Advanced Statistical Inference

BY DR RAJIV SAKSENA
DEPARTMENT OF STATISTICS
UNIVERSITY OF LUCKNOW

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

Parameter and statistic

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

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

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

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

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

Parametric and non-parametric methods

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

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

Likelihood function of sample values

Let $X_1, X_2 \ldots.X_n$ be a random sample of size n taken from the population whose
p.d.f. or p.m.f. is $f(x,\theta)$, θ is the parameter . θ may be single or vector valued. Then, likelihood function of sample values denoted by L or $L(x_1,x_2,\ldots x_n;\theta)$,is defined as

$$L= L(x_1,x_2\ldots..x_n ; \theta)$$
$$=f(x_1,\theta). f(x_2,\theta)\ldots\ldots f(x_n,\theta)$$
$$= \prod_{i=1}^{n} f(x_i,\theta)$$

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

Likelihood function

1. Let $x_1, x_2, \ldots, x_n$ be a random sample from a N (μ,σ^2) .Then

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

and the likelihood function of sample values is

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

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

$$f(x) = e^{-\lambda} \lambda^x / x! \quad , \text{where } 0 < \lambda < \infty , x= 0,1,2,\ldots ,\infty$$

then likelihood function of the sample values is

$$L = (e^{-n\lambda} . \lambda^{\Sigma x})/ x_1! \; x_2 ! \; \ldots.. x_n !$$

3. For a random sample drawn from a binomial population with parameters n and p , i.e. with p.m.f.

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

the likelihood function is

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

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

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

$$f(x_i, \theta) = 1/ \theta \quad , \; 0 < x < \theta \;, \quad 0 < \theta < \infty$$
$$= 0 \quad \text{other wise}$$

the likelihood function is

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

$$= (1/ \theta)^n \;\; ; \; 0 \le x_1 , x_2 ,\ldots, x_n \le \theta.$$

Sampling distribution

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

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

Thus the probability distribution of sample mean is called the sampling distribution of sample mean, and probability distribution of sample variances is called the sampling distribution of sample variance. In the same way we can have sampling distributions of sample proportion, sample median or of any other statistic we want to use.
Further it is also very important to note that the sampling distribution of a statistic is dependent on the size of the population, the size of the sample and on the method by which the units are selected in the sample.

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

$$\overline{X} \sim N(\mu , \sigma^2/n) \ .$$

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

$$\frac{(\overline{X} - \mu)\sqrt{n}}{s} \sim t_{n-1} ,$$

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

3. The distribution of sample median, say m from a normal population N (μ,σ^2) is also a normal and is represented as

$$m \sim N(\mu , (\pi \sigma^2)/2n)$$

Standard error of the statistic

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

Problem of Statistical Inference

The problem of Statistical Inference can be divided into two parts –

1. Estimation of parameters,
2. Testing of hypothesis.

1. Estimation of parameters :

On some occasions our interest will be in such feature as the central tendency or dispersion of the distribution of $X_1, X_2, \ldots X_n$ .In order to make a conjecture about this feature ,we may use some statistics T ,i.e. some measurable function of $X_1, X_2, \ldots X_n$. To be precise if $x_1, x_2, \ldots x_n$ be the available set of observations then we put forward the corresponding value of T , say

$$t = T(x_1, x_2, \ldots x_n),$$

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

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

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

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

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

2. Testing of Hypothesis

In some situations we start with tentative notion about the feature of the distribution that we are interested in. This idea may be suggested to us by some authority (e.g. a manufacturer placing a new product in the market or a leading scientist propounding some new scientific theory) or by the results of the previous investigations conducted in the same field or in a similar field. We may then like to know how tenable or valid the idea is in the light of the observations ($x_1, x_2, \ldots x_n$) . The inference problem is now one of testing a hypothesis about the unknown feature of the distribution. Note that the model used and the hypothesis being tested are both assumptions regarding the probability distribution of $X_1, X_2, \ldots X_n$. However, the hypothesis is an assumption the validity of which is questioned ,but is taken for granted.
In much simpler words any assumption that we make about a population parameter is called a hypothesis and the statistical procedures that are used to test the hypothesis on the basis of sampled observation are covered under the topic testing of hypothesis.
For example a doctor may set up a hypothesis that smoking increases the risk of throat cancer in human beings. To ascertain this he will collect some primary or secondary data and then after some statistical analysis he might approve or disapprove it. This is the problem of testing of hypothesis. Additionally the assumption that we wish to test is called a null hypothesis and the assumption that we accept in case the null hypothesis is rejected is called alternative hypothesis.

specific sample of size n from the population .

