M.Sc. (STATISTICS) (COURSE CODE: 138)

Paper – II: STATISTICAL INFERENCE

SYLLABUS

UNIT – I

Sufficiency, Complete sufficiency, Exponential class, Factorization theorem and Minimal sufficiency.

Unbiasedness, UMVUE, Cramer-Rao Inequality, Rao–Blackuell theorem and Lehmann–Scheffe Theorem.

UNIT – II

Consistency, Efficiency, CAN and CAUN estimators. Moment, ML estimation, Interval estimation.

UNIT – III

Non-randomised and Randomised tests, Critical functions MP tests, Neyman–Pearson Lemma, Monotone likelihood ratio, UMP tests. Likelihood ratio and its properties. Relationship between testing and Interval estimation.

UNIT – IV

Sign, Wilcoxon singned rank, Kolmogoron–Sonirnov, Run, Wilcoxon–Mann–Whitney and Median tests with regard to small and large samples.

UNIT – V

SPRT and its properties, Oe and ASN functions. Application to Binomial, Poisson and Normal distributions.

Note: Two Questions are to be set from each unit

CONTENTS

UNIT-I

SUFFICIENCY AND COMPLETENESS

NOTES

OBJECTIVES

After going through this unit, you should be able to:

- describe sufficiency
- know more about completeness
- define Cramer-Rao Inequality
- explain MVU and Blackwellisation

STRUCTURE

1.1 INTRODUCTION

Many fundamental principles of statistical inference originated from the path-breaking contributions of Sir Ronald Aylmer Fisher in the 1920s. Those concepts are alive, well, and indispensable. The deepest of all statistical concepts is *sufficiency* that originated by Fisher and it blossomed further, again in the hands of Fisher in 1920.

Section 1.2 introduces *Neyman factorization* of a *likelihood function*. In Section 1.3, a notion of *minimal sufficiency* and the fundamental results of Lehmann and Scheffe are discussed. These help in locating the *best* sufficient statistic, if it exists. Section 1.6 introduces the concept of *completeness*.

1.2 SUFFICIENCY

NOTES

We begin with observable *independent and identically distributed* (iid) random variables $X_1,\ldots,X_n$ with a common *probability mass function* (pmf) or *probability density function* (pdf) $f(x)$, $x \in \mathcal{X}$. The sample size n is assumed *known*. Practically speaking, we observe $X_1,\ldots,X_n$ from a *population* whose distribution $f(x)$ is indexed by a *parameter* (or *parameter vector*) θ (or $\boldsymbol{\theta}$) which captures important features of the population. A practical aspect of indexing $f(x)$ with θ (or $\boldsymbol{\theta}$) is that the population distribution would be completely specified once we know θ (or $\boldsymbol{\theta}$).

We let a pmf or pdf be $f(x; \theta)$ or $f(x; \boldsymbol{\theta})$ where the parameter θ (or $\boldsymbol{\theta}$) is fixed but unknown. In the case of a single parameter, we write $\theta \in \Theta$, the parameter space, $\Theta \subseteq \Re$. For example, one may have X distributed as $N(\mu, \sigma^2)$ with μ unknown, $-\infty < \mu < \infty$, but $\sigma(>0)$ is known. Then, $f(x; \theta)$ will be the same as a $N(\mu, \sigma^2)$ pdf with $\theta = \mu$ and the parameter space $\Theta = \Re$. But, if both μ, σ^2 are unknown, then $f(x; \boldsymbol{\theta})$ will be the $N(\mu, \sigma^2)$ pdf with the parameter vector $\boldsymbol{\theta} = (\mu, \sigma^2)$ and $\Theta = \Re \times \Re^+$.

Our quest for gaining information about the unknown parameter θ (or $\boldsymbol{\theta}$) may be considered a core of statistical *inference*. The data $X_1,\ldots,X_n$, of course, have all *information* about θ even though we have not yet specified how to quantify "information". A dataset may be large or small, and data may be nice or cumbersome, but it is ultimately incumbent upon an experimenter to summarize the data so that all interesting features are captured by the summary. That is, ideally a summary should have the exact same "information" about θ as do the original data. Such a summary would be as good as the whole data and it will be called *sufficient* for θ.

Definition 1.1. *An observable real (or vector) valued function $T \equiv T(X_1,\ldots,X_n)$ is called a statistic.*

Some examples of statistics are $\overline{X}$, $X_1(X_2 - X_{n:n})$, $\sum_{i=1}^{n} X_i$, S^2, and so on. As long as numerical evaluation of T, having observed data $X_1 = x_1 \ldots X_n = x_n$, does not involve any unknown entities, T will be called a statistic. Supposing that $X_1,\ldots,X_n$ are iid $N(\mu, \sigma^2)$ where μ is unknown, but σ is known, $T \equiv \overline{X}$ is a statistic, but its standardized form $\sqrt{n}(\overline{X} - \mu)/\sigma$ is not.

Definition 1.2. *A statistic T is called sufficient for an unknown parameter θ if and only if the conditional distribution of the random sample $\mathbf{X} \equiv (X_1,\ldots,X_n)$ given $T = t$ does not involve θ, for all $t \in \mathcal{T} \subseteq \Re$.*

In other words, *given* the value t of a *sufficient statistic T*, *conditionally* there is no more "information" or "juice" left in the original data regarding the unknown parameter θ. Put another way, one may think of **X** trying to tell us a story about θ; but once a sufficient summary T is available, the original story becomes redundant. Observe that **X** is sufficient for θ in this sense. But, we are aiming at a "shorter" summary statistic which has the same information available in **X**.

Definition 1.3. *A vector valued statistic $\mathbf{T} \equiv (T_1,\ldots,T_k)$ with $T_i \equiv T_i(X_1,\ldots,X_n)$, $i = 1, \ldots, k$, is called jointly sufficient for the unknown parameter $\boldsymbol{\theta}$ (or θ) if and only if the conditional distribution of $\mathbf{X} \equiv (X_1,\ldots,X_n)$ given $\mathbf{T} = \mathbf{t}$ does not involve $\boldsymbol{\theta}$ (or θ), for all $\mathbf{t} \in \mathcal{T} \subseteq \Re^k$.*

1.2.1 Neyman Factorization

A dataset consists of $X_1,\ldots,X_n$ from a population with a common pmf or pdf $f(x: \theta)$ where θ is an unknown parameter. The *Neyman Factorization Theorem* is widely used to find sufficient statistics.

Definition 1.4. *Having observed $X_i = x_i$, $i = 1, \ldots, n$, a likelihood function is defined as*

NOTES

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta),\ \theta \in \Theta \tag{1.1}$$

In a discrete case, $L(\theta)$ stands $P_\theta\{X_1 = x_1 \cap \ldots \cap X_n = x_n\}$. In a continuous case, $L(\theta)$ stands for the joint pdf at the observed data $(x_1,\ldots,x_n)$ when θ obtains. One may note that once $\{x_i; i=1,\ldots,n\}$ is observed, there are no random entities in equation (1.1). A likelihood function L(.) is simply a function of θ only.

It is not really essential that X be real valued or iid. In many examples, it will be so. But, if X happens to be vector valued or if it is not iid, then the corresponding joint pmf or pdf of $X_i = x_i$, $i=1,\ldots,n$, would be the likelihood function, $L(\theta)$. We will give examples shortly.

Sample size n is assumed known and fixed before data collection begins.

One may note that θ may be real or vector valued, however, we first pretend that θ is real valued. Fisher discovered the fundamental idea of factorization. Neyman discovered a refined approach to factorize a likelihood function. Halmos and Savage, and Bahadur developed more involved measure-theoretic treatments.

Theorem 1.1. (Neyman Factorization Theorem) *A real valued statistic $T = T(X_1,\ldots,X_n)$ is sufficient for an unknown parameter θ if and only if the following factorization holds:*

$$L(\theta) = g\{T(x_1,\ldots,x_n);\ \theta\}\ h(x_1,\ldots,x_n) \text{ for all } x_1,\ldots,x_n \in X \tag{1.2}$$

where functions $g(.;\theta)$ and $h(.)$ are both non-negative, $h(.)$ is free from θ, and $g(T(.);\theta)$ involves $x_1,\ldots,x_n$ only through $T(x_1,\ldots,x_n)$.

Proof: For simplicity, we provide a proof only in a *discrete* case. Let us write $\mathbf{X} = (X_1,\ldots,X_n)$ and $x = (x_1,\ldots,x_n)$. Let A and B respectively denote the events $\mathbf{X} = x$ and $T(\mathbf{X}) = T(x)$, and observe that $A \subseteq B$.

Only if part: Suppose that T is sufficient for θ. Now, we write

$$\begin{aligned} L(\theta) &= P_\theta\{X = x\} \\ &= P_\theta\{X = x \cap T(X) = T(x)\}, \text{ since } A \subseteq B \\ &= P_\theta\{T(X) = T(x)\}\, P_\theta\{X = x \mid T(X) = T(x)\} \end{aligned} \tag{1.3}$$

Denote $g(T(x_1,\ldots,x_n);\theta) = P_\theta\{T(X) = T(x)\}$ and $h(x_1,\ldots,x_n) = P_\theta\{X = x \mid T(X) = T(x)\}$. Since T is sufficient for θ, by Definition 1.2, the conditional probability $P_\theta\{X = x \mid T(X) = T(x)\}$ cannot involve θ. Thus, $h(x_1,\ldots,x_n)$ so defined may involve only $x_1,\ldots,x_n$. So, the factorization given in equation (1.2) holds.

If part: Suppose that the factorization holds. Let $p(t;\theta)$ be the pmf of T. Observe that $p(t;\theta) = P_\theta\{T(X) = t\} = \sum_{y:T(y)=t} \left\{\prod_{i=1}^{n} f(y_i;\theta)\right\} = \sum_{y:T(y)=t} L(\theta)$. It is easy to see:

$$P_\theta\{X = x \mid T(X) = t\} = 0 \text{ if } T(x) \neq t \tag{1.4}$$

For all $x \in X$ such that $T(x) = t$ and $p(t;\theta) \neq 0$, we can express $P_\theta\{X = x \mid T(X) = t\}$ as:

$$\begin{aligned} L(\theta)/p(t;\theta) &= g(T(x);\theta)\, h(x)/p(t;\theta) \\ &= g(t;\theta)\, h(x) \Big/ \sum_{y:T(y)=t} L(\theta) \\ &= g(t;\theta)\, h(x) \Big/ \sum_{y:T(y)=t} g\big(T(y);\theta\big)\, h(y) \end{aligned}$$

because of factorization in equation (1.2). Since $g(t;\theta) \neq 0$, one has:

$$P_\theta\{X = x \mid T(X) = t\} = g(t;\theta)\, h(x) / \{g(t;\theta)\} \left\{ \sum_{y:T(y)=t} h(y) \right\}$$

$$= h(x) / \left\{ \sum_{y:T(y)=t} h(y) \right\} \equiv q(x) \quad (1.5)$$

with $q(x)$ free from θ. Equations (1.4) and (1.5) complete the proof.

NOTES

In Theorem 1.1, we *do not* demand that $g(T(x_1,\ldots,x_n);\theta)$ is a pmf or pdf of $T(X_1,\ldots,X_n)$. It is essential, however, that $h(x_1,\ldots,x_n)$ must be free from θ.

It should be noted that the splitting of L(θ) may not be unique. Also, there may be different versions of sufficient statistics.

Remark 1.1 *In Theorem 1.1, it was not essential that $X_1,\ldots,X_n$ or θ be real valued. If $X_1,\ldots,X_n$ are iid p-dimensional observations with a common pmf or pdf $f(x;\theta)$, $\theta \in \Theta \subseteq \Re^q$, the Neyman Factorization Theorem will hold:*

Denote the likelihood function, $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$.

A vector valued statistic $\mathbf{T} = (T_1,\ldots,T_k)$ is jointly sufficient for $\theta = (\theta_1,\ldots,\theta_q)$ if and only if $L(\theta) = g(T;\theta)\, h(x_1,\ldots,x_n)$ for all $x_1,\ldots,x_n \in X \subseteq \Re^p$, with both $g(.;\theta)$, $h(.)$ non-negative, $g(.;\theta)$ depending upon x only through **T**, *and $h(.)$ is free from θ.* (1.6)

Example 1.1. Suppose $X_1,\ldots,X_n$ are iid Bernoulli(p) with p unknown, $0 < p < 1$. Here, $X = \{0, 1\}$, $\theta = p$, and $\Theta = (0,1)$. Then,

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \quad (1.7)$$

It matches with factorization in equation (1.2) where $g\left(\sum_{i=1}^{n} x_i; p\right) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$ and

$h(x_1,\ldots,x_n) = 1$ for all $x_1,\ldots,x_n \in \{0,1\}$. So, the statistic $T = \sum_{i=1}^{n} X_i$ is sufficient for p. We could

instead express $L(\theta) = g(x_1,\ldots,x_n;p)\, h(x_1,\ldots,x_n)$ with $g(x_1,\ldots,x_n;p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{n-x_i}$

and $h(x_1,\ldots,x_n) = 1$. So, one could also claim that $X = (X_1,\ldots,X_n)$ was sufficient for p. But, $\sum_{i=1}^{n} X_i$ provides a significantly reduced summary compared with X, the whole data.

Example 1.2. Suppose $X_1,\ldots,X_n$ are iid Poisson(λ) with λ unknown, $0 < \lambda < \infty$. Here, $X = \{0, 1, 2, \ldots\}$, $\theta = \lambda$, and $\Theta = (0, \infty)$. Then,

$$L(\lambda) = \prod_{i=1}^{n} \{e^{-\lambda}\lambda^{x_i}/x_i!\} = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}\left(\prod_{i=1}^{n} x_i!\right)^{-1} \quad (1.8)$$

It matches with factorization in equation (1.2) where $g\left(\sum_{i=1}^{n} x_i;\lambda\right) = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}$ and

$h(x_1,\ldots,x_n) = \left(\prod_{i=1}^{n} x_i!\right)^{-1}$ for all $x_1,\ldots,x_n \in \{0, 1, 2, \ldots\}$. So, the statistic $T = \sum_{i=1}^{n} X_i$ is sufficient for λ. Again, from equation (1.8) one notes that X is sufficient too, but $\sum_{i=1}^{n} X_i$ is a significantly reduced summary.

Example 1.3. Suppose $X_1,\ldots,X_n$ are iid $N(\mu, \sigma^2)$ with μ, σ^2 both unknown, $-\infty < \mu < \infty$, $0 < \sigma < \infty$. Denote $\theta = (\mu, \sigma^2)$, $X = \Re$, and $\Theta = \Re \times \Re^+$. Now,

NOTES

$$L(\theta) = \left\{\sigma\sqrt{2\pi}\right\}^{-n} exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right)\Big/\sigma^2\right\} \qquad (1.9)$$

It matches with factorization in equation (1.2) where

$$g\left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2; \theta\right) = \sigma^{-n} exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right)\Big/\sigma^2\right\}$$

and $\qquad h(x_1,\ldots,x_n) = \left\{\sqrt{2\pi}\right\}^{-n}$

for all $(x_1,\ldots,x_n) \in \Re^n$. So, $T = \left(\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i^2\right)$ is a *jointly* sufficient statistic for (μ, σ^2).

> If T is a (jointly) sufficient statistic for θ, then any statistic U which is *a one-to-one* function of T is (jointly) sufficient for θ.

Example 1.4. (Example 1.3 Continued) We have $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$, $S^2 = (n-1)^{-1}$ $\left\{\sum_{i=1}^{n} X_i^2 - n^{-1}\left(\sum_{i=1}^{n} X_i\right)^2\right\}$. It is clear that the transformation from $T = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ to $U = (\overline{X}, S^2)$ is *one-to-one*. So, we can claim that $(\overline{X}, S^2)$ is jointly sufficient for (μ, σ^2).

> Let T be a sufficient statistic for θ. Consider a statistic T′, a function of T. Then, T′ is not necessarily sufficient for θ.

An arbitrary function of a sufficient statistic *T* need not be sufficient for θ. Suppose that X is distributed as $N(\theta, 1)$ where $-\infty < \theta < \infty$ is an unknown parameter. Obviously, X is sufficient for θ. One may check that $T' = |X|$, a function of X, is *not* sufficient for θ.

> From joint sufficiency of a statistic $T = (T_1,\ldots,T_p)$ for $\theta = (\theta_1,\ldots,\theta_p)$, one should not claim that T_i is sufficient for θ_i, $i = 1,\ldots,p$. Note that T, θ may not even have the same dimension! See Example 1.5.

Example 1.5. (Example 1.4 Continued) We know that $(\overline{X}, S^2)$ is jointly sufficient for (μ, σ^2). So, $(S^2, \overline{X})$ is also jointly sufficient for (μ, σ^2). Should one claim that S^2 is sufficient for μ or $\overline{X}$ is sufficient for σ^2? Of course, not.

Example 1.6. Suppose that $X_1,\ldots,X_n$ are iid Uniform(0, θ), and θ (> 0) is unknown. Here, $X = (0, \theta)$ and $\Theta = \Re^+$. Now,

$$L(\theta) = \prod_{i=1}^{n}\left\{\theta^{-1} I(0 < x_i < \theta)\right\}.$$

$$= \theta^{-n} I(0 < x_{n:n} < \theta) I(0 < x_{n:1} < x_{n:n}) \qquad (1.10)$$

where $x_{n:1}$, $x_{n:n}$ are, respectively, the observed smallest and largest order statistics. The last step in equation (1.10) matches with factorization in equation (1.2) where $g(x_{n:n};\theta) = \theta^{-n} I(0 < x_{n:n} < \theta)$ and $h(x_1,\ldots,x_n) = I(0 < x_{n:1} < x_{n:n})$ for all $x_1,\ldots,x_n \in (0,\theta), \theta > 0$. So, $X_{n:n}$ is sufficient for θ.

It is not crucial that $X_1,\ldots,X_n$ be iid to apply Neyman factorization.

Example 1.7. Let X_1, X_2 be independent with

NOTES

$$f_1(x_1;\theta) = \theta e^{-\theta x_1},\ f_2(x_2;\theta) = 2\theta e^{-2\theta x_2}$$

as the respective pdfs where $\theta(> 0)$ is an unknown parameter and $0 < x_1, x_2 < \infty$. The likelihood function is the joint pdf:

$$L(\theta) = f_1(x_1;\theta) f_2(x_2;\theta) = 2\theta^2 e^{-\theta(x_1 + 2x_2)} \tag{1.11}$$

for $0 < x_1, x_2 < \infty$. The step in equation (1.11) matches with factorization in equation (1.2) so that $T = X_1 + 2X_2$ is a sufficient statistic for θ.

The next result shows a simple way to find sufficient statistics when a pmf or a pdf belongs to an *exponential family*. Its proof follows easily from equation (1.2).

Theorem 1.2 (Sufficiency in an Exponential Family) *Let the random variables $X_1,\ldots,X_n$ be iid with a common pmf or pdf*

$$f(x;\theta) = a(\theta)\, g(x) \exp\left\{\sum_{j=1}^{k} b_j(\theta) R_j(x)\right\}$$

belonging to a k-parameter exponential family. Denote the statistic $T_j = \sum_{i=1}^{n} R_j(X_i),\ j = 1,\ldots,k.$

Then, $T = (T_1,\ldots,T_k)$ *is jointly sufficient for* θ.

Sufficient statistics derived earlier can also be found using Theorem 1.2. We leave these as exercises.

How to verify that a statistic is *not* sufficient for θ?
Discussions follow.

If T is *not* sufficient, then the conditional pmf or pdf of $X_1,\ldots,X_n$ given $T = t$ must involve θ, for some data $x_1,\ldots,x_n$ and t. We follow this route.

Example 1.8 (Example 1.1 Continued) Denote $U = X_1 X_2 + X_3$. The question is whether U is a sufficient statistic for p. Observe that

$$\begin{aligned} &P(U = 0) \\ &= P\{X_1 X_2 = 0 \cap X_3 = 0\} \\ &= P\{[X_1 = 0 \cap X_2 = 0 \cap X_3 = 0] \\ &\quad \cup [X_1 = 0 \cap X_2 = 1 \cap X_3 = 0] \\ &\quad \cup [X_1 = 1 \cap X_2 = 0 \cap X_3 = 0]\} \\ &= (1-p)^3 + 2p(1-p)^2 \end{aligned} \tag{1.12}$$

which reduces to $(1-p)^2(1+p)$. Since $\{X_1 = 1 \cap X_2 = 0 \cap X_3 = 0\}$ is a subset of $\{U = 0\}$, we have

$$\begin{aligned} &P(X_1 = 1 \cap X_2 = 0 \cap X_3 = 0 \mid U = 0) \\ &= P(X_1 = 1 \cap X_2 = 0 \cap X_3 = 0)/P(U = 0) \\ &= p(1-p)^2/\{(1-p)^2(1+p)\} = p/(1+p) \end{aligned}$$

which involves p. So, U is not sufficient for p.

1.3 MINIMAL SUFFICIENCY

We noted earlier that **X** must always be sufficient for θ. But, we aim at reducing the data by means of summary statistics in lieu of considering **X**. We found that Neyman factorization provided sufficient statistics which were substantially "reduced" compared with **X** in a number of examples. As a principle, one should use the "shortest sufficient" summary. Pertinent questions arise: How to define a "shortest sufficient" summary and how to get hold of such a summary?

NOTES

Lehmann and Scheffe developed a mathematical formulation of *minimal sufficiency* and gave a technique to locate minimal sufficient statistics. Lehmann and Scheffe included important follow-ups.

Definition 1.5. *A statistic T is called minimal sufficient for unknown parameter θ if and only if*

(i) T is sufficient for θ, and

(ii) T is minimal or "shortest" in the sense that T is a function of any other sufficient statistic.

Let us think about this concept for a moment. We want to summarize **X** by reducing it to some appropriate statistic such as $\overline{X}$, a median (*M*), or a histogram, and so on. Suppose that in a particular situation, a summary statistic $\mathbf{T} = (\overline{X}, M)$ is *minimal sufficient* for θ. Can we reduce this summary any further? Of course, we can. We may simply look at, for example, $T_1 = \overline{X}$ or $T_2 = M$ or $T_3 = \frac{1}{2}(\overline{X} + M)$. Can T_1 or T_2 or T_3 individually be sufficient for θ? The answer is no, none of these could be sufficient for θ. Because if, for example, T_1 was sufficient for θ, then $T = (\overline{X}, M)$ would have to be a function of T_1. But, *T* cannot be a function of T_1 because we cannot uniquely specify T from the value of T_1 alone. A minimal sufficient summary *T* cannot be reduced any further to another sufficient summary statistic. In this case, a minimal sufficient statistic T is the best sufficient statistic.

1.3.1 Lehmann-Scheffe Approach

The following theorem was proved by Lehmann and Scheffe. This is an essential tool to locate minimal sufficient statistics. Its proof requires some understanding of a correspondence between a statistic and a *partition* it induces on a sample space.

Consider $\mathbf{X} \equiv (X_1,\ldots,X_n)$ with $x = (x_1,\ldots,x_n) \in X^n$. A statistic $\mathbf{T} \equiv \mathbf{T}(X_1,\ldots,X_n)$ is a mapping from X^n onto some space T. For $t \in$ T, let $X_t = \{x{:}x \in X^n$ such that $T(x) = t\}$. These are disjoint subsets of X^n and also $\lambda^n = \lambda_t \in \mathrm{T}\, X_t$. In other words, $\{X_t{:}t \in \mathrm{T}\}$ forms a *partition* of the space λ^n induced by the statistic T.

Theorem 1.3. (Minimal Sufficient Statistic) *Consider* $h(x, y; \theta) = \prod_{i=1}^{n} f\left(x_i;\theta \prod_{i=1}^{n} f(x_{ti};\theta)\right)$,

the ratio of the likelihood functions from equation (1.1) at x and y, x, y ∈ X^n. Suppose that there is a statistic $T \equiv T(X_1,\ldots,X_n) = (T_1,\ldots,T_k)$ such that the following holds:

With arbitrary data points $x = (x_1,\ldots,x_n)$, $y = (y_1,\ldots,y_n)$,

both from X^n, the expression $h(x, y; \theta)$ does not involve θ (1.13)

if and only if $T_i(x) = T_i(y)$, $i = 1,\ldots,k$.

Then, T is a minimal sufficient statistic for ***θ****.*

Proof: We first show that **T** is a sufficient statistic for θ and then we verify that **T** is minimal. For simplicity, let us assume that $f(x; \theta)$ is positive for all $x \in X^n$ and θ.

NOTES

Sufficiency part: Start with $\{X_t : t \in T\}$ which is a partition of X^n induced by T. In X_t, fix an element x_t. If we look at an arbitrary element $x \in X^n$, then this element x belongs to X_t for some unique t so that both x and x_t belong to the same set X_t. So, one has $T(x) = T(x_t)$. Thus, by invoking the "if part" of the statement in equation (1.13), we can claim that $h(x, x_t; \theta)$ is free from θ. Denote $h(x) \equiv h(x, x_t; \theta)$, $x \in X^n$. Hence, we write:

$$\prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} f(x_{ti};\theta)\, h(x) = g(T(x);\theta)\, h(x)$$

with $x_t = (x_{t1}, \ldots, x_{tn})$. Using Neyman factorization, the statistic T is sufficient for θ.

Minimal part: Suppose $U = U(X)$ is another sufficient statistic for θ. Then, by Neyman factorization, we write:

$$\prod_{i=1}^{n} f(x_i;\theta) = g_0(U(x);\theta)\, h_0(x)$$

with some appropriate $g_0(.;\theta)$ and $h_0(.)$. Here, $h_0(.)$ does not involve θ. Now, for any two sample points $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ from X^n such that $U(x) = U(y)$, we obtain:

$$\begin{aligned} h(x, y; \theta) &= \prod_{i=1}^{n} f(x_i;\theta) \Big/ \prod_{i=1}^{n} f(y_i;\theta) \\ &= \{g_0(U(x);\theta)\, h_0(x)\} / \{g_0(U(y);\theta)\, h_0(y)\} \\ &= h_0(x)/h_0(y), \text{ since } g_0(U(x);\theta) = g_0(U(y);\theta) \end{aligned}$$

Thus, $h(x, y; \theta)$ is free from θ. Now, by invoking the "only if" part from equation (1.13), we claim that $T(x) = T(y)$. That is, T is a function of U. Now, the proof is complete.

Example 1.9 (Example 1.1 Continued) With $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, both data points from X, we have:

$$\left\{\prod_{i=1}^{n} f(x_i;\theta)\right\} \Big/ \left\{\prod_{i=1}^{n} f(y_i;\theta)\right\} = \left(p(1-p)^{-1}\right)^{\left\{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i\right\}} \tag{1.14}$$

From equation (1.14), it is clear that $\left(p(1-p)^{-1}\right)^{\left\{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i\right\}}$ would become free from p if and only if $\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0$, that is, if and only if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. Hence, by the theorem of Lehmann-Scheffe, $T = \sum_{i=1}^{n} X_i$ is minimal sufficient for p.

Theorem 1.4. *A statistic which is a one-to-one function of a minimal sufficient statistic is minimal sufficient.*

The following theorem provides a useful tool for finding minimal sufficient statistics within a rich class of statistical models, namely an *exponential family.*

Theorem 1.5. (Minimal Sufficiency in an Exponential Family) *Let $X_1, \ldots, X_n$ be iid with a common pmf or pdf*

$$f(x;\theta) = a(\theta)\, g(x)\, exp\left\{\sum_{j=1} b_j(\theta) R_j(x)\right\} \tag{1.15}$$

belonging to a k-parameter exponential family. Now, let us denote the statistic $T_j = \sum_{i=1}^{n} R_j(X_i)$, $j = 1, \ldots, k$. *Then,* $T = (T_1, \ldots, T_k)$ *is (jointly) minimal sufficient for* θ.

NOTES

1.4 COMPLETENESS

1. The statistic T is said to be complete if for any function $\phi(t)$ such that $E\{\phi(t)\} = 0$ for almost everywhere
$\Rightarrow$ $\phi(t) = 0$
2. The statistic T is said to be complete iff its family of distribution is complete.
3. The statistic T on the family of distribution $\{F_\theta, \theta \in \textcircled{H}\}$ is called complete if for any measurable function $\phi(t)$ such that $E\{\phi(t)\} = 0$ for all $(\forall)$ $\theta \in \textcircled{H}$ where $\textcircled{H}$ is the parametric space.
$\Rightarrow$ $\phi(t) = 0$ for almost everywhere

1.4.1 Bounded

The statistic T is said to be boundedly complete if for any function $\phi(t)$ such that $|\phi(t)| < M$, for some M

and $E\{\phi(t)\} = 0$ for almost everywhere

$\Rightarrow$ $\phi(t) = 0$

Example 1.10. *If* $X \sim b(1, \theta)$, *obtain* $T = \sum_{i=1}^{n} x_i$ *is sufficient and complete statistic for* θ.

Solution. For sufficient statistic:

We have $X \sim b(1, \theta)$

$$\Rightarrow \quad T = \sum_{i=1}^{n} x_i \sim B(n, \theta)$$

$$P(T = t) = p(t) = \binom{n}{t} \theta^t (1-\theta)^{n-t} \quad \ldots(1)$$

The joint p.m.f. of $x_1, x_2, \ldots, x_n$ is given by

$$P(x_1, x_2, \ldots, x_n; \theta) = \theta^{\sum x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}$$

$$= \theta^t (1-\theta)^{n-t} \quad \ldots(2)$$

Thus, the conditional p.m.f. of $x_1, x_2, \ldots, x_n$ given $T = t$ is given by $P(x_1, x_2, \ldots, x_n; \theta / T = t)$

$$= \frac{P(x_1, x_2, \ldots, x_n; \theta)}{p(t)}$$

$$= \frac{\theta^t (1-\theta)^{n-t}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} \quad \text{[See (1) and (2)]}$$

$$= \frac{1}{\left({}^nc_t\right)} \text{ is independent of } \theta$$

Hence, T is sufficient for θ.

For complete statistic:

Since,
$$T = \sum x_i \sim B(n, \theta)$$

$$P(T = t) = p(t) = {}^nc_t \, \theta^t (1-\theta)^{n-t}$$

NOTES

Let $\phi(t)$ be a measurable function of θ, *i.e.*,

$$E\{\phi(t)\} = \sum_{t=0}^{n} \phi(t)\, p(t)$$

$$= \sum_{t} \phi(t) \binom{n}{t} \theta^t (1-\theta)^{n-t}$$

$$= (1-\theta)^n \sum_{t} \phi(t) \left({}^nc_t\right) \left(\frac{\theta}{1-\theta}\right)^t$$

Since, $E\{\phi(t)\} = 0$

$\Rightarrow$ $\phi(t) = 0$

$$\Rightarrow (1-\theta)^n \sum_{t=0}^{n} \phi(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t = 0$$

$$\Rightarrow \quad \sum_{t=0}^{n} \phi(t) \left({}^nc_t\right) \lambda^t = 0, \text{ where } \left(\frac{\theta}{1-\theta}\right) = \lambda$$

or

$$\sum_{t=0}^{n} a(t) \lambda^t = 0, \text{ where } a(t) = \phi(t) \left(n_{c_t}\right) \qquad \ldots(3)$$

$\Rightarrow \quad a(0) + a(1)\,\lambda + a(2)\,\lambda^2 + \ldots + a(n)\,\lambda^n = 0$

This is a polynomial in λ, which is identically zero. This implies all the coefficients are zero.

i.e., $a(t) = 0$

or $\phi(t)\ (\hat{t}) = 0$ [See(3)]

$\Rightarrow$ $\phi(t) = 0$

Thus, $T = \sum_{i=1}^{n} x_i$ is complete statistic for θ

Hence, T is sufficient and complete statistic for θ.

