - 1\right]^2 = \frac{1}{\lambda^2}[x - \lambda]^2$$

$$= \frac{\lambda}{\lambda^2}$$

$$= \frac{1}{\lambda}$$

C–R lower bond is

$$V(\hat{\theta}) \geq \frac{1}{n.\frac{1}{\lambda}} = \frac{1}{\frac{n}{\lambda}} = \frac{\lambda}{n}$$

This is expected since the variance of $\overline{X}$ is $\frac{\sigma^2}{n}$.

In-Text Question

Knowledge of …………………………..has been extended by looking at the Cramer Rao lower bound.

In-Text Answer

Minimum variance unbiased estimator

Example 2

Let $f(x,\theta) = \theta\, \ell^{-\theta x}; 0 < \text{x} < \infty$,; obtain the Cramer-Rao lower bound for the variance of θ

Solution

If $\tau(\theta) = \theta$. We can show that the regularity conditions are satisfied.

$$\tau^1(\theta) = 1. \text{ Hence } \text{V}(\hat{\theta}) \geq \frac{1}{n \in \left(\frac{\partial}{\partial\theta} \text{In f}_{(\text{x},\ \theta)}\right)^2}$$

$$Now \qquad \frac{\partial}{\partial\theta} \text{In f}_{(\text{x},\ \theta)} = \frac{\partial}{\partial\theta} \text{In}\left(\theta\, \ell^{\text{-}\theta\text{x}}\right)$$

$$= \frac{\partial}{\partial\theta}\left[In\, \theta + \text{In } \ell^{\text{-}\theta\text{x}}\right]$$

$$= \frac{\partial}{\partial\theta}\left[In\, \theta \text{ - } \theta\text{x}\right] = \frac{1}{\theta} - x$$

Since T is a biased estimator for λ

$$\text{V}(\hat{\theta}) \geq \frac{[\tau'(\lambda)]^2}{nI(\lambda)} = \frac{[-1/\lambda^2]^2}{n/\lambda^2}$$

$$= \frac{1}{\text{n}\lambda^2}$$

Example 3

What is the bound on the variance of the estimator π for θ from a Bernoulli distribution?

Solution

The p.d.f of the Bernoulli distribution is given as:

$f(x;\pi) = \pi^x (1-\pi)^{1-x}$; for x = 0, 1

and the log likelihood function is

$$\ell = Inf(x;\pi) = \frac{x}{\pi} + \frac{1-x}{1-\pi}(-1) = \frac{x-\pi}{\pi(1-\pi)}$$

$$I = E\left[\frac{\partial}{\partial\theta} Inf(x,\pi)\right]^2 = E\left[\frac{(x-\pi)^2}{\pi^2(1-\pi)^2}\right] = \frac{1}{\pi(1-\pi)}$$

$$I_X = \frac{n}{\pi(1-\pi)}$$

Since $E(X) = \pi$ is unbiased for π. So

$$Var(T) \geq \frac{1}{n\left[\frac{1}{\pi(1-\pi)}\right]} = \frac{\pi(1-\pi)}{n}$$

Example 4

Find the minimum variance for the exponential distribution with p.d.f.

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

Solution

$$L(\theta) = \frac{1}{\theta^n} \exp\left(\frac{-\sum x}{\theta}\right)$$

$$\ell = Inf = -In\theta - \frac{x}{\theta}$$

$$\frac{\partial \ell}{\partial\theta} = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} Inf\right]^2 = E\left[\frac{x-\theta}{\theta^2}\right] = \frac{1}{\theta^2}$$

Since $T = \overline{X}$ and T is unbiased;

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

Therefore, $Var(T) \geq \frac{1}{nI(\theta)} = \frac{1}{\frac{n}{\theta^2}} = \frac{\theta^2}{n}$

Hence, $\hat{\theta}$ is UMVE for θ.

Summary for Study Session 6

In this study session, you have learnt that;

1. Knowledge of minimum variance unbiased estimator has been extended by looking at the Cramer Rao lower bound for minimum variance of an estimator.
2. The regularity conditions under which the C-R inequality is valid were discussed and some examples were given to illustrate the concept

Self-Assessment Questions (SAQs) for Study Session 6

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Let X: (x_1, $x_{2\ldots}$ x_n) be a random sample from a Bernoulli distribution with p.d.f.

$$f(x,\theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & \text{x} = 0,1,2.....; \;\; 0 < \theta < 1. \\ 0 & \text{elsewhere} \end{cases}$$

Obtain the M.L.E. for θ.

2. *Consider* $\text{X}_\text{i} \sim \text{POI}(\theta)$ with p.d.f $f(x,\theta) = \frac{\ell^{-\theta}\theta^x}{x!}$.

Find the MLE of θ.

Notes on SAQ

1. The likelihood function is given by

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

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

$$= \theta^{\sum xi}(1-\theta)^{n-\sum xi}$$

The log likelihood is also given as:

$$In\,\mathrm{L}(\theta) = \sum x_i \;\mathrm{In}\,\theta + \left(\mathrm{n} - \sum \mathrm{x}_\mathrm{i}\right)\mathrm{In}\,(1 - \theta)$$

Differentiating with respect to θ, we have

$$\frac{d}{d\theta} In\,\mathrm{L}(\theta) = \frac{\sum x_i}{\theta} + \frac{n - \sum x_i}{1-\theta}$$

Equating $\frac{d}{d\theta} In\,\mathrm{L}(\theta)$ to zero

$\frac{d}{d\theta} In\,\mathrm{L}(\theta) = 0$, we have

$$\frac{\sum x_i}{\theta} + \frac{\left(n - \sum x_i\right)}{1-\theta} = 0$$

$$\Rightarrow \qquad \hat{\theta} = \frac{\sum x_i}{n}$$

The sample mean is the MLE of θ.

2. The likelihood function is

$$L(\theta) = \frac{\ell^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

while the log likelihood function is also given as

$$In\,\mathrm{L}(\theta) = -n\,\theta + \sum_{i=1}^{n} \mathrm{X}_\mathrm{i}\;\mathrm{In}\,\theta - \mathrm{In}\left(\prod_{i=1}^{n} \mathrm{X}_\mathrm{i} \quad \Big|.\right)$$

Differentiating w.r. t θ and equate to zero to obtain

$$\frac{d}{d\theta}\,\text{In L}(\theta) = \frac{-n}{\theta} + \frac{\sum X_i}{\theta} = 0$$

$$\text{n}\,\theta = \sum_{\text{i=1}}^{\text{n}} \text{X}_\text{i}$$

$$\hat{\theta} = \frac{\sum_{\text{i=1}}^{\text{n}} \text{X}_\text{i}}{\text{n}}$$

the sample mean.

To verify that this is the maximum

$$\frac{d^2}{d\theta^2}\ \text{In L}(\theta) = \sum \frac{X_i}{\theta^2}$$

which is negative when evaluated at? $\bar{x}$, i.e $\frac{\text{- n}}{\bar{\text{x}}} < 0$

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) "*Statistical Inference*" Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey. Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics,* 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 7: Sufficiency and Rao Blackwell Theorem

Introduction

You draw an inference when information from the sample is used to estimate the population. But, are you sure that the information from the sample data is sufficient to draw any kind of inference about the population? Thus, in this session, the idea of sufficiency shall be extended and the above question will be addressed. The Rao Blackwell theorem which provides the criteria for sufficiency shall also be discussed.

Learning Outcomes for Study Session 5

At the end of this study session, you should be able to:

7.1 Explain the concept of sufficiency;

7.2 Discus the Rao Blackwell theorem.

7.1 Concept of Sufficiency

Sufficiency is another very crucial property of a good estimator. The idea of computing a statistic from the sample of size n leads to loss of information in the population to a single value. In other words, in the process of data reduction, some information about the population may be lost.

Sufficiency helps us to determine if the sample statistics T(x) = t (x_1, x_2, ….,x_n) contains all the relevant information about the true value (the parameter) θ. That is, knowing or studying the whole population does not contribute any more to the inference about θ.

Definition *1*

Let $f(x,\theta), \theta \in \Theta$ be a family of distribution of the discrete type. For a random sample x_1, x_2, …, x_n from $f(x,\theta)$, define $T = t(x_1, x_2, ..., x_n)$. Then T is a sufficient statistics for θ. it, for all θ. and all possible sample points,

$$P(X_{oi} = x_1, X_2 = x_2, ...xX_n = x_n)/T = tx_1x_2...x_n \quad)$$

does not involve the parameter θ.

Remark: A distribution free of θ means that the distribution is completely known; hence, corresponding random quality can be generated with random number of generation.

Definition 2

Let x_1, x_2,…,x_n be a random sample of size n from a population with distribution $f(x,\theta), \theta \in \Omega$ .Let y_1, y_2,…,y_n be a one – to one transformation of $y_i = u_i x_i$ with joint p.d.f $g(\underline{y},\theta)$ i.e $f(\underline{x,\theta}) = g(\underline{y,\theta})$, then y_i is said to be a sufficient statistic for θ if the conditional distribution

$$h(y_2, y_3, ..., y_n / y_1, \theta) = \frac{g(\underline{y,\theta})}{g(\underline{y_1,\theta})}$$

is independent of θ in its mathematical form.

Before you discuss the idea of sufficiency, let us review some important concepts.

7.1.1 The score and Information function

One of the important concepts in estimating the variance of estimate is the notion of information in sample and how to measure this information when data obtained through several experiments are combined.

The score: The score function V is defined as the first derivative of the log likelihood function with respect to θ.

$$V = \frac{\delta}{\delta\theta} \log \mathrm{L}(\theta; \mathrm{x}) = \frac{L'(\theta)}{L(\theta)} = \ell'(\theta)$$

Where $\ell(\theta) = \log \mathrm{L}(\theta)$

Properties of V

The proof of the properties of V depends on the regularity condition of the C–R rule. They are:

i. $E(V) = 0$

Proof:

$$E(\mathrm{V}) = E\left(\frac{\delta}{\delta\theta}\ell(\theta)\right) = E\left(\frac{\delta}{\delta\theta} Inf\right) = \int \frac{1}{f} \cdot \frac{\delta}{\delta\theta}. fdx$$

$$= \int \frac{\delta f}{\delta\theta}.dx$$

since $\int f \; dx = V$

Then by differentiating with respect to $\theta,$ we have

$$\int \frac{d}{d\theta} f \; \mathrm{dx} = 0$$

$$\therefore E(\mathrm{V}) = 0$$

Information function (Fisher's Information)

The information in a sample is denoted by $I_x(\theta)$ is measured by: $Var(\mathrm{V})$.

It is also defined as the minus second derivative of the log likelihood function with respect to θ.

$$i.e. \; \mathrm{I_x}(\theta) = Var(\mathrm{V}) = E\left\{\left[\frac{\delta}{\delta\theta} \log \mathrm{f}(\mathrm{X};\theta)\right]^2\right\}$$

Note that:

1. Information is additive over independent experiments. For X and Y independent, we have $I_x(\theta)+I_y(\theta)=I_{X+Y}(\theta)$.
2. Information in a sample is n times the information in a single observation, that is $I_x(\theta)=n\,I_x(\theta)$
3. The information provided by a sufficient statistics T= t(x) is the same as that in the sample X, i.e. $I_T(\theta)=I_x(\theta)$
4. The information in the sample can also be computed by the formula:

$$I_x(\theta)=-E\left(\frac{dV}{d\theta}\right)$$

OR

$$I_x(\theta)=-E\left(\frac{d^2L}{d\theta}\right)$$

5. The regularity conditions will not be half of the uniform distribution for sample size *n = L*

In-Text Question

What is Score Function?

In-Text Answer

$$V=\frac{\delta}{\delta\theta}\log \mathrm{L}(\theta;\mathrm{x})=\frac{L'(\theta)}{L(\theta)}=\ell'(\theta)$$

Where $\ell(\theta)=\log \mathrm{L}(\theta)$

Example 1

If $f(x,\theta)=\frac{1}{\theta}$; $0<\mathrm{x}<\theta$; define the score function and the information function.

Solution

$f(\theta;x)=\frac{1}{\theta}$; $\log \mathrm{L}(\theta;x) = -Log(\theta)$

The score function is $V = \frac{d}{d\theta} \log L(\theta; x) = \frac{-1}{\theta}$

The information function is $I_x(\theta) = E(V) = \frac{-dV}{d\theta} = \int_0^\theta \frac{-1}{\theta} . \frac{1}{\theta} \text{ dx}$

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

Example 2

Compute the information on $\hat{P}$ from *n* Bernoulli trials with probability of success equal to p.

Solution

$$f(x; p) = P^x (1-P)^{1-x}$$

$$\log f(x; p) = x \log p + (1 - x) \log (1 - P)$$

$$V = \frac{d}{dp} \log f(x; p) = \frac{x}{p} - \frac{1-x}{1-p}$$

$$\frac{dV}{dp} = \frac{-x}{p^2} - \frac{(1-x)}{(1-P)^2}$$

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

$$-E\left(\frac{dV}{d\ p}\right) = \frac{1}{pq} \qquad \text{where } q = 1 - p$$

for a sample of size *n*, the information on p is $I_x(p) = \frac{n}{pq}$

Example: Compute the information on λ from a Poisson distribution.