The problem of Statistical Inference can be divided into two parts –
1. Estimation of parameters,
2. Testing of hypothesis.

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

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

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

Point Estimation

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

If we use the value of a statistic to estimate a population parameter, this value is a point estimate of the parameter. For example, in the case of binomial (n, p) population, if we use sample proportion to estimate the parameter p, this estimate is called point estimate because it is single number, or point on the real axis. The statistic, whose value is used as the point estimate of a parameter, is called an estimator.

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

Theory of Point Estimation

Let $X_1, X_2 \ldots\ldots X_n$ be a random sample of size n drawn from the population whose p.d.f. or p.m.f. is f(x, θ), θ is the parameter of the population. We denote the sample observations $x_1, x_2 \ldots\ldots\ldots\ldots x_n$ by $\underline{x}$ i.e.,

$$\underline{x} = (x_1, x_2 \ldots\ldots\ldots\ldots x_n)$$

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

For the purpose of estimation, we make use of some statistic T, a measurable function of sample values. The value of T at $\underline{x}$ is assumed to be

$t=T(\underline{x})$. One may propose to estimate θ by this value t, known as estimate of θ corresponding to the given random sample $\underline{x}$.

To make a distinction, the function T is called the estimator of θ. Thus, estimator is defined as a function of sample values while estimate is the value of estimator for a specific set of sample values $\underline{x}$. One should note here that a statistic is also defined as the function of sample observation. A statistic becomes an estimator as soon as one estimates the unknown parameter θ by it. Thus, every estimator is a statistic but every statistic is not an estimator. A statistic becomes an estimator only when one proposes to estimate an unknown parameter (or a function of unknown parameter) by it.

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

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

(1) Method of moments

(2) Method of maximum likelihood

(3) Method of minimum variance

(4) Method of least squares

(5) Method of minimum chi-square.

These methods give different estimators for the estimation of the same parameter. These methods have been discussed in other blocks/ units in detail. Now the question arises which one of the estimators should be chosen from and why?

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

Properties of a Good Estimator

A very important decision, which an experimenter has to take is to decide which estimator one should choose among a number of possible estimators. The answer is provided by defining various desirable statistical properties of the estimators like unbiased ness, minimum variance, consistency, efficiency and sufficiency to decide which estimator is most appropriate to a given situation.

Here following desirable properties of a good estimator are being discussed-

(1) Unbiased ness

(2) Consistency

(3) Efficiency

(4) Sufficiency

Unbiasedness

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

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

E(T) = θ

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

bias (T) = bias in T = E(T)–θ.

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

Some times, it is noted that

E (T) → θ as n → ∞.

In this case, T is known as asymptotically unbiased estimator of θ.

A very important point about unbiasedness is that unbiased estimators are not unique. That is, there may exist more than one unbiased estimator for a parameter. It is also to be noted that unbiased estimator does not always exists.

Example (a)

If X has a binomial distribution with parameters n and θ, then x/n, the observed proportion of success, is an unbiased estimator of the parameter θ.

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

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

Hence x/n is an unbiased estimate of θ.