Example 1.11. *If $X \sim P(\theta)$, show $T = \sum_{i=1}^{n} x_i$ is complete and sufficient statistic for θ.*

Solution. For sufficient statistic:

We have

$$P(X = x) = p(x) = \frac{e^{-\theta}\theta^x}{x!},\ x = 0, 1, 2\ldots$$

The joint p.m.f. of $x_1, x_2, \ldots, x_n$ is given by

$$P(x_1, x_2, \ldots, x_n; \theta) = \frac{e^{-n\theta}\, \theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

$$= \frac{e^{-n\theta}\theta^t}{\prod_{i=1}^{n} x_i!}$$

Since, $X \sim P(\theta) \Rightarrow \sum_{i=1}^{n} x_i = t \sim P(n, \theta)$

We have

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

The conditional p.m.f. of $x_1, x_2, \ldots, x_n$ is given by $\sum x_i = t$

$$P(x_1, x_2, \ldots, x_n; \theta / T = t = \sum x_i) = \frac{P(x_1, x_2, \ldots, x_n; \theta)}{p(t)}$$

$$= \frac{e^{-n\theta}\theta^t}{\prod_{i=1}^{n} x_i} \times \frac{t!}{e^{-n\theta}(n\theta)^t}$$

$$= \frac{t!}{\prod_{i=1}^{n} x_i \times n^t}$$

which is independent of θ.

Hence, $t = \sum x_i$ is sufficient statistic for θ.

1.4.2 Complete Statistic

Let $\phi(t)$ be a measurable function of t

i.e.,

$$E_\theta\{\phi(t)\} = \sum_{t=0}^{\infty} \phi(t) P[T = t]$$

$$= \sum_{t=0}^{\infty} \phi(t) \frac{e^{-n\theta}(n\theta)^t}{t!}$$

$$= e^{-n\theta} \sum_{t=0}^{\infty} \frac{\phi(t)\, n^t \theta^t}{t!}$$

$$= e^{-n\theta} \sum_{t=0}^{\infty} a(t)\theta^t$$

where $a(t) = \dfrac{\phi(t) \times n^t}{t!}$

Since, $E_\theta\{\phi(t)\} = 0$

$$\Rightarrow \quad e^{-n\theta} \sum_{t=0}^{\infty} a(t)\theta^t = 0$$

$$\Rightarrow \quad \sum_{t=0}^{\infty} a(t)\theta^t = 0$$

This is a power series, which is uniformly zero, then all its coefficient will be zero,

i.e., $a(t) = 0$

$$\Rightarrow \quad \phi(t) \times \frac{n^t}{t!} = 0$$

$$\Rightarrow \quad \phi(t) = 0$$

Thus, $T = \sum x_i$ is complete statistic for θ hence, T is sufficient and complete statistic for θ.

Example 1.12. *If X ~ U(0, θ), show that T = $X_{(n)}$, n^{th} order statistic is complete and sufficient statistic for θ.*

Solution. Sufficient statistic:

We have $$F(x) = \frac{1}{\theta}, 0 \le x \le \theta$$

NOTES

$$\Rightarrow \quad F(x) = \frac{x}{\theta}$$

Then the joint p.d.f. of $x_1, x_2, ..., x_n$ is given by $f(x_1, ..., x_n; \theta) = \frac{1}{\theta^n}$

Note that the p.d.f. of n^{th} order statistic is $f_n(x) = n[F(x)]^{n-1} f(x)$

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

$$= \frac{n}{\theta^n} x^{n-1} \quad 0 \le x \le \theta$$

The conditional distribution of $x_1, x_2, ..., x_n$ given T = $X_{(n)}$ is given by

$$f(x_1, x_2, ..., x_n; \theta / T = X_{(n)}) = \frac{f(x_1, x_2, ..., x_n)}{f_n(x)}$$

$$= \frac{1/\theta^n}{n/\theta^n \, x^{n-1}}$$

is independent of θ. Hence, $X_{(n)}$ is sufficient statistic for θ

Complete statistic:

We have $$f_n(x) = \frac{n}{\theta^n} x^{n-1}, 0 \le x \le \theta$$

Let $\phi(t)$ be a measurable function of T *i.e.,* θ

$$E_\theta\{\phi(t)\} = 0$$

$$\Rightarrow \quad \int_0^\theta \phi(t) f(t) \, dt = 0$$

$$\Rightarrow \quad \int_0^\theta \phi(t) \frac{n}{\theta^n} t^{n-1} \, dt = 0$$

$$\Rightarrow \quad \frac{n}{\theta^n} \int_0^\theta a(t) \, dt = 0$$

where $$a(t) = \phi(t) \, t^{n-1}$$

$$\int_0^\theta a(t) \, dt = 0$$

Let $A = \int_0^\theta a(t) \, dt = 0$ differentiating it w.r.t. θ both sides, we get

$$\frac{dA}{d\theta} = a(\theta)$$

$\Rightarrow \quad a(\theta) = 0 \;\forall\; \theta$

$a(t) = 0 \Rightarrow \phi(t)\, t^{n-1} = 0$

$\Rightarrow \quad \phi(t) = 0$

Thus, $T = X_{(n)}$ is complete statistic for θ.

Hence, T is complete and sufficient for θ

Note:
$$A = \int_0^{u(x)} f(x)\, dx$$

$$\frac{dA}{dx} = f(u)\frac{du}{dx}$$

NOTES

1.5 SOME IMPORTANT TERMS

1.5.1 Laplace Transformation

(*a*) Unilateral Laplace transformation

(*i*) If $\int_0^{\infty} h(x)\, e^{-\theta x}\, dx = 0, \forall\, \theta > 0$

Then $h(x) = 0$

(*ii*) If $\int_{-\infty}^{\infty} h(x)\, e^{-\theta x}\, dx = 0, \forall\, \theta > 0$

Then $h(x) = 0$

(*b*) Bilaterial Laplace transformation

(*i*) If $\int_{-\infty}^{\infty}\int_0^{\infty} h(x, y)\, e^{-\theta_1 x}\, e^{-\theta_2 y}\, dx\, dy = 0, \forall(\theta_1, \theta_2) > 0$

Then $h(x, y) = 0$

Example 1.13. *If $X \sim N(\theta, 1)$, show that $T = \overline{X}$ is complete and sufficient statistic for θ.*

Solution. For sufficient statistics:

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-\theta)^2}, -\infty < x < \infty$$

The joint p.d.f. of $x_1, x_2, \ldots, x_n$ is given by

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

$$= \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\theta)^2}$$

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

$$= g(t, \theta)\, h(x_1, x_2, \ldots, x_n)$$

where
$$g(t, \theta) = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{n}{2}(\bar{x}-\theta)^2}$$

and
$$h(x_1, x_2, \ldots, x_n) = e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Complete statistic:

$$x \sim N(\theta, 1)$$

$$\therefore \quad T = \bar{x} \sim N\left(\theta, \frac{1}{n}\right)$$

The p.d.f. of T

$$f(T = t) = \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-n/2(t-\theta)^2}$$

NOTES

Let $\phi(t)$ be a measurable function of T, *i.e.,*

$$E_\theta\{\phi(t)\} = \int_{-\infty}^{\infty} \phi(t) f(t) dt$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(t) e^{-n/2(t-\theta)^2} dt$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}\theta^2} \int_{-\infty}^{\infty} \phi(t) e^{-\frac{n}{2}t^2} e^{nt\theta} dt$$

Since,

$$E_\theta\{\phi(t)\} = 0$$

$$\Rightarrow \quad \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}\theta^2} \int_{-\infty}^{\infty} \phi(t) e^{-\frac{n}{2}t^2} e^{nt\theta} dt = 0$$

$$\Rightarrow \quad \int_{-\infty}^{\infty} a(t) e^{-nt\theta} dt = 0$$

where $\quad a(t) = \phi(t) e^{-\frac{n}{2}t^2}$

or $\quad \displaystyle\int_{-\infty}^{\infty} a(t) e^{-t\theta^*} dt = 0$

where $\quad \theta^* = -n\theta$

$\Rightarrow \quad a(t) = 0$, by unilateral Laplace transformation

$$\Rightarrow \quad \phi(t) e^{-\frac{n}{2}t^2} = 0$$

$\Rightarrow \quad \phi(t) = 0$ as $e^{-\frac{n}{2}t^2} \neq 0$

Thus, $\quad T = \bar{X}$ is complete statistic for θ.

1.5.2 Estimator and Estimate

An Estimator is a function of **random variables** but the estimate is a function of **observed values.**

1.5.3 Parameter

The population unknown (such as μ and σ^2) is called a parameter *e.g.,*

(*a*) **Normal Distribution :** The p.d.f. is

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

$$x \sim N(\mu, \sigma^2)$$

(p.d.f. means continuous distribution)

In the above distribution μ & σ are parameter.

(*b*) **Uniform or Rectangular Distribution :** The p.d.f. $x \sim U(a, b)$ or $x \sim R(a, b)$ is

(*i*) $f(x) = \frac{1}{b-a}, a \le x \le b$

(*ii*) $f(x) = 1, 0 \le x \le 1$ *i.e.*, $x \sim U(0, 1)$

(*iii*) $f(x) = \frac{1}{b}, 0 \le x \le b, x \sim U(0, b)$

In general $f(x) = \frac{1}{\theta}, 0 \le x \le \theta, x \sim R(0, \theta)$

(*c*) **Two Parameter Negative exponential Distribution:** The p.d.f. is

(*i*) $f(x) = \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x-\mu)}, \mu \le x \le \infty$

It is also called exponential distribution (with parameter μ and σ) used in life testing.

(*ii*) If $\mu = 0$ then p.d.f. is

$$f(x) = \frac{1}{\sigma} e^{-\frac{1}{\sigma x}}, 0 \le x \le \infty$$

or

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

Name of this distribution is single parameter negative exponential or exponential distribution with parameter σ (or θ)

(*d*) **Gamma Distribution:** If $x \sim G(\alpha, p)$ the p.d.f. is

(*i*) $f(x) = \frac{\alpha^P}{\Gamma P} x^{p-1} e^{-\alpha x}, 0 < x < \infty, \alpha, p > 0$

Parameters of this distribution are α and *p*.

(*ii*) If $\alpha = 1$ then $f(x) = \frac{1}{\Gamma P} x^{p-1} e^{-x}, \; x > 0, p > 0$

It is also known as gamma variate distribution with parameter *p*.

If $p = 1$ then $f(x) = \alpha e^{-\alpha x}, x > 0$

is known as exponential distribution with parameter α.

(*e*) **Weibull Distribution:** (Extended form of exponential distribution)

(*i*) The p.d.f. is $f(x) = abx^{b-1} e^{-ax^b}$

is called Weibull distribution with, $x \ge 0$ parameter *a* and *b*

(*ii*) For $b = 1$, $f(x) = a\,e^{-ax}$ *i.e.*, Weibull distribution reduces to exponential distribution with parameter *a*.

(*f*) **Pareto Distribution:** The p.d.f. is

$$f(x) = \theta K^{\theta} x^{-(\theta+1)} = \frac{\theta}{K}\left(\frac{K}{x}\right)^{\theta+1}, K \le x \le \infty$$

It is also called income distribution with parameters θ and K.

(*g*) **Power Distribution:** The p.d.f. is

$$f(x) = \theta a^{-\theta} x^{\theta+1}, 0 \le x \le a$$

It is called power distribution with parameter 'θ' and '*a*'.

For $a = 1$, $f(x) = \theta x^{\theta+1}, 0 \le x \le 1$

is also known as power distribution with parameter θ.

Discrete Cases

(*h*) **Binomial Distribution:** The p.m.f.

$$p(x) = \binom{n}{x} p^x q^{n-x}, \ x = 1, 2, \ldots, n; \ p + q = 1$$

is known as binomial distribution with parameter n, p and is denoted by $x \sim \beta(n, p)$

For $n = 1$ binomial distribution reduces to Bernoulli distribution with parameter p

i.e., $p(x) = p^x q^{1-x}, x = 0, 1$

NOTES

(*i*) **Poisson Distribution:** The p.m.f.

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \ x = 0, 1, 2, \ldots$$

is called poisson distribution with parameter λ

(*j*) **Geometric Distribution:** The p.m.f.

$$p(x) = pq^x = p(1-p)^x, \ x = 0, 1, 2, \ldots; \ p + q = 1, \ 0 < p \le 1$$

is known as geometric distribution with parameter p.

(*k*) **Negative Binomial Distribution:** The p.m.f.

$$p(x) = \begin{pmatrix} r + x - 1 \\ (r-1) \end{pmatrix} p^r q^x, \ x = 0, 1, 2, \ldots,$$

is known as negative binomial distribution with parameter r and p.

For $r = 1$, we have

$$p(x) = pq^x \ x = 0, 1, 2, \ldots$$

which is the p.m.f. of geometric distribution with parameter p.

1.5.4 Estimation of Parameter

Suppose we are given random sample from a population such that its probability distribution (p.d.) has some known form (*i.e.*, normal, exponential uniform, weibull, gamma etc.,) with the parameters of this are not known.

For example: The height of the people of certain area is likely to follow normal distribution, lifetime of certain items follow exponential distribution.

The problem then arises of estimating the parameters of the assumed distribution from the given sample. The estimation problem can be divided into two types:

1. Interval estimation
2. Point estimation

Internal estimation: we obtain an interval in which the true value of the parameter may lie with certain probability.

In point estimation we try to find the exact value of the parameter.

Point Estimation: In order to obtain a point estimate of a parameter we shall use a statistic (estimator).

1.5.5 Statistic

Any function $h(x_1, x_2, \ldots, x_n)$ that does not depend upon any unknown parameter is called a statistic (or sample known are called statistic). Thus $T = \sum_{i=1}^{n} x_i$ is a statistic but $\frac{x - \mu}{\sigma}$ is not a statistic unless μ and σ are known.

Suppose we want to study the average life of bulbs in a consignment of twenty thousand manufactured by a certain firm. For this we take a sample $x_1, x_2, \ldots, x_n$ of n bulbs and measure their average. Suppose its distribution is given by $f(x, \theta)$ (exponential). On the basis of the simple we have to estimate θ. Now as we know the possible estimates of θ are arithmetic mean, mode, median, geometric mean ($\overline{X}$, M_0, $\hat{M}$, G^n), variance etc., but which one should be preferred? To solve this problem certain requirement are to be laid down for a good estimate and those requirement are

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

NOTES

1.6 UNBIASEDNESS

A statistic 'T' (sample known) is said to be an unbiased estimate of parameter θ (population unknown)

If $E(T) = \theta$

If $E(T) = \theta \pm c$ then c is called the bias in T

If $E(T) = a\theta + b$ then T is linearly biased such a bias can be corrected as follows

Suppose $E(T) = a\theta + b$ then

$$E\left(\frac{T-b}{a}\right) = \theta$$

Hence, $\frac{T-b}{a}$ is an unbiased estimate of θ.

Example 1.14. *Show that the sample mean $(\overline{X})$ is an unbiased estimate of population mean μ in sampling from a normal population.*

Solution. For a random sample $x_1, x_2, \ldots, x_n$

The sample mean $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

$$= \frac{1}{n}E\left[\sum_{i=1}^{n} x_i\right]$$

$$= \frac{1}{n}E\left[x_1 + x_2 + \ldots + x_n\right]$$

Since, x_i's are identically independent random variables (*i. i. r. v*)

$$\therefore \quad E(\overline{X}) = \frac{1}{n}\left[E(x_1) + E(x_2) + \ldots + E(x_n)\right]$$

$$= \frac{1}{n}[\mu + \mu + \ldots + \mu]$$

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

$$= \mu$$

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

Hence $\overline{X}$ is unbiased estimate of μ.

Note: If $x \sim N(\mu, \sigma^2)$

i.e., $E(\overline{X}) = \mu$

$V(X) = \sigma^2$

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

NOTES

i.e., $E(\overline{X}) = \mu$

$$V(\overline{X}) = \frac{\sigma^2}{n}$$

We have proved $E(\overline{X}) = \mu$

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n^2} V\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n^2} V[x_1 + x_2 + \ldots + x_n]$$

$$= \frac{1}{n^2}[V(x_1) + V(x_2) + \ldots + V(x_n)]$$

Since x_i's are identical independent random variables (i.i.r.v)

$$V(\overline{X}) = \frac{1}{n^2}[\sigma^2 + \sigma^2 + \ldots + \sigma^2]$$

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

$$= \frac{1}{n}\sigma^2.$$

Example 1.15. *Show that the sample mean is an unbiased estimate of the population mean in sampling from exponential distribution.*

Solution. The p.d.f. of a random variable x is

$$f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, x \geq \theta$$

We have
$$E(X) = \int_0^\infty x f(x) dx$$

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

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

Since, $\int_0^\infty \frac{1}{\theta^P \Gamma(P)} x^{p-1} e^{-x/\theta} dx = 1$ putting $\alpha = \frac{1}{\theta}$ in gamma distribution

$$\int_0^{\infty} x^{p-1} e^{-x/\theta}\, dx = \theta^P \overline{|P}$$

For $P = 2$

$$\int_0^{\infty} x \cdot e^{-x/\theta}\, dx = \theta^2$$

NOTES

$\therefore$ $E(X) = \frac{1}{\theta} \cdot \theta^2$

$\therefore$ $E(X) = \theta$

$$\Rightarrow \quad E(\overline{X}) = E\left(\frac{1}{n}\sum x\right) = \frac{1}{n}\sum_{i=1}^{n} E(X)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \theta$$

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

$$= \theta$$

Hence, sample mean is an unbiased estimate of population mean θ.

Example 1.16. *If $x_1, x_2, \ldots, x_n$ is a random sample from a normal population N (μ, 1), show that $t = \frac{1}{n}\sum_{i}^{n} x_i^2$ is an unbiased estimator of $\mu^2 + 1$ (i.e., $E(t) = \mu^2 + 1$).*

Solution. We are given

$$E(X_i) = \mu \ \forall\ i = 1, 2, \ldots, n$$
$$V(X_i) = 1$$

Since, $V(X_i) = E(X_i^2) - \{E(X_i)\}^2$

$$\Rightarrow \quad E(X_i^2) = V(X_i) + \{E(X_i)\}^2$$

$$= 1 + \mu^2$$

or $E(X_i^2) = \mu^2 + 1$

Therefore, $E(t) = E\left(\frac{1}{n}\sum_{i}^{n} X_i^2\right)$

$$= \frac{1}{n}\sum_{i}^{n} E(X_i^2)$$

$$= \frac{1}{n}\sum_{i}^{n} (\mu^2 + 1)$$

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

$$= \mu^2 + 1$$

Hence, t is an unbiased estimator of $\mu^2 + 1$.

Example 1.17. *If $x_1, x_2, \ldots, x_n$ are random observations on a Bernoulli variable x taking the value '1' is in probability θ and the value 0 in the probability $1-\theta$, show that $\frac{t(t-1)}{n(n-1)}$ is an unbiased estimate of θ^2 where $t = \sum_i^n x_i$.*

NOTES

Solution. Since,

x_i	0	1
p_i	$1-\theta$	θ

We have

$$E(x_i) = \theta \;\forall\; i = 1, 2, \ldots,$$

$$E(x) = \sum x p(x)$$

$$E(x^2) = \sum x^2 p(x)$$

$$E(x_i^2) = \theta$$

$$V(x_i) = E(x_i^2) - \{E(x_i)\}^2$$
$$= \theta - (\theta)^2$$
$$= \theta(1-\theta)$$

Now,

$$E(t) = E\left\{\sum_i^n x_i\right\} = \sum_i^n E(x_i)$$

$$= \sum_i^n \theta = n\theta$$

$$V(t) = E(t^2) - \{E(t)\}^2$$

$$= V\left(\sum_i^n x_i\right)$$

$$= V(x_1 + x_2 + \ldots + x_n)$$

Since x_i's are *i.i.r.v*,

$$V(t) = V(x_1) + V(x_2) + \ldots + V(x_3)$$
$$V(t) = \theta(1-\theta) + \theta(1-\theta) + \ldots + \theta(1-\theta)$$
$$= n\theta(1-\theta)$$
$$E(t^2) = V(t) + \{E(t)\}^2$$
$$= n\theta(1-\theta) + (n\theta)^2$$
$$= n\theta - n\theta^2 + n^2\theta^2$$

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

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

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

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

Hence, $\frac{t(t-1)}{n(n-1)}$ is an biased estimate of θ^2.

Example 1.18. *Show that the sample mean* $(\overline{X})$ *is an unbiased estimate of the parameter of poisson distribution (i.e.,* $E(\overline{X}) = \lambda$*) if* x_i*'s are random variable.*

NOTES

Solution.
$$x \sim P(x)$$
$$E(x) = \theta$$

or
$$x \sim P(\theta)$$

$$E(x) = \frac{1}{n}\sum_{i}^{n} x_i$$

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i$$

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

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

$$= \frac{1}{n}\sum_{i}^{n} \theta$$

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

For a poisson distribution

$$P(x) = \frac{e^{-\theta}\theta^x}{x!},\ x = 0, 1, 2, \ldots$$

$$E(x) = \sum xp(x)$$

$$= \sum_{0} x\frac{e^{-\theta}\theta^x}{x!}$$

$$= \theta\sum_{i}^{\infty}\frac{e^{-\theta}\theta^{x-1}}{(x-1)!} = \theta.1 = \theta.$$

Another method

$$E(x) = e^{-\theta}\sum_{0}^{\infty}\frac{x\cdot\theta^x}{x!}$$

$$= e^{-\theta}\left\{\theta + \theta^2 + \frac{2\theta^2}{2!} + \frac{3\theta^3}{3!} + \ldots\right\}$$

$$= e^{-\theta}\cdot\theta\left[1 + \theta + \frac{\theta^2}{2} + \frac{\theta^3}{3!} + \ldots\right]$$

$$= e^{-\theta}\cdot\theta\, e^{\theta}$$

$$E(x) = \theta.$$

Now define
$$T = \frac{1}{n}\sum_{i}^{n} x_i = \text{sample mean}$$

$$E(T) = \frac{1}{n}\sum_{i}^{n} E(x_i)$$

$$= \frac{1}{n}\sum_{i}^{n} \theta = \frac{1}{n}\cdot n\theta = \theta$$

Hence sample mean is an unbiased estimation of θ.

NOTES

Example 1.19. *Show that* $\frac{x(x-1)}{n(n-1)}$ *is an unbiased estimation of* p^2 *if*

$$P(X = x) = \binom{n}{x} p^x q^{n-x},\ x = 0, 1, 2, \ldots, n.$$

Solution. We know that $E(x) = \sum x\,p(x)$

$$E\left\{\frac{x(x-1)}{n(n-1)}\right\} = \sum_{x=0}^{n} \frac{x(x-1)}{n(n-1)} \cdot P(X = x)$$

$$= \frac{1}{n(n-1)} \sum_{x=0}^{n} x(x-1)\cdot\binom{n}{x} p^x q^{n-x}$$

$$= \frac{1}{n(n-1)} \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= \frac{1}{n(n-1)} \sum_{x=0}^{n} x(x-1)\frac{n(n-1)(n-2)!}{x(x-1)(x-2)!(n-x)!} p^x q^{n-x}$$

$$= p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x} = 1$$

$$= p^2 \cdot 1$$

$$= p^2.$$

Hence, $\frac{x(x-1)}{n(n-1)}$ is unbiased estimation of p^2.

1.7 UNIFORMLY MINIMUM VARIANCE UNBIASED ESTIMATOR

This section gives methodologies for finding a *uniformly minimum variance unbiased estimator* (UMVUE) of a parametric function. The first approach relies upon the celebrated *Cramer-Rao Inequality* (Theorem 1.6). C.R. Rao and H. Cramer independently discovered, under mild regularity conditions, a lower bound for the variance of an unbiased estimator τ(θ). Then, we introduce another fundamental result (Theorem 1.7) which jumped out of the Rao-Blackwell Theorem.

1.7.1 Cramer-Rao Inequality

Lehmann referred to this as "information inequality," a name which was suggested by Savage. Lehmann wrote. "The first version of the information inequality appears to have been given by Frechet". We will continue to refer to this inequality by its commonly used name, the *Cramer-Rao Inequality*, for ease of cross-referencing and searching other sources.

The variance bound, called the *Cramer-Rao lower bound* (CRLB), for unbiased estimators of τ(θ) is appreciated where one can (*i*) derive an explicit expression of CRLB, and (*ii*) easily locate an unbiased estimator of $\tau(\theta)$ whose variance coincides with CRLB. In these situations, one has then found the UMVUE for $\tau(\theta)$!

Consider iid real valued observations $X_1,\ldots,X_n$ with a common p.m.f. or p.d.f. $f(x;\theta)$ where the unknown parameter $\theta \in \Theta \subseteq \Re$ and $x \in X \subseteq \Re$. Denote $\mathbf{X} \equiv (X_1,\ldots,X_n)$. We *pretend* working with a p.d.f. and hence expectations would be written as appropriate (multiple) integrals. In a discrete case, one would simply replace the integrals by summations.

NOTES

Standing Assumptions: The support X does not involve θ and the first partial derivative of $f(x;\theta)$ with respect to θ and integrals with respect to $x = (x_1,\ldots,x_n)$ are interchangeable.

Theorem 1.6. (Cramer-Rao Inequality) *Suppose that $T = T(X)$ is an unbiased estimator of a real valued parametric function $\tau(\theta)$. Assume that $\frac{d}{d\theta}\tau(\theta)$, denoted by $\tau'(\theta)$, is finite for all $\theta \in \Theta$, is finite for all $\theta \in \Theta$. Then, for all $\theta \in \Theta$, under the standing assumptions, we have:*

$$V_\theta(T) \geq \frac{\{\tau'(\theta)\}^2}{nE_\theta\left[\left\{\frac{\partial}{\partial\theta}\left[\log f(X_i;\theta)\right]\right\}^2\right]} \tag{1.16}$$

The expression on the right-hand side (rhs) of this inequality is the Cramer-Rao lower bound.

Proof: Without any loss of generality, assume that $0 < V_\theta(T) < \infty$. We have

$$\tau(\theta) = \int_X \ldots \int_X T(x_1,\ldots,x_n)\prod_{i=1}^{n} f(x_i;\theta)\prod_{i=1}^{n} dx_i \tag{1.17}$$

which implies that $\tau'(\theta)$ is:

$$\frac{d}{d\theta}\left[\int_X \ldots \int_X T(x_1,\ldots,x_n)\prod_{i=1}^{n} f(x_i;\theta)\prod_{i=1}^{n} dx_i\right]$$

$$= \int_X \ldots \int_X T(x_1,\ldots,x_n)\left[\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i;\theta)\right]\prod_{i=1}^{n} dx_i \tag{1.18}$$

Observe that

$$\prod_{i=1}^{n} f(x_i;\theta) = exp\left\{\sum_{i=1}^{n}\log f(x_i;\theta)\right\} \text{ for } x \in X \tag{1.19}$$

so that, we get:

$$\frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n} f(x_i;\theta)\right] = exp\left\{\sum_{i=1}^{n}\log f(x_i;\theta)\right\}\frac{\partial}{\partial\theta}\left\{\sum_{i=1}^{n}\log f(x_i;\theta)\right\} \tag{1.20}$$

$$= \left\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left[\log f(x_i;\theta)\right]\right\}\prod_{i=1}^{n} f(x_i;\theta)$$

Denote $Y = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left[\log f(X_i;\theta)\right]$ and combine equation (1.18) with equation (1.20) to rewrite

$$\tau'(\theta) = \int_X \ldots \int_X T(x_1,\ldots,x_n)\left\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left[\log f(x_i;\theta)\right]\right\}\prod_{i=1}^{n} f(x_i;\theta)\prod_{i=1}^{n} dx_i$$

$$= E_\theta\{TY\} \tag{1.21}$$

One obviously has $\int_X f(x;\theta)\,dx = 1$ so that

$$0 \equiv \frac{d}{d\theta}\int_X f(x;\theta)\,dx = \int_X \left[\frac{\partial}{\partial\theta} f(x;\theta)\right]dx = \int_X \left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right] f(x;\theta)\,dx \quad (1.22)$$

Hence, we write

NOTES

$$E_\theta[Y] = E_\theta\left\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left[\log f(X_i;\theta)\right]\right\} = \sum_{i=1}^{n} E_\theta\left\{\frac{\partial}{\partial\theta}\left[\log f(X_i;\theta)\right]\right\} = 0$$

for all $\theta \in \Theta$, since $X_1,\ldots,X_n$ have identical distributions. Thus, equation (1.21) gives:

$$\tau'(\theta) = E_\theta\{TY\} = Cov_\theta(T, Y)$$

which is rewritten as

$$\{\tau'(\theta)\}^2 = Cov_\theta^2(T, Y) \le V_\theta(T)\,V_\theta(Y) \quad (1.23)$$

by virtue of the Cauchy-Schwartz inequality variance inequality. Recall that Y is a sum of n iid random variables and thus, in view of equation (1.22), we obtain:

$$V_\theta(Y) = nV_\theta\left\{\frac{\partial}{\partial\theta}\left[\log f(X_1;\theta)\right]\right\} = nE_\theta\left[\left\{\frac{\partial}{\partial\theta}\left[\log f(X_1;\theta)\right]\right\}^2\right] \quad (1.24)$$

Now, the inequality from equation (1.16) follows by combining equations (1.23) and (1.24).

Remark 1.2. *One can see that CRLB will be attained by the variance of an unbiased estimator T of $\tau(\theta)$ for all $\theta \in \Theta$ if and only if the equality in equation (1.23) holds, that is if and only if T and Y are linearly related w.p.1. Hence, CRLB will be attained by the variance of T if and only if*

$$T - a(\theta) = b(\theta)\,Y \text{ w.p.1. for all } \theta \in \Theta \quad (1.25)$$

with some fixed real valued function a(.) and b(.).

Remark 1.3. *Combining CRLB and $I_{X_1}(\theta)$, we can immediately restate the Cramer-Rao Inequality as follows:*

$$V_\theta(T) \ge \frac{\{\tau'(\theta)\}^2}{n\,I_{X_1}(\theta)} \quad (1.26)$$

We will interchangeably use the Cramer-Rao Inequality given by equations (1.16) or (1.26).

1.7.2 Lehmann-Scheffe Theorems

In situations encountered in Examples (1.18) and (1.19), neither the Rao-Blackwell Theorem nor the Cramer-Rao Inequality helps in deciding whether W is the UMVUE of $\tau(\theta)$. An alternative approach is needed.