Solution

$$f(x,\theta)=\frac{\ell^{-\lambda}\lambda^{x}}{x!}$$

$$f(\underline{x},\theta)=\frac{\ell^{-n\lambda}\lambda^{\sum xi}}{\pi_i(x!)}$$

and so the log likelihood is

$$\ell=-n\lambda+\sum x_i In\lambda-\prod_{i=1}^{h}(|_n\ \mathrm{x_i}!)$$

this given $\dfrac{\delta\ell}{\delta\lambda}=-n+\dfrac{\sum x_i}{\lambda}$

$$I(\theta)=-E\left(\frac{d\ell}{d\lambda}\right)^2=E\left[-n+\frac{\sum x_i}{\lambda}\right]^2$$

$$=\mathrm{n}^2+E\frac{\left(\sum x_i\right)^2}{\lambda^2}-2\,n\,E\left(\frac{\sum \mathrm{x_i}}{\lambda}\right)$$

but $\left(\sum \mathrm{x_i}\right)^2=\sum x_i{}^2+2\sum x_i\,\mathrm{x_j}$

and $\mathrm{Var}(\mathrm{x})=E\left(X^2\right)-E(X)^2$ $\quad *$

we have

$$\mathrm{I}(\theta)=E\left(\frac{\left(\sum x_i\right)^2}{\lambda^2}-n^2\right)$$

Substituting $*$ in $I(\theta)$ we have

$$I_x(\theta)=\left[nVar(x)+n\,\mu_{\mathrm{x}}^2+2\,\frac{(1(n-1)}{2}\lambda^2\right]\Big/\lambda^2\quad-n^2$$

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

To give

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

$$= \frac{\mathrm{n}}{\lambda}$$

Thus, the Cramer Rao lower bound for variance of an estimator is the inverse of the Fisher's Information.

Example 3

For a random sample X_1, X_2 … X_n from a normal distribution with mean μ and variance σ^2. Find the information for μ and σ^2.

Solution

$$\text{for } \mu : f(x_i \mu) = \left(2\pi\sigma^2\right)^{-1/2} \ell^{-(x_i-\mu)^1 / 2\sigma^2}$$

$$\log f(x_i, \mu) = \frac{-1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}(x_i - \mu)^2$$

$$= \frac{-1}{2}\log 2\pi \frac{-1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(x_i - \mu)^2$$

$$V = \frac{d}{d\mu}\log f = \frac{1}{2\sigma^2}(x_i - \mu)$$

$$= \frac{dV}{d\mu} = \frac{1}{\sigma^2}; cons\tan t$$

$$I_x(\mu) = -E\left(\frac{dV}{d\mu}\right) = \frac{1}{\sigma^2}$$

$$\text{for } \sigma^2 : V = \frac{d}{d\sigma}\log f = \frac{-1}{2\sigma^2} + \frac{(x_i - \mu)^2}{2\sigma^4}$$

$$\frac{dV}{d\sigma^2} = \frac{+1}{2\sigma^4} - \frac{(x_i - \mu)^2}{\left(\sigma^2\right)}^3$$

$$-E\left(\frac{dV}{d\sigma^2}\right)=\frac{-1}{2\sigma^2}+E\frac{(x_i-\mu)^2}{(\sigma^2)^3}$$

$$\frac{-1}{2\sigma^4}+\frac{\sigma^2}{(\sigma^2)^3}$$

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

for a sample for size *n*

$$I_x(\theta)=\frac{n}{2\sigma^4}$$

To find the joint information matrix for $\mu,$ and σ^2, we take the second derivative as

$$\frac{d^2\ell}{d\mu^2}=-\sum_{i=1}^{n}\frac{1}{\sigma^2}=\frac{-n}{\sigma^2} \text{ from previous page from} *$$

$$\frac{d^2\ell}{d\mu^4}=\frac{n}{2\sigma^4}-\sum(x_i-\mu^2)\Big/\sigma^6$$

$$E\left[\frac{d^2\ell}{d\mu^4}\right]=\frac{n}{2\sigma^4}-\frac{n\sigma^2}{\sigma^6}=\frac{-n}{2\sigma^4}$$

Since $E(x_i-\mu)^2=\sigma^2$

also $$E\left\{\frac{d^2\ell}{d\mu\,\delta\sigma^2}\right\}=E\left\{\frac{d^2\ell}{d\sigma^2\;\delta\mu}\right\}=E\left[-\sum_i(x_i-\mu)\Big/\sigma^4\right]=0$$

Thus the matrix of the second derivatives of ℓ is

Example 4

A random sample of size *n* is taken from the Poisson distribution $P(x)$.

Is $\sum_{i=1}^{n} X_i$ sufficient for λ?

Solution

$$X \approx P(\lambda);\ f(x,\lambda) = \frac{\ell^{-\lambda}\lambda^{x}}{x!}$$

and $\sum_{i=1}^{n} X_i \approx P(n\lambda);\ f\left(\sum X_i ; n\lambda\right) = \ell^{-n\lambda}(n\lambda) \Big/ \left(\sum X_i\right)!$

$$f(x_1,\theta) X \ldots\ldots X f(xn,\theta) = \prod_{i=1}^{n} \ell^{-\lambda}\lambda^{x_i} \Big/ x_i!$$

$$\therefore P\left(X_1,\ldots X_n / \sum X_i = t\right) = \frac{\prod_{i=1}^{n} \ell^{-\lambda}\lambda^{x_i} / x_i!}{\ell^{-n\lambda}(n\lambda)^{\sum X_i} \Big/ \left(\sum_{i=1}^{n} X_i\right)!}$$

$$= \frac{\ell^{-n\lambda}\lambda^{\sum X_i} / \left(\prod_{i=1}^{n} X_i\right)}{\ell^{-n\lambda} n^{\sum X_i} \lambda^{\sum X_i} / \left(\sum_{i=1}^{n} X_i\right)!}$$

$$= \frac{\left(\sum_{i=1}^{n} X_i\right)!}{n^{\sum X_i}\left(\prod_{i=1}^{n} X_i\right)}$$

which does not involve (independent of) λ, and so $T = \sum_{i=1}^{n} X_i$ is sufficient for λ.

Rao-Blackwell Theorem

The concept of sufficiency is useful in the construction of unbiased estimator with smaller variance. The Rao Blackwell theorem provides such a method of construction.

Theorem

Let $\hat{\theta}_1$ be a sufficient statistic for θ and $\hat{\theta}_2$ be an unbiased estimator of θ.Then $\varphi(\theta_1) = E(\hat{\theta}_2 / \hat{\theta}_1)$ is also unbiased for θ with $V(\varphi(\theta_1)) \le V(\hat{\theta}_2)$.

Proof

Denote the joint p.d.f of $\hat{\theta}_1$ and $\hat{\theta}_2$ by $f(\hat{\theta}_1, \hat{\theta}_2)$; their

*M*arg*inal* distributions are $g_1(\hat{\theta}_1)$; $g_2(\hat{\theta}_2)$

The Conditional distribution are $h_1(\hat{\theta}_1 / \hat{\theta}_2)$ and $h_2(\hat{\theta}_2 / \hat{\theta}_1)$

Re*quired to* prove that $E[\varphi(\hat{\theta}_1)] = \theta$

$$E(\hat{\theta}_2 / \hat{\theta}_1) = \int \hat{\theta}_2 \; h_2 = (\hat{\theta}_2 / \hat{\theta}_1) \, \partial \hat{\theta}_2$$

$$= \int \hat{\theta}_2 \frac{f(\hat{\theta}_1 \, \hat{\theta}_2)}{g_1(\hat{\theta}_1)} \, \partial \hat{\theta}_2$$

$$= \varphi(\hat{\theta}_1)$$

from the above

$$\varphi(\hat{\theta}_1) g_1(\hat{\theta}_1) = \int \hat{\theta}_2 \, f(\hat{\theta}_1, \hat{\theta}_2) \, \partial \hat{\theta}_2$$

but $E[\varphi(\hat{\theta}_1)] = \int \varphi(\hat{\theta}_1) g_1(\hat{\theta}_1) \partial(\hat{\theta}_1)$

$$= \int \hat{\theta}_2 \, f(\hat{\theta}_1, \hat{\theta}_2) \, \partial \hat{\theta}_2$$

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

To show that $V[\varphi(\hat{\theta}_1)] \leq V(\hat{\theta}_2)$

and $V(\hat{\theta}_2) = E[\hat{\theta}_2 - \theta]^2$

$$Let\ V(\hat{\theta}_2) = E[\hat{\theta}_2 - \varphi(\hat{\theta}_1) + \varphi\hat{\theta}_1 - \theta]^2$$

$$= E(\hat{\theta}_2 - \varphi(\hat{\theta}_1))^2 + E(\varphi(\hat{\theta}_1) - \theta)^2 + 2E(\quad)(\quad)$$

$$= \theta + V[\varphi(\hat{\theta}_1)]$$

where $\theta > 0$

$$V[\varphi(\hat{\theta}_1)] \leq V(\hat{\theta}_2)$$

Completeness

A family of density function $f(x,\theta)$ $\theta \in \Omega$ is complete if $E[\mu(x)] = 0$ $\forall$ $\theta \in \Omega$ implies $\mu(\hat{x}) = 0$ with probability 1 $\forall$ $\theta \in \Omega$ (i.e. at every point of x).

Example

$$Given\ f(x,\theta) = \frac{1}{2\theta} \qquad 0 < \theta < 1$$

$$-\theta < x < \theta$$

Show that $f(x,\theta)$ if $\mu(x) = x$

Solution

$$\int_{-\theta}^{\theta}\frac{1}{2\theta}dx = \left[\frac{x}{2\theta}\right]_{-\theta}^{\theta}$$

$$= \frac{\theta}{2\theta} - \frac{-\theta}{2\theta} = \frac{1}{2} + \frac{1}{2} = 1 \qquad \text{i.e. proper p.d.f.}$$

$Let\ \mu(\mathrm{x}) = x$

$$E(\mu(\mathrm{x})) = \int \mu(x)\ \ \mathrm{f}(\mathrm{x},\theta)dx$$

$$= \int_{-\theta}^{\theta}\frac{\mathrm{x}}{2\theta}dx$$

$$= \frac{\mathrm{x}^2}{4\theta}\ \Big]_{-\theta}^{\theta}$$

$$= \frac{\mathrm{x}^2}{4\theta} - \frac{(-\theta)^2}{4\theta}$$

$$= 0$$

Summary for Study Session 7

In this study session, you have learnt that;

1. The idea of sufficiency of an estimator was extended to the discussion of the score and information functions as they related to the C-R lower bound. The Rao Blackwell theorem was stated and proved. Some examples were given to illustrate the concepts.

2. Sufficiency helps us to determine if the sample statistics T(x) = t (x_1, x_2, ….,x_n) contains all the relevant information about the true value (the parameter) θ. That is, knowing or studying the whole population does not contribute any more to the inference about θ.

Self-Assessment Questions (SAQs) for Study Session 7

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. If $f(x,\theta)=\frac{1}{\theta};\quad 0<\mathrm{x}<\theta$; define the score function and the information function.
2. Compute the information on λ from a Poisson distribution.

Notes on SAQ

1. $f(\theta;x)=\frac{1}{\theta};\qquad \log \mathrm{L}(\theta;x) = -Log(\theta)$

The score function is $\mathrm{V}=\frac{d}{d\theta}\log \mathrm{L}(\theta;x) = \frac{-1}{\theta}$

The information function is $I_x(\theta)=E(V)=\frac{-dV}{d\theta}=\int_0^\theta \frac{-1}{\theta}.\frac{1}{\theta}\ \mathrm{dx}$

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

2.
$$f(x,\theta)=\frac{\ell^{-\lambda}\lambda^x}{x!}$$

$$f(\underline{x},\theta)=\frac{\ell^{-n\lambda}\lambda^{\sum xi}}{\pi_i(x!)}$$

and so the log likelihood is

$$\ell=-n\lambda+\sum x_i In\lambda-\prod_{i=1}^{h}(\ell_n\ \mathrm{x_i}!)$$

this gives $$\frac{\delta\ell}{\delta\lambda}=-n+\frac{\sum x_i}{\lambda}$$

$$I(\theta) = -E\left(\frac{d\ell}{d\lambda}\right)^2 = E\left[-n + \frac{\sum x_i}{\lambda}\right]^2$$

$$= \mathrm{n}^2 + E\frac{\left(\sum x_i\right)^2}{\lambda^2} - 2\,n\,E\left(\frac{\sum \mathrm{x_i}}{\lambda}\right)$$

but $\left(\sum \mathrm{x_i}\right)^2 = \sum x_i{}^2 + 2\sum x_i\,\mathrm{x_j}$

and $\mathrm{Var}(\mathrm{x}) = E\left(X^2\right) - E(X)^2$ $\quad *$

we have

$$\mathrm{I}(\theta) = E\left(\frac{\left(\sum x_i\right)^2}{\lambda^2} - n^2\right)$$

Substituting $*$ in $I(\theta)$ we have

$$I_x(\theta) = \left[nVar(x) + n\,\mu_{\mathrm{x}}^2 + 2\,\frac{(1(n-1)}{2}\,\lambda^2\right] \Big/ \lambda^2 \quad - n^2$$

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

To give

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

$$= \frac{\mathrm{n}}{\lambda}$$

Thus, the Cramer Rao lower bound for variance of an estimator is the inverse of the Fisher's Information.

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) “Statistical Inference” Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974): Introduction to the Theory of Statistics, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 8: Sufficiency and Fisher-Neyman Criterion

Introduction

The process of finding a sufficient statistic from a distribution is lengthy and complicated. In this study session, you shall be introduced to a theorem that simplifies the search for a sufficient statistics, if one exists.

Learning Outcomes for Study Session 8

At the end of this study session, you should be able to:

8.1 Explain the Sufficiency and Fisher-Neyman criterion

8.1 Fisher-Neyman criterion

To find a sufficient statistic, one has to make a choice of an appropriate statistic first, and then the sufficient statistic should be checked for sufficiency by finding the conditional distribution. If the conditional distribution is not free from θ, then one has to start again.

The search for a sufficient statistic can be simplified if one exists by factorisation criterion often referred to as the Fisher-Neyman criterion.

Definition 1

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population with probability density function $f(x,\theta)$, $\theta \in \Omega$ and let $y_1, y_2, \ldots. y_n$ be a one-to-one set of transformed variable $y_i = \mu_i(\underline{x})$. Let the joint density function of $\underline{x}$ be given by

$$f(\underline{x},\theta) = g(\underline{y},\theta)$$

If the conditional distribution of y_1 given as

$$h(y_2 \ldots\ldots y_n / y_1, \theta) = \frac{g(\underline{y}\ \theta)}{g(y_1,\ \theta)}$$

can be expressed as

$$h(y_2 \ldots\ldots y_n / y_1,\ \theta) = \Phi(y_2 \ldots\ldots y_n)$$

a function which is independent of θ in its mathematical form and in its domain, then y_1 is said to be a sufficient statistic for θ.