Example (b)

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

Proof:

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

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

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

Where

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

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

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

$$= \sigma^2$$

Consistency

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

$$\Pr(|T - \theta| \le \varepsilon) \to 1$$

or $$\Pr(|T - \theta| \le \varepsilon) \to 0$$

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

A Sufficient Condition for Consistency

T is a consistent estimator of θ if (i) $E(T) \to \theta$

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

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

$$\Pr\left[|x - E(x)| \le \varepsilon \right] \ge \left[1 - \frac{Var(x)}{\varepsilon^2} \right]$$

Applying it to 'T' we get

$$\Pr\left[|T - E(T)| \le \varepsilon \right] \ge \left[1 - \frac{Var(T)}{\varepsilon^2} \right]$$

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

$$\Pr\left[|T - \theta| \le \varepsilon \right] \ge 1 \text{ for } n \to \infty$$

But probability can never exceed unity, therefore, we write

$$\Pr\left[|T - \theta| \le \varepsilon \right] \ge 1 \text{ as } n \to \infty$$

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

Proved

Mean Square Error (or m.s.e.)

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

The mean square error of an estimator T is defined as

$$\text{m.s.e.}(T) = \text{MSE}(T) = E(T-\theta)^2$$

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

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

(Cross term vanishes)

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

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

m.s.e. (T) = Var(T) (as E(T) = θ)

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

Efficiency

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

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

m.s.e. (T_1) < m.s.e. (T_2).

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

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

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

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

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

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

$$E = \frac{\sigma_1^2}{\sigma_2^2} \qquad \text{This is always} < 1 .$$

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

$$E = \frac{Var\bar{x}}{Var(samplemedian)} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}}$$

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

The minimization of m.s.e. for all θεΩ it self is found to be a difficult task. One may resolve this problem giving insistence to unbiased ness that is, if one confines to the class of unbiased estimators. Minimization of m.s.e. will then amount to the minimization of the variance.

Criterion of unbiased ness has no great merit. It only provides a process of choosing estimators within a mathematically tractable framework.

The criterion of unbiased ness may be deemed defective in cases where biased estimators have smaller m.s.e. than unbiased ones. The question, then, arises why to leave them out of consideration ? In some investigation it becomes necessary to pool the evidence collected from several sources. The evidence may be in nature of an estimate, perhaps with a standard error attached to it. If the estimates are unbiased then a combined estimate may be formed with reduced standard error and with the accumulation of more evidence the true value may be approached. On the other hand, if biased estimates are combined without any indication regarding the magnitude of the bias, then nothing definite can be said about such combined estimates. The bias may actually exceed the standard error at some stage and combined estimate may ever approach the true value.

Minimum Variance Unbiased Estimator (MVUE)

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

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

(ii) $Var(T) < Var(T_1)$ for every $\theta \in \Omega$

Where T_1 is any other estimator of θ satisfying (i).

We know that

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

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

MVUE and CR Inequality

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

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

CR Inequality

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

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

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

For sake of notational simplicity the multiple integral

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

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

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

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

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

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

(iv) T be an unbiased estimator of ψ(θ)

i.e. $E(T) = \psi(\theta)$

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

(vi) $E\left(\partial \log \frac{L}{\partial \theta}\right)^2$ exists and is positive for each θ∈Ω.

Under these assumptions

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

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

We may denote

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

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

Proof

We have $\int_{\underline{x}} L d\underline{x} = 1$

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

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

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

Or $\quad E\left(\frac{\partial \log L}{\partial\theta}\right) = 0$ or $E(Q) = 0 \qquad (1)$

Where $\quad Q = \left(\frac{\partial \log L}{\partial\theta}\right)$

Again $\quad E(T) = \psi(\theta)$

Or $\quad \int_{\underline{x}} T L d\underline{x} = \psi(\theta)$

Differentiating both side partially w.r.t. 'θ' and applying (v)

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

or $\quad \int_{\underline{x}} T\left(\frac{1}{L}\frac{\partial L}{\partial\theta}\right).L d\underline{x} = \psi'(\theta)$

or $\quad E(TQ) = \psi'(\theta) \qquad (2)$

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

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

$$= E\left(\frac{\partial \log L}{\partial\theta}\right)^2 \qquad (3)$$

Also, $\text{Cov}(TQ) = E(TQ) - E(T).E(Q)$

$$= E(TQ) \text{ as } E(Q) = 0$$

$$= \psi'(\theta) \text{ [using (2)]} \qquad (4)$$

We may write $\text{Con}(TQ) = \rho_{TQ}.\left(\sqrt{\text{Var}(T)\text{Var}(Q)}\right)$

[Where ρ_{TQ} is the correlation coefficient between T & Q].