Theorem 1.7. (Lehmann-Scheffe Theorem-I) *Suppose that T is an unbiased estimator of a real valued parametric function $\tau(\theta)$, $\theta \in \Theta \subseteq \Re^k$. Let **U** be a complete (jointly) sufficient statistic for θ. Define $g(u) \equiv E_\theta[T \mid \mathbf{U} = u]$ for $u \in \mathbf{U}$. Then, the statistic $W \equiv g(\mathbf{U})$ is a unique (w.p.1) UMVUE of $\tau(\theta)$.*

Proof: The difference between the Rao-Blackwell Theorem and this theorem is that now **U** is also assumed complete! The Rao-Blackwell Theorem assures us that in order to search for the best unbiased estimator of $\tau(\theta)$, we need to focus on unbiased estimators which are functions of **U** alone. We already know that (*i*) W is a function of **U**, and (*ii*) W is an unbiased estimator of $\tau(\theta)$. Suppose that there is another unbiased estimator of W^* of $\tau(\theta)$ where W^* is also a function of **U**. Define $h(\mathbf{U}) = W - W^*$ and then we have:

$$E_\theta[h(\mathbf{U})] = E_\theta[W - W^*] = \tau(\theta) - \tau(\theta) \equiv 0 \text{ for all } \theta \in \Theta \quad (1.27)$$

Now, use Definition of completeness of a statistic. Since U is a complete statistic, it follows that $h(\mathbf{U}) \equiv 0$ w.p.1. So, $W = W^*$.

In our quest for finding the UMVUE of $\tau(\theta)$, we may not always go through conditioning with respect to a complete sufficient statistic **U**. In many problems, the following alternate result may be directly applicable. Its proof can be easily constructed.

Theorem 1.8. (Lehmann-Scheffe Theorem - II) *Suppose that **U** is a complete sufficient statistic for $\theta \in \Theta \subseteq \Re^k$. Also, suppose that a statistic $W \equiv g(U)$ is an unbiased estimator of a real valued parametric function $\tau(\theta)$. Then, W is a unique UMVUE of $\tau(\theta)$.*

1.8 MVU AND BLACKWELLISATION

Cramer-Rao inequality provides us a technique of finding if the unbiased estimator is also an MVU estimator or not. Here, since the regularity conditions are very strict, its applications become quite restrictive. More-over MVB estimator is not the same as an MVU estimator since the Cramer-Rao lower bound may not always be attained. More-over, if the regularity conditions are violated, then the least attainable variance may be less than the Cramer-Rao bound. In this section we shall discuss *how to obtain MVU estimator from any unbiased estimator through the use of sufficient statistic*. This technique is called Blackwellisation after D. Blackwell. The result is contained in the following Theorem due to C.R. Rao and D. Blackwell.

Theorem 1.9. (Rao-Blackwell Theorem) *Let X and Y be random variables such that $E(Y) = \mu$ and $Var(Y) = \sigma_y^2 > 0$*

Let $E(Y \mid X = x) = \phi(x)$. Then (i) $E[\phi(X)] = \mu$, and (ii) $Var[\phi(X)] \leq Var(Y)$.

Proof: Let $f_{XY}(x, y)$ be the joint p.d.f. of random variables X and Y, $f_1(.)$ and $f_2(.)$ the marginal p.d.f.'s of X and Y respectively and $h(y \mid x)$ be the conditional p.d.f. of Y for given $X = x$ such that $h(y \mid x) = \{f(x, y)/f_1(x)\}$.

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot h(y|x)\, dy = \int_{-\infty}^{\infty} y \cdot \frac{f(x,y)}{f_1(x)}\, dy$$

$$= \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y\, f(x,y)\, dy = \phi(x), \text{ (say)} \quad \ldots(1.28)$$

$$\Rightarrow \quad \int_{-\infty}^{\infty} y\, f(x,y)\, dy = \phi(x) \cdot f_1(x) \quad \ldots(1.29)$$

From (1.28) we observe that the conditional distribution of Y given $X = x$ does not depend on the parameter μ. Hence X is sufficient statistic for μ. Also

$$E\{\phi(X)\} = E\{E(Y \mid X)\} = E(Y) = \mu, \quad \ldots(1.30)$$

which establishes part(*i*) of the theorem.

Now,
$$Var(Y) = E[Y - E(Y)]^2 = E[Y - \mu]^2 = E[Y - \phi(X) + \phi(X) - \mu]^2$$
$$= E[Y - \phi(X)]^2 + E[\phi(X) - \mu]^2 + 2E[\{Y - \phi(X)\}\{\phi(X) - \mu\}] \quad \ldots(1.31)$$

The product term gives

$$E[\{Y - \phi(X)\}\{\phi(X) - \mu\}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{(y - \phi(x))\}\{\phi(x) - \mu\}\, f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{(y-\phi(x))\}\{\phi(x)-\mu\}\, f_1(x)\, h(y|x)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\{\phi(x)-\mu\}\left[\int_{-\infty}^{\infty}\{y-\phi(x)\}\, h(y|x)\,dy\right]dx$$

NOTES

But $\int_{-\infty}^{\infty}\{y-\phi(x)\}\, h(y|x)\,dy = 0$ $\quad [\because \quad E(Y \mid X = x) = \phi(x)]$

$\therefore$ $E[Y-\phi(X)\}\{\phi(X)-\mu\}] = 0$ Substituting in (1.31), we get

$$\text{Var}(Y) = E\{Y-\phi(X)\}^2 + \text{Var}\{\phi(X)\} \qquad \ldots(1.32)$$

$\Rightarrow$ $$\text{Var } Y \geq \text{Var}[\phi(X)] \qquad [\because \quad E(Y-\phi(X))^2 \geq 0]$$

$\therefore$ $$\text{Var}\{\phi(X)\} \leq \text{Var } Y, \qquad \ldots(1.33)$$

which completes the proof of the theorem.

Remark 1.4. *From (1.32), it is obvious that the sign of equality holds in (1.33) iff*

$$E\{Y-\phi(X)\}^2 = 0 \quad \Rightarrow \quad Y-\phi(X) = 0, \text{ almost surely}$$

i.e., $\quad$ *iff* $P\{x, y) : y - \phi(x) = 0\} = 1 \qquad \ldots(1.34)$

Remark 1.5. *Here we have proved the theorem for continuous r.v.'s. The result can be similarly proved for discrete case, replacing integration by summation.*

Example 1.20. *Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta, 1)$. Obtain MVUE of θ.*

Solution. It can be easily proved that the statistic:

$$T = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i \text{, is complete sufficient statistic for } \theta.$$

Consider $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{T}{n} = g(T)$, (say), Since $\overline{X}_n = g(T)$, is unbiased estimator of θ, $\overline{X}_n$ is MVUE of θ.

Example 1.21. *Let $X_1, X_2, \ldots, X_n$ be a random sample from $U[0, \theta]$ population. Obtain MVUE for θ.*

Solution. We have seen that in sampling from $U(0, \theta)$ population, the statistic: $T = X_{(n)} = \max_{1 \leq i \leq n} (X_i)$ is sufficient for θ.

Also $$E(T) = E[X_{(n)}] = \left(\frac{n}{n+1}\right)\theta \quad \Rightarrow \quad E\left\{\frac{(n+1)\,T}{n}\right\} = \theta$$

Hence,

$$[(n+1)\,T/n] = [(n+1)\,X_{(n)}/n] \text{ is an MVU estimator of } \theta.$$

Example 1.22. *Given:* $f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta, \theta > 0 \\ 0, & \text{elsewhere} \end{cases} \qquad \ldots(1)$

compute the reciprocal of $nE\left[\left\{\frac{\partial \log f(x, \theta)}{\partial \theta}\right\}^2\right]$ *and compare this with the variance of $(n+1)$ Y_n/n, where Y_n is the largest item of a random sample of size n from this distribution. Comment on the result.*

Solution. $\log f(x,\theta) = -\log\theta \Rightarrow \frac{\partial}{\partial\theta}\log f = -\frac{1}{\theta}$ or $n\,\mathrm{E}\left(\frac{\partial}{\partial\theta}\log f\right)^2 = n\,\mathrm{E}\left(\frac{1}{\theta^2}\right) = \frac{n}{\theta^2}$

Hence, reciprocal of $n\,\mathrm{E}\left[\left\{\frac{\partial}{\partial\theta}\log f(x,\theta)\right\}^2\right] = \frac{\theta^2}{n}$ ...(2)

For the rectangular population (1), the pdf of nth order statistic (the largest sample observation), Y_n is : $g(y) = n\cdot\{F(y,\theta)\}^{n-1}\cdot f(y,\theta)$,

where $\quad F(x,\theta) = P(X \le x) = \int_0^x f(u)\,du = \int_0^x \frac{1}{\theta} = \frac{x}{\theta}$

$$g(y) = n\left(\frac{y}{\theta}\right)^{n-1}\frac{1}{\theta} = \frac{n}{\theta^n}y^{n-1}; 0 \le y < \theta$$

$$E\left(Y_n^r\right) = \int_0^\theta y^r\cdot g(y)\,dy = \frac{n}{\theta^n}\int_0^\theta y^{r+n-1}\,dy = \frac{n\theta^r}{n+r}$$

Taking $r = 1$ and 2; $E(Y_n) = \frac{n\theta}{n+1}$; $E\left(Y_n^2\right) = \frac{n\theta^2}{n+2}$ ...(3)

Now, $\mathrm{E}\left(\frac{n+1}{n}\cdot Y_n\right) = \frac{n+1}{n}\mathrm{E}\left(Y_n\right) = \theta$ [Using 3]

$\Rightarrow \quad (n+1)Y_n/n$ is an unbiased estimator of θ.

$$\mathrm{Var}\left(\frac{n+1}{n}Y_n\right) = \left(\frac{n+1}{n}\right)^2\cdot Var\left(Y_n\right) = \left(\frac{n+1}{n}\right)^2\left\{EY_n^2 - \left(EY_n\right)^2\right\}$$

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

$\Rightarrow \quad \mathrm{Var}\left(\frac{n+1}{n}\cdot Y_n\right) \le 1/\left\{n\,\mathrm{E}\left(\frac{\partial}{\partial\theta}\log f\right)^2\right\}$. Hence $(n+1)\,Y_n/n$ is an MVUE.

Remark 1.6. *This example illustrates that if the regularity conditions underlying Cramer-Rao inequality are violated, then the least attainable variance may be less than the Cramer-Rao lower bound.*

1.9 SUMMARY

- The deepest of all statistical concepts is *sufficiency* that originated from Fisher (1920), and it blossomed further, again in the hands of Fisher (1922).
- Lehmann and Scheffe (1950) developed a mathematical formulation of *minimal sufficiency* and gave a technique to locate minimal sufficient statistics.
- An Estimator is a function of **random variables** but the estimate is a function of **observed values.**
- Cramer-Rao inequality provides us a technique of finding if the unbiased estimator is also an MVU estimator or not.

1.10 GLOSSARY

- **Statistic :** An observable real (or vector) valued function $T \equiv T(X_1, \ldots, X_n)$.
- **Poisson distribution :** A random variable X is said to follow poisson distribution if it assumes non-negative values and its p.m.f. given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{\lfloor x}, x = 0, 1, 2, \ldots$$

where λ is the parameter. This distribution has only one parameter.

NOTES

- **Probability :** If a random experiment or a trial results in *'n'* exhaustive, mutually exclusive and equally likely outcomes (or cases), out of which *'m'* are favourable to the occurrence of an event E, then probability *'p'* of occurrence (or happening) of E, usually denoted by P(E), is given by

$$p = P(E) = \frac{\text{Number of favourable cases}}{\text{Total number of exhaustive cases}} = \frac{m}{n}$$

- **Blackwellisation :** Technique to obtain MVU estimator from any unbiased estimator through the use of sufficient statistic.

NOTES

1.11 REVIEW QUESTIONS

1. What do you understand by Point Estimation ? Define the following terms and give one example for each :
   (*i*) Consistent Statistic ; (*ii*) Unbiased Statistic ;
   (*iii*) Sufficient Statistic ; and (*iv*) Efficiency.
2. What do you understand by Point Estimation ? When would you say that estimate of a parameter is good ?
3. Define sufficient statistic. Let $x_1, x_2, ..., x_n$ be a random sample from N (μ, σ^2) population. Find sufficient estimators for μ and σ^2.
4. State and prove Facotorization Theorem.
5. Define Minimum Variance Unbiased Estimator. Prove that minimum variance unbiased estimator is essentially unique.
6. How is Cramer-Rao inequality useful in obtaining MVUE ? Derive this inequality.
7. State and prove Cramer-Rao inequality.
8. State and prove Rao-Blackwell Theorem and explain its significance in point estimation.
9. Let $x_1, x_2, ..., x_n$ be a random sample from a population with p.d.f.

$$f(x, \theta) = \theta\, x^{\theta - 1} \,;\, 0 < x < 1, \theta > 0.$$

   Show that $t_1 = \prod_{i=1}^{n} x_i$, is sufficient for θ.
10. Let $x_1, x_2, ..., x_n$ be a random sample from N $(\theta, 1)$. Obtain MVUE for θ.
11. Let $x_1, x_2, ..., x_n$ be a random sample from U $(0, \theta)$ population. Obtain MVUE for θ.

1.12 FURTHER READINGS

Goon, A.M., Gupta, M.K. and Das Gupta. B. An outline of statistical theory. Volume II, the world press, Calcutta, 1980.

Huntsberger, D.V. and Billingsley, P. Elements of Statistical Inference, Allyn and Bacon, London, 1977.

Rao, C.R., Linear Statistical Inference and its Applications, Wiley Eastern, New Delhi, 2nd ed. 1973.

Walker, H.M. and Lev, J., Statistical Inference, Henry Holt Company New York, 1953.

Gupta, S.C. and Kapoor, V.K., fundamentals of Mathematical Statistics, Sultan Chand & Sons, New Delhi.

UNIT-II

CONSISTANT ESTIMATOR

NOTES

OBJECTIVES

After going through this unit, you should be able to:

- explain consistant estimator
- define efficiency
- describe interval estimation
- give brief account on likelihood

STRUCTURE

2.1 INTRODUCTION

One of the main objectives of statistics in to draw inferences about a population from the analysis of a sample drawn from the population. Two important problems in statistical inference are estimation and testing of hypothesis.

The theory of estimation was founded by prof. R.A. Fisher in a series of fundamental papers round about 1930.

Consistency is one of the characteristic of estimators. It is a property concerning the behaviour of an estimator for indefinitely large values of the sample size n, *i.e.*, $n \to \infty$. Nothing is regarded of its behaviour for finite n. In this chapter, we will discuss about the resistant estimator, efficiency, minimum variance unbiased estimator (M.V.U.E). Interval estimation, estimators in multivariate case and also the fundamental theorem Cramer-Rao Inequality.

NOTES

2.2 CONSISTANT ESTIMATOR

We know that every estimate is a function of observed value and sample size. For therewith *the increase in the sample the error in the estimate likely reduces.* In view of this facts a sequence $\{t_n\}$ of estimators is said to be consistance if it approaches to the true value of the parameter with probabily 1 as sample size increases.

Mathematically, we call a sequence of a estimator $\{t_n\}$ to be consistant for a estimating θ

if $$\lim_{n \to \infty} P\{|t_n - \theta| > \epsilon\} = 0; \forall \epsilon > 0$$

or

equivalently

$$\lim_{n \to \infty} P\{|t_n - \theta| \leq \epsilon\} = 1; \forall \epsilon > 0$$

It is the definition of consistant estimator.

Example 2.1. *Show that if random sample from a normal population of a sample mean is consistant estimator for the population* μ.

or

Given a random sample from normal population with mean μ *and variance* σ^2. *Show that the sample mean is a consistant estimator of* μ.

Solution. Let $X_1, \ldots, X_n$ be a random sample drawn from a normal population with mean μ and S.D. σ *i.e.*, $X \sim N(\mu, \sigma^2)$

$$\Rightarrow \qquad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

or $$(\bar{X} - \mu) \sim N\left(0, \frac{\sigma^2}{n}\right)$$

Then according to the condition,

$$\begin{aligned} & P\{|t_n - \theta| \leq \varepsilon\} \\ &= P\{|\bar{x} - \mu| \leq \varepsilon\} \\ &= P\{|Z| \leq \varepsilon\} \\ &= P\{-\epsilon \leq Z \leq \varepsilon\} \end{aligned}$$

Since $$Z = \bar{X} - \mu \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\bar{X} - \mu)^2}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$f(\bar{X}) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}(\bar{X}-\mu)^2}$$

$$f(\bar{X}-\mu) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}z^2}$$

$$= \int_{-\epsilon}^{\epsilon} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}\cdot z^2} dz$$

NOTES

Define a new variate $\quad Y = \dfrac{\sqrt{n}\,Z}{\sigma}$

$$dY = \frac{\sqrt{n}}{\sigma} dZ$$

$$dz = dy \frac{\sigma}{\sqrt{n}}$$

$$\text{Limit} = -\frac{\sqrt{n}}{\sigma}\varepsilon \text{ and } +\frac{\sqrt{n}}{\sigma}\varepsilon$$

$$\Rightarrow \quad \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \int_{-\frac{\sqrt{n}}{\sigma}\varepsilon}^{\frac{\sqrt{n}}{\sigma}\varepsilon} e^{-\frac{n}{2\sigma^2}\cdot\frac{\sigma^2}{n}Y} \cdot \frac{\sigma}{\sqrt{n}} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\frac{\sqrt{n}}{\sigma}\varepsilon}^{\frac{\sqrt{n}}{\sigma}\varepsilon} e^{-\frac{1}{2}y^2} \cdot dy$$

Taking limit on both sides as $n \to \infty$

$$\lim_{n\to\infty} P\{t_n - \theta| \le \varepsilon\} = \lim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon\frac{\sqrt{n}}{\sigma}}^{\varepsilon\frac{\sqrt{n}}{\sigma}} e^{-\frac{1}{2}y^2} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = 1$$

Hence sample mean is a consistant estimator of μ.

Theorem 2.1. *If a statistic t_n is such that*

$$E(t_n) = \theta$$

$$V(t_n) \to 0 \text{ as } n \to \infty$$

Then $\{t_n\}$ is a consistant estimator of θ.

Proof: From Chebyshev's inequality, we have

$$P[\,|t_n - \theta| > \varepsilon] \le E\,\frac{\{t_n - \theta\}^2}{\varepsilon^2}$$

$$= \frac{1}{\varepsilon^2}\left[E\{t_n - E(t_n) + E(t_n) - \theta\}\right]^2$$

$$= \frac{1}{\varepsilon^2}\left[E\{t - E(t_n)\}^2 + 2E\{t_n - E(t_n)\}\{E(t_n) - \theta\} + E\{E(t_n) - \theta\}^2\right]$$

$$= \frac{1}{\varepsilon^2}\left[V(t_n) + 0 + E\{E(t_n) - \theta\}^2\right]$$

Hence letting $n \to \infty$ and using the hypothesis of the theorem we find that

$$\lim_{n\to\infty} P\left[|t_n - \theta| \ge \varepsilon\right] \le 0$$

but the probability of any given can not be less than zero. Hence it must be zero which shows that $\{t_n\}$ is consistant.

NOTES

Example 2.2. *Prove that in sample from a normal with N (μ, σ^2) the sample mean is consistant estimate of μ.*

Solution. If $$X \sim N(\mu, \sigma^2)$$

Then $$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

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

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

Thus as $n \to \infty$

i.e., $$\lim_{n \to \infty} E\left(\bar{X}\right) = \mu$$

$$\lim_{n \to \infty} V\left(\bar{X}\right) = 0 \text{ as } \lim_{n \to \infty} \frac{1}{n} \to 0$$

Hence $\bar{x}$ is a consistant estimator of μ.

Example 2.3. *Show that sample variance is biased but consistant estimator of σ^2 if $\bar{x} \sim N\left(N, \frac{\sigma^2}{N}\right)$. Hence or otherwise find an unbiased estimator of σ^2.*

If $x \sim \chi^2$ *or* $x \sim \chi^2_{n-1}$

$E(x) = n$ $\qquad E(x) = n - 1$

$V(x) = 2n$ $\qquad V(x) = 2(n-1)$

Solution. We have

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

If $$x \sim \chi^2_n \text{ then } \frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$$

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

or $$E(s^2) = \frac{n-1}{n}\sigma^2 \qquad \text{... (2.1)}$$

$\Rightarrow$ $E(s^2)$ is not equal to σ^2.

This shows that s^2 is not unbiased estimate for σ^2

$$\Rightarrow \quad \frac{ns^2}{\sigma^2}$$

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

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

$\Rightarrow$ $\frac{n}{n-1}s^2$ is an unbiased estimator of σ^2.

In order to show that s^2 (sample variance) is a consistant estimator, we need $V(s^2)$ for which $E(s^2)^2$ is an unbiased

$$V(s^2) = E(s^4) - [E(s^2)]^2$$

$$E(s^4) = E(s^2)^2 = \int \left(s^2\right)^2 f\left(s^2\right) ds^2$$

NOTES

where

$$f(s^2) = \frac{1}{2^{\frac{n-1}{2}}\sqrt{\frac{n-1}{2}}}\left(\frac{ns^2}{\sigma^2}\right)^{\frac{n-1}{2}-1} e^{-\frac{ns^2}{2\sigma^2}} \cdot \frac{n}{\sigma^2}$$

$$\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$f(x^2) = \sum\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2_n$$

$$\sum\left(\frac{x_i - \bar{X}}{\sigma}\right) \sim \chi^2_{n-1}$$

$$s^2 = \frac{1}{n}\sum\left(x_i - \bar{X}\right)^2$$

$$\frac{\sum\left(x_i - \bar{X}\right)^2}{\sigma^2} = \frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$E(s^4) = \frac{1}{2^{\frac{n-1}{2}}\sqrt{\frac{n-1}{2}}}\int_0^\infty s^4\left(\frac{ns^2}{\sigma^2}\right)^{\frac{n-3}{2}} e^{-\frac{ns^2}{2\sigma^2}} \frac{n}{\sigma^2}\, ds^2$$

Now define

$$x = \frac{ns^2}{\sigma^2} \Rightarrow s^2 = \frac{\sigma^2}{n}x \Rightarrow s^4 = \frac{\sigma^4}{n^2}x^2$$

$$\frac{\sigma^2}{n}\, dx = d(s^2)$$

$$E(s^4) = \frac{1}{2^{\frac{n-1}{2}}\sqrt{\frac{n-1}{2}}}\int_0^\infty \frac{\sigma^4}{n^2} x^2 (x)^{\frac{n-3}{2}} e^{-\frac{1}{2}x} \frac{n}{\sigma^2}\cdot\frac{\sigma^2}{n}\, dx$$

$$= \frac{\sigma^4}{n^2\, 2^{\frac{n-1}{2}}\sqrt{\frac{n-1}{2}}}\int_0^\infty x^{\frac{n+1}{2}} e^{-\frac{1}{2}x}\, dx$$

$$= \frac{\sigma^4}{n^2\, 2^{\frac{n-1}{2}}\sqrt{\frac{n-1}{2}}}\int_0^\infty x^{\frac{n+3}{2}-1} e^{-\frac{1}{2}x}\, dx \qquad \text{... (2.2)}$$

since $$f(x) = \frac{\alpha^p}{\sqrt{p}} x^{p-1} e^{-\alpha x} \qquad \textbf{(Gamma function)}$$

$$\int f(x)\,dx = 1 \Rightarrow \frac{\alpha^p}{\sqrt{p}}\int_0^\infty x^{p-1} e^{-\alpha x}\, dx = 1$$

$$\int_0^\infty x^{p-1} e^{-\alpha x}\, dx = \frac{\sqrt{p}}{\alpha^p}.$$

From (2.2),

$$E(s^4) = \frac{\sigma^4}{n^2\, 2^{\frac{n-1}{2}}\sqrt{\frac{n-1}{2}}} \cdot \frac{\sqrt{n+\frac{3}{2}}}{\left(\frac{1}{2}\right)^{n+\frac{3}{2}}}$$

$$= \frac{\sigma^4}{n^2\, 2^{\frac{n-1}{2}}}\, 2^{\frac{(n+3)}{2}-\frac{(n-1)}{2}} \cdot \frac{\left(\frac{n+1}{2}\right)\left(\frac{n-1}{2}\right)\sqrt{\frac{n-1}{2}}}{\sqrt{\frac{n-1}{2}}}$$

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

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

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

NOTES

Now,

$$V(s^2) = E(s^2)^2 - \{E(s^2)\}^2$$

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

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

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

Now expected value of sample variance

$$E(s^2) = \frac{n-1}{n}\sigma^2 = \left(1 - \frac{1}{n}\right)\sigma^2 = \sigma^2 \text{ as } n \to \infty$$

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

$$= 2\left(\frac{1}{n} - \frac{1}{n^2}\right)\sigma^4 = 0 \text{ as } n \to 0$$

$\Rightarrow$ Sample variance (s^2) is a consistant estimator of σ^2.

2.3 EFFICIENCY

If T_1 and T_2 are two consistant estimator of θ then T_1 is said to be more efficient then T_2 if variance of $(T_1) \leq$ Variance (T_2) $\forall$ *n i.e.*, $V(T_1) \leq V(T_2)$.

2.4 MINIMUM VARIANCE UNBIASED ESTIMATOR (M.V.U.E.)

An estimator is said to minimum variance unbiased estimator if it is unbiased and more efficient

i.e., $E(T) = 0, V(T) \leq V(T')$

where T′ is any other unbiased estimator of θ for all values of x.

If T_1 and T_2 are two estimators with variances σ^2 & σ^2 respectively then efficiency of T_2 over T_1 is defined as

$$e = \frac{V(T_1)}{V(T_2)}.$$

Theorem 2.2. *Minimum variance unbiased estimator (M.V.U.E.) is unique.*

i.e., if T_1 and T_2 are two (M.V.U.E.) for a parameter θ then $T_1 = T_2$.

Proof : We are given $E(T_1) = \theta = E(T_2)$

and $V(T_1) = V(T_2) = \sigma^2$

Because T_1 and T_2 are M.V.U.E.

Define $$T = \frac{T_1 + T_2}{2}$$

$$E(T) = \frac{1}{2}\{E(T_1) + E(T_2)\}$$

$$= \frac{1}{2}[\theta + \theta] = \theta \quad \text{... (2.3)}$$

$\Rightarrow$ T is an unbiased estimator of θ

$$V(T) = V\left(\frac{T_1 + T_2}{2}\right) = \frac{1}{4}V(T_1 + T_2)$$

NOTES

$$= \frac{1}{4}\left[V(T_1) + V(T_2) + 2\,\text{Cov.}(T_1, T_2)\right]$$

$$= \frac{1}{4}\left(\sigma^2 + \sigma^2 + 2\rho\sigma^2\right) \quad \text{... (2.4)}$$

where $\rho = \dfrac{\text{Cov.}(T_1, T_2)}{\sigma_{T_1} \cdot \sigma_{T_2}} \Rightarrow \text{Cov.}(T_1 T_2) = \rho\,\sigma_{T_1}\,\sigma_{T_2}$

$$= \rho\sigma^2$$

From (2.4) $V(T) = \dfrac{2\sigma^2}{4}(1+\rho) = \dfrac{\sigma^2}{2}(1+\rho)$

Since T_1 and T_2 are M.V.U.E.

$$V(T) \geq \sigma^2$$

$$\frac{\sigma^2}{2}(1+\rho) \geq \sigma^2$$

$$1 + \rho \geq 2$$

$$\rho \geq 1$$

since $\rho > 1$ we must have $\rho = 1$.

$\Rightarrow$ T_1, T_2 must have a linear relation of the form

$$T_2 = a + bT_1 \text{ where } a \text{ and } b \text{ are constants.}$$

Taking Expectation on both the sides, we get

$$E(T_2) = a + bE(T_1)$$

$$\theta = a + b\theta \quad \text{...(2.5)}$$

also $V(T_2) = b^2 V(T_1)$

$$\sigma^2 = b^2\sigma^2$$

$$b^2 = 1 \quad \Rightarrow \quad b = \pm 1$$

for $b = +1, a = 0 \quad \Rightarrow \quad T_1 = T_2$

for $b = -1, 2\theta = a$

Remark 2.1. *If we take $b = -1$, there will be a negative correlation between T_1 and T_2, hence $b = +1$.*

Example 2.4. *If T_1 and T_2 be two unbiased estimator of a parameter θ with variance σ_1^2, σ_2^2 and correlation ρ. Find best unbiased linear combination of the two estimators. Also find its variance.*

Solution. Where $E(T_1) = E(T_2) = \theta$

$$V(T_1) = \sigma_1^2, V(T_2) = \sigma_2^2$$

$$\text{Cov.}(T_1, T_2) = \rho\sqrt{\sigma_1^2\,\sigma_2^2} = \rho\,\sigma_1\,\sigma_2.$$

Let T be linear combination T_1 and T_2 given by

$$T = \alpha T_1 + \beta T_2 \quad \text{... (2.6)}$$

where α and β are constants.

In order that T is an unbiased estimator of θ we must have $E(T) = \theta$.

Taking Expectation of both side in equation (2.6), we get

$$E(T) = \alpha E(T_1) + \beta E(T_2)$$

$$\theta = \alpha\theta + \beta\theta$$

$$\alpha + \beta = 1 \quad \text{... (2.7)}$$

In order that T have minimum variance, we proceed as follows :

$$\begin{aligned} V(T) &= V(\alpha T_1 + \beta T_2) \\ &= V(\alpha T_1) + V(\beta T_2) + 2\alpha\beta \,\text{Cov.}(T_1, T_2) \\ &= \alpha^2 \sigma_1^2 + \beta^2 \sigma_2^2 + 2\alpha\beta\, \sigma_1 \sigma_2 \rho \qquad \text{... (2.8)} \end{aligned}$$

$$\frac{d}{d\alpha} V(T) = 0$$

NOTES

$$2\alpha\sigma_1^2 + 2\beta\rho\sigma_1\sigma_2 = 0 \qquad \text{...(2.9)}$$

$$\frac{d}{d\beta} V(T) = 0$$

$$2\beta\sigma_2^2 + 2\alpha\rho\sigma_1\sigma_2 = 0 \qquad \text{...(2.10)}$$

Subtracting equation (2.9) by (2.10), we get

$$= 2\alpha\sigma_1^2 + 2\rho\sigma_1\sigma_2(\beta - \alpha) - 2\beta\sigma_2^2 = 0$$

$$\alpha\sigma_1^2 + \rho\sigma_1\sigma_2(\beta - \alpha) - \beta\sigma_2^2 = 0$$

or $\alpha\sigma_1^2 + \rho\sigma_1\sigma_2\beta - \rho\sigma_1\sigma_2\alpha - \beta\sigma_2^2 = 0$

$$\alpha\left(\sigma_1^2 - \rho\sigma_1\sigma_2\right) = \beta\left(\sigma_2^2 - \rho\sigma_1\sigma_2\right)$$

$$\frac{\alpha}{\left(\sigma_2^2 - \rho\sigma_2\sigma_1\right)} = \frac{\beta}{\left(\sigma_1^2 - \rho\sigma_1\sigma_2\right)}$$

$$\frac{\alpha}{\left(\sigma_2^2 - \rho\sigma_2\sigma_1\right)} = \frac{\beta}{\left(\sigma_1^2 - \rho\sigma_1\sigma_2\right)} = \frac{\alpha + \beta}{\left(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\right)} = \frac{1}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \text{ from (2.7)}$$

So that the value of α

$$\frac{\alpha}{\left(\sigma_2^2 - \rho\sigma_2\sigma_1\right)} = \frac{1}{\left(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\right)}$$

$$\alpha = \frac{\sigma_2^2 - \rho\sigma_2\sigma_1}{\left(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\right)}$$

$$\beta = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\left(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\right)}$$

which is the least value of α and β, T given by (2.6) is the best combination which unbiased of T_1 and T_2 and its variance is given by (2.8).