Explanation

Let $T = t(x_1, x_2 \ldots\ldots x_n)$ be a statistic from a random sample of size n from the distribution $f(\underline{x}, \theta)$, and if subsequent transformation

$\underline{y} = y_1, \ldots, y_n$ and $y_i = \mu_i(x)$, $i = 1, 2, \ldots, n$ is possible,

Then, the statistic T is sufficient for θ if $f(x_1, x_2, \ldots, x_n, \theta / y_1, \theta) = H(\underline{x})$

where $g(y_1, \theta)$ is the marginal distribution of y_1 and $H(\underline{x})$ is a non-negative function that does not involve θ .

Alternatively

One can say that T(x) is sufficient for θ , if and only if

$$\frac{L(y;\theta)}{g_1(y_1;\theta)} = k(x_1, x_2, \ldots, x_n)$$

where $k(x_1, x_2, \ldots, x_n)$ is free from θ .

Example 1

Let $x_1, x_2, \ldots, x_n$ denote a random sample of size n from a Bernoulli distribution with p.d.f.

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

Consider $y_1 = x_1 + x_2 \ldots\ldots\ldots + x_n$

The distribution of y_1 is binomial with parameter n, θ

$$g_1(y_1, \theta) = \binom{n}{y_1} \theta^{y_1} (1-\theta)^{n-y_1}$$

The joint p.d.f. of x_1, x_2 …..x_n is

$$L(X;\theta) = \prod_{\tau=1}^{n} f(\underline{x_i, \theta}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

The marginal distribution function of y_1 given as

$$g_1(y_1, \theta) = \binom{n}{\sum x_i} \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Let there be a function $H_{(x)}$ defined as:

$$f(y_2 \ldots\ldots\ldots y_n / (y_1, \theta)) = H_{(x)} = \frac{\prod_{\tau=i}^{n} f(\underline{x}_i, \theta)}{g_1(y_i, \theta)}$$

$$= \frac{\theta^{\sum \mathrm{xi}}(1-\theta)^{n-\sum xi}}{\binom{n}{\sum xi} \theta^{\sum \mathrm{xi}}(1-\theta)^{n-\sum xi}}$$

$$= \frac{1}{\binom{\mathrm{n}}{\sum \mathrm{xi}}} \qquad \text{(a function independent of } \theta\text{)}$$

$$\equiv \phi(\mathrm{y}_2, y_3 \ldots\ldots\ldots\ldots y_n)$$

$\prod_{\psi=i}^{n} f(\underline{x}, \theta) = g_1(y_1, \theta)\mathrm{H}(\mathrm{x})$; where $H_{(x)}$ is the conditional density function of y_1.

Thus, y_1 is said to be a sufficient statistic for θ. It is often necessary to obtain marginal distribution g_1 (y_1,θ).

Example 2

For a random sample of size n from an exponential distribution with p.d.f $f(x, \lambda) = \lambda e^{-\lambda x}$, show that the sample total is sufficient for the exponential parameter λ

Solution

A form of the exponential distribution is

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

The sample total is sufficient if

$$\frac{\prod_{i=1}^{n} f(x_i;\lambda)}{f_T t(x_1,...,x_n)} \qquad \text{does not involve } \lambda .$$

The distribution of sample total is Gamma with n and λ as parameters. It can be shown using the moment generating function.

The conditional distribution becomes

$$f(x_1,...,x_n/\lambda) = \frac{\prod_{i=1}^{n} \lambda e^{-\lambda x}}{\lambda^n (\sum x_i)^{n-1} e^{-\lambda \sum x_i} / \Gamma(n)}$$

$$= \frac{\Gamma(n)}{(\sum_i x_i)^{n-1}}$$

which does not contain λ, showing that the sample total is sufficient for λ.

The Theorem

Let x_1, x_2 ,….,x_n be a random sample of size n from a population with function f (x, θ), then the statistic T(x) is sufficient for θ if and only if the joint density function $\prod_{i=1}^{n} f(\underline{x},\theta)$ of x_1, x_2 ,…,x_n is factored as

$$f(x_1,...,x_n;\theta) = k_1(t(x_1,...,x_n);\theta)k_2(x_1,...,x_n)$$

Where k_1 is the sampling distribution of the statistic T(x) and k_2 is a non-negative function that is free from the parameter θ and depend only on x_1, x_2 ,…,x_n only through the function $t(x_1,x_2 ,…,x_n)$.

Proof

Given $\hat{\theta}_1 = \mu_1(x_1.......x_n)$

$$\overset{2}{\theta} = \mu_2\,(x_1.......\ x_n)$$

:

$$\hat{\theta}_n = \mu_n\,(x_1.......\ x_n)$$

by transformation

$$x_1 = w_1\left(\hat{\theta}_1, \hat{\theta}_2..........\hat{\theta}_n\right)$$

:

$$x_n = w_n\left(\hat{\theta}_1, \hat{\theta}_2..........\hat{\theta}_2\right)$$

But

$$\mathrm{f}(\underline{\mathrm{x}}, \theta) = f\left[(w_1 w_2 ....w_n)|J|\right]$$
$$= \mathrm{g}\left(\hat{\theta}_1, \hat{\theta}_2 ....\hat{\theta}_n)/\theta\right)$$

and

$$\mathrm{g}\left(\hat{\theta}_2, \hat{\theta}_3........\hat{\theta}_n / \theta_1\right) = k_1\left(\hat{\theta}_1, \theta\right) k_2\left(\hat{\theta}_1, \theta_2..........\hat{\theta}_n|J|\right)\right)$$

$$= \int\int\int \mathrm{g}\left(\hat{\theta}_1, \theta_2..........\hat{\theta}, \theta\right)\partial\hat{\theta}_1\,\partial\hat{\theta}_2.......\partial\hat{\theta}_{\mathrm{n}}$$

$$g\left(\hat{\theta}_1, \theta\right) = k_1\left(\hat{\theta}_1, \theta\right)\Phi\left(\hat{\theta}_1\right)$$

$$\therefore k_1\left(\hat{\theta}_1, \theta\right) = \frac{g\left(\hat{\theta}_1, \theta\right)}{\Phi\left(\hat{\theta}_1\right)}$$

then

$$f(\underline{x}, \theta) \equiv g\left(\hat{\theta}_1,........\hat{\theta}_1\right) = k_1\left(\hat{\theta}_1, \theta\right)\mathrm{k}_2(x)$$

$$= \frac{\mathrm{g}_1\left(\hat{\theta}_1, \theta\right)\mathrm{k}_2(x)}{\Phi\left(\hat{\theta}_1\right)}$$

$$= \quad k_1\left(\hat{\theta}_1, \theta\right)\int\int_{k=\in} k_2\left\{\hat{\theta}_1..........\hat{\theta}_n\right\}\ \partial\hat{\theta}_1.......\partial\hat{\theta}_{\mathrm{n}}$$

$$= \quad k_1\left(\hat{\theta}_1, \theta\right)\Phi\left[\hat{\theta}_1\right]$$ Note that $\Phi\left(\hat{\theta}_1\right)$ is free from θ

$f(\underline{x},\theta)=k_1(\hat{\theta}_1,\theta)k_2(\underline{x})$

nothing that
$\underline{x}=(x_1,x_2.........x_n)$

Then we can write $k_2(x)/\Phi(\hat{\theta}_1)\equiv H(\underline{x})$

Since we can write

$$g(\underline{\underline{\theta}},\theta)=g_1(\hat{\theta}_1,\theta)\mathrm{q}(\mathrm{x})$$

$$where\ \mathrm{q}(\underline{\mathrm{x}})=\frac{k_2(\underline{x})}{H(\underline{x})}$$

In view of the Fisher-Neyman Criterion, the above indicates sufficiency of $\hat{\theta}_1$ for θ.

Example 3

Let X: ($X_1, X_2, \ldots, X_n$) be a random sample from a distribution with p.d.f.

$$\mathrm{f}(\mathrm{x},\lambda)=\begin{cases}\lambda x^{\lambda-1} & ; \quad 0<\mathrm{x}<1\\ 0 & ; \quad \text{elsewhere}\end{cases}$$

Show that $T=x_1,x_2..................x_n$ is sufficient for θ or λ

Solution

Given that $\mathrm{f}(\mathrm{x},\lambda)=\begin{cases}\lambda x^{\lambda-1} & ; \quad 0<\mathrm{x}<1\\ 0 & ; \quad \text{elsewhere}\end{cases}$

Let $(\underline{\mathrm{x}})=(x_1,x_2..................x_n)$

the joint p.d.f. is

$$\mathrm{f}(\underline{\mathrm{x}},\lambda)=\lambda^n(x_1,x_2..............x_n)^{\lambda-1}$$

thus

$$\mathrm{f}(\underline{\mathrm{x}},\lambda)=\frac{\lambda^n(x_1,x_2..............x_n)^{\lambda}}{(x_1,x_2............x_n)}$$

Put $\quad T = x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n$

$$= \lambda^n T^{\lambda} \frac{1}{(x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n)}$$

$$= k_1(\hat{\theta}; \lambda) k_2(\underline{x})$$

where $\quad k_1(\hat{\theta}; \lambda) = \lambda^n T^{\lambda} \quad$ and $\; k_2(\underline{x}) = \dfrac{1}{(x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n)}$

$\therefore T = x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n$ is sufficient for θ or λ

Example 4

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from $N(\theta, \sigma^2)$; $\infty < \theta < \infty$ where the variance σ^2 is known.

Show that $\overline{X}$ is sufficient for μ

Solution

Let $f(x, \theta) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \; \exp\left\{-\dfrac{1}{2\sigma^2}(x-\mu)^2\right\}$ and the joint density function be defined as:

$$f(\underline{x}, \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Using the Fisher-Neyman criterion

$$\sum(x_i-\mu)^2=\sum[x_i-\bar{x}+\bar{x}-\mu]^2;\ \text{where} \qquad \bar{x}=\frac{1}{n}\sum x_i$$

$$=\sum(x_i-\bar{x})^2+\sum(\bar{x}-\mu)^2-2\sum(x_i-\bar{x})\ (\bar{x}-\mu)$$

$$=\sum(x_i-\bar{x})^2+n(\bar{x}-\mu)^2+0$$

$$\therefore\ f(\underline{x},\mu)=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right)\right\}$$

$$=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)\right\}\left\{-\frac{1}{2\sigma^2}\sum(x_i-\bar{x})^2\right\}$$

$$=k_1(\bar{x},\mu)\,k_2(\underline{x})$$

where $\quad k_1(\bar{x},\theta)=\exp\left\{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)\right\}$ $\quad$ and $k_2(\underline{x})=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\bar{x})^2\right\}$

The first expression on the R.H.S. involves μ , while the second expression does not involve μ . This implies that $\bar{X}$ is sufficient for μ .

Summary for Study Session 8

1. In this study session, the Fisher - Neyman factorisation criterion was explained. The theorem was stated and proved. Some examples were also given to illustrate the application of the theorem.
2. Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population with probability density function $f(x,\theta)$, $\theta\in\Omega$ and let $y_1, y_2, \ldots, y_n$ be a one-to-one set of transformed variable $y_i=\mu_i(\underline{x})$. Let the joint density function of $\underline{x}$ be given by

   $f(\underline{x},\theta)=g(\underline{y},\theta)$

Self-Assessment Questions (SAQs) for Study Session 8

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. For a random sample of size n from an exponential distribution with p.d.f $f(x,\lambda)=\lambda e^{-\lambda x}$, show that the sample total is sufficient for the exponential parameter λ.
2. Let X: ($X_1, X_2, \ldots, X_n$) be a random sample from a distribution with p.d.f.

$$f(x,\lambda)=\begin{cases}\lambda x^{\lambda-1} & ; \quad 0<x<1\\ 0 & ; \quad \text{elsewhere}\end{cases}$$

Show that $T = x_1, x_2 \ldots\ldots\ldots\ldots\ldots x_n$ is sufficient for θ or λ

3. Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from $N(\theta,\sigma^2); \quad \infty<\theta<\infty$ where the variance σ^2 is known.

Show that $\overline{X}$ is sufficient for μ

Notes on SAQ 8

1. A form of the exponential distribution is

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

The sample total is sufficient if

$$\frac{\prod_{i=1}^{n} f(x_i;\lambda)}{f_T t(x_1,\ldots,x_n)} \qquad \text{does not involve } \lambda.$$

The distribution of sample total is Gamma with n and λ as parameters. It can be shown using the moment generating function.

The conditional distribution becomes

$$f(x_1,\ldots,x_n/\lambda) = \frac{\prod_{i=1}^{n} \lambda e^{-\lambda x}}{\lambda^n (\sum x_i)^{n-1} e^{-\lambda \sum x_i} / \Gamma(n)}$$

$$= \frac{\Gamma(n)}{(\sum_i x_i)^{n-1}}$$

which does not contain λ, showing that the sample total is sufficient for λ.