So $\{\text{Cov}(TQ)\}^2 < \text{Var}(T). \text{Var}(Q)$

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

$$\geq \frac{\{\psi'(\theta)\}^2}{E\left(\frac{\partial \log L}{\partial \theta}\right)^2} \qquad \text{[using (3) and (4)]}$$

Proved.

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

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

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

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

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

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

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

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

Example 1

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

In sampling from a normal population, the sample mean $\bar{x}$ is also normally distributed with mean μ and variance σ^2/n.

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

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

$\bar{x}$, thus, satisfies the conditions for consistency of the estimator and therefore is a consistent estimator for population mean μ.

Example 2

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

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

Solution: Here we have

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

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

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

$$\text{Var}(t) = \text{Var}\,\frac{x_1 + 2x_2 + x_3}{4}$$

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

efficiency of t relative to $\bar{x} = \frac{\text{var}(t)}{\text{var}(\bar{x})}$

$$= \frac{\frac{\sigma^2}{3}}{\frac{6\sigma^2}{16}} = \frac{8}{9}$$

Example 3

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

(a) $wx_1 + (1-w)x_2$ value $0 \le w \le 1$ is an unbiased estimator of μ

(b) the variance of this estimator is a minimum

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

Solution

a. Let T $= w\,x_1 + (1 – w)\,x_2$

E (T) $= E(x_1) + (1 – w)\,E(x_2)$

$= w\,\mu + (1 – w)\,\mu = \mu$.

Hence T is an unbiased estimator of μ.

b. $\text{Var}(T) = w^2\,\text{Var}(x_1) + (1 – w)^2\,\text{Var}(x_2)$

$= w^2\,\frac{\sigma^2{}_1}{n} + (1 – w)^2\,\frac{\sigma_2^2}{n}$

If Var (T) is minimum, then d/dw (Var(T)) = 0

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

d/dw (Var(T)) = 0 gives

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

i.e $w\,(\sigma_1^2 + \sigma_2^2) = \sigma_2^2$

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

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

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

Example 4

X_1, X_2 and X_3 is a random sample of size 3 from a population with mean μ and variance σ^2. T_1, T_2 and T_3 are the estimators used to estimate mean value μ where

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

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

I. Are T_1 and T_2 unbiased estimators?

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

III. With the value of λ is T_3 a consistent estimator? and

IV. Which is the best estimator?

Solution Since X_1, X_2, X_3, is a random sample from a population with mean μ and varianceσ^2

$E(X_1) = \mu$, $Var(X_1) = \sigma^2$ and $Cov(X_1, X_2) = 0$ $i \neq j = 1,2,\ldots,n$.

I. $E(T_1) = E(X_1) + E(X_2) - E(X_3)$

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

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

$E(T_2) = E(2X_1) + E(3X_3) - E(X_4)$

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

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

II. T_3 is an unbiased estimator $\Rightarrow E(T_3) = \mu$

$1/3\ \{\lambda E(X_1) + E(X_2) + E(X_3)\} = \mu$

i.e. $1/3\ (\lambda + 2)\ \mu = \mu$

i.e. $\lambda = 1$

III. With $\lambda = 1$, $T_3 = \frac{X_1 + X_2 + X_3}{3} = \bar{x} =$ Sample Mean

Sample mean i.e. T_3 is a consistent estimator of μ.

(IV) We have

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

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

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

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

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

Example 5

If $x = \frac{1}{2}(x_1 + x_2)$, where x_1 and x_2 are most efficient estimators with variance S^2, then show that $Var(x) = \frac{1+\rho}{2} = S^2$, where ρ is the correlation coefficient between x_1 and x_2.