Example 2.5. *If T_1 is M.V.U.E. and T_2 is any other unbiased estimator with variance $\frac{\sigma^2}{e}$ where e is efficiency. Then show that correlation coefficient between T_1 and T_2 is $\sqrt{e}$.*

Solution. We are given that

$$V(T_1) = \sigma_1^2 = \sigma^2 \text{ (say)}, \; V(T_2) = \frac{\sigma^2}{e}$$

and $\quad e = \dfrac{V(T_1)}{V(T_2)} = \dfrac{V(T_1)}{\sigma^2/e}$

$$V(T_1) = \frac{\sigma^2}{e} \cdot e = \sigma^2$$

and $\quad E(T_1) = E(T_2) = \theta$

Consider the best linear combination of T_1 and T_2

i.e., $$T = \alpha T_1 + \beta T_2 \qquad ...(2.11)$$

We have

$$\alpha = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

$$= \frac{\frac{\sigma^2}{e} - \rho\sigma\frac{\sigma}{\sqrt{e}}}{\sigma^2 + \frac{\sigma^2}{e} - 2\rho\sigma\frac{\sigma}{\sqrt{e}}}$$

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

$$\alpha = \frac{1-\rho\sqrt{e}}{1+e-2\rho\sqrt{e}} = \frac{1-\rho\sqrt{e}}{D}$$

where $D = 1 + e - 2\rho\sqrt{e}$

Similarly, $\beta = \dfrac{e-\rho\sqrt{e}}{D}$

Substituting it first we get the best unbiased estimator

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

$$= \frac{\left(1-\rho\sqrt{e}\right)T_1 + \left(e-\rho\sqrt{e}\right)T_2}{D}$$

The minimum variance

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

$$= \frac{1}{D^2}\left[\left(1-\rho\sqrt{e}\right)^2 V(T_1) + \left(e-\rho\sqrt{e}\right)^2 V(T_2) + 2\left(1-\rho\sqrt{e}\right)\left(e-\rho\sqrt{e}\right)\text{Cov.}(T_1\,T_2)\right]$$

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

Since, $$P = \frac{\text{Cov.}\,T_1\,T_2}{\sqrt{V(T_1)\,V(T_2)}}$$

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

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

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

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

NOTES

$$= \frac{\sigma^2}{D^2}\left[D - D\rho^2\right]$$

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

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

NOTES

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

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

or $$\frac{V(T)}{\sigma^2} = \frac{1-\rho^2}{\left(1-\rho^2\right)+\left(\sqrt{e}-\rho\right)^2} \le 1 \qquad \text{...(2.12)}$$

Since T_1 is M.V.U.E., $\qquad V(T) \not< \sigma^2$

Therefore $\qquad V(T) = \sigma^2$

From (2.12) $$\frac{1-\rho^2}{\left(1-\rho^2\right)+\left(\rho-\sqrt{e}\right)^2} = 1$$

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

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

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

$$\Rightarrow \qquad \rho = \sqrt{e}$$

2.5 LIKELIHOOD

The joint p.d.f. of the sample values is known as its likelihood function. In other words, if we take random sample of size n from a population following p.d.f. $f(x, \theta)$, then

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

is called the likelihood function.

If the observations are independent each having the same distribution $f(x, \theta)$ then

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

2.6 CRAMER-RAO INEQUALITY (C.R. INEQUALITY)

Theorem 2.3. *If t is unbiased estimator of some function of θ say $g(\theta)$ and the following regularity condition are hold.*

1. *Likelihood function is differentiable twice w.r.t. θ.*
2. *Limit of integration do not depend upon θ.*
3. *Differentiation under the sign of integrals is possible, then*

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

or $$V(t) \geq \frac{[g'(\theta)]^2}{I(\theta)}$$

where $I(\theta)$ is the information on θ.

NOTES

Proof : We know that

$$\iint\cdots\int f(x_1,\theta), f(x_2,\theta)\cdots f(x_n,\theta)\, dx_1\cdots dx_n = 1$$

or $$\iint\cdots\int L\, dx = 1, \text{ where } L = f(x_1, \theta)\ldots f(x_n, \theta) \text{ and } dx = dx_1, \ldots, dx_n$$

Differentiating it w.r.t. θ, we get

$$\iint\cdots\int \frac{d}{d\theta} L\, dx = 0$$

or $$\iint\cdots\int \frac{1}{L}\left(\frac{dL}{d\theta}\right) L\, dx = 0$$

or $$\int\cdots\int\left(\frac{d}{d\theta}\log L\right) dx\, L = 0$$

$$E\left[\frac{d}{d\theta}\log L\right] \times 1 = 0 \qquad \ldots(2.13)$$

Also $$\iint\cdots\int t\cdot L\, dx = g(\theta)$$

$$E(x) = \int x\, p(x)\, dx$$

$$\therefore \quad E(T) = \theta$$

$$\int f(t)\cdot t\cdot dt = \theta$$

$$\int t\cdot L\, dx = g(\theta)$$

Since t is unbiased estimator of $g(\theta)$.

Differentiate it w.r.t. θ, we get

$$\iint\cdots\int t\left(\frac{dL}{d\theta}\right) dx = g'(\theta)$$

where $$g'(\theta) = \frac{d}{d\theta} g(\theta)$$

or $$\int\cdots\int t\frac{1}{L}\left(\frac{dL}{d\theta}\right) L\, dx = g'(\theta)$$

or $$\iint\cdots\int t\left(\frac{d}{d\theta}\log L\right) L\, dx = g'(\theta)$$

$$E\left[t\frac{d}{d\theta}\log L\right] = g'(\theta) \qquad \ldots(2.14)$$

Now $$\text{Cov.}\left(t, \frac{d}{d\theta}\log L\right) = E\left(t\frac{d}{d\theta}\log L\right)$$

Solution. We have $f(x, \theta) = \dfrac{e^{-\theta}\theta^x}{x!}$

$$\log f(x, \theta) = -\theta + x \log \theta - \log x\,!$$

$$\frac{\partial}{\partial\theta} \log f(x, \theta) = -1 + \frac{x}{\theta} = \frac{x-\theta}{\theta}$$

NOTES

$$\mathrm{E}\left[\frac{\partial}{\partial\theta} \log f(x, \theta)\right]^2 = \mathrm{E}\left[\frac{x-\theta}{\theta}\right]^2$$

$$= \frac{1}{\theta^2}\mathrm{E}(x-\theta)^2 \text{ in Poisson distribution } \{\mathrm{V}(x) = m = \theta\}$$

$$= \frac{1}{\theta^2}\mathrm{V}(x) = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

Since mean and variance are same.

$$\mathrm{I}(\theta) = n\mathrm{E}\left[\frac{\partial}{\partial\theta} \log f(x, \theta)\right]^2 = \frac{n}{\theta}$$

now $g(\theta) = \theta$

$g'(\theta) = 1.$

Now according to C.R. inequality

$$\mathrm{V}(t) \geq \frac{1}{\mathrm{I}(\theta)} = \frac{\theta}{n} \qquad ...(2.18)$$

But
$$\mathrm{V}(\bar{x}) = \mathrm{V}\left[\frac{1}{n}\sum \mathrm{X_L}\right] = \frac{1}{n^2}\sum \mathrm{V}(x)$$

$$= \frac{1}{n^2}\sum \theta = \frac{n\theta}{n^2} = \frac{\theta}{n}$$

Hence C.R. lower bound is attained.

$\Rightarrow$ $\bar{\mathrm{X}}$ is an efficient or minimum variance unbiased estimate.

Example 2.8. *If* $x \sim \mathrm{N}(\mu, \sigma^2)$

Show that $\bar{x}$ *is efficient for* μ *or M.V.U.E. or* $\bar{x}$ *attain CRLB or M.V.U.E.*

Solution. We have $f(x, \mu) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

Taking log

$$\log f(x, \mu) = \log c - \frac{(x-\mu)^2}{2\sigma^2}$$

Differentiate it w.r.t. μ

$$\frac{1}{f(x, \mu)} f'(x, \mu) = \frac{-2(x-\mu)\times(-1)}{2\sigma^2}$$

$$\frac{d}{d\mu} \log f(x, \mu) = \frac{(x-\mu)}{\sigma^2}$$

$$\mathrm{E}\left[\frac{d}{d\mu} \log f(x, \mu)\right]^2 = \frac{\mathrm{E}(x-\mu)^2}{\sigma^4} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Hence CR inequality is

$$V(t) \geq \frac{1}{n \cdot \frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

But $$V(\bar{x}) = \frac{\sigma^2}{n}$$

Hence C.R.L.B. is attained

$\Rightarrow$ $\bar{x}$ is efficient estimator or M.V.U.E.

NOTES

Example 2.9. $f(x) = \theta^x (1-\theta)^{1-x}$ *where* $x = 0, 1$

Show that $\bar{x}$ *is M.V.U.E.*

Solution. $$\log f(x) = x \log \theta + (1-x) \log (1-\theta)$$

Differentiating w.r.t. θ

$$\frac{\partial}{\partial \theta} \log f(x) = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

$$= \frac{(1-\theta)x - \theta(1-x)}{\theta(1-\theta)}$$

$$\frac{\partial}{\partial \theta} \log f(x) = \frac{x - \theta x - \theta + \theta x}{\theta(1-\theta)} = \frac{(x-\theta)}{\theta(1-\theta)}$$

Taking Expectation on both the side

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

$$= \frac{1}{\theta^2(1-\theta)^2} V(x)$$

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

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

$$I(\theta) = nE\left[\frac{\partial}{\partial \theta} \log f(x, \theta)\right]^2 = \frac{n}{\theta(1-\theta)}$$

$$V(t) \geq \frac{1}{I(\theta)} = \frac{1}{n/\theta(1-\theta)}$$

$$V(t) = \frac{\theta(1-\theta)}{n}$$

But $V(\bar{x})$ in Bernoulli distribution $= V\left(\frac{1}{n}\sum x\right) = \frac{1}{n^2}\sum V(x)$

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

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

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

Hence $\bar{x}$ is efficient estimator.

NOTES

2.7 INTERVAL ESTIMATION

In interval estimation, we find an interval in which the value of the parameter may lie with certain true probability.

For example, suppose we are measuring the height of persons of a certain locality. If we says that the height of 90% people of that area lie in between 150 – 165 cm, we have found interval (150, 165) with probability 0.9 to represent the height. The *point estimation* procedure would have given some specific value say 160 cm to be the average height.

1. **Random Interval**

   An interval in which end points are random variable is called random interval. *e.g.*, $(x, 2x)$ is random interval, (1, 5) is not random internal.

2. **Confidence Interval**

   An interval in which parameter lies is called confidence interval.

3. **Confidence Co-efficient**

   It is the measure of assurance that the parameter lies in confidence interval *e.g.*,

   $$P(T_1 < \theta < T_2) = 1 - \alpha$$

   (T_1, T_2) are confidence interval, $1 - \alpha$ is confidence co-efficient (level).

2.7.1 Shortest Confidence Interval

We have

$$P[T_1 \le \theta \le T_2] = 1 - \alpha$$

Suppose $\quad P[\theta \le T_1] = \alpha_1$

and $\quad P[\theta \ge T_2] = \alpha_2$

s.t $\quad \alpha_1 + \alpha_2 = \alpha$

For given $\quad \alpha_1 \ge 0, \alpha_2 \ge 0$

Their may be infinite combination for $\alpha_1 + \alpha_2 = \alpha$ and hence infinite (T_1, T_2). It is therefore suggested that interval which has shortest confidence interval *i.e.*, for which $T_2 - T_1$ least, such a confidence interval is called shortest confidence interval.

Formulation

Let $x_1, x_2, \ldots\ldots, x_n$ be a random sample of size n from $f(x, \theta)$, where form of $f(x, \theta)$ is known [$f(x, \theta)$ is binominal, poisson ...] but θ is unknown. We want to estimate θ (parameter). For this define two statistics

$$T_1 = \theta_1(x_1, \ldots\ldots, x_n) = \text{lower confidence limit}$$

$$T_2 = \theta_2(x_1, \ldots\ldots, x_n) = \text{upper confidence limit}$$

and $\quad P[T_1 < \theta < T_2] = 1 - \alpha$

then (T_1, T_2) is called 100 $(1 - \alpha)$% confidence interval for θ. This $(1 - \alpha)$ is called confidence co-efficient.

We note that it does not mean that the probability of θ would lie (T_1, T_2) is $(1 - \alpha)$ for this would employ that θ is a random variable [which is not a case when θ is fixed though unknown and constant].

Therefore, either the specified interval contains or does not contains θ. If it contains, the probability is 1 if it does not contain, the probability is 0. If 0 is not a random variable (T_1, T_2) is random interval. If we take a large number of random sample we shall get a corresponding large number of interval and it is interpreted to mean that on the average 100 $(1 - \alpha)$% of these intervals are accepted to include θ.

Example 2.10. *Find a confidence interval for μ of a normal population for*

(i) σ is known

(ii) σ is unknown.

Solution. In order to find confidence interval for μ we take sample mean as it is minimum variance unbiased estimator.

NOTES

(*i*) If

$$x \sim N(\mu, \sigma^2)$$

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

Take $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$ so that we have to solve

$$\int_{-\infty}^{A} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} d\bar{x} = \frac{\alpha}{2}$$

$$\int_{B}^{\infty} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} d\bar{x} = \frac{\alpha}{2}$$

Define

$$Y = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$dy = \frac{1}{\sigma/\sqrt{n}} d\bar{x} \quad \Rightarrow \quad d\bar{x} = \frac{\sqrt{n}}{\sigma} dy$$

$$\Rightarrow \int_{-\infty}^{\frac{A-\mu}{\sigma/\sqrt{n}}} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2}} \cdot dy \frac{\sigma}{\sqrt{n}}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{A-\mu}{\sigma/\sqrt{n}}} e^{-\frac{y^2}{2}} \cdot dy = \frac{\alpha}{2}$$

And

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{B-\mu}{\sigma/\sqrt{n}}}^{\infty} e^{-\frac{y^2}{2}} dy = \frac{\alpha}{2} \qquad \text{...(2.19)}$$

Also we may write

$$\left.\begin{aligned} \int_{-\infty}^{\frac{-Z\alpha}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy &= \frac{\alpha}{2} \\ \text{and} \quad \int_{\frac{Z\alpha}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy &= \frac{\alpha}{2} \end{aligned}\right\} \qquad \text{...(2.20)}$$

From the table of standard normal distribution we have from (2.19) and (2.20)

$$\frac{A-\mu}{\sigma/\sqrt{n}} = -Z\alpha/2, \quad A \Rightarrow \mu - \frac{\sigma}{\sqrt{n}} Z\alpha/2$$

$$\frac{B-\mu}{\sigma/\sqrt{n}} = Z\alpha/2, \quad B \Rightarrow \mu + \frac{\sigma}{\sqrt{n}} Z\alpha/2$$

$$\Rightarrow P\left[\mu - Z\alpha/2 \frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu + Z\alpha/2 \frac{\sigma}{\sqrt{n}}\right] = \alpha$$

$$\Rightarrow \quad P\left[\mu - 1.96 \frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu + 1.96 \frac{\sigma}{\sqrt{n}}\right] = 0.05$$

But $$\mu - Z\alpha/2\frac{\sigma}{\sqrt{n}} \le \bar{x}$$

$$\Rightarrow \quad \mu \le \bar{x} + Z\alpha/2\frac{\sigma}{\sqrt{n}} \quad ...(2.21)$$

and $$\bar{x} \le \mu + Z\alpha/2\frac{\sigma}{\sqrt{n}}$$

NOTES

$$\Rightarrow \quad \mu \ge \bar{x} - Z\alpha/2\frac{\sigma}{\sqrt{n}} \quad ...(2.22)$$

Now combining (2.21) and (2.22), we find

$$\left(\bar{x} - Z\alpha/2\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z\alpha/2\frac{\sigma}{\sqrt{n}}\right)$$

or $$\left(\bar{x} - Z\alpha/2\frac{\sigma}{\sqrt{n}}, \bar{x} + Z\alpha/2\frac{\sigma}{\sqrt{n}}\right)$$

$\Rightarrow \left(\bar{x} \pm Z\alpha/2\frac{\sigma}{\sqrt{n}}\right)$ is the interval of 100 $(1 - \alpha)$% confidence interval when σ is known.

(ii) When σ is unknown, then

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

In that case we find A and B such that

$$P\{A < T < B\} = 1 - \alpha$$

$$\int_A^B f(t) \cdot dt = 1 - \alpha$$

where $f(t)$ is the p.d.f. of student's t-distribution with $(n - 1)$ degree of freedom.

Noting that E $[B - A]$ is minimum, when

$$B = -A$$

A and hence B can be obtained from the table of t distribution Table values are available for n = 2, 3, ..., 30 for large sample of distribution convert into normal.

Once we have obtained A and B, the confidence interval

$$A < \frac{\bar{x} - \mu}{s/\sqrt{n}} < B$$

$$\Rightarrow \quad \frac{As}{\sqrt{n}} < \bar{x} - \mu \text{ and } \bar{x} - \mu < \frac{Bs}{\sqrt{n}}$$

$$\mu < \bar{X} - \frac{As}{\sqrt{n}} \text{ and } \mu > \bar{x} - \frac{Bs}{\sqrt{n}}$$

$$\bar{x} - \frac{Bs}{\sqrt{n}} < \mu < \bar{x} - \frac{As}{\sqrt{n}}$$

$$\left\{\bar{x} - t\alpha/2\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\alpha/2\frac{s}{\sqrt{n}}\right\}$$

when σ is not known.

2.8 METHOD OF ESTIMATION

However we have not discussed as yet a procedure with which we can find a reasonable estimator. Several such procedure exist. We shall discuss the following method for finding the estimator.

1. Maximum likelihood
2. Minimum chi-square
3. Method of moments
4. Method of variance

NOTES

2.8.1 Maximum Likelihood Method

The maximum likelihood method is to obtain that value of parameter (in terms of sample value) which maximize the likelihood function of the sample values.

For this we solve $$\frac{\partial L}{\partial \theta} = 0 \text{ and check } \left.\frac{\partial L^2}{\partial \theta^2}\right|_{\theta=\hat{\theta}} < 0$$

or $$\frac{\partial \log L}{\partial \theta} = 0.$$

Thus instead of $\frac{\partial L}{\partial \theta} = 0$, we generally solve

$$\frac{\partial \log L}{\partial \theta} = 0 \text{ and check that } \left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta=\hat{\theta}} < 0.$$

Example 2.11. *Find a Maximum Likelihood Estimate (M.L.E.) for θ from a sample obtained from a poission distribution. Show that this estimator is efficient and M.V.U.E.*

Solution. We have $$f(x, \theta) = \frac{e^{-\theta}\,\theta^x}{x!}, \text{ where } x = 0, 1, 2, \ldots$$

The likelihood function $$L = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\,\theta^{x_i}}{x_i!}$$

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

Taking log on both sides,

$$\log L = -n\theta + \sum_{i=1}^{n} x_i \log \theta - \log c$$

where $$c = \prod_{i=1}^{n} x_i\,!$$

$$\frac{\partial \log L}{\partial \theta} = 0 \Rightarrow -n + \sum_{i=1}^{n} \frac{x_i}{\theta} - 0 = 0$$

$$(1-\theta)\,n-\theta\sum x_i = 0$$

$$n-n\theta-\theta\sum x_i = 0$$

$$n=\theta\left(\sum_{i=1}^{n} x_i+n\right)$$

NOTES

or $$\hat{\theta}=\frac{n}{n\bar{x}+n}=\frac{1}{\bar{x}+1}$$

Differentiating (2.24) again

$$-\frac{n}{\theta^2}+\frac{\sum x_i}{(1-\theta)^2}=0.$$

(*iv*) We have $$f(x)=\frac{\alpha^{\beta}}{\overline{|\beta}}\,x^{\beta-1}\,e^{-\alpha x}$$

The likelihood function

$$\mathrm{L}=\frac{\alpha^{n\beta}}{\left(\overline{|\beta}\right)^n}\left\{\prod_{i=1}^{n} x_i^{\beta-1}\right\}e^{-\alpha\sum_{i=1}^{n}x_i}$$

In order to estimate α, we must know β otherwise we need to estimate α and simultaneously. If β is known, then

$$\log \mathrm{L}=n\beta\log\alpha-n\log\left(\overline{|\beta}\right)+(\beta-1)\sum_{i=1}^{n}\log x_i-\alpha\sum_{i=1}^{n}x_i$$

$$\frac{\partial}{\partial\alpha}\log \mathrm{L}=0\Rightarrow\frac{n\beta}{\alpha}-\sum_{i=1}^{n}x_i=0$$

$$\frac{n\beta}{\sum_{i=1}^{n}x_i}=\alpha$$

$$\frac{n\beta}{n\bar{x}}=\alpha$$

$$\frac{\beta}{\bar{x}}=\alpha$$

$$\hat{\alpha}=\frac{\beta}{\bar{x}}$$

(*v*) If $$f(x)=\beta x^{\beta-1}\,e^{-x^{\beta}}$$

The likelihood

$$\mathrm{L}=\beta^n\prod_{i=1}^{n}x_i^{\beta-1}\,e^{-\sum_{i=1}^{n}x_i^{\beta}}$$

$$\log \mathrm{L}=n\log\beta+(\beta-1)\sum_{i=1}^{n}\log x_i-\sum_{i=1}^{n}x_i^{\beta}$$

$$\frac{\partial}{\partial\beta}\log \mathrm{L}=0$$

$$\Rightarrow \quad \frac{n}{\beta} + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^{\beta} \log x_i$$

It is not possible to solve this equation for β, hence M.L.E. is not available.

Note : When the range of variable depends on parameter θ (say), then general method of differentiation fails and in this case arguments are required.

NOTES

Example 2.13. *Find the M.L.E of the following*

(i) $f(x) = \frac{1}{\theta}, 0 \le x \le \theta$

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

Solution.

(i)
$$f(x) = \frac{1}{\theta}, 0 \le x \le \theta$$

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

$$\log L = -n \log \theta$$

$$\frac{\partial}{\partial \theta} \log L = \frac{-n}{\theta} = 0$$

$$\theta = 0$$

Since $0 \le x \le \theta$, $\frac{1}{\theta}$ will be maximum if x takes its maximum that is M.L.E. of $\theta = \max(x_i)$.

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

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

$$\log L = -\sum_{i=1}^{n}(x_i - \theta)$$

$$\frac{\partial \log L}{\partial \theta} = \theta = 0$$

We want to maximize $e^{-\sum_{i=1}^{n}(x_i - \theta)}$ and this will be maximum if $x - \theta$ is minimum that is if put $\theta \le x_1 \le x_2 \le \ldots\ldots\ldots, \le x_n$ then $\theta = x_1$ is M.L.E.

2.8.1.1 Properties of M.L.E.

1. The M.L.E. is unique
2. It is consistent
3. Most efficient.
4. It is sufficient if any exists.

2.8.2 Method of Minimum Chi-square

For a qualitative sampling, we distribute the sample observations into certain categories. Suppose there are K categories in which the population can be distributed and Let $\pi_1(\theta)$, $\ldots, \pi_k(\theta)$ be the probabilities that an observation will lie in $I^{st}, \ldots, k^{th}$ category.

The likelihood function for a random sample is given by

$$F(x) = \frac{n!}{n_1!\ldots\ldots n_k!}\prod_{i=1}^{k}\pi^{n_i}(\theta) = L$$

where n_i is the number of observation in i^{th} category and $\sum_{i=1}^{k} n_i = n_i \sum_{i=1}^{k} \pi_i(\theta) = 1$. This

NOTES

distribution is known as the multinomial distribution.

χ^2 method is used only when we have a multinomial distribution whose all probabilities $\pi_1(\theta)$ .... $\pi_k(\theta)$ depends on an unknown parameter θ and the data is in the form of frequency n_1 ...., n_k which minimise

$$\chi^2 = \sum_{i=1}^{k}\left[\frac{[n_i - n\pi_i(\theta)]^2}{n\pi_i(\theta)}\right]$$

$$= \sum_{i=1}^{k}\frac{n_i^2}{n\pi_i(\theta)} - n$$

where, n_i = observed frequency
$n\pi_i(\theta)$ = expected frequency

Since $\pi_i(\theta)$, probability (an observed falling in i^{th} cell), the number of observation out of n is expected to fall in the i^{th} cell is $E(x_i) = n\pi_i(\theta)$.

In actual practice, we solve the equation

$$\frac{\partial t}{\partial \theta} = 0 \text{ and then } \hat{\theta} \text{ is minimum } \chi^2 \text{ estimate of } \theta \text{ if } \left.\frac{\partial^2 \chi^2}{\partial \theta^2}\right|_{\theta=\hat{\theta}} > 0$$

i.e.,

$$\sum_{i=1}^{k}\frac{n_i^2}{\pi_i^2(\theta)}\frac{\partial}{\partial\theta}\pi_i(\theta) = 0.$$

2.8.3 Modified minimum χ^2 method of estimation

We shall define χ^2 as

$$\chi^2_{\text{mod}} = \sum_{i=1}^{k}\left[\frac{\{n_i - n\pi_i(\theta)\}^2}{n_i}\right]$$

$$= \sum_{i=1}^{k}\frac{n^2\pi_i^2(\theta)}{n_i} - n$$

or

$$\frac{\partial}{\partial\theta}\chi^2_{\text{mod}} = \sum_{i=1}^{k}\frac{n^2\pi_i(\theta)}{n_i}\frac{\partial}{\partial\theta}\pi_i(\theta) = 0$$

The solution of χ^2_{mod} estimator is

$$\frac{\partial}{\partial\theta}\chi^2_{\text{mod}} = 0$$

Let $\hat{\theta}$ be the solution of χ^2_{mod}.

If $\left.\frac{\partial^2}{\partial\theta^2}\chi^2_{\text{mod}}\right|_{\theta=\hat{\theta}} > 0$ then $\hat{\theta}$ is the minimum modified χ^2 estimator of θ.

NOTES

Example 2.14. *In a rectangular distribution R (0, θ) let there be two cells with respective probabilities* $\pi_1(\theta) = \frac{a}{\theta}$ *and* $\pi_2(\theta) = 1 - \frac{a}{\theta}$ *and let* n_1 *and* n_2 *be observed frequency. Obtain minimum* χ^2 *estimator of* θ.

Solution. We have $n = n_1 + n_2$

Since $$\frac{\partial}{\partial\theta}\chi^2 = \sum_{i=1}^{2}\frac{n_i^2}{\pi_i(\theta)}\frac{\partial}{\partial\theta}\pi_i(\theta) = 0$$

or $$\frac{n_1^2}{\pi_1^2(\theta)}\frac{\partial}{\partial\theta}\pi_1(\theta) + \frac{n_2^2}{\pi_2^2(\theta)}\frac{\partial}{\partial\theta}\pi_2(\theta) = 0$$

$$= \frac{n_1^2}{a^2/\theta^2}\frac{\partial}{\partial\theta}\frac{a}{\theta} + \frac{n_2^2}{\left(1-\frac{a}{\theta}\right)^2}\frac{\partial}{\partial\theta}\left(1-\frac{a}{\theta}\right) = 0$$

$$\frac{n_1^2\,\theta^2}{a^2}\left(-\frac{a}{\theta^2}\right) + \frac{n_2^2\,\theta^2}{(\theta-a)^2}\frac{a}{\theta^2} = 0$$

$$= -\frac{n_1^2}{a} + \frac{an_2^2}{(\theta-a)^2} = 0$$

$$\frac{a(n-n_1)^2}{(\theta-a)^2} - \frac{n_1^2}{a} = 0 \quad \text{since } n = n_1 + n_2$$

$$\frac{a^2(n-n_1)^2 - n_1^2(\theta-a)^2}{a(\theta-a)^2} = 0$$

$$a^2n^2 + a^2n_1^2 - 2ann_1 - n_1^2\theta^2 - n_1^2a^2 + 2a\theta n_1^2 = 0$$

$$a^2n^2 - n_1^2\theta^2 + 2a\theta n_1^2 - 2a^2nn_1 = 0$$

or $$-\theta^2n_1^2 + 2an_1^2\theta + a^2n(n-2n_1) = 0$$

$$\theta^2n_1^2 - 2an_1^2\theta + a^2n(n_2-n_1) = 0$$

$$\theta = \frac{2an_1^2 \pm \sqrt{4a^2n_1^4 + 4a^2n_1^2n_2^2 - 4a^2n_1^4}}{2n_1^2}$$

$$= \frac{an_1^2 \pm an_1n_2}{n_1^2}$$

$$= \frac{an_1 \pm an_2}{n_1}$$

$$= a\frac{(n_1 \pm n_2)}{n_1}$$

⇒ $$\hat{\theta} = \frac{a}{n_1}n = ap, \text{ where } p = \frac{n}{n_1}$$

and $$\text{also } \hat{\theta} = \frac{a}{n_1}(n_1 - n_2) = a\left(1 - \frac{n_2}{n_1}\right)$$

2.8.4 Method of Moments

Let $x_1, x_2, ..., x_n$ be i.i.d.r.v from $f(x, \theta_1, ..., \theta_k)$. Let m_r' and μ_r' denote the r^{th} sample moment and population moment respectively

$$\mu_r' = \int x^r f(x, \theta_1, \cdots\cdots \theta_k)\, dx$$

NOTES

and $m_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r$. Clearly μ_r' will be function of one or more parameters $\theta_1, \theta_2, ..., \theta_k$.

The method of moments for estimating parameters consist of solving the equations and

$$m_r' = \mu_r'(\theta_1 \cdots\cdots \theta_k)$$

and there by obtaining the solutions $\hat{\theta}_1, \hat{\theta}_2, \cdots\cdots \hat{\theta}_k$ in terms of $m_1' \cdots\cdots m_k'$.

Remark 2.2. *We can apply this technique only when the system of equation*

$$m_1' = \mu_1'(\theta_1 \cdots\cdots \theta_k)$$

$$m_2' = \mu_2'(\theta_1 \cdots\cdots \theta_k)$$

$$\vdots \quad \vdots \quad \vdots$$

$$m_k' = \mu_k'(\theta_1 \cdots\cdots \theta_k)$$

Explicitly yields $\theta_1, \theta_2, ..., \theta_k$ in terms of $m_1', \cdots\cdots m_k'$. There may be situations when the above system of equations do not admit explicit expression for θ_1 ...... θ_k in terms of $m_1' \cdots\cdots m_k'$ and then the moments method fails.

2.8.4.1 Properties

1. The estimators are consistant.
2. The estimators are asymtotically normal but not efficient in general.
3. It is less efficient then the M.L.E.