2. Given that $f(x,\lambda) = \begin{cases} \lambda x^{\lambda-1} & ; \quad 0 < x < 1 \\ 0 & ; \quad \text{elsewhere} \end{cases}$

Let $(\underline{x}) = (x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n)$

the joint p.d.f. is

$$f(\underline{x},\lambda) = \lambda^n (x_1, x_2 \ldots\ldots\ldots\ldots x_n)^{\lambda-1}$$

thus

$$f(\underline{x},\lambda) = \frac{\lambda^n (x_1, x_2 \ldots\ldots\ldots\ldots x_n)^{\lambda}}{(x_1, x_2 \ldots\ldots\ldots\ldots x_n)}$$

Put $T = x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n$

$$= \lambda^n T^{\lambda} \frac{1}{(x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n)}$$

$$= k_1(\hat{\theta};\lambda) k_2(\underline{x})$$

where $k_1(\hat{\theta};\lambda) = \lambda^n T^{\lambda}$ and $k_2(\underline{x}) = \dfrac{1}{(x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n)}$

$\therefore T = x_1, x_2 \ldots\ldots\ldots\ldots\ldots\ldots x_n$ is sufficient for θ or λ

3. Let $f(x,\theta) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\dfrac{1}{2\sigma^2}(x-\mu)^2\right\}$ and the joint density function be defined as:

$$f(\underline{x},\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Using the Fisher-Neyman criterion

$$\sum(x_i-\mu)^2 = \sum[x_i-\bar{x}+\bar{x}-\mu]^2; \text{ where } \quad \bar{x} = \frac{1}{n}\sum x_i$$

$$= \sum(x_i-\bar{x})^2 + \sum(\bar{x}-\mu)^2 - 2\sum(x_i-\bar{x})(\bar{x}-\mu)$$

$$= \sum(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2 + 0$$

$$\therefore f(\underline{x},\mu) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right)\right\}$$

$$= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)\right\}\left\{-\frac{1}{2\sigma^2}\sum(x_i-\bar{x})^2\right\}$$

$$= k_1(\bar{x},\mu)\,k_2(\underline{x})$$

where $k_1(\bar{x},\theta) = \exp\left\{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)\right\}$ and $k_2(\underline{x}) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\bar{x})^2\right\}$

The first expression on the R.H.S. involves μ, while the second expression does not involve μ. This implies that $\bar{X}$ is sufficient for μ.

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) "*Statistical Inference*" Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992): *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 9: Exponential Class of Distribution

Introduction

Here, the exponential family of distribution shall be defined. You will be shown how you can obtain a sufficient statistic from a distribution that belongs to the exponential family. The complete sufficient statistic shall also be defined.

Learning Outcomes for Study Session 9

At the end of this study session, you should be able to:

9.1 identify the distribution that belongs to the regular exponential family of distribution;

9.1 Exponential Family Distribution

A family $f(x;\theta),\ \theta \in \Omega$ is said to be a regular exponential family, if we can write $f(x;\theta)$ as

$$f(x,\theta) = \begin{cases} q(\theta)\mathrm{S}(\mathrm{x})\exp\left[\sum \mathrm{p}(\theta)k(x)\right], & \mathrm{a} < \mathrm{x} < \mathrm{b}, \quad \theta \in \Omega \\ 0, & \text{elsewhere} \end{cases}$$

Or alternatively,

$$f(x,\theta) = \begin{cases} Exp\left\{\mathrm{P}(\theta)\mathrm{k}(\mathrm{x}) + q(\theta) + S(x)\right\}; & \mathrm{a} < \mathrm{x} < \mathrm{b}, \quad \theta \in \Omega \\ 0, & \text{elsewhere} \end{cases}$$

where

1. a, b are independent of parameter θ,
2. $p(\theta)$ and $q(\theta)$ are continuous function of θ,
3. $k(x)$ and $\mathrm{S(X)}$ are continuous function of X; $\mathrm{x} \in (a,b)$,
4. $k'(x) \neq 0$

Complete Sufficient Statistic

If a sufficient statistic $\hat{\theta}$ for θ is such that $\hat{\theta} = \sum_i K(x_i)$ from the regular exponential family expressible as $f(x,\theta) = \exp\{k(x)\,P(\theta) + q(\theta) + S(x)\}$ then $\hat{\theta}$ is complete sufficient.

Example 1

Consider a Bernoulli distribution with p.d.f.

$$f(x,\theta) = \theta^x (1-\theta)^{1-x} \qquad x = 0, 1,$$

show that Bernoulli distribution belongs to a regular class of exponential distribution.

Solution

Given the function: $f(x,\theta) = \theta^x (1-\theta)^{1-x} \qquad x = 0, 1$ expressible as

$$In\, f(x,\theta) = x\, In\, \theta + In(1-\theta) - x\, In\,(1-\theta)$$

$$f(x,\theta) = \exp\left\{x\, In \frac{\theta}{1-\theta} - In(1-\theta)\right\}$$

$$k(x) = x$$
$$P(\theta) = In\left(\frac{\theta}{1-\theta}\right)$$
$$q(\theta) = In(1-\theta)$$

Domain (0, 1) of x is independent of θ.

$P(\theta)$ is a continuous function of θ,

k(x) is non trivial.

Hence, Bernoulli distribution belongs to the regular exponential family.

Example 2

Consider the Binomial distribution p.d.f.

$f(x,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$, $x=0,1,2,...,n$, show that $f(x,\theta)$ is a regular exponential family.

Solution

Given $f(x,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$ $\quad x=0,1,2,...,n$

Express $f(x,\theta)$ as $\exp\{S_{(x)}+P(\theta)k_{(x)}+q(\theta)\}$.

That is

$$\text{In } f(x,\theta)=\text{In}\binom{n}{x}+x\text{In }\theta+(n-x)\text{In }(1-\theta)$$

$$=\text{In}\binom{n}{x}+x\text{ In}\left(\frac{\theta}{1-\theta}\right)+n\text{ In}(1-\theta)$$

$$\therefore f(x,\theta)=\exp\left\{\text{In}\binom{n}{x}+x\text{ In}\left(\frac{\theta}{1-\theta}\right)+n\text{ In}(1-\theta)\right\}$$

$$=\exp\{S(x)+P(\theta)k(x)+q(\theta)\}.$$

Where $P(\theta)=\text{In}\left(\frac{\theta}{1-\theta}\right)$

$q(\theta)=n(1-\theta)$

$S(x)=\text{In}\binom{n}{x}$ and

$K(x) = x$

Hence binomial distribution is of discrete exponential family.

Example 3

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from the distribution with p.d.f.

$f(x,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\ell^{-1/2\sigma^2(x-\mu)^2}$ with mean μ and variance σ^2. Show that it belongs to the exponential family and find a joint sufficient statistic for μ and σ^2.

Solution

Given $X\sim N(\mu,\sigma^2)$ and the p.d.f

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \ell^{-1/2\sigma^2 (x-\mu)^2} \qquad -\infty < x < \infty$$

Express $f(x, \mu, \sigma^2)$ as $\exp\{S_{(x)} + P(\theta)k_{(x)} + q(\theta)\}$.

we have

$$= \exp\left\{ In\frac{1}{\sigma\sqrt{2\pi}} - \frac{(x^2 - 2\mu x + \mu^2)}{2\sigma^2} \right\}$$

$$= \left\{ In\frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2\sigma^2} + \frac{2\mu x}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} \right\}$$

$$= -\frac{1}{2} In 2\pi - In\frac{1}{2\sigma^2} - x^2 \frac{1}{2\sigma^2} + x\frac{\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}$$

where

$$S_{(x)} = -\frac{1}{2} In 2\pi$$

$$k_1(x) = x, \ldots\ldots\ldots\ldots k_2(x) = x^2$$

$$P(\mu, \sigma^2) = \mu/2\sigma^2$$

$$q_1(\mu, \sigma^2) = \mu^2/2\sigma^2, , , , , , , , q_2(\mu, \sigma^2) = \frac{1}{2\sigma^2}$$

Hence, the normal distribution is of continuous exponential distribution.

We can show that the joint sufficient statistic for μ and σ^2 is $T(x) = \left(\sum_i K_1(x_i), \sum_i K_2(x_i^2) \right)$

is

$$T(x) = \left(\sum_i X_i, \sum_i x_i^2 \right)$$

Remember

A complete sufficient statistic $\hat{\theta}$ for θ is such that $\hat{\theta} = \sum k(x_i)$ from the exponential family of $f(x,\theta) = \exp\{k_{(x)} P(\theta) + q(\theta) + S(x)\}$.

Example 4

Let x_1, x_2 , …,x_n be a random sample of size n from the distribution with p.d.f.

$$f(x,\lambda) = \frac{\ell^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots\ldots\ldots\ldots.$$

Is the statistic $\hat{\lambda} = \sum x_i$ complete sufficient

Solution

Examine if $f(x,\lambda)$ is expressible as:

$$f(x,\lambda) = \exp\{P(\lambda)k(x) + q(\lambda) + S(x)\}$$

i.e.

$$\text{Inf}(x,\lambda) = \exp\left\{In \frac{(\ell^{-\lambda}\lambda^x)}{x!}\right\}$$

$$= \exp\{-\lambda + x\,\text{In}\lambda - \text{In}(x!)\}$$

where

$P(\lambda) = In\,\lambda$

$k(x) = x$

$q(\lambda) = -\lambda$

$S(x) = -In\,(x!)$

i.e

$\hat{\lambda} = \sum x_i = \sum k(x_i)$

Hence, the family of poisson distribution for $\theta \in (0,\infty)$ is a regular exponential family and since $\hat{\lambda} = \sum k(x_i)$; $\hat{\lambda}$

Hence $\hat{\lambda}$ is complete sufficient for λ

Example 5

Check whether the Laplace distribution with p.d.f. $f(x;\theta)=\frac{1}{2}\exp-|x-\theta|, \quad -\infty < x < \infty, i\infty < \theta < \infty$

with parameter θ is a regular exponential family.

Solution

This can not be written in the form $\exp\{S_{(x)} + P(\theta)k_{(x)} + q(\theta)\}$

Therefore, Laplace distribution does not belong to the regular exponential family.

Location Parameter

A parameter θ of a distribution $f(x,\theta)$ is a location parameter if

$$f(x,\theta) = f_0(x-\theta)$$

A parameter is however scale parameter if

$$f(x,\theta) = \frac{1}{\theta} \quad f_0\left(x/\theta\right)$$

where $f_0(Z)$ is a p.d.f that is free of unknown parameter including θ.

Example 6

Consider $f(x,\mu) = \frac{1}{\sigma\sqrt{2\pi}}\ell^{-}\frac{(x-\mu)^2}{2\sigma^2} \qquad -\infty < x < \infty$

Is a location parameter?

Solution

$$f(x,\mu) = \frac{1}{\sigma\sqrt{2\pi}}\ell^{-}\frac{(x-\mu)^2}{2\sigma^2}$$

regard σ^2 as a constant

$$f(x,\mu) = f_0(x-\theta)$$

μ is a location parameter

However $f(x,\sigma^2)$ can not be expressed as

$$f\left(x,\sigma^2\right), \neq f_0\left(x - \sigma^2\right)$$

Hence σ^2 is not a location parameter.

Pitmans Estimator

Let X_1, X_2............X_n denote a r.s. from the density $f(x,\theta)$ where θ is a location parameter and Ω is the real line. Then the estimator

$$\hat{\theta} = \frac{\int \theta \, f(\underline{x},\theta)\, d\theta}{\int f\,(\underline{x},\theta)\, d\theta}$$

defines the pitman estimator

Example 7

For Normal distribution $\left[\left(i.e.\ X \sim N\left(\mu,\sigma^2\right)\right]\right.$ where μ is a location parameter, obtain the Pitman estimator for μ .

Solution

Given

$$f(x,\theta) = \frac{1}{\sigma\sqrt{2\pi}}\, \ell^{-}\, \frac{(x-\mu)^2}{2\sigma^2}$$

$$f(\underline{x},\theta) = f(x_1, x_2 ....x_n;\theta)$$

$$= \frac{1}{\sigma^n (2\pi)^{n/\partial}} \quad \ell^{-}\, \frac{\sum(x-\mu)^2}{2\sigma^2}$$

So the Pitman estimator for μ is

$$\hat{\theta} = \tau(x_1, x_2 ....x_n) = \frac{\int \mu \frac{1}{\left(\sigma^2 2\pi\right)^{n/2}}\, \ell^{-}\, \frac{\sum(x-\mu)^2}{2\sigma^2}\, d\mu}{\frac{1}{(\sigma\, 2\pi)^{n/2}}\, \frac{\sum(x-\mu)^2}{2\sigma^2}\, d\mu}$$

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

and $\sum(x-\mu)^2 = \sum x^2 - 2\mu\sum x + n - \mu^2$

We have

$$\hat{\theta} = \int \frac{\mu\left(2\pi\sigma^2\right)^{-n/2} \exp\left\{\frac{1}{2\sigma^2}\left(\sum x^2 - 2n\bar{x} - \mu + n\mu^2\right\}d\mu}{\left(2\pi\sigma^2\right)^{-n/2} \exp\left\{\frac{1}{2\sigma^2}\left(\sum x^2 - 2n\bar{x} - \mu + n\mu^2\right\}d\mu}$$

But $\ell^{\frac{-1}{2\sigma^2}}\sum x^2$ is common to be with numeration and denomination and is independent of the variable of integrator.

$$\therefore \hat{\theta} = \int \frac{\mu \exp\left\{\frac{1}{2\sigma^2}\left(n - \mu^2 - 2n\mu\bar{x} + \sigma^2\right)\right\}d\mu}{\int \exp\left\{\frac{1}{2\sigma^2}\left(n - \mu^2 - 2n\mu\bar{x}\right)\right\}d}$$

$$= \int \frac{\mu \exp\left\{\frac{-n}{2\sigma^2}\left(\mu^2 - 2\mu\bar{x} + \bar{x}^2\right)d\mu\right.}{\exp\left\{\frac{-n}{2\sigma^2}\left(\mu^2 - 2\mu\bar{x} + \bar{x}^2\right)d\mu\right.}$$

$$= \int \frac{\mu\, \ell \frac{-n}{2\sigma^2}\left(\mu - \bar{x}\right)d\mu}{\int \ell \frac{-n}{2\sigma^2}\left(\mu - \bar{x}\right)^2 d\mu}$$

$$= \frac{\int \ell^{-} \frac{\left(\bar{x} - \mu^1\right)}{2\,\sigma^2/n}\, d\mu}{\int \ell^{-} \frac{\left(\bar{x} - \mu^1\right)}{2\,\sigma^2/n}\, d\mu}$$

$$= \bar{x}$$

So $\bar{x}$ which is Pitman estimator of the location parameter μ .