Solution

Since both x_1 and x_2 are most efficient estimators

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

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

$$= \frac{S^2 + S^2 + 2\rho S^2}{4} = \frac{2S^2 + 2S^2 + \rho}{4}$$

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

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

The C-R inequality provides the lower bound for the variance of an unbiased estimator of $\psi(\theta)$ and states that

$$V(t) \geq \frac{(\psi'(\theta))^2}{E\left(\frac{\partial \log L}{\partial \theta}\right)^2}$$

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

Sufficiency and Factorization Theorem

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

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

Now we will study the concept of sufficiency in detail.

Sufficiency

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

It is this objective that leads to the concept of sufficient statistics. The principle of sufficiency is a principle for reducing or condensing the original random sample to a few statistics which may than be used for the purpose of drawing inference about the parent population characterized by θ. Loosely speaking, sufficiency amounts to replacing the sample observations X_1. X_2, ……X_n by few statistics T_1, T_2,…. T_K and thus discarding information, which is not relevant to θ and retaining every thing that is essential.

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

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

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

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

Example

Let $\left(X_1, X_2, \ldots, X_n\right)$ be a random sample from a Bernoulli population with parameter 'p', 0<p<1

(i.e) $$X_i = \begin{cases} 1 \text{ with probability p} \\ 0 \text{ with probability (1 - p)} \end{cases}$$

Let $T = \sum X_i$

Hence $P(T = K) = \left({}^nC_k\right)p^k(1-p)^{n-k}$

The conditional distribution of $\left(X_1, X_2, \ldots, X_n\right)$ given T, is

$$P\left(X_1 \cap X_2 \cap X_3 \cap \ldots \cap X_n \cap T = K\right)$$

$$= \frac{P\left(X_1 \cap X_2 \cap X_3 \cap \ldots \cap X_n \cap T = K\right)}{P(T = K)}$$

$$= \frac{p^k(1-p)^{n-k}}{\left({}^nC_k\right)p^k(1-p)^{n-k}} = \frac{1}{{}^nC_k}$$

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

Remark: It can be quite tedious to check whether a statistic is sufficient for a given parameter by using the above definition based on the determination of conditional distribution of sample values given the statistic.

To overcome this difficulty Neyman and Fisher developed a method of examining sufficiency of a statistic known as Neyman -Fisher Factorization theorem.

Neyman-Fisher Factorization Theorem

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

Mathematically, T is sufficient statistic of θ iff

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

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

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

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

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

Example:

The statistic $\overline{X}$ is a sufficient estimator of the mean of a normal population whit mean μ and variance σ^2 (μ unknown, σ^2 known).

The likelihood function of sample values based on a random sample of size n is

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

We may write -

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

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

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

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

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

Here, the first factor on the right hand side depends only on sample mean and the population mean μ , and the second factor dose not involve μ. Therefore, according to the factorization theorem, sample mean is a sufficient statistic for normal population mean μ with the known variance σ^2.

Important about sufficiency

1. The original sample $X_1, X_2, \ldots X_n$ is always a sufficient statistic.
2. A sufficient estimator is always a consistent estimator.
3. A sufficient estimator may or may not be an unbiased one.
4. A sufficient estimator is the most efficient one if a sufficient estimator exists.

The Koopman's form of the distribution

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

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

where h(θ) and a(θ) are the functions of the parameter θ only and $g(x)$ and $\{\psi(x)\}$ are functions of the sample observations only.

Binomial , Poisson ,Normal, Exponential are some examples of this type of distributions.

The Invariance Property of A Sufficient Estimator

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

COMPLETE SUFFICIENT STATISTICS AND RAO-BLACKWELL THEOREM

The concept of sufficiency has already been discussed in earlier. The principle of sufficiency plays a very important role in various models of statistical inference. Here in this unit another important concept that of complete family of distributions has been discussed.

Consider the statistic T based on a random sample of size n say $X_1, X_2, \ldots X_n$ with joint distribution depending upon $\theta \in \Theta$. The distribution of T itself will in general depend upon θ . Let $\{f_\theta(t)\}$ be the family of distributions related to T.