Example 2.15. *$x \sim N(\theta, \sigma^2)$ find estimator of θ and σ^2 by method of moments.*

Solution. We have

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x; \theta, \sigma^2)\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^r e^{-\frac{1}{2\sigma^2}(x-\theta)^2}\, dx$$

$$\int x f(x)\, dx = E(x) = \theta \; ; \int f(x) = 1$$

$\Rightarrow$ $$\mu_1' = \theta,\ \mu_2' = \theta^2 + \sigma^2$$

Hence, $$\theta = \mu_1'$$

and $$\sigma^2 = \mu_2' - \theta^2 \quad \Rightarrow \quad \mu_2' - (\mu_1')^2$$

Since $$m_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

$\Rightarrow$ $$m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} = \mu_1'$$

$$m_2' = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \mu_2'$$

Therefore, $\hat{\theta} = m_1' = \bar{x}$

and $\hat{\sigma}^2 = m_2' - (m_1')^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

NOTES

Theorem 2.4. *If a sufficient estimator exists, it is a function of the Maximum Likelihood Estimator.*

Proof: If $t = t(x_1, x_2, ..., x_n)$ is a sufficient estimator of θ, then Likelihood Function can be written as $L = g(t, \theta)\, h(x_1, x_2, x_3, ..., x_n \mid t)$, where $g(t, \theta)$ is the density function of t and θ and $h(x_1, x_2, ..., x_n \mid t)$ is the density function of the sample, given t, and is independent of θ.

$$\therefore \quad \log L = \log g(t, \theta) + \log h(x_1, x_2, ..., x_n \mid t)$$

Differentiating w.r. to θ, we get : $\frac{\partial \log L}{\partial \theta} = \frac{\partial}{\partial \theta} \log g(t, \theta) = \psi(t, \theta)$, (say), ...

which is a function of t and θ only.

M.L.E. of θ is given by $\frac{\partial \log L}{\partial \theta} = 0 \Rightarrow \psi(t, \theta) = 0$

$$\therefore \quad \hat{\theta} = \eta(t) = \text{Some function of sufficient statistic}$$

$$\Rightarrow \quad \hat{t} = \xi(\hat{\theta}) = \text{Some function of M.L.E.}$$

Hence the theorem.

2.9 CAN AND CAUN ESTIMATORS : MULTIPARAMETER CASE

Let $T = (T_1, ..., T_m)'$ be a vector valued estimator which is consistent for a vector parameter $\theta = (\theta_1, ..., \theta_m)'$. If there exists a sequence $a_n > 0$ and $a_n \to \infty$ such that the vector valued r.v. $a_n(T - \theta)$ has in the limit a non-singular proper m—dimensional normal distribution then we say that T is CAN for θ. Thus if $a_n(T - \theta) \xrightarrow{d} N^{(m)}(0, \Lambda(\theta))$ where $\Lambda(\theta)$ is a symmetric positive definite matrix then T is said to have asymptotic normal distribution with mean vector θ and variance covariance matrix $\frac{\Lambda(\theta)}{a_n^2}$. This is denoted by $T \sim AN^{(m)}\left(\theta, \frac{\Lambda(\theta)}{a_n^2}\right)$. As in case of $m = 1$, usually the choice of $a_n = \sqrt{n}$ will suffice.

One consequence of the above definition is that each T_i is CAN for θ_i with $AV(T_i) = \frac{\lambda_{ii}(\theta)}{n}$

and any linear combination $T' = \sum_{i=1}^{m} l_i T_i$ is CAN for $\sum_{i=1}^{m} l_i \theta_i$ with $AV(T') = \frac{1}{n} l' \Lambda(\theta) l$.

As univariate CLT for i.i.d.r.v. with finite variance is helpful is constructing CAN estimators for a real parameter θ, in a similar way multivariate CLT (MVCLT) for i.i.d.r.v. with finite p.d. variance-covariance matrix is useful to obtain CAN estimators for

$\theta = (\theta_1 \ldots\ldots \theta_m)'$. Let $Z = (Z_1, \ldots\ldots Z_m)'$ be such that $E(Z) = \mu$ and $M_z = \Lambda$ which we assume to be p.d. Let $[Z_i]_1^n$ be i.i.d.r.v.s. distributed as Z and let $\bar{Z} = (\bar{Z}_1, \ldots\ldots, \bar{Z}_m)'$ be the mean vector of the sample *i.e.*, $\bar{Z}_r = \frac{1}{n}\sum_{i=1}^{n} Z_{ri}$, $r = 1, 2, \ldots\ldots, m$. Then MVCLT asserts that

NOTES

$\sqrt{n}(\bar{Z} - \mu) \xrightarrow{d} N^{(m)}(0, \Lambda)$ or $\bar{Z} - AN^{(m)}\left(\mu, \frac{1}{n}\Lambda\right)$. We now illustrate this technique by way of a few examples.

Example 2.16. Let $(X_1, \ldots\ldots, X_n)$ be i.i.d. $N(\mu, \sigma^2)$. Then $E(X_i) = \mu$, $E(X_i^2) = \mu^2 + \sigma^2$. Let $Z = (Z_1, Z_2)'$ where $Z_1 = X_1$ and $Z_2 = X_1^2$. Then A in general is given by $\begin{pmatrix} \mu_2' - \mu_1'^2 & \mu_3' - \mu_1'\mu_2' \\ \mu_3' - \mu_1'\mu_2' & \mu_4' - \mu_2'^2 \end{pmatrix}$ and for the normal distribution $A = \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\sigma^2\mu^2 \end{pmatrix}$. Therefore by MVCLT, we have for samples from $N(\mu, \sigma^2)$.

$$\begin{pmatrix} m_1' \\ m_2' \end{pmatrix} - AN^{(2)}\left[\begin{pmatrix} \mu \\ \mu^2 + \sigma^2 \end{pmatrix}, \frac{1}{n}\begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\sigma^2\mu^2 \end{pmatrix}\right]$$

where m_1' and m_2' are first two raw sample moments.

Example 2.17. Let the p.m.f. of (X, Y) be given by

$$f(x, y, \lambda, p) = \binom{x}{y} p^y (1-p)^{x-y} e^{-\lambda}\frac{\lambda^x}{x!}, 0 < p < 1, 0 < \lambda,$$

$$y = 0, 1, 2, \ldots\ldots, x;\ x = 0, 1, 2, \ldots..$$

Then $E(X) = \lambda$, $E(Y) = \lambda p$ and $M_z = \begin{pmatrix} \lambda & \lambda p \\ \lambda p & \lambda p \end{pmatrix}$. This follows from the fact that $X \sim P(\lambda)$, $Y \sim P(\lambda p)$ and $E(XY) = E[XE(Y \mid X)] = E(XXp) = E(X^2 p) = (\lambda + \lambda^2)p$. Therefore,

$$\begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix} \sim AN^{(2)}\left[\begin{pmatrix} \lambda \\ \lambda p \end{pmatrix}, \frac{1}{n}\begin{pmatrix} \lambda & \lambda p \\ \lambda p & \lambda p \end{pmatrix}\right].$$

Example 2.18. Let $Z = (Z_1, \ldots\ldots, Z_k)'$ be a multinomial r.v. in k-cells with $f(z_1, \ldots\ldots, z_k) = p_1^{z_1}, p_2^{z_2}, \ldots\ldots p_x^{z_k}$, $z_i = 0$ or 1, $\sum_{i=1}^{k} z_i = 1$ and $0 < p_i < 1$, $\Sigma p_i = 1$. Then $E(Z_i) = p_i$ and $V(Z_i) = \lambda_{ii} = p_i(1 - p_i)$ and $Cov(Z_i Z_j) = -p_i p_j$. Thus,

$$M_z = \begin{pmatrix} p_1(1-p_1) & \cdots & -p_1 p_k \\ \cdots & & \cdots \\ -p_k p_1 & \cdots & p_k(1-p_k) \end{pmatrix}$$

However as $\Sigma z_i = 1$, M_z is singular and has rank $(k-1)$. We therefore consider only $(k-1)$ of z_i's say $(z_1, \ldots\ldots, z_{k-1})$ so that $z_k = (1 - z_1 \ldots\ldots, -z_{k-1})$, and $p_k = (1 - p_1 - p_2, \ldots\ldots, -p_{k-1})$.

NOTES

Then by MVCLT as applied to $\bar{Z} = \left(\bar{Z}_1, \cdots\cdots, \bar{Z}_{k-1}\right)'$ the vector of $(k-1)$ relative cell frequencies in $(k-1)$ cells will have the asymptotic normal distribution with mean vector $(p_1, \ldots\ldots, p_{k-1})'$ and asymptotic variance covariance matrix $\frac{1}{n}\Lambda$ where $\lambda_{ii} = p_i(1-p_i)$ and $\lambda_{ij} = -p_i p_j$, $i = 1, 2, \ldots\ldots, k-1$; $j = 1, 2, \ldots\ldots, k-1$.

Similar establishes invariance of the CAN property under continuous differentiable transformation we can show that if $T \sim AN^{(m)}\left(\theta, \Lambda(\theta)/a_n^2\right)$ and if $\psi = (\psi_1, \ldots\ldots, \psi_k)'$ are such that $\frac{\partial \psi_i}{\partial \theta_j}$ are continuous for $i = 1, 2, \ldots\ldots, k$ and $j = 1, 2, \ldots\ldots, m$ then $\psi(T) \sim AN^{(k)}$ $\left(\psi(\theta), \frac{G\Lambda G'}{a_n^2}\right)$ where

$$G = \begin{pmatrix} \frac{\partial \psi_1}{\partial \theta_1} & \cdots\cdots & \frac{\partial \psi_1}{\partial \theta_m} \\ \vdots & & \vdots \\ \frac{\partial \psi_k}{\partial \theta_1} & \cdots\cdots & \frac{\partial \psi_k}{\partial \theta_m} \end{pmatrix}$$

provided $G \Lambda G'$ is positive definite. Uses Taylor series expansion of each component of the vector ψ. Thus $a_n(\psi(T) - \psi(\theta)) = G\, a_n(T-\theta) + a_n R$ where $a_n R \xrightarrow{P} 0$ and as $a_n(T-\theta) \xrightarrow{d} N^{(m)}(0, \Lambda)$, $G\, a_n(T-\theta) \xrightarrow{d} N^{(k)}(0, G\Lambda G')$. Hence, we have theorem. Let $T \sim AN^{(m)}\left(\theta, \Lambda(\theta)/a_n^2\right)$ and $\psi = (\psi_1, \ldots\ldots, \psi_k)'$ and $G = \left(\left(\frac{\delta \psi_r}{\delta \theta_s}\right)\right)$ such that $G\Lambda G'$ is p.d. then $\psi(T) \sim AN^{(k)}\left(\psi(\theta), G\Lambda G'/a_n^2\right)$.

Example 2.19. In example 2.15, we showed that $(m_1', m_2')'$ is $AN^{(2)}\left(\mu_1', \mu_2'\right)', \Lambda/n$. Note that here $(\theta_1, \theta_2)' = (\mu_1', \mu_2')' = (\mu, \sigma^2 + \mu^2)' \neq (\mu, \sigma^2)'$.

Now suppose we want to obtain CAN estimators of μ and σ^2. Then we take $\psi_1 = \mu_1' = \mu$, $\psi_2 = \mu_2' - \mu_1'^2 = \sigma^2$. Then

$$G = \begin{pmatrix} 1 & 0 \\ -2\mu_1' & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2\mu & 1 \end{pmatrix}$$

Therefore $\psi_1(m_1', m_2') = m_1' = \bar{X}$ and $\psi_2(m_1', m_2') = m_2' - m_1'^2 = \frac{S^2}{n} = m_2$ the second central moment of the sample and we have

$$\begin{pmatrix} \bar{X} \\ S^2/n \end{pmatrix} \sim AN^{(2)}\left[\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \frac{G\Lambda G'}{n}\right]$$

By straightforward calculations we can show that

$$G\Lambda G' = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}.$$

If we use the general expression for Λ given in Example 2.15 namely,

$$\Lambda = \begin{pmatrix} \mu_2' - \mu_1'^2 & \mu_3' - \mu_2'\mu_2' \\ \mu_3' - \mu_1'\mu_2' & \mu_4' - \mu_2'^2 \end{pmatrix}$$

NOTES

and $G = \begin{pmatrix} 1 & 0 \\ -2\mu_1' & 1 \end{pmatrix}$ then $G\Lambda G' = \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}$

where μ_2, μ_3, μ_4 are central moments of the p.d.f. $[f(x, \theta), \theta \in \Omega]$. Note that if the p.d.f. is symmetric about mean or even if just $\mu_3 = 0$ we have $G\Lambda G' = \text{diag}\left(\mu_2, \mu_4 - \mu_2^2\right)$ and $\bar{X}$ and $\frac{S^2}{n}$ would be asymptotically normal with off diagonal elements zero or equivalently $\bar{X}$, and $\frac{S^2}{n}$ would be asymptotically independent. Further observe that $\left(\bar{X}, \frac{S^2}{n}\right)'$ is a solution of the moment equations $m_1' = \mu$ and $m_2' = \mu^2 + \sigma^2$ and $\frac{\partial(\mu, \mu^2 + \sigma^2)}{\partial(\mu, \sigma^2)} = \begin{pmatrix} 1 & 0 \\ 2\mu & 1 \end{pmatrix} = G^{-1}.$

2.10 SUMMARY

- The joint p.d.f. of the sample values is known as its likelihood function.
- In interval estimation, we find an interval in which the value of the parameter may lie with certain true probability.
- Interval which has shortest confidence interval *i.e.*, for which $T_2 - T_1$ least, such a confidence interval is called shortest confidence interval.
- The maximum likelihood method is to obtain that value of parameter (in terms of sample value) which maximize the likelihood function of the sample values.
- The estimators are asymtotically normal but not efficient in general.
- If a sufficient estimator exists, it is a function of the Maximum Likelihood Estimator.

2.11 GLOSSARY

- **Efficiency :** If T_1 and T_2 are two consistant estimator of θ then T_1 is said to be more efficient then T_2 if variance of $(T_1) \leq (T_2)$ $\forall n$.
- **Random Interval :** An interval in which end points are random variable.
- **Confidence Interval :** An interval in which parameter lies is called confidence interval.
- **Confidence co-efficient :** It is the measure of assurance that the parameter lies in confidence interval.

2.12 REVIEW QUESTIONS

1. Define consistant estimator. Show that the sample variance is a biased but consistant estimator of σ^2, when a random sample is taken from a N (μ, σ^2) population.
2. Define CAN and CAUN estimators.

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

   (*i*) $t_1 = \dfrac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$, (*ii*) $t_2 = \dfrac{x_1 + x_2}{2} + x_3$,

   (*iii*) $t_3 = \dfrac{2x_1 + x_2 + \lambda x_3}{3}$,

   where λ is such that t_3 is an unbiased estimator of μ. Find λ. Are t_1 and t_2 unbiased ? State giving reasons, the estimator which is the best among t_1, t_2 and t_3.
4. Describe the method of maximum likelihood estimation and state its important properties.
5. Find the MLE of the variance for a normal populations if mean is known.
6. State and explain the principle of maximum likelihood for estimating of population parameter.
7. Define likelihood function for a random sample drawn from a discrete population.
8. Write a note on interval estimation. Obtain 95% confidence interval for the mean of a normal distribution when σ is (*i*) known and (*ii*) unknown.
9. Describe the method of moments for estimating the parameters. What are the properties of the estimates obtained by this method ?

NOTES

2.13 FURTHER READINGS

Goon, A.M., Gupta, M.K. and Das Gupta. B. An outline of statistical theory. Volume II, the world press, Calcutta, 1980.

Huntsberger, D.V. and Billingsley, P. Elements of Statistical Inference, Allyn and Bacon, London, 1977.

Rao, C.R., Linear Statistical Inference and its Applications, Wiley Eastern, New Delhi, 2nd ed. 1973.

Walker, H.M. and Lev, J., Statistical Inference, Henry Holt Company New York, 1953.

Gupta, S.C. and Kapoor, V.K., fundamentals of Mathematical Statistics, Sultan Chand & Sons, New Delhi.

NOTES

UNIT-III

TESTING OF HYPOTHESIS

OBJECTIVES

After going through this unit, you should be able to:

- define statistical hypothesis
- explain optimum test under different situations
- describe likelihood ratio test

STRUCTURE

3.1 INTRODUCTION

The main problems in statistical inference can be broadly classified into two areas:

(*i*) The area of estimation of population parameter(s) and setting up of confidence intervals for them, *i.e.*, the area of *point and interval estimation* and

(*ii*) Tests of statistical hypothesis.

The first topic has already been discussed in Unit II. In this unit we shall discuss:

(*a*) The theory of testing of hypothesis initiated by J. Neyman and E.S. Pearson,

(*b*) Sequential analysis propounded by A. Wald.

In Neyman-Pearson theory, we use statistical methods to arrive at decisions in certain situations where there is lack of certainty on the basis of a sample whose size is fixed in advance while in Wald's sequential theory the sample size is not fixed but it regarded as a random variable. Before taking up a detailed discussion of the topics in (*a*) and (*b*), we shall explain below certain concepts which are of fundamental importance.

3.2 STATISTICAL HYPOTHESIS—SIMPLE AND COMPOSITE

A statistical hypothesis is some statement or assertion about a population or equivalent about the probability distribution characterising a population, which we want to verify on the basis of information available from a sample. If the statistical hypothesis specifies the population completely then it is termed as a *simple statistical hypothesis* otherwise it is called a *composite statistical hypothesis.*

For example, if $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal population with mean μ and variance σ^2, then the hypothesis $H_0 : \mu = \mu_0, \sigma^2 = \sigma_0^2$ is a simple hypothesis, whereas each of the following hypotheses is a composite hypothesis:

(*i*) $\mu = \mu_0$, (*ii*) $\sigma^2 = \sigma_0^2$,

(*iii*) $\mu < \mu_0, \sigma^2 = \sigma_0^2$, (*iv*) $\mu > \mu_0, \sigma^2 = \sigma_0^2$

(*v*) $\mu = \mu_0, \sigma^2 < \sigma_0^2$, (*vi*) $\mu = \mu_0, \sigma^2 > \sigma_0^2$,

(*vii*) $\mu < \mu_0, \sigma^2 > \sigma_0^2$.

A hypothesis which does not specify completely 'v' parameters of a population is termed as a *composite hypothesis with* v *degrees of freedom.*

3.2.1 Test of a Statistical Hypothesis

A test of a statistical hypothesis is a two-action decision problem after the experimental sample values have been obtained, the two-actions being the acceptance or rejection of the hypothesis under consideration.

3.2.2 Null Hypothesis

In hypothesis testing, a statistician or decision-maker should not be motivated by prospects of profit or loss resulting from the acceptance or rejection of the hypothesis. He should be completely impartial and should have no brief for any party or company nor should be allow his personal views to influence the decision. *Much, therefore, depends upon how the hypothesis is framed*. For example, let us consider the 'light-bulbs' problem. Let us suppose that the bulbs manufactured under some standard manufacturing process have an average life of μ hours and it is proposed to test a new procedure for manufacturing light bulbs. Thus, we have two populations of bulbs, those manufactured by standard process and those manufactured by the new process. In this problem the following three hypotheses may be set up:

(*i*) New process is better than standard process.

(*ii*) New process is inferior to standard process.

(*iii*) There is no difference between the two processes.

The first two statements appear to be biased since they reflect a preferential attitude to one or the other of the two processes. Hence the best course is to adopt *the hypothesis of no difference*, as stated in (*iii*). This suggests that the *statistician should take up the neutral or null attitude regarding the outcome of the test*. His attitude should be on the null or zero line in which the experimental data has the due importance and complete say in the matter. *This neutral or non-committal attitude of the statistician or decision-maker before the sample observations are taken is the keynote of the null hypothesis.*

Thus in the above example of light bulbs if μ_0 is the mean life (in hours) of the bulbs manufactured by the new process then the null hypothesis which is usually denoted by H_0, can be stated as follows: $H_0 : \mu = \mu_0$.

As another example let us suppose that two different concerns manufacture drugs for inducing sleep, drug *A* manufactured by first concern and drug B manufactured by second concern. Each company claims that its drug is superior to that of the other and it is desired to test which is a superior drug *A* or B? To formulate the statistical hypothesis let *X* be a random variable which denotes the additional hours of sleep gained by an individual when drug *A* is given and let the random variable Y denote the additional hours of sleep gained when drug *B* is used. Let us suppose that X and *Y* follow the probability distributions with means μ_X and μ_Y respectively. Here our *null hypothesis would be that there is no difference between the effects of two drugs*. Symbolically, $H_0 : \mu_X = \mu_Y$.

NOTES

3.2.3 Alternative Hypothesis

It is desirable to state what is called an *alternative hypothesis* in respect of every statistical hypothesis being tested because the acceptance or rejection of null hypothesis is meaningful only when it is being tested against a rival hypothesis which should rather be explicitly mentioned. *Alternative hypothesis* is usually denoted by H_1. For example, in the example of light bulbs, alternative hypothesis could be $H_1 : \mu > \mu_0$ or $\mu < \mu_0$ or $\mu \neq \mu_0$. In the example of drugs, the alternative hypothesis could be $H_1 : \mu_X > \mu_Y$ or $\mu_X < \mu_Y$ or $\mu_X \neq \mu_Y$.

In both the cases, the first two of the alternative hypothesis give rise to what are called *'one tailed'* tests and the third alternative hypothesis results in *'two tailed'* tests.

3.2.3.1 Important Remarks

1. In the formulation of a testing problem and devising a 'test of hypothesis' the roles of H_0 and H_1 are not at all symmetric. In order to decide which one of the two hypothesis should be taken as null hypothesis H_0 and which one as alternative hypothesis H_1, the intrinsic difference between the roles and the implifications of these two terms should be clearly understood.
2. If a particular problems cannot be stated as a test between two simple hypothesis, *i.e.*, simple null hypothesis against a simple alternative hypothesis, then the next best alternative is to formulate the problem as the test of a simple null hypothesis against a composite alternative hypothesis. In other words, one should try to structure the problem so that null hypothesis is simple rather than composite.
3. Keeping in mind the potential looses due to wrong decisions (which may or may not be measured in terms of money), the decision maker is somewhat conservative in holding the null hypothesis as true unless there is a strong evidence from the experimental sample observations that it is false. To him, the *consequences of wrongly rejecting a null hypothesis seem to be more severe than those of wrongly accepting it*. In most of the cases, the statistical hypothesis is in the form of a claim that a particular product or product process is superior to some existing standard. The null hypothesis H_0 in this case is that there is no difference between the new product or productions process and the existing standard. In other words, null hypothesis nullifies this claim. The rejection of the null hypothesis wrongly which amounts to the acceptance of claim wrongly involves huge amount of pocket expenses towards a substantive overhaul of the existing set-up. The resulting loss is comparatively regarded as more serious than the opportunity loss in wrongly accepting H_0 which amounts to wrongly rejecting the claim, *i.e.*, in sticking to the less efficient existing standard. In the light-bulbs problem discussed earlier, suppose the research division of the concern, on the basis of the limited experimentation, claims that its brand is more effective than that manufactured by standard process. If in fact, the brand fails to be more

effective the loss incurred by the concern due to an immediate *obsolescence* of the product, decline of the concern's image, etc., will be quite serious. On the other hand, the failure to bring out a superior brand in the market is an opportunity loss and is not a consideration to be as serious as the other loss.

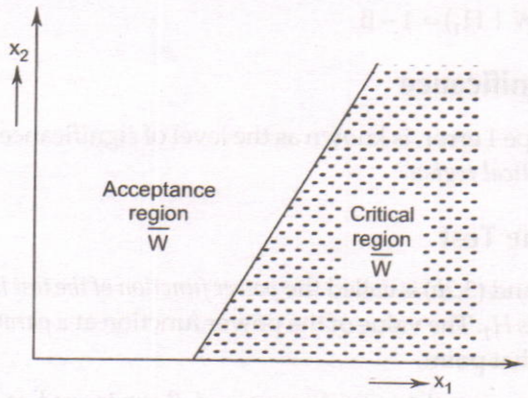
3.2.4 Critical Region

NOTES

Let $x_1, x_2, \ldots, x_n$ be the sample observations denoted by O. All the values of O will be aggregate of a sample and they constitute a space, called the *sample space*, which is denoted by S.

Since the sample values $x_1, x_2, \ldots, x_n$ can be taken as a point in n-dimensional space, we specify some region of the n-dimensional space and see whether this point lies within this regions or outside this region. We divide the whole sample space S into two disjoint parts W and S – W or $\overline{W}$ or W′. The null hypothesis H_0 is rejected if the observed sample point falls in W and if it falls in W′ we reject H_1 and accept H_0. *The region of rejection of H_0 when H_0 is true is that region of the outcome set where H_0 is rejected if the sample point falls in that region and is called critical region.* Evidently, the size of the critical region is α, the probability of committing type 1 error (discussed below).

Suppose if the test is based on a sample of size 2, then the outcome set or the sample space is the first quadrant in a two-dimensional space and a test criterion will enable us to separate our outcome set into two complementary subsets, W and $\overline{W}$. If the sample point falls in the subset W, H_0 is rejected, otherwise H_0 is accepted. This is shown in the adjoining diagram:



3.2.5 Two Types of Errors

The decision to accept or reject the null hypothesis H_0 is made on the basis of the information supplied by the observed sample observations. The conclusion drawn on the basis of a particular sample may not always be true in respect of the population. The four possible situations that arise in any test procedure are given in the following table.

		Decision From Sample	
		Reject H_0	**Accept H_0**
True State	H_0 True	Wrong (Type 1 Error)	Correct
	H_0 False (H_1 True)	Correct	Wrong (Type II Error)

From the above table it is obvious that in any testing problem we are liable to commit two types of errors.

Errors of Type I and Type II. The error of rejecting H_0 (accepting H_1) when H_0 is true is called *Type 1 error* and the error of accepting H_0 when H_0 is false (H_1 is true) is called

Type II error. The probabilities of type I and type II errors are denoted by α and β respectively. Thus

$$\alpha = \text{Probability of type I error}$$
$$= \text{Probability of rejecting } H_0 \text{ when } H_0 \text{ is true.}$$
$$\beta = \text{Probability of type II error}$$
$$= \text{Probability of accepting } H_0 \text{ when } H_0 \text{ is false.}$$

Symbolically:

NOTES

$$P(x \in W \mid H_0) = \alpha, \text{ where } x = (x_1, x_2, \ldots, x_n) \Rightarrow \int_W L_0\, dx = \alpha \quad \ldots(3.1)$$

where L_0 is the likelihood function of the sample observations under H_0 and $\int dx$ represents the n-fold integral $\int\int \ldots \int dx_1\, dx_2 \ldots dx_n$.

Again $$P(x \in \overline{W} \mid H_1) = \beta \Rightarrow \int_{\overline{W}} L_1\, dx = \beta \quad \ldots(3.2)$$

where L_1 is the likelihood function of the sample observations under H_1. Since

$$\int_W L_1\, dx + \int_{\overline{W}} L_1\, dx = 1,$$

we get $$\int_W L_1\, dx = 1 - \int_{\overline{W}} L_1\, dx = 1 - \beta \quad \ldots(3.2a)$$

$$\Rightarrow \quad P(x \in W \mid H_1) = 1 - \beta \quad \ldots(3.2b)$$

3.2.6 Level of Significance

α, the probability of type 1 error, is known as the level of significance of the test. It is also called the *size of the critical region*.

3.2.7 Power of the Test

$1 - \beta$, defined in (3.2*a*) and (3.2*b*) is called the *power function of the test hypothesis H_0 against the alternative hypothesis H_1*. The value of the power function at a *parameter* point is called the *power of the test* at that point.

Remark 3.1. *In quality control terminology, α and β are termed as producer's risk and consumer's risk, respectively.*

Remark 3.2. *An ideal test would be the one which properly keeps under control both the types of errors. But since the commission of an error of either type is a random variable, equivalently an ideal test should minimize the probability of both the types of errors, viz., α and β. But unfortunately, for a fixed sample size n, α and β are so related (like producer's and consumer's risk in sampling inspection plans), that the reduction in one results in an increase in the other. Consequently, the simultaneous minimizing of both the errors is not possible. Since type I error is deemed to be more serious than the type II error the usual practice is to control α at a predetermined low level and subject to this constraint on the probabilities of type I error, choose a test which minimizes β or maximizes the power function $1 - \beta$. Generally, we choose $\alpha = 0.05$ or 0.01.*

3.3 STEPS IN SOLVING TESTING OF HYPOTHESIS PROBLEM

The major steps involved in the solution of a 'testing of hypothesis' problem may be outlined as follows:

1. Explicit knowledge of the nature of the population distribution and the parameter(s) of interest, *i.e.*, the parameter(s) about which the hypotheses are set up.
2. Setting up of null hypothesis H_0 and the alternative hypothesis H_1 in terms of the range of the parameter values each one embodies.
3. The choice of a suitable statistic $t = t(x_1, x_2, ..., x_n)$ called the *test statistic*, which will best reflect upon the probability of H_0 and H_1.
4. Partitioning the set of possible values of the test statistic t into two disjoint sets W (called the *rejection region* or *critical region*) and $\overline{W}$ (called the *acceptance region*) and framing the following test:
   (*i*) Reject H_0 (*i.e.*, accept H_1) if the value of t falls in W.
   (*ii*) Accept H_0 if the value of t falls $\overline{W}$.
5. After framing the above test, obtain experimental sample observations, compute the appropriate test statistic and take action accordingly.

3.4 OPTIMUM TEST UNDER DIFFERENT SITUATIONS

In any testing problems the first two steps, *viz.*, the form of the population distribution, the parameter(s) of interest and the framing of H_0 and H_1 should be obvious from the description of the problem. The most crucial step is the choice of the *best test*, *i.e.*, the best static '*t*' and the critical region W where *by best test we mean one which in addition to controlling α at any desired low level has the minimum type II error β or maximum power $1-\beta$, compared to β of all other tests having this α.* This leads to the following definition.

3.4.1 Most Powerful Test (MP Test)

Let us consider the problem of testing a simple hypothesis : $H_0 : \theta = \theta_0$

against a simple alternative hypothesis: $H_1 : \theta = \theta_1$

Definition 3.1. *The critical region W is the most powerful (MP) critical region of size α (and the corresponding test a most powerful test of level α) for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ if*

$$P(x \in W \mid H_0) = \int_W L_0 \, dx = \alpha \qquad \ldots(3.3)$$

and

$$P(x \in W \mid H_1) \geq P(x \in W_1 \mid H_1) \qquad \ldots(3.3a)$$

for every other critical region W_1 satisfying (3.3).

3.4.2 Uniformly Most Powerful Test (UMP Test)

Let us now take up the case of testing a simple null hypothesis against a composite alternative hypothesis, *e.g.*, of testing $H_0 : \theta = \theta_0$

against the alternative $H_1 : \theta \neq \theta_0$

In such a case, for a predetermined α, the best test for H_0 is called the *uniformly most powerful test of level α.*

Definition 3.2. *The region W is called uniformly most powerful (UMP) critical region of size α [and the corresponding test as uniformly most powerful (UMP) test of level α] for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ i.e., $H_1 : \theta = \theta_1 \neq \theta_0$ if*

$$P(x \in W \mid H_0) = \int_W L_0 \, dx = \alpha \qquad \ldots(3.4)$$

and

$$P(x \in W \mid H_1) \geq P(x \in W_1 \mid H_1) \text{ for all } \theta \neq \theta_0 \qquad \ldots(3.4a)$$

whatever the region W_1 satisfying (3.4).