Therefore, μ is UMVUE of $\hat{\theta}$ i.e. the estimator that is best among location invariant estimators is also best among unbiased estimators.

Summary for Study Session 9

You have defined the exponential family of distribution. You established that some distributions actually belong to the regular exponential family. Complete sufficient statistic was also defined and established for some distributions. The concept of Pitman estimator was also introduced to you

Self-Assessment Questions (SAQs) for Study Session 9

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Consider a Bernoulli distribution with p.d.f.

$$f(x,\theta) = \theta^{x}(1-\theta)^{1-x} \qquad x = 0, 1,$$

show that Bernoulli distribution belongs to a regular class of exponential distribution.

2 Let x_1, x_2 ,…,x_n be a random sample of size n from the distribution with p.d.f.

$f(x,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \ell^{-1/2\sigma^2 (x-\mu)^2}$ with mean μ and variance σ^2. Show that it belongs to the exponential family and find a joint sufficient statistic for μ and σ^2.

Notes on SAQ

1 Given the function: $f(x,\theta)=\theta^x(1-\theta)^{1-x}$ $x=0,1$ expressible as

$$In\,f(x,\theta)=x\,In\,\theta+In(1-\theta)-x\,In\,(1-\theta)$$

$$f(x,\theta)=\exp\left\{x\,In\frac{\theta}{1-\theta}-In(1-\theta)\right\}$$

$$k(x)=x$$
$$P(\theta)=In\left(\frac{\theta}{1-\theta}\right)$$
$$q(\theta)=In(1-\theta)$$

Domain (0, 1) of x is independent of θ.

$P(\theta)$ is a continuous function of θ,

k(x) is non trivial.

Hence, Bernoulli distribution belongs to the regular exponential family.

2. Given $X\sim N(\mu,\sigma^2)$ and the p.d.f

$$f(x,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\,\ell^{-1/2\sigma^2(x-\mu)^2}\qquad -\infty<x<\infty$$

Express $f(x,\mu,\sigma^2)$ as $\exp\{S_{(x)}+P(\theta)k_{(x)}+q(\theta)\}$

we have

$$= \exp\left\{ In\frac{1}{\sigma\sqrt{2\pi}} - \frac{\left(x^2 - 2\mu x + \mu^2\right)}{2\sigma^2}\right\}$$

$$= \left\{ In\frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2\sigma^2} + \frac{2\mu x}{2\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\}$$

$$= -\frac{1}{2}In2\pi - In\frac{1}{2\sigma^2} - x^2\frac{1}{2\sigma^2} + x\frac{\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}$$

where

$$S_{(x)} = -\frac{1}{2}In2\pi$$

$$k_1(x) = x, \ldots\ldots\ldots\ldots k_2(x) = x^2$$

$$P(\mu, \sigma^2) = \mu / 2\sigma^2$$

$$q_1(\mu, \sigma^2) = \mu^2 / 2\sigma^2 \text{ ,,,,,,,, } q_2(\mu, \sigma^2) = \frac{1}{2\sigma^2}$$

Hence, the normal distribution is of continuous exponential distribution.

We can show that the joint sufficient statistic for μ and σ^2 is $T(x) = \left(\sum_i K_1(x_i), \sum_i K_2(x_i^2)\right)$ is

$$T(x) = \left(\sum_i X_i, \sum_i x_i^2\right)$$

Remember

A complete sufficient statistic $\hat{\theta}$ for θ is such that $\hat{\theta} = \sum k(x_i)$ from the exponential family of $f(x,\theta) = \exp\{\mathrm{k}_{(x)}\,\mathrm{P}(\theta) + q(\theta) + S(x)\}$.

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) "Statistical Inference" Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag, New York.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Study Session 10: Interval Estimation

Introduction

Since point estimators - $\hat{\theta}'s$ do not give the level of reliability, it might be desirable to allow for some 'degree of error' in the estimation of the 'true value' parameter θ. This session therefore is devoted to the methods and techniques of obtaining the interval estimates of a parameter value.

Learning Outcomes for Study Session 8

At the end of this study session, you should be able to:

10.1 Discus the confidence interval (C.I) for μ in a normal distribution.

10.1 Interval Estimation

A point estimate on its own does not convey any indication of unreliability in the estimation of the parameter θ. But the point estimator $\hat{\theta}$ together with its standard error $S.E(\hat{\theta})$ provides some degree of reliability. This idea is incorporated into Confidence Interval, a range of values within which we are 'fairly confident' that the true (unknown) value of the parameter θ lies.

The length and location of the interval are random variables and the level of certainty that θ will actually fall within the limits would have to be determined using the level of significance (α). The object of Confidence Interval is to generate narrow interval which includes θ with high degree of probability.

Definition

A $100(1-\alpha)\%$ confidence interval estimation for θ is an interval within which we are $100(1-\alpha)\%$ sure will contain the parameter θ.

Properties

The following are the desirable properties of interval estimation:

1. Confidence coefficient should be as high as possible i.e
   a. Pr (interval estimator will enclose the parameter) = $1-\alpha$
2. Confidence coefficient should be expressed as a percentage i.e
   $100(1-\alpha)\%$ is the probability that the interval estimator contains the parameter.
3. Margin of error should be as small as possible i.e $e = Z_{\alpha/2}\dfrac{S.E(\hat{\theta})}{\sqrt{n}}$ should be as small as possible.

Pivotal Quantity

This is a very useful method of construction of a Confidence Interval for a population quartile.

A pivotal quantity Q is of the form $Q = \dfrac{\hat{\theta}-\theta}{\sqrt{1/nI(\theta)}}$ ~ N(0,1) where $\hat{\theta}$ is the estimator, θ is the parameter value and $I(\theta)$ is the observed information (remember!).

The pivotal quantity has two properties:

3. It is a function of the sample X_1, X_2, …,X_n and the target parameter θ,
4. The probability distribution of the pivotal quantity is free of θ.

Theorem

Suppose $T = t(x)$ is a reasonable point estimate of θ. Let $P(t,\theta)$ be a pivotal quantity from a known probability distribution function, then for a specified α, $(0<\alpha<1)$ and constants a and b such that (a<b)

$$P[a < P(T,\theta) < b] = 1-\alpha \quad .$$

So, given T, the above inequality is solved for θ to contain a region of θ-values which is a confidence region (usually an interval) for θ corresponding to observe T – values. This re-arrangement result in equation of the form $\Pr\left(\hat{\theta}_1 < \theta < \hat{\theta}_2\right) = 1-\alpha$ for some specified probability $1-\alpha$ where α is the level of confidence and $100(1-\alpha)$ is called the degree of confidence for θ. The end-points of the interval $\hat{\theta}_1$ and $\hat{\theta}_2$ are called the upper and lower *confidence limits* respectively.

Confidence interval estimate for Population Mean:

[Large sample size (n>30)]

Let $X_1, X_2, \ldots, X_n$ be a random sample of values taken from a population with an unknown population mean μ and a known population variance σ^2 .

From our knowledge of point estimation, we know that the sampling distribution of sample mean $\overline{X}$ from a random samples of size n forms a normal population with the mean μ is $\mu_{(\overline{X})} = \mu$ and the variance σ^2 is $\sigma^2_{(\overline{X})} = \dfrac{\sigma^2}{n}$.

For specified value of α, we define a pivotal quantity Z for μ as:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \sim \quad N(0,1)$$

Then we can write

$$\Pr\left(|Z| < z_{\alpha/2}\right) = \alpha$$

where $\alpha/2$ is the integral of the standard normal density function

$$P(z) = \int \frac{1}{2\pi} \, \ell^{-\frac{z^2}{2}} \, dz = \alpha/2$$

Therefore, we can express the possible size of error in our estimation as

$$\Pr\left(\left|\overline{X} - \mu < Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right|\right) = 1 - \alpha$$

Re-arranging the above, we have:

$$\Pr\left(\overline{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

It follows that the $(1-\alpha)100\%$ confidence interval for the mean of the population (μ) is

$$\left(\overline{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad ; \quad \overline{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

Where $\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $\bar{X} + Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}$ are the lower and upper confidence limits respectively?

Therefore, a $100(1-\alpha)\%$ confidence interval estimate for μ is given as:

$$\bar{X} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

and $B = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ is the margin of error while $W = 2B$ is the width of the confidence interval.

In-Text Question

What is Pivotal Quantity?

In-Text Answer

A pivotal quantity Q is of the form $Q = \frac{\hat{\theta} - \theta}{\sqrt{1/nI(\theta)}}$ ~ N(0,1) where $\hat{\theta}$ is the estimator, θ is the parameter value and $I(\theta)$ is the observed information (remember!).

Note that:

1. Confidence interval for μ could be based on the sample median or mid-range with mean deviation from the median as the measure of variation.
2. Even though the confidence interval for mean is based on the assumption of normal population and known variance σ_1^2 by virtue of central limit theorem, the results can be used for random sample from non-normal population provided *n* is sufficiently large i.e. $n \geq 30$.

Example 1

To estimate the average time required for certain repairs, an automobile manufacturer had a random sample of 40 mechanics, tuner in the performance of task. If it took them on the average 24.05 minutes with a standard deviation of 2.68 minutes to complete the repairs, what can the manufacture assert with 95% confidence about the maximum error, if he uses $\bar{X} = 24.05$ minutes as an estimate of the actual mean time required to perform the given repairs?

Solution

$n = 40, \quad \sigma = 2.68, \quad (1 - \alpha)100\% = 95\%$

$\overline{X} = 24.05, \quad \mu = 23.05 \therefore \ \alpha = 0.05$

$$If \quad \Pr\left(\left| \overline{X} - \mu \right| < Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

Maximum error is

$$Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$1.96 . \frac{2.68}{\sqrt{40}} = 0.8306$$

The 95% confidence interval is

$$\Pr\left[\overline{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha$$

$$\Pr\left[24.05 - 0.8306 < \mu < 24.05 + 0.8306 \right] = 0.95$$

$$i.e. \left(23.22 < \mu < 24.88 \right)$$

Example 2

It is desired to estimate the mean number of unoccupied seats per flight, μ for a major airline. A random sample of $n = 225$ flights shows that the sample mean is 11.6 and the standard deviation is 4.1.

3. Give a 95% bound on the error in the estimation of CI, also obtain the width of the interval
4. Find a 90% and 99% confidence interval for μ
5. Interpret the CI found in (2) above.

Solution

We have $100(1-\alpha)\% = 95\%$ implies $\alpha = 0.05$ and $Z_{\alpha/2} = Z_{0.025} = 1.96$ (from the standard normal table).

1. The bound on the error is $B = 1.96\left(\frac{4.1}{\sqrt{225}} \right) = 0.5357$ and the width is

$$W = 2B = 2\left[1.96\left(\frac{4.1}{\sqrt{225}} \right) \right] = 1.0714$$

2. $100(1-\alpha)\% = 90\%$ implies that $\alpha = 0.10$ and $Z_{\alpha/2} = Z_{0.05} = 1.645$

The CI for μ is $\bar{X} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$$11.6 \pm 1.645\left(\frac{4.1}{\sqrt{225}}\right)$$

$$11.6 \pm 0.45 = (11.15\ ;\ 12.05)$$

$100(1-\alpha)\% = 99\%$ implies that $\alpha = 0.01$ and $Z_{\alpha/2} = Z_{0.005} = 2.58$

The CI for μ is $\bar{X} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$$11.6 \pm 2.58\left(\frac{4.1}{\sqrt{225}}\right)$$

$$11.6 \pm 0.5923 = (11.01\ ;12.19)$$

3. The intervals contain μ with probability 0.90 and0.99 respectively.

OR

If repeated sampling is used, the 90% and 99% of CI constructed would respectively contain μ .

Confidence interval estimate for Population Mean:

[Small sample size (n<30)]

Let X_1, X_2, …,X_n be a random sample of values taken from a population with an unknown population mean μ and an unknown population variance σ^2 .

When the sample size is small i.e. $n \le 30$, and σ is unknown, we make use of the (Pivotal quantity) standardised variable defined as

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

where S i.e. the sample standard deviation.

The standardised random variably has the t-distribution with *n-1* degrees of freedom.

Thus, the confidence interval for T is $\Pr\left(|T| < t_{\alpha/2, n-1}\right) = 1-\alpha$

and the confidence interval for population mean μ is given as

$$\left(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

The above is referred to as the small sample confidence interval for population mean for (μ).

Thus, a $100(1-\alpha)\%$ confidence interval estimate for μ is given as:

$$\overline{X} \pm t_{\alpha/2,n-1}\left(\frac{S}{\sqrt{n}}\right)$$

and $\mathrm{B} = t_{\alpha/2,n-1}\left(\frac{S}{\sqrt{n}}\right)$ is the margin of error while 2B is the width of the confidence interval.

We could see from the above that the pivotal quantity can be applied to both large and small sample sizes.

Example 3

The weights of 10 tubers of yam randomly selected from a farm are 9.8, 9.9, 10.3, 10.4, 10.3, 10.2, 9.7, 10.1, and 9.8 kilogram's. Find the (i) 95% and 99% confidence interval for the population mean μ of all the yams in the farm.

Solution

n =10, i.e n is small, assuming a normal population with student t- distribution.

$$\overline{X} = \frac{9.8+9.8+...+9.8}{10} = 10.06 \;; \qquad S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = 0.246$$

1. $\alpha = 0.05, \quad \alpha/2 = 0.025$

$$\overline{X} \pm t_{\alpha/2,n-1}\left(\frac{S}{\sqrt{n}}\right)$$

$$10.06 \pm t_{0.025,9}\left(\frac{0.246}{\sqrt{9}}\right) = 10.06 \pm 2.26\left(\frac{0.246S}{\sqrt{9}}\right) = 10.06 \pm 0.0185$$

The 95% CI is

$(10.042 \;;\; 10.245)$

2. $\alpha = 0.01, \qquad \alpha/2 = 0.005$

$$\overline{X} \pm t_{\alpha/2,n-1}\left(\frac{S}{\sqrt{n}}\right)$$

$$10.06 \pm t_{0.005,9}\left(\frac{0.246}{\sqrt{9}}\right) = 10.06 \pm 3.25\left(\frac{0.246S}{\sqrt{9}}\right) = 10.06 \pm 0.2265$$

The 99% CI is

$(9.794 \;\; ; \;\; 10.326)$

Determination of Sample Size

Sometimes, it may be desirable to estimate the sample necessary to achieve a specified level of precision or to obtain the sample size that would increase or decrease the margin of error of an estimate. To achieve this for a the population mean, the sample size n can be obtained for the formula for the margin of error :

$n \cong \frac{(Z_{\alpha/2})^2 \sigma^2}{B}$, where $B = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ or a predetermined error level.