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

$$E(\phi(T) = 0 \quad \Rightarrow \quad \phi(T) = 0 \quad \text{almost everywhere (for all } \theta \in \Theta)$$

i.e. $$E[\phi(T)] = \int \phi(t) dF_\theta(t) = 0 \Rightarrow \phi(t) = 0$$

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

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

.

Some illustrations

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

$$= \theta^x (1-\theta)^{1-x} \quad \text{if } x = 0,1$$
$$= 0 \quad \text{, otherwise ,}$$

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

Now T has a binomial p.m.f.

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

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

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

where $a(t) = {}^nC_t . \psi(t)$.

Hence ,

$$E_\theta \psi(\mathrm{T}) = 0 \text{ for all } \theta \text{ such that } 0 < \theta < 1$$
$$\Rightarrow \mathrm{a}(0)(1-\theta)^{\mathrm{n}} + a(1)\theta(1-\theta)^{n-1} + \dots\dots\dots + \mathrm{a(n)}\theta^{\mathrm{n}} = 0$$
$$\text{for all } \theta \text{ such that } 0 < \lambda < \infty \text{ where } \lambda = \theta/(1-\theta).$$

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

$\psi(t) = 0$, for t = 0,1,2......n , i.e. for all the values of T with non zero probabilities (for all 0 < θ <1) .
Hence T is a complete statistic. In other words, the binomial family of distributions is complete .

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

$$f_\theta(x) = \left\{ \frac{\exp(-\theta)\,\theta^{x}}{x!} \right\} \text{ if } x = 0,1,2.....$$
$$= 0 \qquad \text{other wise}$$

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

we have already seen that $T = \sum X_i$ is a sufficient statistic for θ . Again, T is also distributed in the Poisson form with parameter nθ i.e. with p.m.f.

$$f_\theta(t) = \frac{\exp(-n\theta)\,.\,(n\theta)^t}{t!} \text{ if } t = 0,1,2,.....$$
$$= 0 \qquad \text{other wise}$$

Hence

$$E_\theta(\psi(T)) = \sum_{t=0}^{\infty} \psi(t)\,.\,\frac{\exp(-n\theta)\,.(n\theta)^t}{t!}$$
$$= \exp(-n\theta) \sum_{t=0}^{\infty} a(t)\,.\theta^t \quad \text{, say ,}$$

where $a(t) = \dfrac{\psi(t)\,.n^t}{t!}$

consequently , $E_\theta \psi(t) = 0$ for all θ such that $0 < \theta < \infty$

$$\Rightarrow \sum_{t=0}^{\infty} a(t).\theta^t = 0 \text{ for all } \theta \text{ such that } 0 < \theta < \infty$$

however , it is known from algebra that a convergent power series which is identically zero must have all the coefficients equal to zero . As such ,

a(t) = 0 for t=0,1,2,….

i.e. $\psi(t) = 0$ for t = 0,1,2.......

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

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

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

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

If $E_\theta[\psi(T)] = 0 \ \forall\ \theta \in \Theta$,then

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

for $-\infty < \theta < \infty$. However the left hand side is the bilateral Laplace transform of the function $\psi(t)\exp(-nt^2 / 2\sigma^2)$ .From the unicity theorem of this type of transform it follows that

$$\psi(t)\exp(-nt^2 / 2\sigma^2) \quad = 0 \ \ (\text{ all } \ \theta \in \Theta\ )$$

$$\psi(t) = 0 \qquad \text{for all } \theta \in \Theta$$

Hence the family of normal distributions with known variance is complete .

Complete Sufficient Statistic

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

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

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

$\overline{X}$ = sample mean , is a complete sufficient statistic.

Example 2 : For a binomial distribution B(n,p) , i.e. $f(x) = {}^{n}C_x . p^x . (1-p)^{n-x}$, X/n is a complete sufficient statistic for p .

Rao –Blackwell Theorem

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

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

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

$\phi(T) = E[U/T]$,which is independent of θ .