3.5 NEYMAN J. AND PEARSON, E.S. LEMMA

This Lemma provides the most powerful test of simple hypothesis against a simple alternative hypothesis. The theorem, known as Neyman-Pearson Lemma, will be proved for density function $f(x, \theta)$ of a single continuous variate and a single parameter. However, by regarding x and θ as vectors, the proof can be easily generalized for any number of random variables $x_1, x_2, \ldots, x_n$ and any number of parameters $\theta_1, \theta_2, \ldots, \theta_k$. The variables $x_1, x_2, \ldots, x_n$ occurring in this theorem understood to represent a random sample of size n from the population where density function is $f(x, \theta)$. The lemma is concerned with a simple hypothesis $H_0 : \theta_0$ and a simple alternative $H_1 : \theta = \theta_1$.

NOTES

Theorem 3.1. (Neyman-Pearson Lemma) *Let $k > 0$, be a constant and W be a critical region of S such that*

$$W = \left\{ x \in S : \frac{f(x, \theta_1)}{f(x, \theta_0)} > k \right\}$$

$$\Rightarrow \qquad W = \left\{ x \in S : \frac{L_1}{L_0} > k \right\} \qquad \ldots(3.5)$$

and

$$\overline{W} = \left\{ x \in S : \frac{L_1}{L_0} < k \right\} \qquad \ldots(3.5a)$$

where L_0 and L_1 are the likelihood functions of the sample observation $x = (x_1, x_2, \ldots, x_n)$ under H_0 and H_1 respectively. Then W is the most powerful critical region of the test hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$.

Proof: We are given

$$P(x \in W \mid H_0) = \int_W L_0 \, dx = \alpha \qquad \ldots(3.6)$$

The power of the region is

$$P(x \in W \mid H_1) = \int_W L_1 \, dx = 1 - \beta, \text{ (say)}. \qquad \ldots(3.6a)$$

In order to establish the lemma, we have to prove that there exists no other critical region, of size less than or equal to α, which is more powerful than W. Let W_1 be another critical region of size $\alpha_1 \le \alpha$ and power $1 - \beta_1$ so that we have

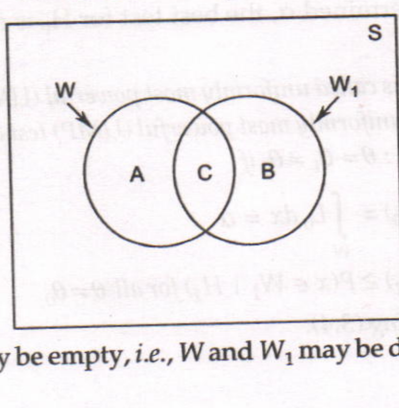
$$P(x \in W_1 \mid H_0) = \int_{W_1} L_0 \, dx = \alpha_1 \qquad \ldots(3.7)$$

and

$$P(x \in W_1 \mid H_1) = \int_{W_1} L_1 \, dx = 1 - \beta_1 \qquad \ldots(3.7a)$$

Now we have to prove that $1 - \beta \ge 1 - \beta_1$

Let $\qquad W = A \cup C$ and $W_1 = B \cup C$



(C may be empty, *i.e.*, W and W_1 may be disjoint).

If $\alpha_1 \le \alpha$, we have

$$\int_{W_1} L_0\, dx \le \int_{W} L_0\, dx$$

$$\Rightarrow \quad \int_{B\cup C} L_0\, dx \le \int_{A\cup C} L_0\, dx$$

$$\Rightarrow \quad \int_{B} L_0\, dx \le \int_{A} L_0\, dx$$

$$\Rightarrow \quad \int_{A} L_0\, dx \ge \int_{B} L_0\, dx \qquad \ldots(3.8)$$

Since $A \subset W$,

$$\Rightarrow \quad \int_{A} L_1\, dx > k\int_{A} L_0\, dx \ge k\int_{B} L_0\, dx \qquad \ldots(3.8a)$$

Also [3.5(*a*)] implies

$$\frac{L_1}{L_0} \le k \,\forall\, x \in \overline{W}$$

$$\Rightarrow \quad \int_{\overline{W}} L_1\, dx \le k\int_{\overline{W}} L_0\, dx$$

This result also holds for any subset of $\overline{W}$, say $\overline{W} \cap W_1 = B$. Hence

$$\int_{B} L_1\, dx \le k\int_{B} L_0\, dx \le \int_{A} L_1\, dx \qquad \text{[From (3.8a)]}$$

Adding $\int_{C} L_1\, dx$ to both sides, we get

$$\int_{W_1} L_1\, dx \le \int_{W} L_1\, dx \quad \Rightarrow \quad 1-\beta \ge 1-\beta_1$$

Hence the Lemma.

Remark 3.3. *Let W defined in (3.5) of the above theorem be the most powerful critical region of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, and let it be independent of $\theta_1 \in \Theta_1 = \Theta - \Theta_0$, where Θ_0 is the parameter space under H_0. Then we say that C.R. W is the UMP CR of size α for testing : $H_0 : \theta = \theta_0$, against $H_1 : \theta \in \Theta_1$.*

3.5.1 Unbiased Test and Unbiased Critical Region

Let us consider the testing of $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$: *The critical region W and consequently the test based on it is said to be unbiased if the power of the test exceed the size of the critical region, i.e., if*

$$\text{Power of the test} \ge \text{Size of the C.R.} \qquad \ldots(3.9)$$

$$\Rightarrow \quad 1-\beta \ge \alpha$$

$$\Rightarrow \quad P_{\theta_1}(W) \ge P_{\theta_0}(W)$$

$$\Rightarrow \quad P[x : x \in W \mid H_1] \ge P[x : x \in W \mid H_0] \qquad \ldots(3.9a)$$

In other words, the critical region W is said to be unbiased if

$$P_{\theta}(W) \ge P_{\theta_0}(W), \forall \theta (\ne \theta_0) \in \Theta \qquad \ldots(3.9b)$$

Theorem 3.2. *Every most powerful (MP) or uniformly most powerful (UMP) critical region (CR) is necessarily unbiased.*

(a) *If W be an MPCR of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, then it is necessarily unbiased.*

(b) *Similarly if W be UMPCR of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta_1$, then it is also unbiased.*

Proof: (*a*) Since W is the MPCR of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, by Neyman-Pearson Lemma, we have; for $\forall\, k > 0$,

$$W = \{x : L(x, \theta_1) \geq k\, L(x, \theta_0)\} = \{x : L_1 \geq k\, L_0\}$$

NOTES

and $$W' = \{x : L(x, \theta_1) < k\, L(x, \theta_0)\} = \{x : L_1 < k\, L_0\},$$

where k is determined so that the size of the test is α *i.e.,*

$$P_{\theta_0}(W) = P[x \in W \mid H_0] = \int_W L_0\, dx = \alpha \qquad \text{...(3.10)}$$

To prove that W is unbiased, we have to show that:

Power of $$W \geq \alpha \quad \textit{i.e.,} \quad P_{\theta_1}(W) \geq \alpha \qquad \text{...(3.11)}$$

We have: $$P_{\theta_1}(W) = \int_W L_1\, dx \geq k \int_W L_0\, dx = k\alpha$$

$$[\because \text{ On W}, L_1 \geq k\, L_0 \text{ and Using (3.10)}]$$

i.e., $$P_{\theta_1}(W) \geq k\alpha, \forall\, k > 0 \qquad \text{...(3.12)}$$

Also

$$1 - P_{\theta_1}(W) = 1 - P(x \in W \mid H_1) = P(x \in W' \mid H_1) = \int_{W'} L_1\, dx$$

$$< k \int_{W'} L_0\, dx = k\, P(x : x \in W' \mid H_0) \quad [\because \text{ On } W', L_1 < k\, L_0]$$

$$= k[1 - P(x : x \in W \mid H_0)] = k(1 - \alpha) \qquad [\text{Using (3.10)}]$$

i.e., $$1 - P_{\theta_1}(W) \leq k(1 - \alpha), \forall\, k > 0 \qquad \text{...(3.13)}$$

Case (*i*) $k \geq 1$, if $k \leq 1$ then from (*iii*), we get

$$P_{\theta_1}(W) \geq k\alpha \geq \alpha$$

$\Rightarrow$ W is unbiased CR.

Case (*ii*) $0 < k < 1$. If $0 < k < 1$, then from (*iv*), we get:

$$1 - P_{\theta_1}(W) < 1 - a \;\Rightarrow\; P_{\theta_1}(W) \geq \alpha \;\Rightarrow\; \text{W is unbiased C.R.}$$

Hence MP critical region is unbiased.

(*b*) If W is UMPCR of size a then also the above proof holds if for θ_1 we write θ, such that $\theta \in \Theta_1$. So we have

$$P_\theta(W) > \alpha, \forall\, \theta \in \Theta_1 \;\Rightarrow\; \text{W is unbiased CR.}$$

3.5.2 Optimum Regions and Sufficient Statistics

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population with p.m.f. or p.d.f. $f(x, \theta)$, where the parameter θ may be a vector. Let T be a sufficient statistic for θ. Then by Factorization Theorem,

$$L(x, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = g_\theta(t(x)) \cdot h(x) \qquad \text{...(3.14)}$$

where $g_\theta(t(x))$ is the marginal distribution of the statistic $T = t(x)$.

By Neyman-Pearson Lemma, the MPCR for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is given by:

$$W = \{x : L(x, \theta_1) \geq k\, L(x, \theta_0)\}, \forall\, k > 0 \qquad \text{...(3.15)}$$

From (3.14) and (4.15), we get

$$W = \{x : g_{\theta_1}(t(x)) \cdot h(x) \geq k \cdot g_{\theta_0}(t(x)) \cdot h(x)\}, \ \forall k > 0$$

$$= \{x : g_{\theta_1}(t(x)) \geq k \cdot g_{\theta_0}(t(x))\}, \ \forall k > 0$$

Hence if $T = t(x)$ is sufficient statistic for θ then the MPCR for the test may be defined in terms of the marginal distribution of $T = t(x)$, rather than the joint distribution of $x_1, x_2, \ldots, x_n$.

Example 3.1. *Given the frequency function:*

$$f(x, \theta) = \begin{cases} \dfrac{1}{\theta}, 0 \leq x \leq \theta \\ 0, \text{elsewhere} \end{cases}$$

and that you are testing the null hypothesis $H_0 : \theta = 1$ against $H_1 : \theta = 2$, by means of a single observed value of x. What would be the sizes of the type 1 and type II errors, if you choose the interval (i) $0.5 \leq x$, (ii) $1 \leq x \leq 1.5$ as the critical regions? Also obtain the power function of the test.

Solution. Here we want to test $H_0 : \theta = 1$, against $H_1 : \theta = 2$

(*i*) Here $W = \{x : 0.5 \leq x\} = \{x : x \geq 0.5\}$

and $\overline{W} = \{x : x \leq 0.5\}$

$$\alpha = P\{x \in W \mid H_0\} = P(x \geq 0.5 \mid \theta = 1\} = P\{0.5 \leq x \leq \theta \mid \theta = 1\}$$

$$= P\{0.5 \leq x \leq 1 \mid \theta = 1\} = \int_{0.5}^{1} [f(x, \theta)]_{\theta=1} dx = \int_{0.5}^{1} 1 \cdot dx = 0.5$$

Similarly, $\beta = P\{x \in \overline{W} \mid H_1\} = P(x \leq 0.5 \mid \theta = 2\}$

$$= \int_{0}^{0.5} [f(x, \theta)]_{\theta=2} dx = \int_{0}^{0.5} \frac{1}{2} dx = 0.25$$

Thus the sizes of type I and type II errors are respectively $\alpha = 0.5$ and $\beta = 0.25$ and power function of the test $= 1 - \beta = 0.75$

(*ii*) $W = \{x : 1 \leq x \leq 1.5\}$

$$\alpha = P\{x \in W \mid \theta = 1\} = \int_{1}^{1.5} [f(x, \theta)]_{\theta=1} dx = 0,$$

since under $H_0 : \theta = 1, f(x, \theta) = 0$, for $1 \leq x \leq 1.5$.

$$\beta = P\{x \in \overline{W} \mid \theta = 2\} = 1 - P\{x \in W \mid \theta = 2\}$$

$$= 1 - \int_{1}^{1.5} [f(x, \theta)]_{\theta=2} dx = 1 - \left|\frac{x}{2}\right|_{1}^{1.5} = 0.75$$

$\therefore$ Power Function $= 1 - \beta = 1 - 0.75 = 0.25$

Example 3.2. *If $x \geq 1$ is the critical region for testing $H_0 : \theta = 2$ against the alternative $\theta = 1$, on the basis of the single observation from the population.*

$$f(x, \theta) = \theta \exp(-\theta x), 0 \leq x < \infty,$$

obtain the values of type I and type II errors.

Solution. Here $W = \{x : x \geq 1\}$ and $\overline{W} = \{x : x < 1\}$ and $H_0 : \theta = 2, H_1 : \theta = 1$

$$\alpha = \text{Size of Type I error} = P[x \in W \mid H_0] = P[x \geq 1 \mid \theta = 2]$$

Using Neyman-Pearson Lemma, B.C.R. for K > 0, is given by

$$\frac{\beta_1^n \exp\left\{-\beta_1 \sum_{i=1}^{n}(x_i-\gamma_1)\right\}}{\beta_0^n \exp\left\{-\beta_0 \sum_{i=1}^{n}(x_i-\gamma_0)\right\}} \geq k$$

NOTES

$$\Rightarrow \quad \left(\frac{\beta_1}{\beta_0}\right)^n \exp\left\{-\beta_1 \sum_{i=1}^{n}(x_i-\gamma_1)+\beta_0 \sum_{i=1}^{n}(x_i-\gamma_0)\right\} \geq k$$

$$\Rightarrow \quad \left(\frac{\beta_1}{\beta_0}\right)^n \exp\left[-\beta_1\, n(\bar{x}-\gamma_1)+\beta_0\, n(\bar{x}-\gamma_0)\right] \geq k$$

$$\Rightarrow \quad n\log(\beta_1/\beta_0) - n\bar{x}(\beta_1-\beta_0) + n\beta_1\gamma_1 - n\beta_0\gamma_0 \geq \log k$$

(since $\log x$ is an increasing function of x).

$$\Rightarrow \quad \bar{x}(\beta_1-\beta_0) \leq \left\{\gamma_1\beta_1 - \gamma_0\beta_0 - \frac{1}{n}\log k + \log\left(\frac{\beta_1}{\beta_0}\right)\right\}$$

$$\therefore \quad \bar{x} \leq \frac{1}{\beta_1-\beta_0}\left\{\gamma_1\beta_1 - \gamma_0\beta_0 - \frac{1}{n}\log k + \log\left(\frac{\beta_1}{\beta_0}\right)\right\} \text{ provided } \beta_1 > \beta_0.$$

Example 3.3. *Examine whether a best critical region exists for testing the m hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$ for the parameter θ of the distribution:*

$$f(x, \theta) = \frac{1+\theta}{(x+\theta)^2},\ 1 \leq x < \infty$$

Solution.

$$\prod_{i=1}^{n} f(x_i, \theta) = (1+\theta)^n \prod_{i=1}^{n} \frac{1}{(x_i+\theta)^2}$$

By Neyman-Pearson Lemma, the B.C.R. for $k > 0$, is given by

$$(1+\theta_1)^n \prod_{i=1}^{n} \frac{1}{(x_i+\theta_1)^2} \geq k(1+\theta_0)^n \prod_{i=1}^{n} \frac{1}{(x_i+\theta_0)^2}$$

$$\Rightarrow \quad n\log(1+\theta_1) - 2\sum_{i=1}^{n}\log(x_i+\theta_1) \geq \log k + n\log(1+\theta_0) - 2\sum_{i=1}^{n}\log(x_i+\theta_0)$$

$$\Rightarrow \quad 2\sum_{i=1}^{n}\log\left(\frac{x_i+\theta_0}{x_i+\theta_1}\right) \geq \log k + n\log\left(\frac{1+\theta_0}{1+\theta_1}\right)$$

Thus the test criterion is $\sum_{i=1}^{n}\log\left(\frac{x_i+\theta_0}{x_i+\theta_1}\right)$, which cannot be put in the form of a function of the sample observations, not depending on the hypothesis. Hence no. B.C.R. exists in this case.

3.6 LIKELIHOOD RATIO TEST

Neyman-Pearson Lemma based on the magnitude of the ratio of two probability density functions provides best test for testing simple hypothesis against simple alternative hypothesis. The best test in any given situation depends on the nature of the population distribution and the form of the alternative hypothesis being considered. In this section we shall discuss a general method of test construction called the *Likelihood Ratio* (L.R.). Test introduced by Neyman and Pearson for testing a hypothesis, simple or composite, against a simple or composite alternative hypothesis. This test is related to the maximum likelihood estimates.

Before defining the test, we give below some notations and terminology.

Parameter Space. Let us consider a random variable X with p.d.f. $f(x, \theta)$. In most common applications, though not always, the functional form of the population distribution is assumed to be known except for the value of some unknown parameter(*s*) θ which may take any value on a set Θ. This is expressed by writing the p.d.f. in the form $f(x, \theta), \theta \in \Theta$. The set Θ, which is the set of all possible values of θ is called the *parameter space*. Such a situation gives rise not to one probability distribution but a family of probability distributions which we write as $\{f(x, \theta) = \theta \in \Theta\}$. For example, if $X \sim N(\mu, \sigma^2)$, then the parameter space is:

$$\Theta = \{(\mu, \sigma^2): -\infty < \mu < \infty, 0 < \sigma < \infty\}$$

In particular, for $\sigma^2 = 1$, the family of probability distributions is given by

$$\{N(\mu, 1); \mu \in \Theta\}, \text{ where } \Theta = \{\mu : -\infty < \mu < \infty\}$$

In the following discussion we shall consider a general family of distributions:

$$\{f(x : \theta_1, \theta_2, \ldots, \theta_k) : \theta_i \in \Theta, i = 1, 2, \ldots, k\}$$

The null hypothesis H_0 will state that the parameters belong to some subspace Θ_0 of the parameter space Θ.

Let $x_1, x_2, \ldots, x_n$ be a random sample of size $n > 1$ from a population with p.d.f. $f(x, \theta_1, \theta_2, \ldots, \theta_k)$, where Θ, the parameter space is the totality of all points that $(\theta_1, \theta_2, \ldots, \theta_k)$ can assume. We want to test the null hypothesis:

$$H_0 : (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta_0$$

against all alternative hypothesis of the type:

$$H_1 : (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta - \Theta_0$$

The likelihood function of the sample observations is given by

$$L = \prod_{i=1}^{n} f(x_i ; \theta_1, \theta_2, \ldots, \theta_k) \qquad \ldots(3.16)$$

According to the principle of maximum likelihood, the likelihood equation for estimating any parameter θ_i is given by

$$\frac{\partial L}{\partial \theta_i} = 0, \ (i = 1, 2, \ldots, k) \qquad \ldots(3.17)$$

Using (3.17), we can obtain the maximum likelihood estimates for the parameters $(\theta_1, \theta_2, \ldots, \theta_n)$ as they are allowed to vary over the parameter space Θ and the subspace Θ_0. Substituting these estimates in (3.16), we obtain the maximum values of the likelihood function for variation of the parameter in Θ and Θ_0 respectively. Then the criterion for the likelihood ratio test is defined as the quotient of these two maxima and is given by

$$\lambda = \lambda(x_1, x_2, \ldots, x_n) = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\underset{\theta \in \Theta_0}{Sup}\, L(x, \theta)}{\underset{\theta \in \Theta}{Sup}\, L(x, \theta)} \qquad \ldots(3.18)$$

where $L(\hat{\Theta}_0)$ and $L(\hat{\Theta})$ are the maxima of the likelihood function (3.16) with respect to the parameters in the regions Θ_0 and Θ respectively.

The quantity λ is a function of the sample observations only and does not involve parameters. Thus λ being a function of the random variables, is also a random variable. Obvious $\lambda > 0$. Further

$$\Theta_0 \subset \Theta \Rightarrow L(\Theta_0) \le L(\Theta) \Rightarrow \lambda \le 1$$

Hence, we get $0 \le \lambda \le 1$ ...(3.19)

The critical region for testing H_0 (against H_1) is an interval

$$0 < \lambda < \lambda_0 \quad \text{...(3.20)}$$

NOTES

where λ_0 is some number (< 1) determined by the distribution of λ and the desired probability of type I error, *i.e.*, λ_0 is given by the equation:

$$P(\lambda < \lambda_0 \mid H_0) = \alpha \quad \text{...(3.21)}$$

For example, if $g(.)$ is the p.d.f. of λ then λ_0 is determined from the equation:

$$\int_0^{\lambda_0} g(\lambda \mid H_0)\, d\lambda = \alpha \quad \text{...(3.21a)}$$

A test that has critical region defined in (3.20) and (3.21) is a *likelihood ratio test* for testing H_0.

Remark 3.4. *Equations (3.20) and (3.21) define the critical region for testing the hypothesis H_0 by the likelihood ratio test. Suppose that the distribution of λ is not known but the distribution of some function of λ is known, then this knowledge can be utilized as given in the following theorem.*

Theorem 3.3. *If λ is the likelihood ratio for testing a simple hypothesis H_0 and if $U = \phi(\lambda)$ is a monotonic increasing (decreasing) function of λ then the test based on U is equivalent to the likelihood ratio test. The critical region for the test based on U is:*

$$\phi(0) < U < \phi(\lambda_0), \mid [\phi(\lambda_0) < U < \phi(0)] \quad \text{...(3.22)}$$

Proof: The critical region for the likelihood ratio test is given by $0 < \lambda < \lambda_0$, where λ_0 is determined by

$$\int_0^{\lambda_0} g(\lambda \mid H_0)\, d\lambda = \alpha \quad \text{...(3.23)}$$

Let $U = \phi(\lambda)$ be a monotonically increasing function of λ. Then (3.23) gives

$$\alpha = \int_0^{\lambda_0} g(\lambda \mid H_0)\, d\lambda = \int_{\phi(0)}^{\phi(\lambda_0)} h(u \mid H_0)\, du$$

where $h(u \mid H_0)$ is the p.d.f of U when H_0 is true. Here the critical region $0 < \lambda < \lambda_0$ transforms to $\phi(0) < U < \phi(\lambda_0)$. However if $U = \phi(\lambda)$ is a monotonic decreasing function of λ, then the inequalities are reversed and we get the critical region as $\phi(\lambda_0) < U < \phi(0)$.

2. If we are testing a simple null hypothesis H_0 then there is a unique distribution determined for λ. But if H_0 is composite, then the distribution of λ may or may not be unique. In such a case the distribution of λ may possibly be different for different parameter points in Θ_0 and then λ_0 is to be chosen such that

$$\int_0^{\lambda_0} g(\lambda \mid H_0)\, d\lambda \le \alpha \quad \text{...(3.24)}$$

for all values of the parameters in Θ_0.

However, if we are dealing with large samples, a fairly satisfactory situation to this testing of hypothesis problem exists as stated (without proof) in the following theorem.

Theorem 3.4. *Let $x_1, x_2, \ldots, x_n$ be a random sample from a population with p.d.f. $f(x; \theta_1, \theta_2, \ldots, \theta_k)$ where the parameter space Θ is k-dimensional. Suppose we want to test the composite hypothesis*

$$H_0 : \theta_1 = \theta_1', \theta_2 = \theta_2', \ldots, \theta_r = \theta_r' ; r < k$$

where $\theta_1', \theta_2', \ldots, \theta_r'$ are specified numbers. When H_0 is true, $-2 \log_e \lambda$ is asymptotically distributed as chi-square with r degrees of freedom, i.e., under H_0,

$$-2 \log \lambda \sim \chi_{(r)^2}, \text{ if } n \text{ is large.} \qquad \ldots(3.25)$$

Since $0 \leq \lambda \leq 1$, $-2 \log_e \lambda$ is an increasing function of λ and approaches infinity when $\lambda \rightarrow 0$, the critical region for $-2 \log \lambda$ being the right hand tail of the chi-square distribution. Thus at the level of significance 'α', the test may be given as follows:

Reject H_0 if $-2 \log_e \lambda > \chi_{(r)^2} (\alpha)$

where $\chi_{(r)^2} (\alpha)$ is the upper α-point of the chi-square distribution with r.d.f. given by:

$$P\left[\chi^2 > \chi_{(r)^2} (\alpha)\right] = \alpha,$$

otherwise H_0 may be accepted.

3.6.1 Properties of Likelihood Ratio Test

Likelihood Ratio (L.R.) test principle is an intuitive one. If we are testing a simple hypothesis H_0 against a simple alternative hypothesis H_1 then the LR principle leads to the same test as given by the Neyman-Pearson lemma. This suggests that LR test has some desirable properties, specially large sample properties.

In LR test, the probability of type I error is controlled by suitably choosing the cut off point λ_0. LR test is generally UMP if an UMP test at all exists. We state below, the two asymptotic properties of LR tests.

1. Under certain conditions, $-2 \log_e \lambda$ has an asymptotic chi-square distribution.
2. Under certain assumptions, LR test is consistent.

3.7 SUMMARY

- If the statistical hypothesis specifies the population completely then it is termed as a *simple statistical hypothesis* otherwise it is called a *composite statistical hypothesis.*
- A test of a statistical hypothesis is a two-action decision problem after the experimental sample values have been obtained, the two-actions being the acceptance or rejection of the hypothesis under consideration.
- *This neutral or non-committal attitude of the statistician or decision-maker before the sample observations are taken is the keynote of the null hypothesis.*
- The value of the power function at a *parameter* point is called the *power of the test* at that point.
- This Lemma provides the most powerful test of simple hypothesis against a simple alternative hypothesis. The theorem, known as Neyman-Pearson Lemma, will be proved for density function $f(x, \theta)$ of a single continuous variate and a single parameter.
- *Every most powerful (MP) or uniformly most powerful (UMP) critical region (CR) is necessarily unbiased.*
- Neyman-Pearson Lemma based on the magnitude of the ratio of two probability density functions provides best test for testing simple hypothesis against simple alternative hypothesis.

3.8 GLOSSARY

- **Statistical Hypothesis :** A statistical hypothesis is some statement or assertion about a population or equivalent about the probability distribution characterising a population, which we want to verify on the basis of information available from a sample.
- **Test Statistic :** The choice of a suitable statistic $t = t(x_1, x_2, \ldots, x_n)$, which will best reflect upon the probability of H_0 and H_1.
- **Best Test :** One which in addition to controlling α at any desired low level has the minimum type II error β or maximum power $1-\beta$, compared to β of all other tests having this α.

NOTES

3.9 REVIEW QUESTIONS

1. Define randomised and non-randomised tests.
2. What are simple and composite statistical hypothesis ?
3. Explain the following terms :
   (*i*) Critical function ; (*ii*) Most powerful test ; (*iii*) Uniformly most powerful test ; and (*iv*) Level of significance.
4. State and prove Neyman-Pearson Lemma.
5. Define Likelihood Ratio Test. Under what circumstance would you recommend this test ?
6. Define uniformly most powerful (UMP) tests. What is uniformly most powerful critical region (UMPCR)?
7. Let $x_1, x_2, \ldots\ldots, x_n$ be a random sample from
   $$f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \; ; x \geq 0, \theta > 0$$
   Obtain UMP size α test for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$.
8. Define interval estimation. What is the relationship between testing and interval estimation ?

3.10 FURTHER READINGS

Goon, A.M., Gupta, M.K and Das Gupta. An outline of statistical theory, Volume II, the world press, Calcutta, 1980.

Lehman, E.L., Testing Statistical Hypothesis, Wiley Eastern, New Delhi, 1976.

Meyer, P.L., Introductory Probability and Statistical Applications, Addison-Wesley Publishing Company, Philippines, 1965.

Gupta, S.C. and Kapoor, V.K., fundamentals of Mathematical Statistics, Sultan Chand & Sons, New Delhi.

UNIT-IV

NON-PARAMETRIC OR DISTRIBUTION-FREE TESTS

NOTES

OBJECTIVES

After going through this unit, you should be able to:

- explain assumptions of non-parametric test
- describe one-sample non-parametric tests
- demonstrate Friedman's test

STRUCTURE

4.1 INTRODUCTION

The statistical methods of inference which were discussed in the preceding chapters make certain assumptions about the populations from which the samples are drawn. For example, the assumptions may be that the populations are normally distributed, have the same variance etc. Population values, as we have seen, are known as parameters; the statistical tests which make assumptions about the parameters are called "parametric tests." In other words, a "parametric statistical test" is a test whose model specifies certain conditions about the parameters of the population from which the sample are drawn.

In many biological investigations the research worker may not know that nature of the distribution or other required values of the population. On many occasions the number of observation made for the study may not be sufficient to the test the assumptions regarding the population. Also in many instances, when the observation is represented by a numerical figure, the scale of the measurement may not be really numerical. Some biological measurements are necessarily crude early stages of investigation. Grades of severity of an illness and ranks given to anaesthetic agents for their effectiveness are examples of data which, though expressed numerically, do not possess the characteristics necessary for arithmetical processess. The parametric tests may not be suitable in such situations.

NOTES

A "non-parametric (N.P.) or distribution-free test" is a test that does not depend on the particular form of the basic frequency function from which the samples are drawn. The other words, non-parametric test does not make any assumption regarding the form of the population. This chapter contains only some popular non-parametric tests.

4.2 ASSUMPTIONS OF NON-PARAMETRIC TESTS

The certain assumptions associated with non-parametric tests are:

(*i*) Sample observations are independent.

(*ii*) The variable under study is continuous.

(*iii*) p.d.f. is continuous.

(*iv*) Lower order moments exists.

Obviously, these assumptions are fewer and much weaker than those associated with parametric tests.

4.3 ADVANTAGES AND DISADVANTAGES OF NON-PARAMETRIC TESTS

Advantages of non-parametric tests

(*i*) Non-parametric tests are readily comprehensible, very simple and easy to apply and do not require complicated sample theory.

(*ii*) No assumption is made about the form of the frequency function of the parent population from which the sampling is done.

(*iii*) Probability statements obtained from most non-parametric tests are exact probabilities.

(*iv*) If samples are of sizes as small as 6, there is no alternative to using a non-parametric test unless the nature of the population distribution is precisely known.

(*v*) Non-parametric tests are available to deal with the data which are given in ranks or whose seemingly numerical scores have the strength of ranks. For instance, no parametric test can be applied if the scores are given in grades such as A^+, A^-, B, A, B^+ etc.

Disadvantages of non-parametric tests

(*i*) Non-parametric tests can be used only if the measurements are nominal or ordinal. In other words, if all the assumptions of a statistical model are satisfied by the data and if the measurements are of required strength, then the non-parametric tests are wasteful of data.

(*ii*) There is no non-parametric tests for testing interactions in the analysis of variance.

(*iii*) Tables of critical values may not be easily available.

4.4 ONE-SAMPLE NON-PARAMETRIC TESTS

A random sample of size n is drawn from a population and the sample values are arranged in order of magnitude and ranked accordingly, if need be. Different tests evolved in one sample case for the test of hypothesis are discussed here. These tests lead us to decide whether the sample has come from a particular population. Also, we test whether the median of the population is equal to a known value or not. Such tests are classified as tests for goodness of fit like the chi-square test.