Note

3. If the sample size n is increased, the margin of error decreases, and the width of the CI decreases.
4. In the absence of data, σ is sometimes approximated by $\frac{R}{4}$ where R is the range.

Example 4

Suppose you want to construct a 99% CI for μ so that W = 0.05. You are told that preliminary data ranges from 13.3 to 13.7. What sample size should you choose?

Solution

Data summary: $\alpha = 0.01$; R = 13.7 – 13.3 = 0.4;

So $\sigma = 0.4/4 = 0.1$

Now B = W/2 = 0.05/2 = 0.025

Therefore,

$$n \cong \frac{(2.58)^2 0.1^2}{0.025^2} = 106.50$$

Example 5

If it is known that the height of men in a particular community is normally distributed with mean 170cm and a standard deviation of 6.35cm.How large a sample should be taken in order to be 99% sure that the sample mean does not differ from the true mean by more that 3.7cmin absolute term?

Solution

Data summary: $1-\alpha = 0.99$; $\alpha = 0.01$; $Z_{\alpha/2} = 2.58$ and

$$\mu = 170; \quad \sigma = 6.35 \quad \text{and } \left|\overline{X} - \mu\right| = 3.7$$

Then

$$n \cong \frac{(2.58)^2 (6.35)^2}{3.7^2} = 19.6 \cong 20$$

n = 20

Example 6

Let X_1, X_2, …,X_n be a random sample from $N(\mu,1)$. What sample size can half the length of a confidence interval, at $\alpha = 0.05$?

Solution

$B = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = 1.96\left(\frac{1}{\sqrt{n}}\right) = \frac{1.96}{\sqrt{n}}$, therefore W= 2B = 2(1.96)

$$\sqrt{n}\, 0.5 = 3.29$$

$$0.25\, n = 15.3664$$

that is

$\therefore$ n = 4X(15.3664)

n $\cong$ 62

Confidence Interval estimate for Single Binomial Population

Given a Bernoulli sequence of length n, with probability of success p in a single trial, we might be interested in estimating the CI for p. The parameter of a binomial population is population proportion P, which can be estimated by sample proportion p and the standard error is $\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$, therefore:

A $100(1-\alpha)\%$ confidence interval estimate for p is given as:

$$\hat{p} \pm Z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\,\hat{q}}{n}}\right)$$

where $\hat{q} = 1 - \hat{p}$

and $B = Z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\,\hat{q}}{n}}\right)$ is the margin of error while $W = 2B$ is the width of the confidence interval.

Example 7

In a sample of 20 Bernoulli trials, 6 successes were observed. Obtain a 95% confidence interval for p.

Solution

$\hat{p} = 6/20 = 0.3, \quad Z_{\alpha/2} = 1.96$

We have $\hat{p} \pm Z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\,\hat{q}}{n}}\right)$

$$0.3 \pm 1.96\left(\sqrt{\frac{0.3(0.7)}{20}}\right)$$

$$0.3 \pm 0.2008$$

$$(0.0992 \;;\; 0.5008)$$

Example 8

A random sample of n = 484 voters in a community produced x = 257 votes in favour of candidate A.

1. What is the point estimate of p and its margin of error?
2. Find a 90% confidence Interval for P.

Solution

1. $\hat{p} = \frac{x}{n} = \frac{257}{484} = 0.531$

   Since the level of significance is not given, we assume $\alpha = 5\%$.

   therefore $B = 1.96\left(\sqrt{\frac{0.531(0.469)}{484}}\right) = 0.044$

2. The 90% CI is

   $0.531 \pm 1.645\left(\sqrt{\frac{o.531(0.469)}{484}}\right)$

   0.531 ± 0.037

   $(0.494 \quad ; \quad 0.568)$

Sample size estimate for a binomial population

The estimate of sample size is

$$n \cong \frac{(Z_{\alpha/2})^2 \, \hat{p} \, \hat{\mathrm{q}}}{B^2}$$

Example 9

Suppose you want to provide an accurate estimate of customers preferring one brand of coffee over another. You need to construct a 95% CI for p so that B = 0.015.You are told that the preliminary data shows $\hat{p} = 0.35$ .What sample size should you choose? Use $\alpha = 5\%$.

Solution

Data summary: $\alpha = 0.05, \quad \hat{\mathrm{p}} = 0.35, \quad \mathrm{B} = 0.015$

$$n \cong \frac{(Z_{\alpha/2})^2 \, \hat{p} \, \hat{\mathrm{q}}}{B^2}$$

$$n \cong \frac{(1.96)^2 0.35(0.65)}{0.015^2} = 3884.28$$

$$= 3{,}885$$

So n = 3885

Confidence Interval estimate for Population Variance

We have obtained the CI for population mean μ from a sample $N(\mu, \sigma^2)$, when σ^2 is known or unknown. Now we shall obtain CI for σ^2 when μ is known or unknown.

When μ is known

Let X_1, X_2, …,X_n be a random sample from . If μ is known, say $\mu = \mu_0$

$\sum_{i=1}^{n} \left(\frac{X_i - \mu_0}{\sigma} \right)^2$ is the sum of squares of standard normal variates and hence it has a chi-square distribution with n degrees of freedom (d.f) irrespective of the value of σ^2.

Using the χ^2 - tables we can get, a and b such that

$$\Pr\left[a < \sum_{i=1}^{n} \left(\frac{X_i - \mu_0}{\sigma} \right)^2 \le b \right] = 1 - \alpha$$

One such pair (a,b) is obtained by choosing a as the lower $\alpha/2$ point and b as the upper $\alpha/2$ point of the Chi-square distribution. i.e

$$a = \chi^2_{n,(1-\alpha/2)}, \quad b = \chi^2_{n,(\alpha/2)}.$$

From the above we have:

$$\Pr\left[\chi^2_{n,(1-\alpha/2)} < \sum_{i=1}^{n}\left(\frac{X_i-\mu_0}{\sigma}\right)^2 \le \chi^2_{n,(\alpha/2)}\right] = 1-\alpha$$

$$\Pr\left[\frac{1}{\chi^2_{n,(1-\alpha/2)}} \ge \frac{\sigma}{\sum_{i=1}^{n}(X_i-\mu_0)^2}\sum_{i=1}^{n}\left(\frac{X_i-\mu_0}{}\right)^2 \ge \frac{1}{\chi^2_{n,(\alpha/2)}}\right] = 1-\alpha \qquad (**)$$

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{n,(1-\alpha/2)}} \ge \sigma \ge \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{n,(\alpha/2)}}\right] = 1-\alpha$$

Thus

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{n,(1-\alpha/2)}} \;\ge\; \sigma \;\ge\; \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{n,(1-\alpha/2)}}\right)$$

is the CI for σ^2 at $(1-\alpha)$ level.

When μ is known

In this case, we may use the statistic $\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2$ where $\overline{X}$ is the sample mean, which has the Chi-square distribution with (n-1) degrees of freedom instead of $\sum_{i=1}^{n}\left(\frac{X_i-\mu_0}{\sigma}\right)^2$ in (**) above.

Proceeding as above, CI for σ^2 is given by

$$\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\chi^2_{(n-1),(1-\alpha/2)}} \geq \sigma \geq \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\chi^2_{(n-1),(1-\alpha/2)}} \right).$$

Example10

A random sample of size 50 from a normal population gave a sample variance of 4. Find a 90% CI for the population variance σ^2.

Solution

Since $S^2 = 4, \quad n = 30, \quad nS^2 = 120$. Choosing $\alpha = 1, \quad \alpha/2 = 0.05, \quad 1\text{-}\alpha/2 = 0.95$

from a chi-square table

$\chi^2_{29,0.05} = 42.5569, \quad \chi^2_{29,0.95} = 17.7084$

Therefore a 90% CI for σ^2 is

$$\left(\frac{120}{42.5569}, \quad \frac{120}{17.7084} \right)$$

$$(2.8189, \quad 6.7764)$$

We can observe that 6.7764 is the 90%upper confidence bound and 2.8981 is the 90% lower confidence bound for σ^2.

Confidence interval estimate for Two Quantitative Populations

It may be desired to estimate the sum or difference of means from two populations.

Let there be two samples of sizes n_1 and n_2 with their respective means $\overline{X_1}$, $\overline{X_2}$ and sample standard deviations s_1 and s_2 from two populations 1 and 2.

The estimator of difference of means (for example) is $(\mu_{\overline{x_1}} - \mu_{\overline{x_2}}) = (\mu_1 - \mu_2)$ and the standard error of difference of means is $S.E(\overline{X_1} - \overline{X_2}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Therefore, the confidence interval for difference of means is defined as:

C.I is $$(\overline{X}_1 - \overline{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

The margin of error is $B = Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

and the formula for estimating the sample size in the two populations is

$$n \cong \frac{(Z_{\alpha/2})^2(\sigma_1^2 + \sigma_2^2)}{B^2}$$

Confidence interval estimate for Two Binomial Populations

It may be desired to estimate the sum or difference of proportions from two populations.

Let there be two random samples: sample 1 and sample 2 of sizes n_1 and n_2 with their respective proportions $\hat{p}_1 = \frac{x_1}{n_1}$, $\hat{p}_2 = \frac{x_2}{n_2}$ with unknown parameters P_1 and P_2 .

The point estimator of difference in proportions (for example) is $(\mu_{\hat{p}_1} - \mu_{\hat{p}_2}) = (P_1 - P_2)$ and the standard error of difference of means is $S.E(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Therefore, the confidence interval for difference of proportions is defined as:

C.I is $$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

The margin of error is $B = Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

and the formula for estimating the sample size in the two populations is

$$n \cong \frac{(Z_{\alpha/2})^2(\hat{p}_1\hat{q}_1 + \hat{p}_2\hat{q}_2)}{B^2}$$

For unknown parameters, the formula is

$$n \cong \frac{(Z_{\alpha/2})^2(0.5)}{B^2}$$

Summary for Study Session 10

1. In this study session, you have introduced to you the interval estimation technique, CI for the population mean, for proportion and CI for the population variance.
2. The formula for estimating the sample size necessary to obtain a specified level of precision was also derived. Some relevant examples were also given to illustrate the procedures. You were also given the procedure for interval estimation of difference and sum of means and proportions of two populations

Self-Assessment Questions (SAQs) for Study Session 10

Having completed this study session, you can mea sure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Suppose you want to provide an accurate estimate of customers preferring one brand of coffee over another. You need to construct a 95% CI for p so that B = 0.015.You are told that the preliminary data shows $\hat{p} = 0.35$.What sample size should you choose? Use $\alpha = 5\%$.
2. A random sample of size 50 from a normal population gave a sample variance of 4. Find a 90% CI for the population variance σ^2.

Notes on SAQ

1. Data summary: $\alpha = 0.05, \quad \hat{\text{p}} = 0.35, \quad \text{B} = 0.015$

$$n \cong \frac{(Z_{\alpha/2})^2 \, \hat{p}\, \hat{\text{q}}}{B^2}$$

$$n \cong \frac{(1.96)^2 \, 0.35(0.65)}{0.015^2} = 3884.28$$

$$= 3{,}885$$

So n = 3885

2. Since $S^2 = 4, \quad n = 30, \quad nS^2 = 120$. Choosing $\alpha = 1, \quad \alpha/2 = 0.05, \quad 1 - \alpha/2 = 0.95$ from a chi-square table

$\chi^2_{29,0.05} = 42.5569, \quad \chi^2_{29,0.95} = 17.7084$

Therefore a 90% CI for σ^2 is

$$\left(\frac{120}{42.5569}, \quad \frac{120}{17.7084} \right)$$

$$(2.8189, \quad 6.7764)$$

We can observe that 6.7764 is the 90%upper confidence bound and 2.8981 is the 90% lower confidence bound for σ^2.

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) "Statistical Inference" Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

El-Taha, Mohammed (2003) "Introduction to Probability and Statistics-Lecture Note"An unpublished lecture notes, Department of Mathematics and Statistics, University of Southern Maine, Potland.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Murison, Bob (2008) "Statistical Inference Part II" An unpublished lecture note, School of Maths. Stats and Comp., University of New England, Armidale, NSW, 2351; http://turing.une.edu.au/

Olubusoye, O.E, Olaomi, J.O and Shittu, O.I (2002) *Statistics for Engineering, Physical and Biological Sciences* A Divine Touch Publications, Ibadan,

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Shagodoyin, D.K, Olubusoye, O.E, Shittu, O.I and Adepoju A.A (2002) *Statistical Theory and Methods* Joytal Printing Press, Ibadan,

Study Session 11: Testing of Hypothesis

Introduction

In the previous study session, you learnt the methods and procedures for estimation of parameters from observed data. Various criteria for evaluating a good estimator were extensively discussed.

In this study session, you will be taught another important aspect of statistical inference- Hypothesis testing. It involves testing the validity of a statistical statement about a stochastic model and its parameters on the basis of sample data. This session will provide sufficient theoretical background for standard tests.

Learning Outcomes for Study Session 11

At the end of this study session, you should be able to:

11.1 Explain the concepts of a statistical hypothesis test.

11.1 Test of Hypothesis

Definition: A statistical hypothesis is a statement or assertion about the stochastic model followed by one or more random variables in particular about the parameter(s) of the probability distribution. Examples of a statistical hypothesis statements such as 'The score of third year students in STA 321 follows the normal distribution and that the average score is greater than 55%'; 'the distribution of lifetime of electric fuses produced by a company is exponential with average life of at least 350 hours.