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

Then , $\phi(T)$ is itself an unbiased estimator of r(θ) and

$$\text{Var}(\phi(T)) \leq \text{Var}(U)$$

Proof :

We have ,

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

E(U) may be written as

E(U) = E $\{E(U/T)\}$ (By the theory of conditional expectation)

= E $(\phi(T))$ (As $\phi(T)$ = $\{E(U/T)\}$)

Hence , we have

E(U) = E $(\phi(T))$ = $r(\theta)$

Which shows that $\phi(T)$ is an unbiased estimator of r(θ)

Further ,we may write Var (U) as

$$\text{Var}(U) = \text{Var}\{E(U/T)\} + E\{V(U/T)\}$$

(By the theory of conditional Expectation).

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

Hence

$$E\ \{V(U/T)\} \geq \quad 0$$

So that we have $\quad$ Var (U) $\geq$ Var $\{E(U/T)\}$

$\geq$ Var ($\phi(T)$) as $\phi(T)$ = $\{E(U/T)\}$

Hence Proved.

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

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

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

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from N (μ, σ^2) , μ known and σ^2 unknown .

We wish to find out MVUE of σ^2. We know that $T = \sum_{i=1}^{n} (x-\mu)^2$ is a complete sufficient statistic of σ^2. Moreover $\sum_{i=1}^{n} \left(\frac{x-\mu}{\sigma}\right)^2$ follows a chi square distribution with n degrees of freedom .

Hence $\quad E\ \sum_{i=1}^{n} \left(\frac{x-\mu}{\sigma}\right)^2 = n$

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

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

Or $\quad E(S_0) = \sigma^2$ where $S_0 = T/n$

This shows that S_0 is MVUE of σ^2.

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

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

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

$$r(\theta) = P_r(X = m) \qquad (\text{ when m is known })$$

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

Let us define a r.v. U such that

$$U = 1 \quad \text{if} \quad X_i = m$$

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

Then ,

$$E(U) = \frac{e^{-\theta}.\theta^m}{m!} = 1 \times P_r(X_1 = m) + 0 \times P_r(X_1 \neq m)$$

$$= r(\theta)$$

Implying U is an unbiased estimator of $r(\theta)$.

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

$$p(t) = P_r(T = t) = \frac{e^{-n\theta}.(n\theta)^t}{t!}$$

$$t = 0,1,2,....$$

Let us consider

$$\phi(T) = E(U/T = t)$$

$$= P_r\left[X_1 = m \Big/ T = \sum_{i=1}^{n} X_i = t \right]$$

$$= \frac{\Pr.\left[X_1 = m, \sum_{i=2}^{n} X_i = t - m \right]}{\Pr\left[T = \sum_{i=1}^{n} X_i = t \right]}.$$

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

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

Thus $\phi(T) = {}^T C_m . \dfrac{(n-1)^{T-m}}{n^T}$ is an unbiased estimate of $r(\theta) = \dfrac{e^{-\theta}.\theta^m}{m!}$.

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

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

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

$$f(x,\theta) = \begin{cases} 1/\theta, 0 < x \le \theta;\ \theta > 0 \\ 0 \quad \text{otherwise} \end{cases}$$

Our problem is to find out MVUE of θ .

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

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

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

Hence , $E(T) = \int_0^\infty t\ f(t)dt$

$$= \int_0^{\theta} \text{t} \, \frac{nt^n}{\theta^n} dt$$

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

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

Or $\frac{n+1}{n} E[T] = \theta$

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

Or $E[\phi(T)] = \theta$

Where $\phi(T) = \frac{n+1}{n} T$

As T is complete and sufficient statistic of θ and $E[\phi(T)] = \theta$, therefore $\phi(T)$ is MVUE of θ .

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

If U is an unbiased estimator of r(θ), then one may improve upon U by forming the new estimator $\phi(T)$ for r(θ), based on U and sufficient statistics T. This estimator $\phi(T)$ is unbiased for r(θ) and has smaller variance (or mean squared error) than U . This process of finding a new improved estimator in the sense of smaller variance , starting from an unbiased estimator is called "Blackwellisation" .

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