NOTES

4.4.1 Kolmogorov-Smirnov (K-S) Test

The K-S test will be applicable when the variable has a continuous distribution. A random sample $X_1, X_2, \ldots, X_n$ of size n is drawn from an unknown population having the cumulative distribution function (c.d.f.) $F(x)$. Let the ordered valued be $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. The K-S test is based on the Glivenko-Cantelli theorem which states that the step function $S_n(x)$, with jumps occurring at the values of the ordered statistics $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ for the sample, approaches the true distribution for all x. Making use of this theorem, the comparison of the empirical distribution function $S_n(x)$ of the sample for any value x is made with the population c.d.f. under H_0(The empirical (sample) distribution function of an ordered random sample $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ of size n, denoted $S_n(x)$ for all real x, is the proportion of sample values which do not exceed x. $S_n(x)$ is a set function which increases by $1/n$ at the jump points which are the values of the ordered sample.

Symbolically, $S_n(x)$ is defined as:

$$S_n(x) = \begin{cases} 0 & \text{if} \quad x \le X_{(1)} \\ K/n & \text{if} \quad X_{(K)} \le x < x_{K+1} \quad \text{for } K = 1, 2, \ldots, n-1 \\ 1 & \text{if} \quad x \ge x_{(n)} \end{cases}$$

$S_n(x)$ is also known as the statistical image of the population distribution function $F_0(x)$. This comparison is made by defining the distance between the two cumulative distribution functions which is taken as the supremum (supremum (sup) means least upper bound. Similarly, infimum (inf) means greatest lower bound) of the absolute deviations *i.e.*, Sup $| S_n(x) - F_0(x) |$ over all x. The hypothesis for the test of goodness of fit is, $H_0 : F(x) = F_0(x)$ V_s $H_1 : F(x) \ne F_0(x)$, where F_0 is a completely specified continuous distribution.

To test H_0, the actual numerical difference $| S_n(x) - F_0(x) |$ is used in K-S test. Since this difference depends on x, the K-S statistic is taken to be the supremum of such differences, *i.e.*,

$$\boxed{D_n = \underset{\text{over all } x}{Sup} \left| S_n(x) - F_0(x) \right|},$$

where D_n is known as the K-S statistic. Under H_0, the statistic D_n has a distribution which is independent of the c.d.f. $F(x)$ that defines H_0. The statistic D_n is distribution-free. To decide about H_0, the test criterion is, reject H_0 if D_n (max $| S_n(x) - F_0(x) |$), exceeds the tabulated value for given n and prefixed significance level α. Otherwise, H_0 is accepted. The critical values of D_n for prefixed α are given in Appendix A.

Example 4.1. *On tossing five coins 192 times, the frequencies of 0 to 5 heads are:*

No. of heads	0	1	2	3	4	5
Frequency	6	26	73	66	14	7

Use Kolmogorov-Smirnov statistic to test the hypothesis that the coin is unbiased.

Solution. The null hypothesis under K-S test is,

$$H_0 : F(x) = F_0(x) \ V_S \ H_1 : F(x) \ne F_0(x).$$

The hypothetical frequencies are calculated with the help of the binomial function, ${}^nC_x\, p^x q^{n-x}$. Here $n = 5, p = 0.5$

Thus, the frequencies for $x = 0, 1, 2, 3, 4, 5$ are

$$f_0 = {}^5C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{5-0} \times 192 = 6,\ f_1 = {}^5C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 \times 192 = 30,$$

and similarly, $f_2 = 60, f_3 = 60, f_4 = 30, f_5 = 6$. Actual and theoretical frequencies, the sample or empirical c.d.f. theoretical c.d.f. and the differences are tabulated as follows:

NOTES

No. of heads	0	1	2	3	4	5
Actual frequencies	6	26	73	66	14	7
Empirical c.d.f.	$\frac{6}{192}$	$\frac{32}{192}$	$\frac{105}{192}$	$\frac{171}{192}$	$\frac{185}{192}$	$\frac{192}{192}$
Theoretical frequencies	6	30	60	60	30	6
Theoretical c.d.f.	$\frac{6}{192}$	$\frac{36}{192}$	$\frac{96}{192}$	$\frac{156}{192}$	$\frac{186}{192}$	$\frac{192}{192}$
Difference: D_n	0	$\frac{4}{192}$	$\frac{9}{192}$	$\frac{15}{192}$	$\frac{1}{192}$	0

$$\text{Max. } D_n = \frac{15}{192} = 0.078.$$

Suppose the prefixed level of significance, $\alpha = 0.1$. The critical value of D_n from Appendix A is $\frac{122}{\sqrt{192}} = 0.088$. Since the value of statistic D_n does not exceed the critical value 0.088, H_0 is accepted. This means that the sample c.d.f. is similar to the hypothetical distribution function.

4.4.2 Run Test

In biostatistical theory, it has normally been assumed that a sample drawn from a population is random. Whether the assumption of randomness is true or not needs verification. The run test is one device to test randomness. Before discussing the test, it is essential to discuss the runs.

4.4.2.1 Run (Definition)

A run is defined as a sequence of like symbols which are followed and preceded by other kinds of symbols or no symbol at either side. For clarity, a vertical line is drawn in between two consecutive sequences of symbols to mark the runs. A sequence of symbols exhibiting a pattern of symbols is usually indicative of lack of randomness. For instance, looking at a queue of persons waiting for the bus at a bus-stop, we observe the following sequence of males (M) and females (F).

M | F | M | F | M | F | M | F | M | F | M | F | M | F

This sequence shows a definite pattern, *i.e.*, the males and females are standing alternately. This sequence clearly shows a lack of randomness. Again a sequence of type,

MMMMMMMM | FFFFFFF

Shows clustering of males and females which is again indicative of lack of randomness. From this, it may be deduced that an excessive number of runs or too few runs for a given set of symbols provide the basis for non randomness.

It is not always that we have sequences of symbols. Any data at hand can be converted into runs. Take the deviation of each observation of the set from the median (any other constant) and denote the positive and zero difference by a and negative difference by b. In this way we get a sequence of symbols a and b. For example, the marks of 15 students are 55, 52, 43, 49, 36, 61, 44, 47, 67, 78, 63, 57, 41, 28, 50. On taking the deviations from 50, the following sequence is obtained.

$$aa \mid bbb \mid a \mid bb \mid aaaa \mid bb \mid a$$

The above sequence has seven runs. Positive and negative signs may also be taken in place of a and b taking zero as positive quantity, to mark the runs.

4.4.2.2 Test for Randomness

The null hypothesis: The symbols a and b occur in random order in the sequence against the alternative, H_1 : symbols a and b do not occur in random order, can be tested by the run test. Let the sample of size n contains n_1 symbols of one type, say a, and n_2 symbols of the other type, say b. Thus, $n = n_1 + n_2$. Also suppose the number of runs of symbol a are r_1 and that of symbol b are r_2, suppose $r_1 + r_2 = r$. In order to perform a test of hypothesis based on the random variable R, we need to know the probability distribution of R under H_0. The probability distribution function of R is given by

$$f_R(r) = 2\binom{n_1-1}{r/2-1}\binom{n_2-1}{r/2-1}\bigg/\binom{n_1+n_2}{n_1}$$

when r is even.

For r even, the number of runs of both types must be the same, *i.e.*, $r_1 = r_2 = r/2$.

Again

$$f_R(r) = \left\{\binom{n_1-1}{\frac{r-1}{2}}\binom{n_2-1}{\frac{r-3}{2}} + \binom{n_1-1}{\frac{r-3}{2}}\binom{n_2-1}{\frac{r-1}{2}}\right\}\bigg/\binom{n_1+n_2}{n_1}$$

where r is odd.

For r odd, $r_1 = r_2 \pm 1$. In this situation, the sum is taken over two pairs of values, $r_1 = \left(\frac{r-1}{2}\right)$ and $r_2 = \left(\frac{r+1}{2}\right)$ and vice versa.

To decide about H_0, the critical number of runs are obtained from tables given in Appendix A. These Tables provide the lower and upper critical values of the number of runs at level 0.05 respectively. If the number of runs in the sample lies between these critical values, the hypothesis H_0 is accepted, otherwise rejected. Rejecting H_0 means that the data are not in random order.

Example 4.2. *The marks of 15 students are 55, 52, 43, 49, 36, 61, 44, 47, 67, 78, 63, 57, 41, 28, 50.*

Test whether the observations occurs in random order at 5% level of significance.

Solution. Null hypothesis, H_0 : the observations occur in random order, against H_1 : the observations do not occur in random order.

On taking the deviation from 50, the following sequence is obtained:

$$aa \mid bbb \mid a \mid bb \mid aaaa \mid bb \mid a$$

Here $\quad n_1 = 8, n_2 = 7, n = 15$

$\quad r_1 = 4, r_2 = 3, r = 7$

For $\alpha = 0.05$, $n_1 = 8$, $n_2 = 7$, the lower critical value = 4 from the Appendix A.

For $\alpha = 0.05, n_1 = 8, n_2 = 7$, the upper critical value = 13 from the Appendix A. The number of runs in the sample is 7 which lies between 4 and 13. It means that the observations occur in random order with 5% Level of significance.

4.5 TWO-SAMPLE NON-PARAMETRIC TESTS

NOTES

In this section, we will discuss the some important two-sample non-parametric tests, *viz.* Kolmogorov-Smirnov test, Mann-Whitney-Wilcoxon U-test, Wald-Wolfwitz run test, Median test and Sign test.

4.5.1 Kolmogorov-Smirnov Two-Sample Test

It is an extension of the one-sample K-S test as applied to a two-sample problem. In the one-sample test, the empirical distribution was compared with the population c.d.f.. But in the two-sample K-S test, the two empirical distributions of two samples are compared and a decision is taken on the basis of the distance between these empirical distributions.

Let two random samples of sizes n_1 and n_2 be drawn from two continuous F_1 and F_2 respectively. Also, it is assumed that the two samples are independent. The observations are taken at least on ordinal scale. If these assumptions hold good, K-S two-sample test can be performed in the following manner. Let the empirical distribution functions be given by $S_{n_1}(x)$ and $S_{n_2}(x)$.

The hypothesis,

$$H_0 : F_1(x) = F_2(x) \text{ for all } x$$
$$V_S\, H_1 : F_1(x) \neq F_2(x) \text{ for some } x$$

can be tested by the K-S statistic

$$\boxed{D_{n_1, n_2} = \max_x \left| S_{n_1}(x) - S_{n_2}(x) \right|}$$

In a real problem, obtain the cumulative step functions of two samples and find the maximum difference. Compare this value with the critical value of D_{n_1, n_2} given in Appendix A. This gives the acceptance limits of D_{n_1, n_2}, for $n_1, n_2 \leq 15$ at two significance levels $\alpha = 0.05$ and $\alpha = 0.01$. Reject H_0 if calculated D_{n_1, n_2} exceeds the tabulated value, otherwise accept H_0.

For large values of n_1, $n_2 > 15$, acceptance limits of D_{n_1, n_2} are calculated by the approximate formulae given below the table for various levels of significance α. The decision about H_0 is taken in the same way as for small samples.

If there is any prior information whether

$$F_1(x) > F_2(x)$$

or

$$F_1(x) < F_2(x),$$

the one-tailed K-S test will be applied. Now we calculate either

$$D^+_{n_1, n_2} = \max_x \left| S_{n_1}(x) - S_{n_2}(x) \right|$$

or

$$D^-_{n_1, n_2} = \max_x \left| S_{n_2}(x) - S_{n_1}(x) \right|$$

as the case may be and perform the test in the usual way.

Example 4.3. *A sample of 26 male patients and another of 25 female patients suffering from respiratory tuberculosis (TB) are randomly selected. The frequency distributions according to age of males and females are given in next page:*

	Age group							
	0–5	5–15	15–25	25–35	35–45	45–55	55–65	above 65
Males suffering from TB	1	1	3	2	2	4	6	7
Females suffering from TB	1	2	4	3	6	4	2	3

Test whether the age distribution of males and females in respect of susceptibility to respiratory tuberculosis is the same by Kolmogorov Smirnov test.

NOTES

Solution. The c.d.f and the differences $\left|S_{n_1}(x) - S_{n_2}(x)\right|$ are shown simultaneously.

Age group	Males suffering from TB	c.d.f. $\{S_{n_1}(x)\}$	Females suffering from TB	c.d.f $\{S_{n_2}(x)\}$	Difference $\left|S_{n_1}(x) - S_{n_2}(x)\right|$
0–5	1	1/26	1	1/25	$\frac{1}{25 \times 26}$
5–15	1	2/26	2	3/25	$\frac{28}{25 \times 26}$
15–25	3	5/26	4	7/25	$\frac{57}{25 \times 26}$
25–35	2	7/26	3	10/25	$\frac{85}{25 \times 26}$
35–45	2	9/26	6	16/25	$\frac{191}{25 \times 26}$
45–55	4	13/26	4	20/25	$\frac{195}{25 \times 26}$
55–65	6	19/26	2	22/25	$\frac{97}{25 \times 26}$
above 65	7	26/26	3	25/25	0

The hypothesis, that the age distribution of males and females in respect of susceptibility to respiratory tuberculosis is the same, can be tested by K-S statistic.

$$D_{n_1, n_2} = \max_x \left|S_{n_1}(x) - S_{n_2}(x)\right|$$

From the above table,

$$D_{n_1, n_2} = \frac{195}{25 \times 26} = 0.300$$

The acceptance limit for $n_1 = 26, n_2 = 25$ at the level of significance $\alpha = 0.05$ with the help of formula given in Appendix A is

$$= 1.36 \sqrt{\frac{n_1 + n_2}{n_1\, n_2}} = 1.36 \times \sqrt{\frac{26 + 25}{26 \times 25}} = 0.381$$

Since the calculated difference D_{n_1, n_2} does not exceed the acceptance limit, H_0 is accepted. It means that the male and female populations have the same proportion of respiratory TB in different age groups.

4.5.2 Mann-Whitney-Wilcoxon U-Test

The usual two-sample situation is one in which the experimenter wishes to compare the effects of two treatments. In case of small samples, we have used *t*-test under the assumption that the population distribution was normal. The normality assumption may be doubtful or it may be difficult to verify the normality. In these situations we can use Mann-Whitney U-Test. It is non-parametric analogue to the usual *t*-test.

The non-parametric test used for two independent samples case was originally proposed by Wilcoxon and studied by Mann and Whitney.

Let x_i $(i = 1, 2, \ldots, n_1)$ and y_j $(j = 1, 2, \ldots, n_2)$ be independent ordered samples of size n_1 and n_2 from the populations with p.d.f. $f_1(.)$ and $f_2(.)$ respectively. We want to test the hypothesis $H_0 : f_1(.) = f_2(.)$, V_S $H_1 : f_1(.) \neq f_2(.)$.

NOTES

Mann-Whitney test is based on the pattern of the x's and y's in the combined ordered sample. Let T denote the sum of ranks of the y's in the ordered sample. The test statistic U is then defined in terms of T as follows:

$$\boxed{U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T}$$

If T is significantly large or small then $H_0 : f_1(.) = f_2(.)$ is rejected. The problem is to find the distribution of T under H_0. Unfortunately, it is very troublesome to obtain the distribution of T under H_0. However Mann and Whitney have obtained the distribution T for small n_1 and n_2, have found the moments of T in general and shown that T is asymptotically normal. It has been established that under H_0, U is asymptotically normally distributed as $N(\mu, \sigma^2)$, where

$$\mu = E(U) = \frac{n_1 n_2}{2}$$

$$\sigma^2 = Var(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Hence

$$Z = \frac{U - \mu}{\sigma} \sim N(0, 1),$$

asymptotically and normal test can be used. The approximation if fairly good if both n_1 and n_2 are greater than 8.

Example 4.4. *The following are the scores of certain randomly selected students at mid term (MT) and final examinations.*

MT scores X	55	57	72	90	57	74	
Final scores Y	80	76	63	58	56	37	75

Test the hypothesis that the distribution of scores at two occasions is the same by Mann-Whitney U-test.

Solution. The null hypothesis H_0 that the distribution of scores at two occasions is the same against H_1, *i.e.*,

$$H_0 : f_1(.) = f_2(.) \; V_S \; H_1 : f_1(.) \neq f_2(.)$$

The combined scores in a sequence in increasing order of magnitude are,

	37	55	56	57	57	58	63	72	74	75	76	80	90
	Y	X	Y	X	X	Y	Y	X	X	Y	Y	Y	X
Rank of Y:	1		3			6	7			10	11	12	

T = sum of rank's of Y = 1 + 3 + 6 + 7 + 10 + 11 + 12 = 50.

The U-Statistic is

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T$$

Here $\quad n_1 = 6, n_2 = 7$, and $T = 50$,

$$U = 6 \times 7 + \frac{7 \times 8}{2} - 50 = 20$$

From Table given in Appendix A for $n_2 = 7, n_1 = 6$ and U = 20, the probability is 0.473. Suppose that the test has been performed at 5% level of significance $\alpha = 0.05$. The tabulated probability is greater than 0.05, hence we infer that the distribution of scores at two occasions is the same.

NOTES

4.5.3 Wald-Wolfowitz Run Test

This test is used for testing of two populations. To avoid any confusion between one-sample run test and this test procedure, we will consider two random samples m and n respectively, instead of sizes n_1 and n_2. Let the two independent samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ combine into a single sequence of ordered statistics. Assuming that the sample have come from continuous probability distribution, a unique sequence of ordered statistics is always possible as, theoretically, no ties should occur. In the combined sequence some identification mark must be put with observations of one type so as to identify whether an observation belongs to X-sample or Y-sample. The definition of runs remains the same as in the one-sample case. Let the combined sequence of order statistics with $m = 5$ and $n = 4$ be

XX | Y | X | YY | XX | Y.

In this case, there are 3 runs of X's and 3 runs of Y's. In all we have 6 runs. From this it may be concluded that the two populations are identical as the total number of runs r is quite large. If the populations are different, we expect that r will be small as the two samples are not well mixed. Here we test,

$$H_0 : f_1(.) = f_2(.) \ V_S \ H_1 : f_1(.) \neq f_2(.)$$

We define a random variable U, the total number of runs in pooled ordered arrangement of m X's and n Y's. As already mentioned, too few runs tend to reject H_0 in favour of H_1 as this indicates that either most of the X's are greater than or less than the Y's. The Wald-Wolfowitz run test of nominal α has a critical region given by

$$U \leq r_\alpha$$

where r_α is chose to be the largest integer such that $P(U \leq r_\alpha) \leq \alpha$ under H_0.

Since X and Y constitute a completely random sequence under H_0 as the letters a and b do in the case of single random samples, the probability distribution of U is exactly the same as given in RUN TEST (section 4.4.2). In this case, $m = n_1$ and $n = n_1$. The only difference in the Wald-Wolfowitz Run Test from the one-sample run test is that only the one-sided test is being used here. Tables of critical values of r at level of significance α are prepared by Swed and Eisenhart.

For large $m, n > 10$, a normal approximation may be made assuming that $m/(m+n)$ and $n/(m+n)$ remain constant as $m + n \to \infty$.

The mean and variance of U are given as,

$$E(U) = \frac{2mn}{m+n} + 1,$$

$$\text{Var}(U) = \frac{2mn(2mn - n - m)}{(m+n)^2 (m+n-1)}$$

and we can use the normal test

$$\boxed{Z = \frac{U - E(U)}{\sqrt{Var(U)}}} \sim N(0, 1), \text{ asymptotically.}$$

The decision about H_0 is taken in the usual way.

Example 4.5. *The following are the rates of flow of a certain gas through two soil samples collected from two different places.*

Sample X	*23*	*27*	*19*	*24*	*22*	*30*		
Sample Y	*21*	*29*	*34*	*32*	*26*	*28*	*36*	*26*

Test the hypothesis that the two populations of soil types are the same with respect to the rates of flow through the soils by Wald-Wolfowitz run test.

NOTES

Solution. The null hypothesis that the populations of soil types are the same with respect to the rates of flow through the soils, *i.e.*, symbolically,

$$H_0 : f_1(.) = f_2(.) \text{ V}_S \, H_1 : f_1(.) \neq f_2(.)$$

The sequence of combined samples with the clearly marked runs is,

19 | **21** | 22, 23, 24 | **26, 26** | 27 | **28, 29** | 30 | **32, 34, 36** |

In this case $m = 6$, $n = 8$ and the number of runs $r = 8$.

The probability for $r = 8$ (even) is given by

$$f(r) = \frac{2\binom{5}{3}\binom{7}{3}}{\binom{14}{6}} = 0.233.$$

Supposing the predecided level of significance $\alpha = 0.1$. Since the probability $f(r)$ is greater than 0.1, we accept H_0 which means that the two soils have identical distributions in respect of rates of flow of a certain gas.

4.5.4 Median Test

This test is attributed to Westenberg and Mood. Median test is a statistical procedure for testing if two independent ordered samples differ in their central tendencies. In other words, it gives information if two independent samples are likely to have been drawn from the populations with the same median.

As in 'Run Test' let $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$ be the independent ordered samples from the populations with p.d.f.'s $f_1(.)$ and $f_2(.)$ respectively. The measurements must be at least ordinal. Let $z_1, z_2, \ldots, z_{n_1+n_2}$ be the combined ordered sample. Let m_1 be the number of x's and m_2 the number of y's exceeding the median value M(say), of the combined sample.

Then under the null hypothesis that the samples come from the same population or from different populations with the same median, *i.e.*, $H_0 : f_1(.) = f_2(.)$, the joint distribution of m_1 and m_2 is the hypergeometric distribution with probability function

$$p(m_1, m_2) = \frac{\binom{n_1}{m_1}\binom{n_2}{m_2}}{\binom{n_1+n_2}{m_1+m_2}}$$

If $m_1 < n_1/2$, then the critical region corresponding to the size of type 1 error α, is given by $m_1 < m_1'$ is computed from the equation

$$\sum_{m_1=1}^{m_1'} p(m_1, m_2) = \alpha$$

The distribution of m_1 under H_0 is also hypergeometric with

$$E(m_1) = \frac{n_1}{2}, \text{ if } N = n_1 + n_2 \text{ is even}$$

$$= \frac{n_1}{2}\frac{N-1}{N}, \text{ if N is odd}$$

and
$$Var(m_1) = \frac{n_1 n_2}{4(N-1)}, \text{ if N is even}$$

$$= \frac{n_1 n_2 (N+1)}{4N^2}, \text{ if N is odd.}$$

NOTES

This distribution is most of the times quite inconvenient to use. However for large samples, we may regard m_1 to be asymptotically normal and use normal test, *viz.*,

$$\boxed{Z = \frac{m_1 - E(m_1)}{\sqrt{Var(m_1)}}} \sim N(0, 1) \text{ asymptotically.}$$

Remark 4.1. *The observations m_1 and m_2 can be classified into the following 2×2 contingency table.*

	Sample I	*Sample II*	*Total*
No. of observations > M	m_1	m_2	$m_1 + m_2$
No. of observations < M	$n_1 - m_1$	$n_2 - m_2$	$n_1 + n_2 - m_1 - m_2$
Total	n_1	n_2	$n_1 + n_2 = N$

If frequencies are small, we can compute the exact probabilities from $p(m_1, m_2)$ rather than approximate them. However, if frequencies are large, we may use χ^2-Test with 1 d.f. (for a 2×2 contingency table) for testing H_0. The approximation is fairly good if both n_1 and n_2 exceed 10.

Remark 4.2. *Median test is sensitive to the differences in location between $f_1(x)$ and $f_2(y)$ but not to differences in their shapes. Thus if $f_1(x)$ and $f_2(y)$ have the same median, we would expect $H_0 : f_1(.) = f_2(.)$ to be accepted ordinarily even though their shapes are quite different.*

Remark 4.3. *Generally, the median test makes the correct decision with a little more assurance than does the sign test (section, 4.5.5) but not as decisively as the t-test.*

Example 4.6. *The data of 10 plots each, under two treatments are as given below:*

(Treat. 1) X	46	45	32	42	39	48	49	30	51	34
(Treat. 2) Y	44	40	59	47	55	50	47	71	43	55

Test the hypothesis of equality of median response under two treatments by median test.

Solution. The hypothesis,

$$H_0 : f_1(x) = f_2(y) \ V_S \ H_1 : f_1(x) \neq f_2(y)$$

Arranging the data in ascending order, we have

30, 32, 34, 39, **40**, 42, **43**, **44**, 45, 46, **47**, **47**, 48, 49, **50**, 51, **55**, **55**, **59**, **71**

In the above ordered statistics, the observations belonging to treatment 2 are marked just to differentiate them from observations under treatment 1. The median of the combined data is 46.5. Here we have 10 observations on the left of the median, and 10 observations on the right of the median, *i.e.*, $m_1 + m_2 = 10$. Also $n_1 = 10, n_2 = 10, n = 20, m_1 = 7, m_2 = 3$.

The probability is

$$p(m_1, m_2) = \frac{\binom{10}{7}\binom{10}{3}}{\binom{20}{10}} = \frac{120 \times 120}{19 \times 17 \times 4 \times 13 \times 11} = 0.078.$$

NOTES

Take the prefixed level of significance $\alpha = 0.05$. Since $p(m_1, m_2)$ is greater than 0.05, H_0 is accepted. This means that the treatments are equally effective with regard to their median effects.

4.5.5 Sign Test

Consider a situation where it is desired to compare two things or materials under various set of conditions. An experiment is thus conducted the following circumstances.

(*i*) When there are pairs of observations on two things being compared.

(*ii*) For any given pair, each of the two observations is made under similar extraneous conditions.

(*iii*) Different pairs were observed under different conditions.

Condition (*iii*) implies that the differences $d_i = x_i - y_i, i = 1, 2, \ldots, n$ have different variance and thus the student's *t*-test (paired) invalid, which would have otherwise been used unless there was obvious non-normality. So, in such a case we use the 'Sign Test', named so since it is based on the signs (plus or minus) of the deviations $d_i = x_i - y_i$. No assumptions are made regarding the parent population. The only assumptions are:

(*i*) Measurements are such that the deviations $d_i = x_i - y_i$, can be expressed in terms of positive or negative signs.

(*ii*) Variables have continuous distribution.

(*iii*) d_i's are independent.

The paired differences or matched-sample approach is a good experimental design for identifying differences in two populations.

Let $(x_i, y_i), i = 1, 2, \ldots, n$ be n paired sample observations drawn from the two populations with p.d.f.'s $f_1(x)$ and $f_2(y)$. We want to test the null hypothesis $H_0 : f_1(x) = f_2(y)$. To test H_0 consider $d_i = x_i - y_i, (i = 1, 2, \ldots, n)$. When H_0 is true, x_i and y_i constitute a random sample of size 2 from the same population. Since the probability that the first of the two sample observations exceeds the second is same as the probability that the second exceeds the first and since hypothetically the probability of a tie is zero, H_0 may be restated as:

$$H_0 : P[X - Y > 0] = \frac{1}{2} \text{ and } P[X - Y < 0] = \frac{1}{2}$$

Let us define

$$U_i = \begin{cases} 1, & \text{if } x_i - y_i > 0 \\ 0, & \text{if } x_i - y_i < 0 \end{cases}$$

The U_i is a Bernoulli variate with $p = P(x_i - y_i > 0) = \frac{1}{2}$.

Since U_i's are independent, $U = \sum_{i=1}^{n} U_i$, the total number of positive deviations, is a Binomial variate with parameters n and p. Let the number of positive deviations be K.

NOTES

Then

$$P(U \leq K) = \sum_{r=0}^{K} \binom{n}{r} p^r q^{n-r}, \left(p = q = \frac{1}{2} \text{ under } H_0\right)$$
$$= \left(\frac{1}{2}\right)^n \sum_{r=0}^{K} \binom{n}{r} = p'(\text{say}).$$

If $p' \leq 0.05$, we reject H_0 at 5% level of significance and if $p' > 0.05$, we conclude that the data do not provide any evidence against the null hypothesis, which may, therefore, be accepted.

For large samples, ($n \geq 30$), we may regard U to be asymptotically normal with, (under H_0)

$$E(U) = np = \frac{n}{2} \quad \text{and} \quad Var(U) = npq = \frac{n}{4}$$

$$\therefore \qquad z = \frac{U - E(U)}{\sqrt{Var(U)}} = \frac{U - \frac{n}{2}}{\sqrt{n/4}}$$

is asymptotically N(0, 1) and we may use normal test.

4.6 THE KRUSKAL-WALLIS TEST FOR DIFFERENCES IN MORE THAN TWO POPULATIONS

When more than 2 samples are considered, the Kruskal-Wallis test can be applied to test whether they all belong to the same population. This test is one-way analysis of variance by ranks, and is useful for deciding whether K-independent samples are from different populations. This test is the most efficient of the non-parametric tests for K independent samples.

If there are K samples, and varying number n_i is studied in each sample, let n = total studied = $\sum_{i=1}^{K} ni$.

Let all the values from all the K samples be combined and ranked in a single series. The smallest value is replaced by rank 1, the next to smallest by rank 2, and the largest by rank n, where n is equal to the total number of independent observations in the K samples. When this has been done, the sum of the ranks in each sample (column) is found. The Kruskal-Wallis test determines whether these sums of ranks are so disparate that they are not likely to have come from samples which were all drawn from the same population. It can be shown that if the K samples actually are from the same population or from identical populations, the H (the statistic used in the Kruskal-Wallis test and defined by the formula given below) is distributed as chi-square with d.f. = K – 1, provided that the sizes of the various K samples are not too small. That is

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(n+1)$$

is distributed approximately as chi-square with (K – 1) d.f., for sample sizes sufficiently large (at least 5 in each group). For smaller sample sizes, special tables are available. When ties occur between two or more values, each value is given the mean of the ranks for which it is tied. Since the value of H is some what influenced by ties, one may wish to correct for ties in computing H. To correct for the effect of ties, H is computed by the formula by the formula given and then divided by

$$1 - \frac{\Sigma T}{n^3 - n}$$

where $T = t^3 - t$ (when t is the number of tied observations in a tied group of values).

The effect of correcting for ties is to increase the value of H and thus to make the result more significant than it would have been if uncorrected. Therefore, if one is able to reject the null hypothesis without making the correction, one will be able to reject it at an even more stringent level of significance if the correction is used.

NOTES

Example 4.7. *In a study of cerebrovascular disease, patients from 3 socio-economic backgrounds were thoroughly investigated. One characteristic measured was the diastolic blood pressure (mm/Hg). Is there any reason to believe that the 3 groups differ with respect to this characteristic?*

Group A	*Group B*	*Group C*
100	92	81
103	97	102
89	88	86
78	84	83
105	90	99
	95	

Solution. Null hypothesis, H_0 : There is no difference among the diastotic blood pressures of the 3 groups.

There are 16 values. Pooling all these values, arranging them from lowest to highest, ranking them and putting them back into the 3 groups, we find the following ranks in each group:

A	*B*	*C*
13	9	2
15	11	14
7	6	5
1	4	3
16	8	12
	10	
$R_1 = 52$	$R_2 = 48$	$R_3 = 36$

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{16(16+1)} \left[\frac{(52)^2}{5} + \frac{(48)^2}{6} + \frac{(36)^2}{5} \right] - 3(16+1)$$

$$= 52.2 - 51.0 = 1.2.$$

H is distributed as a χ^2 with 2 d..f. and for the above value of H, P is > 0.5, *i.e.*, the differences observed can occur by chance alone more than 50% of the time. Therefore, the null hypothesis is accepted.