Every statistical statement involves a decision between two statistical hypotheses, true or null hypothesis and the alternative hypothesis.

Null hypothesis is always a statement of either ‘no effect’ or ‘status quo’. It is symbolised by H_0.

Alternative hypothesis is a statement, which supports the statement (of interest) we are trying to test. It is symbolised by H_1.

The two hypothesis divides a parameter space into two sub-spaces Ω_0 and Ω_1 such that $\Omega_0 \cup \Omega_1 = \Omega$ and $\Omega_0 \cap \Omega_1 = \Phi$ (empty set). Either Ω_0 or Ω_1 corresponds to only one of the sub-spaces H_1 or H_0 and not both. Therefore H_1 and H_0 are complimentary hypothesis.

Simple and Complimentary Hypothesis

A statistical hypothesis which determines the probability distribution completely is known as a *simple hypothesis,* e.g. $H_0 : \theta = \theta_1$ or $H_1 : \theta = 45$

A *composite hypothesis* does not determine the distribution completely, e.g $H_1 : \theta \geq \theta_1$ or $H_1 : \theta < \theta_1$.

Either or both of H_1 and/or H_0 could be simple or composite statistical hypothesis.

Statistical Test

After formulating H_0 and H_1, the next problem is to decide upon their acceptance or rejection in the light of a random sample drawn from the distribution. The decision is either to accept H_0 or reject H_0. The selection of either decision is made on the basis of statistical test and the observed random sample.

Definition

A statistical test for testing H_0 against H_1 is a rule which splits a parameter space S into two exhaustive and disjoint sub-sets C and C* such that, if the observed sample point $X' = (x_1, x_2, ..., x_n)$ fall in C* we reject H_0 and if X falls in C we accept H_0.

If $X \cong f(x, \theta)$ a statistical hypothesis is a statement about the distribution of X. (It is an assertion as claim about the distribution of one or more random variables). If the hypothesis completely specifies f (x, θ), then it is a simple hypothesis.

Definition

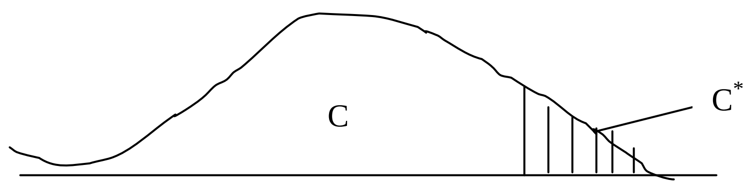
Critical region is the subsets of the parameter space S symbolized by C* is one in which in accordance with a test leads to the rejection of H_0 of the test. This is the subset of the parameter space that leads to the rejection of the hypothesis under consideration.

That is the rejection region is

$C^* = \{(x_1, x_2 \ldots\ldots\ldots xn / \overline{X} \geq k) \text{ for some k}\}$

And the acceptance region is define as

$C = \{(x_1, x_2 \ldots\ldots\ldots xn / \overline{X} < k)\}.$



In-Text Question

……………is always a statement of either 'no effect' or 'status quo'

In-Text Answer

Null hypothesis

11.1.1 Types of error

One is likely to commit two kinds of error while accepting (rejecting) Ho on the basis of sample data.

Type I Error: This means rejecting Ho on the basis of sample data when in fact Ho is true. It is defined as:

$$\alpha = P(\text{Re}\ ject \quad Ho / Ho \quad \text{is} \quad \text{true})$$

This is referred to as the *size of test* or *significance level.*

Type II Error: This means accepting Ho on the basis of sample data when in fact H_1 is true (i.e Ho is false). It is defined as:

$$\beta = P(Accepting \quad Ho / H_1 \quad \text{is} \quad \text{true})$$

The test situation can be summarised in a table as follows:

Error Table

Decision	**Accept Ho**	**Reject Ho**
Accept Ho	Correct decision	***Type II error***
Reject Ho	***Type I error***	Correct decision

However, there is a trade-off between α and β. An attempt to minimise Type I error leads to an increase in the size of Type II error increases or vise – versa.

Example 1

Let X be the number of defective items found in a random sample of size n items from a production lot that follows a poisson distribution. Given the hypothesis

$$H_0 = \lambda = 1 \qquad \text{and the critical region} \qquad C^* = \{x : x > 3\}$$
$$H_1 = \lambda = 0.5 \qquad \qquad C = \{x : x < 3\}$$

Obtain the size of Type I and Type II errors.

Solution

The poisson distribution is given as

$$f(x,\lambda) = \frac{\ell^{-\lambda}\lambda^x}{x!} \qquad \text{x} = 0, 1, 2, ...$$

$$Under \ \text{H}_0 : \ \text{f}(\text{x},\lambda) = \frac{\ell^{-1}1^x}{x!}$$

$$Under \ \text{H}_1 : \ \text{f}(\text{x},\lambda) = \frac{\ell^{-0.5}(0.5)^x}{xx!!}$$

To obtain P (type I error) i.e. (x > 3/H_0 is true) the size of test

$$\Pr(x > 3) = \sum_{x=4}^{\infty} f(x, \lambda)$$

$$= 1 - \sum_{x=0}^{\infty} \frac{\ell^{-1}(1)^x}{x!}$$

$$= 1 - \left[\ell^{-1} + \ell^{-1} + \ell^{-1}/2 + \ell^{-1}/6\right]$$

$$= 1 - 0.9811$$

$$= 0.0189$$

To obtain Pr (Type II error) i.e. H_1 is true but we accept H_0 i.e. $x \leq 3$

$$\Pr\{Type \text{ II error}\} = \Pr(x \leq 3 / H_1 \text{ is true})$$

$$= \sum_{x=0}^{3} \frac{\ell^{-0.5}(0.5)^x}{x!}$$

$$= \ell^{-0.5} + \ell^{-0.5}(0.5) + \frac{\ell^{-0.5}(0.5)^x}{2} + \left(\frac{\ell^{0.5}0.3}{6}\right)$$

$$= 0.9974$$

In-Text Question

What is a Type I Error?

In-Text Answer

This means rejecting Ho on the basis of sample data when in fact Ho is true

The Power of Test

This is the probability of rejecting Ho when H_1 is true, i.e

$$P(\theta) = P(\text{Rejecting} \quad Ho / H_1 \quad \text{is} \quad \text{true})$$

this is equivalent to

$$P(\theta) = 1 - P(\text{type II error}) = 1 - \beta$$

This also implies the probability of taking the right decision. The value of the power function at a parameter point is called power of the test at that point.

This value (p-value) is usually generated by statistical software to enable an experimenter carry out a test without recourse to statistical tables. The best test is usually the one with the maximum power.

Example 2

Obtain the power of test for the problem in example 1 above.

Size of Type II error is denoted by β.

Power of test is 1– β

$$P(\theta) = \Pr\{\text{Rejecting } H_0 / H_1 \text{ is true}\}$$

$$P(\theta) = 1 - Type \text{ II error}$$

from the above example

$$\begin{aligned} &1 - \beta \\ &= 1 - 0.9974 \\ &= 0.0026 \end{aligned}$$

Example 3

Let $X \sim Exp(\theta)$, given the hypothesis

$$\begin{aligned} H_0 &: \theta = 2, \\ H_1 &: \theta = 1, \end{aligned} \quad \text{and the } Critical \text{ region} = \{x : x > 1\}$$

Obtain the power of the test.

Solution

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

$$\alpha = \int_1^\infty 2\ell^{-2x} dx = 0.135 \ \{P(\text{reject } H_0 / H_0 istue\}$$

$$\beta = \int_0^1 \ell^{-x} dx = 0.632 \ \{P(\text{accept } H_0 / H_1 istrue\}$$

$$\begin{aligned} P[\theta] &= \Pr[rejecting \text{ } H_0 / H_1 \text{ is true}] \\ &= 1 - 0.632 \\ &= 0.368 \end{aligned}$$

Comparison of Test

Two tests may be of the same size but of different powers. The more powerful test is the one with the greater chance of detecting when alternative hypothesis is true, (i.e. the test with greatest power).

Theorem

A test H_0: $\theta = \theta_0$ versus H_1: $\theta = \theta_1$ based on critical region C^* is said to be a most powerful test of size $\propto$ if:

1. $P_{c^*}(\theta_0) = \propto$ and $P_c(\theta_0) = \propto$
2. $P_{c^*}(\theta_0) > P_c(\theta_1)$ for any other critical region C.

Example 4: Given that a random variable X has a binomial distribution of the form:

$$f(x,\theta) = \binom{6}{x}\theta^x(1-\theta)^{6-x} \qquad .... x = 0, 1, 2, ...$$

$$H_0 : \theta = 1/2$$

$$H_1 : \theta = 2/3$$

Suppose two critical regions are defined as follows:

$C_1 = \{x : x \geq 5\}$ Test I

$C_2 = \{x : x \leq 1\}$ Test II

2. Determine whether the sizes of the tests are the same.
3. Which is the more powerful test?

Solution

X	0	1	2	3	4	5	6
H_0: $\theta_{(x)}$	$^1/_{64}$	$^6/_{64}$	$^{15}/_{64}$	$^{20}/_{64}$	$^{15}/_{64}$	$^6/_{64}$	$^1/_{64}$
H_1: $\theta_{(x)}$	$^1/_{729}$	$^{12}/_{729}$	$^{60}/_{729}$	$^{160}/_{729}$	$^{240}/_{729}$	$^{192}/_{729}$	$^6/_{729}$

Test I

$$For\ \text{test I : P(type I error)} = \propto = 7/64 = \frac{6+1}{64}$$

$$For\ \text{test II : P(type I error)} = \propto = 7/64 = 6+1/64$$

i.e. The two tests are of the same size.

$$Test\ \text{I} : \beta_1\ \text{P(type II error)} = P\{x < 5/H_1 \text{ is true}\}$$

$$Test\ \text{I} : \beta_2 = P\{x < 1/H_1 \text{ is true}\}$$

$$Test\ \text{I : Power of test} = 1 - \beta_1 = \frac{6}{729} + \frac{192}{729} = 0.35$$

$$Test\ \text{II : Power of test} = 1 - \beta_2 = \frac{1}{729} + \frac{12}{729} = 0.16$$

$\therefore$ Test I is more powerful that test II because it has greater power.

Example 6

Given the p.d.f of a random variable X as:

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

Suppose a single observation is taken from the above distribution and in testing H_0: θ = 2 versus H_1: θ = 3,

1. Determine the size of Type I and Type II errors, if the intervals 1<x<1.5 is the chosen critical region.
2. Calculate the power of the test.

Solution

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

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

The critical region: 1 < x < 1.5

1. $\alpha = \Pr[\text{Type I error}]$

$$= \Pr[1 < x < 1.5 / H_0 : \theta = 2]$$

$$= \frac{1}{2}\int_1^{1.5} dx$$

$$= \frac{1}{2}\Big[x\Big]_1^{1.5} = \tfrac{1}{2}[1.5 - 1]$$

$$= \tfrac{1}{2}(0.5)$$

$$= 0.25$$

$\beta = \Pr[\mathit{Type}\ \text{II error}]$

$$= \Pr[0 < x < 1 \quad \text{or} \ 1.5 < x < \theta / H_1 : \theta = 3]$$

$$= \left[\frac{1}{\theta}\int_0^1 dx + \frac{1}{\theta}\int_{1.5}^{1\theta} dx / H_1 : \theta = 3\right]$$

$$= \frac{1}{3}\,x\Big|_0^1 + \frac{1}{3}\,x\Big|_{1.5}^{\theta}$$

$$= \frac{1}{3}\,x\,1 + \frac{1}{3}[3 - 1.5]$$

$$= \frac{1}{3} + \tfrac{1}{2} = \tfrac{5}{6}$$

$$= 0.833$$

2. $P(\theta) = 1 - \beta$

$$= 1 - 0.833$$

$$= 0.167$$

Likelihood Ratio Test

Given r. s. (x_1 x_2 …. X_n), we can obtain the distribution $L(\theta_0) = f(\underset{-}{x}, \theta / H_0)$ and $L(\theta_1) = f(\underset{-}{x}; \theta / H_1)$.

the ratio $\dfrac{f(\underset{-}{x}, \theta / H_0)}{f(\underset{-}{x}, \theta / H_1)} \geq$ for $\underset{-}{x} \in C$

$$\frac{f(\underset{-}{x}, \theta / H_0)}{f(\underset{-}{x}, \theta / H_1)} < k \text{ for } \underset{-}{x} \neq C$$

Most Powerful Test

A statistical test of level α, whose power under H_1 is maximum is called the most powerful test. The critical region that corresponds to the test is called the best critical region.

The following lemma describes a method for determining the most powerful test when Ho and H_1 are simple hypothesis.

Neyman-Pearson Lemma

Let X_1, X_2 ……X_n be a random sample of size n from f (x, θ), where θ_0 and θ_1 are two possible values of θ. Denote the likelihood function $L(\theta) = L(\theta, x_1, x_2 ..... x_n) = f(x_1\ \theta), f(x_2\ \theta) ...... v(x_n\ \theta)$.

Let k be a positive constant k and C, a subset of the sample space S the test corresponding to C will be the most powerful test of level α for testing $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, if the following conditions are satisfied:

a. $P[(X_1\ X_2 ...., X_n) \in C; \theta_0] = \propto$

b. $\frac{L(\theta_0)}{L(\theta_1)} \leq k \quad \text{for } x_1, x_2 .... x_n \in C$

c. $\frac{L(\theta_0)}{L(\theta_1)} \geq k \quad \text{for } x_1, x_2 .... x_n \in C^1$

Then there is a best critical region of size $\propto$ for testing the simple null hypothesis

H_0: $\theta = \theta_0$ against the simple alternative hypothesis H_1: $\theta = \theta_1$.

Proof

We shall offer proof for the continuous random variable, but something similar can be given for the discrete case also.

The probability of a sample falling in the C when H_0 is true is $\int_C L(\theta_0) = \alpha$

Suppose there is another critical region D. such that $\int_C L(\theta_0) = \int_D L(\theta_0) = \alpha$

If C is the best critical region

then $\int_C L(\theta_0) = \int_D L(\theta_0) = 0$

$$= \int_{CnD^1} L(\theta_0) + \int_{CnD} L(\theta_0) - \int_{CnD} L(\theta_0) - \int_{C^1nD} L(\theta_0) = 0$$

and hence $\int_{CnD^1} L(\theta_0) - \int_{C^1nD} L(\theta_0) = 0$

by condition (b) $K\,\mathrm{L}(\theta_1) \geq L(\theta_0)$ at each point in C and therefore in CnD^1 thus therefore at each point of C^1nD

$$K \int_{CnD^1} \mathrm{L}(\theta_1) \geq \int_{CnD^1} \mathrm{L}(\theta_0)$$

similarly by condition (c) $K\,\mathrm{L}(\theta_1) \leq L(\theta_0)$ at each point in C^1 and therefore in C^1nD thus we obtain

$$K \int_{C1nD} \mathrm{L}(\theta_1) \geq \int_{C^1nD} \mathrm{L}(\theta_0) \qquad \text{i.}$$

Therefore, collecting the like terms on the l.h.s of (i) and r.h.s of (ii), we have

$$0 = \int_{CnD^1} L(\theta_0) - \int_{C^1nD} L(\theta_0) \leq k\left\{ \int_{CnD^1} L(\theta_1) - \int_{C^1nD} L(\theta_1) \right\}$$

$$i.e.\, 0 \leq \mathrm{k}\left\{ \int_{CnD^1} L(\theta_1) + \underline{\int_{CnD} L(\theta_1)} - \int_{CnD} L(\theta_1) - \underline{\int_{C^1nD} L(\theta_1)} \right\}$$

or equivalently

$$0 \leq k\left\{ \int_C L(\theta_1) - \int_D L(\theta_1) \right\}$$

$$\Rightarrow \int_C L(\theta_1) \geq \int_D L(\theta_1)$$

i.e. $P(c, \theta_1) \geq P(D, \theta_1)$, since this is true for every critical region D of size $\propto$, C is the best critical region of size α.

Example 7

Let X_1, X_2 ….. X_n denote a.r.s. from X where $X \approx N(\theta;\ 1)$. Construct a best test for H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$ when $\theta_1 < \theta_0$

Solution

$$f(x,\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}(x-\theta)^2\right\}$$

$$L\left(\underset{-}{x},\theta\right) = \prod_{i=1}^{n} f(x_i,\theta) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}\sum (x_i-\theta)^2\right]$$

$$L(\theta_0) = L\left(\underset{-}{x},\theta_0\right) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}\sum (x_i-\theta_0)^2\right]$$

$$L_1 = L\left(\underset{-}{x},\theta_1\right) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}\sum (x_i-\theta_1)^2\right]$$

The best critical region of size $\propto$ is set such that

$$\frac{L(\theta_0)}{L(\theta_1)} \le k \Rightarrow \frac{\exp\left\{-\tfrac{1}{2}\sum_{i}^{n}(x_i-\theta_1)^2\right\}}{\left\{-\tfrac{1}{2}\sum_{i}^{n}(x_i-\theta_{10})^2\right\}} \le k$$

where k is chosen so that $p\{\bar{x} \in C/H_0 \text{ is true}\} = \propto$ the inequality can be written as

$$\exp\left[\tfrac{1}{2}\sum_{i}^{n}(x_i-\theta_0)^2 - \tfrac{1}{2}\sum_{1}^{n} x_i-\theta)^2\right] \le k$$

taking logarithm we have

$$-\left[\sum_{1}^{n}(x_i-\theta_0)^2 - \sum_{1}^{n}(x_1-\theta_1)^2\right] 2Lnk$$

which can be simplified into

$$\sum x_i - n\theta_0^2 + 2\theta_0 \sum x_i + n\theta_1^2 - 2\theta_i \sum x_i \le 2/k$$

$$\sum x_i + n\theta_0^2 - 2\theta_0 \sum_i^n x_i - \sum_i^n x_i^2 - n\theta_1^2 + 2\theta_1 \sum x_i \ge 2/k$$

$$-\Rightarrow -\left[2(\theta_1 - \theta_2)\sum_1^n x_i + n\left(\theta_1^2 - \theta_0^2\right)\right] \le 2/k$$

$$\frac{\sum_1^n x_i}{n} \ge \frac{21nk - n\left(\theta_1^2 - \theta_0^2\right)}{2n(\theta_1 - \theta_0)}$$

$$\bar{x} \ge \frac{2Ink}{2n(\theta_1 - \theta_0)} + \tfrac{1}{2}(\theta_1 - \theta_0)$$

$$\bar{x} \ge \left[\frac{-Ink}{n(\theta_1 - \theta_0)} + \tfrac{1}{2}(\theta_1 - \theta_0)\right]$$

The best test is rejecting H_0 if $\bar{x} \underset{-}{C} x_0$ where $x_0 = \dfrac{Ink}{n(\theta_1 - \theta_0)} + \tfrac{1}{2}(\theta_1 - \theta_0)$ and k is chosen such that

$$\Pr\left(\bar{x} \underset{-}{C} x_0 \text{ where } H_0 \textit{ is } \text{true}\right) = \propto$$

Example 8

$$\textit{Let } \underset{-}{x} \textit{ denote } \text{a.r.s from f}(x, \lambda) = \frac{\ell^{-\lambda}\lambda^x}{x!}$$

$$\textit{Under } H_0 : \lambda = 1$$

$$H_1 : \lambda = 0.5$$

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{\ell^{-n}(1)^{\sum x_i} \big/ (x_1!x_2!.........x_n!)}{\ell^{-0.5n}(0.5)^{\sum x_i} / (x_1!x_2!.......x_n!)}$$

$$= \frac{\ell^{-n} - \ell^{0.5n}}{(0.5)^{\sum xi}} = \frac{\ell^{-0.5n}}{(0.5)^{\sum xi}} \le k$$

$$= \frac{1}{\ell^{-0.5n}(0.5)^{\sum xi}} \le k$$

taking logarithm

$$0.5m - \sum xiIn0.5 \leq Ink$$

$$-\sum_{1}^{n} xi \geq \frac{Ink - 0.5n}{In0.5}$$

Divide through by n

$$\bar{x} \geq \frac{Ink}{nIn0.5} - 0.5$$

Example 9

Let x_1, x_2… x_n be a random sample from a normal distribution $N(\mu,64)$,

1. show that $c = \{(x_1, x_2, ..., x_n) : \overline{X} \leq c\}$ is the best critical region for $Ho : \mu = 70$ against $H_1 : \mu = 66$.
2. Also find n and c so that $\alpha = 0.05$ and $\beta = 0.05$ approximately.

Solution

1.

$H_0 : \mu = 70; \quad H_1 : \mu = 66$

$$\frac{L(70)}{L(66)} = \frac{(128)^{-n/2} \exp\left\{-\frac{1}{128}\sum (x_i - 70)^2\right\}}{(128)^{-n/2} \exp\left\{-\frac{1}{128}\sum (x_i - 66)^2\right\}}$$

$$= \exp - \frac{1}{128}\left[\sum (x_i - 77)^2 - \sum (x_i - 66)^2\right]$$

$$\Rightarrow \exp\left[-\frac{1}{128}\left(-8\sum x_i + n\,70^2 - n\,66^2\right)\right] \leq k$$

$$\Rightarrow 8\sum x_i - 544n \leq 128 \text{ In k}$$

$$\Rightarrow \sum x_i - 68n \leq 16 \text{ In k}$$

$$\therefore \frac{\sum \text{x}}{\text{n}} - 68 \leq \frac{16}{n}$$

$$\Rightarrow \bar{\text{x}} \leq 68 + \frac{16}{n} \text{In k}$$

$$\bar{\text{x}} \leq \text{C}$$

$$\text{where C} = 68 + \frac{16}{n} \text{In k}$$

According to Nyman Pearson Lemma, the best critical region is

$$C = \{(x_1, \text{x}_2.....x_n)\bar{\text{x}} \leq \text{C}\}$$

2. To find n and C so that $\alpha = 0.05$ and $\beta = 0.05$

$$\alpha = P(\bar{X} \leq C : \mu = 70)$$

$$= \text{P}\left[\frac{\bar{\text{X}} - 70}{\frac{8}{\sqrt{n}}} \leq \frac{\bar{\text{X}} - 80}{\frac{8}{\sqrt{n}}}; \mu = 70\right]$$

$$= \phi\left(\frac{\text{C} - 70}{>}\right)$$

$$\beta = \Pr(\bar{X} \leq C : \mu = 70)$$

$$= \text{P}\left[\frac{\bar{\text{X}} - 66}{\frac{8}{\sqrt{n}}} \leq \frac{\text{C} - 66}{\frac{8}{\sqrt{n}}}; \mu = 66\right]$$

$$= 1 - \phi\left(\frac{\text{C} - 66}{\frac{8}{\sqrt{n}}}\right)$$

$$\phi\left(\frac{\text{C} - 70}{\frac{8}{\sqrt{n}}}\right) = 0.05 \Rightarrow \frac{\text{C} - 70}{\frac{8}{\sqrt{n}}} = 1.645 \quad \text{........................(i)}$$

$$\phi\left(\frac{\text{C} - 66}{\frac{8}{\sqrt{n}}}\right) = 0.95 \Rightarrow \frac{\text{C} - 66}{\frac{8}{\sqrt{n}}} = -1.645 \quad \text{........................(ii)}$$

Dividing (1) by (2). We have

$$\frac{C-70}{C-66}=\frac{1.645}{-1.645}=1$$
$$C-70=-C+66$$
$$2C=66+70$$
$$C=68$$

To find n, substitute in (ii)

$$\frac{68-70}{\frac{8}{\sqrt{n}}}=-1.645$$
$$\frac{-2\sqrt{n}}{8}=-1.645$$
$$\therefore \quad n=43$$

Summary for Study Session 11

1. In this study session, you were introduced to the methods and techniques for carrying out test of hypothesis of the parameters of the stochastic models to you.
2. The types of errors, concept of best test and critical region were discussed. An important theorem for determining the best test was also given. Useful examples were given to better explain the concepts

Self-Assessment Questions (SAQs) for Study Session 11

Having completed this study session, you can measure how well you have achieved its Learning Outcomes by answering these questions. You can check your answers with the Notes on the Self-Assessment Questions at the end of this Module.

SAQ (Testing learning outcomes)

1. Let X be the number of defective items found in a random sample of size n items from a production lot that follows a poisson distribution. Given the hypothesis

$H_0=\lambda=1$, $H_1=\lambda=0.5$ and the critical region $C^*=\{x:x>3\}$, $C=\{x:x<3\}$

Obtain the size of Type I and Type II errors.

Notes on SAQ

1. The poison distribution is given as

$$f(x, \lambda) = \frac{\ell^{-\lambda} \lambda^x}{x!} \qquad \mathrm{x} = 0, 1, 2, ...$$

$$\textit{Under } \mathrm{H_0} : \mathrm{f(x}, \lambda) = \frac{\ell^{-1} 1^x}{x!}$$

$$\textit{Under } \mathrm{H_1} : \mathrm{f(x}, \lambda) = \frac{\ell^{-0.5} (0.5)^x}{xx!!}$$

To obtain P (type I error) i.e. (x > 3/H_0 is true) the size of test

$$\Pr(x > 3) = \sum_{x=4}^{\infty} f(x, \lambda)$$

$$= 1 - \sum_{\mathrm{x}=0}^{\infty} \frac{\ell^{-1}(1)^x}{x!}$$

$$= 1 - \left[\ell^{-1} + \ell^{-1} + \ell^{-1}/2 + \ell^{-1}/6\right]$$

$$= 1 - 0.9811$$

$$= 0.0189$$

To obtain Pr (Type II error) i.e. H_1 is true but we accept H_0 i.e. x ≤3

$$\Pr\{\textit{Type}\ \mathrm{II\ error}\} = \Pr(x \le 3 / H_1 \text{ is true})$$

$$= \sum_{x=0}^{3} \frac{\ell^{-0.5}(0.5)^x}{x!}$$

$$= \ell^{-0.5} + \ell^{-0.5}(0.5) + \frac{\ell^{-0.5}(0.5)^x}{2} + \left(\frac{\ell^{0.5} 0.3}{6}\right)$$

$$= 0.9974$$

References

Bain, L.J.and Max, Engelhardt (1989) *Introduction to Probability and Mathematical Statistics.*

Bhat, B.R and Shah, D.K (1997) "Statistical Inference" Unpublished lecture note in the Department of statistics, University of Botswana, Gaborone, Botswana.

El-Taha, Mohammed (2003) "Introduction to Probability and Statistics-Lecture Note"An unpublished lecture note, Department of Mathematics and Statistics, University of Southern Maine, Portland.

Freund, J.E (1992) *Mathematical Statistics* 5th edition, Prentice Hall, New Jersey.

Graybill, Mood and Boes (1974) *Introduction to the Theory of Statistics*, 3rd Edition, International Student Edition, McGraw-Hill series in probability and statistics.

Kalbfleisch, J.G (1985) *Probability and Statistical Inference* Vol.2, Second Edition, Springer-Verlag,New York.

Murison, Bob (2008) "Statistical Inference Part II" An unpublished lecture note, School of Maths. Stats and Comp., University of New England, Armidale, NSW, 2351; http://turing.une.edu.au/

Olubusoye, O.E, Olaomi, J.O and Shittu, O.I (2002) *Statistics for Engineering, Physical and Biological Sciences*. A Divine Touch Publications, Ibadan.

Robert, V. Hogg and Tannis E.A. (1988) *Probability and Statistical Inference.*

Shagodoyin, D.K, Olubusoye, O.E, Shittu, O.I and Adepoju A.A (2002) *Statistical Theory and Methods.* Joytal Printing Press, Ibadan.