4.7 FRIEDMAN'S TEST

This is a non-parametric test for K-related samples of equal size, say *n*, parallel to two way analysis of variance. Here we have K-related samples of size *n* arranged in n blocks and K columns in a two-way table as given in next page:

NOTES

Blocks	Samples (Treatments) 1	2 .... K	Block Totals
1	R_{11}	R_{12} .... R_{1K}	K(K + 1)/2
2	R_{21}	R_{22} .... R_{2K}	K(K + 1)/2
3	R_{31}	R_{32} .... R_{3K}	K(K + 1)/2
⋮	⋮	⋮ .... ⋮	⋮
⋮	⋮	⋮ .... ⋮	⋮
n	R_{n1}	R_{n2} .... R_{nK}	K(K + 1)/2
Column totals	R_1	R_2 .... R_K	$\frac{nK(K+1)}{2}$

where R_{ij} is the rank of the observation belonging to sample *j* in block *i* for $j = 1, 2, ..., K$ and $i = 1, 2, ..., n$. This is the same situation in which there are K treatments and each treatment is replicated *n* times. Here it should be carefully noted that the observations in a block receive ranks from 1 to K. The smallest observation receives rank 1 at its place and the largest observation receives rank K at its place. Intermediary observations receive ranks accordingly. Hence, the block totals are constant and equal to $\frac{K(K+1)}{2}$, the sum of K integers.

The null hypothesis H_0 to be tested is that all the K samples have come from identical populations. In case of experimental design, the null hypothesis H_0 is that there is no difference between K treatments. The alternative hypothesis H_1 is that at least two samples (treatments) differ from each other.

Under H_0, the test statistic is,

$$F = \frac{12}{nK(K+1)} \sum_{j=1}^{K} R_j^2 - 3n(K+1)$$

The statistic F is distributed as chi-square with (K – 1) d.f. At level α, reject H_0 if $F \geq \chi^2_{\alpha, K-1}$, otherwise H_0 is accepted.

Example 4.8. *The iron determinations (ppm) in five pea-leaf samples, each under three treatments, are as given below:*

Samples no. (Blocks)	*Treatments* 1	2	3
1	*591*	*682*	*72*
2	*818*	*591*	*863*
3	*682*	*636*	*773*
4	*499*	*625*	*909*
5	*648*	*863*	*818*

Test the hypothesis that the iron content in leaves under three treatments is the same by Friedman's test.

Solution. H_0 : The iron content in leaves under three treatments is the same, Vs H_1 : at least two of them have different effect. The ranks of the observations in each block are given in next page:

NOTES

Samples no. (Blocks)	*Treatments*			*Block totals*
	1	2	3	
1	1	2	3	6
2	2	1	3	6
3	2	1	3	6
4	1	2	3	6
5	1	3	2	6
Column totals	7	9	14	30

The Friedman's test statistic is

$$F = \frac{12}{nK(K+1)} \sum_{j=1}^{K} R_j^2 - 3n(K+1)$$

$$= \frac{12}{5 \times 3(3+1)} (7^2 + 9^2 + 14^2) - 3 \times 5 \times (3+1) = 65.2 - 60.0 = 5.2.$$

For $\alpha = 0.05$, the table value of $\chi^2_{0.05,\,2} = 5.99$.

Since the calculated valued of F is less than the table valued 5.99, the null hypothesis is accepted. This means that there is no difference in the iron content of pea leaves due to the treatments.

4.8 SUMMARY

- A "non-parametric (N.P.) or distribution-free test" is a test that does not depend on the particular form of the basic frequency function from which the samples are drawn.
- Non-parametric tests are readily comprehensible, very simple and easy to apply and do not require complicated sample theory.
- Non-parametric tests are available to deal with the data which are given in ranks or whose seemingly numerical scores have the strength of ranks.
- Generally, the median test makes the correct decision with a little more assurance than does the sign test.
- When more than 2 samples are considered, the Kruskal-Wallis test can be applied to test whether they all belong to the same population. This test is one-way analysis of variance by ranks, and is useful for deciding whether K-independent samples are from different populations.

4.9 GLOSSARY

- **Parametric Statistical Test :** It is a test whose model specifies certain conditions about the parameters of the population from which the sample are drawn.
- **Run :** It is defined as a sequence of like symbols which are followed and preceded by other kinds of symbols or no symbol at either side.
- **Median Test :** It is a statistical procedure for testing if two independent ordered samples differ in their central tendencies.
- **Friedman's Test :** This is a non-parametric test for K-related samples of equal size, say n, parallel to two way analysis of variance.

4.10 REVIEW QUESTIONS

1. Explain the term 'distribution- free methods'.
2. Explain the main difference between the parametric and the non-parametric approaches to the theory of statistical inference.

3. Write a short essay on the uses of non-parametric tests.
4. Which nonparametric tests are substitutes for the analysis of variance? Describe their methodology.
5. A die was rolled 132 times and the following results were obtained:

No. of spots	1	2	3	4	5	6
Frequency	12	24	38	32	16	10

Use Kolmogorov-Smirnov statistic to test hypothesis that the die is unbiased.

6. Two quality control laboratories independently collected samples of 25 articles form a number of sales depots and tested them. The number of defective per sales depot were as follows :

Lab A	9	3	1	3	0	7	2	11
Lab B	12	6	6	4	8	5	4	

Test the hypothesis that the two laboratories have samples from the same lot by (*i*) Median test (*ii*) Mann-Whitney U-test (*iii*) Wald-Wolfowitz runs test.

7. On tossing a coin 15 times, the following sequences of heads (H) and tails (T) was obtained:

T T H H H T H T H H H T T H H

Test whether the coin is unbiased by the Run test.

8. The grain yield of paddy (t/ha) with four different levels of nitrogen from a completely randomised design experiment are given below. Test whether there is any significant difference between the effects of different nitrogen levels by Kruskal-Wallis test.

Treatment 1 yield	*Treatment 2 yield*	*Treatment 3 yield*	*Treatment 4 yield*
5.1	4.4	6.0	3.2
5.4	4.9	5.8	3.6
5.3	4.8	5.3	3.1
4.7		6.1	

9. School children taking coaching in three private schools secured the following scores out 100.

No. of children	*Schools*		
	1	*2*	*3*
1	33	32	55
2	38	15	68
3	39	87	27
4	48	32	88
5	58	22	46
6	70	63	52
7	61	56	76
8	41	57	
9	45	44	
10	49		

Test the hypothesis that the students studying the three private schools have identical distribution of marks by applying the (*i*) Median test (*ii*) Kruskal Wallis test, at significance level, $\alpha = 0.1$.

10. The quantity of serum albumin (gms) per 100 ml in lepers under three different drugs and the control groups was as follows:

Group No.	Control	Drug I	Drug II	Drug III
1	4.30	3.65	3.05	3.90
2	4.00	3.60	4.10	3.10
3	4.10	2.70	4.20	3.20
4	3.80	3.15	3.70	4.20
5	3.30	3.75	3.60	3.00
6	4.50	2.95	4.80	3.40

NOTES

Apply Friedman's test to confirm whether the content of serum albumin in different groups of persons is the same.

4.11 FURTHER READINGS

Agarwal, B.L., Basic statistics, New Age International Publishers, New Delhi.

Bradley, J.V., Distribution free Statistical Tests, Prentice Hall, Englewood cliffs, 1968.

Fraser, D.A.S., Nonparametric Methods in Statistics, John Wiley, New York, 1957.

Wilcoxon, F., "Probability Tables for Individual Comparisons by Ranking Methods". Biometrics, 1, 80–83, 1945.

UNIT- V

SEQUENTIAL ANALYSIS

NOTES

OBJECTIVES

After going through this unit, you should be able to:

- explain sequential probability ratio test
- describe operating characteristic function of SPRT
- give brief account on average sample number

STRUCTURE

5.1 INTRODUCTION

We have seen that in Neyman-Pearson theory of testing of hypothesis, n, the sample size is regarded as a fixed constant and keeping α fixed, we minimize β. But in the sequential analysis theory propounded by A. *Wald* n, the sample size is not fixed but is regarded as a random variable whereas both α and β are fixed constants.

5.2 SEQUENTIAL PROBABILITY RATIO TEST (SPRT)

The best known procedure in sequential testing is the *Sequential Probability Ratio Test* (SPRT) developed by A. Wald discussed below:

Suppose we want to test the hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$ for a distribution with p.d.f. $f(x, \theta)$. For any positive integer m, the likelihood function of a sample $x_1, x_2, \ldots, x_m$ from the population with p.d.f. (p.m.f.) $f(x, \theta)$ is given by:

$L_{1m} = \prod_{i=1}^{m} f(x_i, \theta_1)$ when H_1 is true, and by $L_{0m} = \prod_{i=1}^{m} f(x_i, \theta_0)$ when H_0 is true, and the likelihood ratio λ_m is given by:

NOTES

$$\lambda_m = \frac{L_{1m}}{L_{0m}} = \frac{\prod_{i=1}^{m} f(x_i, \theta_1)}{\prod_{i=1}^{m} f(x_i, \theta_0)} = \prod_{i=1}^{m} \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}, \ (m = 1, 2, \ldots) \quad \ldots(5.1)$$

The SPRT for testing H_0 against H_1 is defined as follows:

At each stage of the experiment (at the mth trial for any integral value m), the likelihood ratio λ_m, $(m = 1, 2, \ldots)$ is computed.

(*i*) If $\lambda_m \geq A$, we terminate the process with the rejection of H_0
(*ii*) If $\lambda_m \leq B$, we terminate the process with the acceptance of H_0, and
(*iii*) If $B < \lambda_m < A$, we continue sampling by taking an additional observation. ...(5.2)

Here A and B ($B < A$) are the constants which are determined by the relation

$$A = \frac{1-\beta}{\alpha}, B = \frac{\beta}{1-\alpha} \quad \ldots(5.3)$$

where α and β are the probabilities of type I error and type II error respectively.

From computational point of view, it is much convenient to deal with $\log \lambda_m$ rather than λ_m,

5.2.1 Deviation of Test Statistics for Sample Vs Sample Hypothesis

Let $F(x, \theta)$ be the p.d.f./p.m.f. of a r.v. X. let $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

Thus $F(x, \theta_0)$ and $F(x, \theta_1)$ are the distribution of X under H_0 and H_1 respectively. For any positive integral value in the probability that sample $x_1, \ldots\ldots, x_m$ is obtained as

$$p_{1m} = F(x_1, \theta_1)\, F(x_2, \theta_1) \ldots\ldots, F(x_m, \theta_1) \text{ under } H_1.$$
$$p_{0m} = F(x_1, \theta_0)\, F(x_2, \theta_0) \ldots\ldots, F(x_m, \theta_0) \text{ under } H_0.$$

Since the sample hypothesis is tested vs a sample alternative. Using the likelihood ratio test, we have

$$\lambda_m = p_{1m}/p_{0m}$$

As a basis for deciding between H_0 and H_1. For fixed sample size,

H_0 is accepted if $\lambda_m < k$

H_0 is rejected if $\lambda_m \geq k$.

A sequential that can be constructed by extending the fixed sample size method to include a region of continuing sampling.

In the fixed sample size test, one could accept H_0 if λ_m is small and accept H_1 if λ_m is large. Thus in sequential test being two numbers A and B are chosen and divided successive observations are taken for $m = 1, 2, \ldots\ldots$, as long as $B < \lambda_m < A$.

If $\lambda_m \geq A$, the process is terminated with the rejection of H_0. If $\lambda_m \leq B$, the process terminated with the acceptance of H_0. For practical purpose '$\log \frac{p_{1m}}{p_{0m}}$' is more easy to handling then that if 'p_{1m}/p_{0m}' that is

$$\log \frac{p_{1m}}{p_{0m}} = \sum_{i=1}^{m} \log \frac{F(x_i, \theta_1)}{F(x_i, \theta_0)} = \sum_{i=1}^{m} z_i$$

where $$z_i = \log \frac{F(x_i, \theta_1)}{F(x_i, \theta_0)}$$

The SPRT reduce to

(I) Accept H_0 if $\sum_{i=1}^{m} z_i \le \log B$

(II) Reject H_0 if $\sum_{i=1}^{m} z_i \ge \log A$

(III) Continue if $\log B < \sum_{i=1}^{m} z_i < \log A$.

NOTES

5.2.2 Determination of A and B

Let α and β be the strength of the test, α = size of the type (I) error and β = size of the type (II) error.

$$\alpha = p\,\{\text{Reject } H_0 / H_0\}$$

$$\alpha = \sum_{m=1}^{\infty} p\left\{x \in R_{1m}/H_0\right\}$$

$$\alpha = \sum_{m=1}^{\infty} \int_{R_{1m}} P_{0m}\, dx_1 \ldots\ldots dx_m$$

$$\lambda_m = p_{1m}/p_{0m} \ge A$$

or $$p_{0m} \le p_{1m} / A$$

∴ $$\alpha \le \frac{1}{A} \sum_{m=1}^{\infty} \int_{R_{1m}} p_{1m}\, dx$$

or

$$\alpha \le \frac{1}{A} \sum_{m=1}^{\infty} P\left\{x \in R_{1m}/H_1\right\}$$

$$\alpha \le \frac{1}{A}\, p\left\{\text{Reject } H_0/H_1\right\}$$

$$\alpha \le \frac{1}{A}(1-\beta)$$

$$\boxed{A \le \frac{1-\beta}{\alpha}} \qquad \ldots(5.4)$$

$$1-\alpha = 1 - P\,\{\text{Reject } H_0/H_0\}$$

$$= P\,\{\text{Accept } H_0/H_0\}$$

$$= \sum_{m=1}^{\infty} P\left\{x \in R_{0m}/H_0\right\}$$

$$= \sum_{m=0}^{\infty} \int_{R_{0m}} p_{0m}\, dx_1, dx_2, \ldots\ldots, dx_m$$

$$\lambda_m = \frac{p_{1m}}{p_{0m}} \le \beta$$

$\therefore$ $$p_{0m} \ge p_{1m}/\beta$$

NOTES

$\therefore$ $$1-\alpha \ge \frac{1}{B} \sum_{m=1}^{\infty} \int p_{1m}\, dx_1, \ldots\ldots, dx_m$$

or $$1-\alpha \ge \frac{1}{B} \sum_{m=1}^{\infty} p\,\{x \in R_{0m}/H_1\}$$

or $$1-\alpha \ge \frac{1}{B} p\,\{\text{Accept } H_0/H_1\}$$

$$1-\alpha \ge \frac{1}{B}\beta$$

$\Rightarrow$ $$\boxed{B \ge \frac{\beta}{1-\alpha}} \quad \ldots(5.5)$$

In actual practice

$$A \simeq 1 - \frac{\beta}{\alpha} \text{ and } B \simeq \frac{\beta}{1-\alpha}.$$

5.2.3 Illustrative Examples Based on the Application to Probability Distribution

Example 5.1. *If $X \sim N(\theta, \sigma^2)$, σ^2 is known, solve SPRT for testing*

$$H_0: \theta = \theta_0 \text{ vs } \theta = \theta_1$$

$$\{(I)\ \theta_1 > \theta_0 \text{ and } (II)\ \theta_1 < \theta_0\}$$

Solution. We have

$$F(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$

Now,

$$z_i = \log \frac{F(x, \theta_1)}{F(x, \theta_0)}$$

$$z_i = \log \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2} \Bigg/ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta_0)^2} \right\}$$

$$z_i = \log \left\{ e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2 - (x-\theta_0)^2} \right\}$$

$$z_i = \log \left[e^{-\frac{1}{2\sigma^2}\{x^2 + \theta_1^2 - 2\theta_1 x - x^2 - \theta_0^2 + 2\theta_0 x\}} \right]$$

$$= \log \left[e^{-\frac{1}{2\sigma^2}\{\theta_1^2 - \theta_0^2 - 2(\theta_1 - \theta_0)x\}} \right]$$

$$= \left(\frac{\theta_1 - \theta_0}{\sigma^2} \right) x - \frac{\theta_1^2 - \theta_0^2}{2\sigma^2}$$

Now, taking Σ both sides,

$$\sum_{i=1}^{m} z_i = \frac{(\theta_1 - \theta_0)}{\sigma^2} \sum_{i=1}^{m} x_i - m\left(\frac{\theta_1^2 - \theta_0^2}{2\sigma^2}\right) \quad ...(5.6)$$

Case I. If $\theta_1 > \theta_0 \Rightarrow \theta_1 - \theta_0 > 0$

Reject H_0 if $\sum_{i=1}^{m} z_i \geq \log A$, we have

NOTES

$$\left(\frac{\theta_1 - \theta_0}{\sigma^2}\right) \sum_{i=1}^{m} x_i - m\left(\frac{\theta_1^2 - \theta_0^2}{2\sigma^2}\right) \geq \log A \quad ...(5.7)$$

$$\frac{(\theta_1 - \theta_0)}{\sigma^2} \sum_{i=1}^{m} x_i \geq \log A + \frac{m\left(\theta_1^2 - \theta_0^2\right)}{2\sigma^2}$$

or
$$\sum_{i=1}^{m} x_i \geq \frac{\sigma^2 \log A}{\theta_1 - \theta_0} + m\left(\frac{\theta_1 + \theta_0}{2}\right)$$

⇒
$$\sum_{i=1}^{m} x_i \geq a + mc \quad ...(5.8)$$

Now accept H_0 if $\sum_{i=1}^{m} z_i \leq \log B$

We have from (5.6)

$$\frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^{m} x_i - m \frac{\left(\theta_1^2 - \theta_0^2\right)}{2\sigma^2} \leq \log B$$

or
$$\frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^{m} x_i \leq \log B + m \frac{\theta_1^2 - \theta_0^2}{2\sigma^2}$$

$$\sum_{i=1}^{m} x_i \leq \sigma^2 \frac{\log B}{\theta_1 - \theta_0} + m \frac{\theta_1 + \theta_0}{2}$$

$$\sum_{i=1}^{m} x_i \leq b + mc \quad ...(5.9)$$

If $\log B \leq \sum_{i=1}^{m} z_i \leq \log A$

We have
$$b + mc \leq \sum_{i=1}^{m} x_i \leq a + mc \quad ...(5.10)$$

Case II. If $\theta_1 < \theta_0 \Rightarrow \theta_1 - \theta_0 \leq 0$

Now, reject H_0 if $\sum_{i=1}^{m} z_i \geq \log A$, we have from (5.6),

$$\frac{\theta_1 - \theta_0}{\sigma^2} \sum_{i=1}^{m} x_i - m\left(\frac{\theta_1^2 - \theta_0^2}{2\sigma^2}\right) \geq \log A$$

$$\sum_{i=1}^{m} x_i \leq \frac{\sigma^2 \log A}{\theta_1 - \theta_0} + m\left(\frac{\theta_1 - \theta_0}{2}\right) \geq \log A$$

$$\Rightarrow \qquad \sum_{i=1}^{m} x_i \leq a + mc \qquad ...(5.11)$$

NOTES

Now accept H_0 if $\sum_{i=1}^{m} x_i \leq \log B$

$$\sum_{i=1}^{m} x_i \geq b + mc \qquad ...(5.12)$$

From (5.11) and (5.12), we have

$$a + mc \leq \sum_{i=1}^{m} x_1 \leq b + mc.$$

Example 5.2. *If $X \sim B(1, \theta)$. Develop SPRT for testing $H_0 \cdot \theta = \theta_0$ against $H_1 \cdot \theta = \theta_1$ $(\theta_1 > \theta_0)$.*

Solution. We have,

$$F(x, \theta) = \theta^x (1-\theta)^{1-x}$$

Now,

$$z_i = \log \frac{F(x, \theta_1)}{F(x, \theta_0)}$$

$$= \log \frac{\theta_1^x (1-\theta_1)^{(1-x)}}{\theta_0^x (1-\theta_0)^{(1-x)}}$$

$$= \log \left(\frac{\theta_1 (1-\theta_0)}{\theta_0 (1-\theta_1)}\right)^x \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

$$= x \log \left[\frac{\theta_1 (1-\theta_0)}{\theta_0 (1-\theta_1)}\right] + \log \frac{(1-\theta_1)}{(1-\theta_0)}$$

Taking $\sum_{i=1}^{m}$ both sides, we get

$$\sum_{i=1}^{m} z_i = \sum_{i=1}^{m} x_i \log \left[\frac{\theta_1 (1-\theta_0)}{\theta_0 (1-\theta_1)}\right] + m \log \left(\frac{1-\theta_1}{1-\theta_0}\right) \qquad ...(5.13)$$

If $\theta_1 > \theta_0$, then

$$1 - \theta_1 < 1 - \theta_0$$

$$\therefore \qquad \log \left[\frac{\theta_1}{\theta_0} \cdot \frac{(1-\theta_0)}{(1-\theta_1)}\right] > 0,$$

$$0 < \left(\frac{1-\theta_0}{1-\theta_1}\right)$$

(I) Reject H_0, if $\sum_{i=1}^{m} z_i \geq \log A$, we have

From (5.13),

$$\sum_{i=1}^{m} x_i \log \left(\frac{\theta_1}{\theta_0} \cdot \frac{1-\theta_0}{1-\theta_1}\right) + m \log \left(\frac{1-\theta_1}{1-\theta_0}\right) \geq \log A$$

NOTES

or $$\sum_{i=1}^{m} x_i \geq \log A \Big/ \log\left(\frac{\theta_1}{\theta_0}\cdot\frac{1-\theta_0}{1-\theta_1}\right) - \frac{m\log\left(\dfrac{1-\theta_1}{1-\theta_0}\right)}{\log\left(\dfrac{\theta_1}{\theta_0}\cdot\dfrac{1-\theta_0}{1-\theta_1}\right)}$$

or $$\sum_{i=1}^{m} x_i \geq a - mc \qquad ...(5.14)$$

(II) Accept H_0, if $\sum_{i=1}^{m} z_i \leq \log B$.

$$\Rightarrow \quad \sum_{i=1}^{m} x_i \leq b - mc \qquad ...(5.15)$$

Continue, if $\log B < \sum_{i=1}^{m} z_i < \log A$

From (5.14) and (5.15), we get

$$\Rightarrow \quad b - mc \leq \sum_{i=1}^{m} x_i \leq a - mc.$$

Example 5.3. *If $X \sim P(\theta)$. Develop SPRT for testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ [(I) $\theta_1 > \theta_0$ and (II) $\theta_1 < \theta_0$]*

Solution. We have,

$$F(x, \theta) = \frac{e^{-\theta}\,\theta^x}{\lfloor x}$$

Now,
$$z_i = \log \frac{F(x, \theta_1)}{F(x, \theta_0)}$$

$$= \log \frac{e^{-\theta_1}\cdot\theta_1^x/\lfloor x}{e^{-\theta_0}\cdot\theta_0^x/\lfloor x}$$

$$= \log \frac{e^{-\theta_1}\,\theta_1^x}{e^{-\theta_0}\,\theta_0^x}$$

$$= \log e^{-(\theta_1-\theta_0)}\left(\frac{\theta_1}{\theta_0}\right)^x$$

$$= x\log\left(\frac{\theta_1}{\theta_0}\right) - (\theta_1 - \theta_0)$$

Now, taking $\sum_{i=1}^{m}$ both sides

$$\sum_{i=1}^{m} z_i = \sum_{i=1}^{m} x_i \log\left(\frac{\theta_1}{\theta_0}\right) - m(\theta_1 - \theta_0) \qquad ...(5.16)$$

Case I. If $\theta_1 > \theta_0$

$$\Rightarrow \quad \frac{\theta_1}{\theta_0} > 1$$

Taking log, we get

$$\log\left(\frac{\theta_1}{\theta_0}\right) > 0$$

Reject H_0 if $\sum_{i=1}^{m} z_i \geq \log A$

NOTES

We have, from (5.16)

$$\sum_{i=1}^{m} x_i \log\left(\frac{\theta_1}{\theta_0}\right) - m(\theta_1 - \theta_0) \geq \log A$$

$$\Rightarrow \quad \sum_{i=1}^{m} x_i \geq \frac{\log A}{\log\left(\frac{\theta_1}{\theta_0}\right)} + \frac{m(\theta_1 - \theta_0)}{\log\left(\frac{\theta_1}{\theta_0}\right)}$$

$$\text{or} \quad \sum_{i=1}^{m} x_i \geq a + mc \qquad \text{...(5.17)}$$

If Accept H_0, then $\sum_{i=1}^{m} z_i \leq \log B$

$$\Rightarrow \quad \sum_{i=1}^{m} x_i \leq b + mc \qquad \text{...(5.18)}$$

Continue if $\quad \log B < \sum_{i=1}^{m} z_i < \log A$

From (5.17) and (5.18),

$$\Rightarrow \quad b + mc < \sum_{i=1}^{m} x_i < a + mc.$$

Case II. If $\theta_1 < \theta_0$

$$\Rightarrow \quad \frac{\theta_1}{\theta_0} < 1$$

$$\log\left(\frac{\theta_1}{\theta_0}\right) < 0$$

Reject H_0, if $\sum_{i=1}^{m} z_i > \log A$

$$\Rightarrow \quad \sum_{i=1}^{m} x_i \leq a + mc \qquad \text{...(5.19)}$$

Accept H_0 if $\sum_{i=1}^{m} z_i \leq \log B$

$$\Rightarrow \quad \sum_{i=1}^{m} x_i \geq b + mc \qquad \text{...(5.20)}$$

If $$\log B < \sum_{i=1}^{m} z_i < \log A$$

From (5.19) and (5.20), we get

$$a + mc < \sum_{i=1}^{m} x_i < b + mc.$$

NOTES

5.3 OPERATING CHARACTERISTIC (O.C.) FUNCTION OF SPRT

The O.C. function L(θ) is defined as

L(θ) = Probability of accepting $H_0 : \theta = \theta_0$ when θ is the true value of the parameter.

and since the power function

P(θ) = Probability of rejecting H_0 where θ is the true value, we get

$$L(\theta) = 1 - P(\theta) \qquad \ldots(5.21)$$

The O.C. function of a SPRT for testing $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$, in sampling from a population with density function $f(x, \theta)$ is given by

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}} \qquad \ldots(5.22)$$

where for each value of θ, the value of $h(\theta) \neq 0$ is to be determined so that

$$E\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} = 1 \qquad \ldots(5.23)$$

where the constant A and B have already been defined in (5.3). It has been proved that under very simple conditions on the nature of the function $f(x, \theta)$, there exists a unique value of $h(\theta) \neq 0$ such that (5.23) is satisfied.

5.3.1 Derivation of O.C. Function

Theorem 5.1. Statement *If* $z = log \dfrac{F(x_i, \theta_1)}{F(x_i, \theta_0)}$ *and* $P(|z| > 0)$, *then*

$$L(\theta) = \frac{1 - A^h}{B^h - A^h}$$

where h is the root of M (t) = 1

Proof : From the fundamental identity, we have

$$E[e^{S_n t}\{M(t)\}^n] = 1$$

Since h is the root of M (t) = 1

then $$M(h) = 1, \text{therefore at } t = h$$

$$E[e^{S_n t}] = 1$$

Note that $E(B) = \sum_{i=1}^{k} P(A)\, E(B/A_i)$

$$E(e^{S_n t}) = P(Sn \geq \log A)\, E\left[e^{S_n t}/S_n \geq \log A\right]$$
$$+ P(Sn \leq \log B)\, E(e^{S_n h}/S_n \leq \log B) + P(\log B \leq S_n \leq \log A)$$
$$+ E(e^{S_n h}/\log B \leq S_n \leq \log A)$$

Now, we have

$$P(\log B < S_n < \log A) \to 0 \text{ as } n \to \infty$$

(because SPRT terminate with probability 1)

Also, consider the boundry cases

Reject H_0 if $S_n = \log A$

Accept H_0 if $S_n = \log B$

Then,

NOTES

$$E[e^{S_n h}] = P(S_n = \log A)\, E[e^{S_n h}/S_n = \log A]$$
$$+ P(S_n = \log B)\, E(e^{S_n h}/S_n = \log B)$$
$$= P(S_n = \log A)\, E(e^{h \log A}] + P(S_n = \log B) \cdot E[e^{h \log B}]$$
$$1 = P(\text{Reject } H_0/\theta)\, A^h + P(\text{Accept } H_0/\theta)\, B^h$$
$$1 = (1 - L(\theta)]\, A^h + L(\theta)\, B^h$$
$$1 = A^h - A^h L(\theta) + B^h L(\theta)$$
$$1 = A^h - L(\theta)[A^h - B^h]$$

$\Rightarrow$ $$L(\theta) = 1 - A^h / B^h - A^h$$

where $$A = \frac{1-\beta}{\alpha} \text{ and } B = \frac{\beta}{1-\alpha}$$

5.4 AVERAGE SAMPLE NUMBER (A.S.N.)

The sample size n in sequential testing is a random variable which can be determined in terms of the true density function $f(x, \theta)$. The A.S.N. function for the S.P.R.T. for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is given by

$$E(n) = \frac{L(\theta)\log B + [1 - L(\theta)]\log A}{E(Z)} \quad \text{...(5.24)}$$

where $$Z = \log\left(\frac{f(x, \theta_1)}{f(x, \theta_0)}\right), A = \frac{1-\beta}{\alpha}, B = \frac{\beta}{1-\alpha} \quad \text{...(5.25)}$$

5.4.1 Derivation of A.S.N. Function

Theorem 5.2. *Let $X \sim F(x, \theta)$ and we want to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$. Let $z_1, ..., z_n$ be iid random variables defined as $Z = \log \frac{F(x, \theta_1)}{F(x, \theta_0)}$ and n is the number of observation required for reaching a terminal decision, then expected value*

$$E_\theta(n) = \frac{[1 - L(\theta)]\log A + L(\theta)\log B}{E(z)}$$

Proof : Since $S_n = \sum_{i=1}^{n} z_i$, we have

$$E(S_n) = P(S_n \geq \log A)\, E(S_n / S_n \geq \log A)$$
$$+ P(S_n \leq \log B)\, E_\theta(S_n / S_n \leq \log B)$$
$$+ P(\log B < S_n < \log A)\, E[S_n / \log B < S_n < \log A]$$

If the SPRT terminates at the n^{th} step, then either $S_n \geq \log A$, *i.e.*, Reject H_0

or $S_n \leq \log B$, *i.e.*, Accept H_0.

In these two cases, we take approximation

Reject H_0 if $S_n = \log A$

Accept H_0 if $S_n = \log B$

Therefore,

$$E(S_n) = P(S_n = \log A)\log A + P(S_n \leq \log B)\log B$$
$$= P(\text{Reject } H_0/\theta)\log A + P(\text{Accept } H_0/\theta)\log B$$

From Abrahum lemmma

$$E_\theta(n)\,E(z) = \{1 - L(\theta)\}\log A + L(\theta)\log B$$

$$\Rightarrow \quad \boxed{E_\theta(n) = \frac{\{1 - L(\theta)\}\log A + L(\theta)\log B}{E(z)}}$$

5.5 APPLICATION TO BINOMIAL, POISSON AND NORMAL DISTRIBUTION

Example 5.4. *Find OC and ASN functions for $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.*

(I) $X \sim N(\theta, \sigma^2)$, *$\sigma^2$ is known*

(II) $X \sim N(\theta, \sigma^2)$, *$\sigma$ is known*

$$H_0 : \sigma = \sigma_0 \text{ vs } H_1 : \sigma = \sigma_1$$

(III) $X \sim P(\theta)$

(IV) $X \sim B(1, \theta)$

Solution.

(I) For O.C. function, we have

$$F(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$

Now,

$$\frac{F(x, \theta_1)}{F(x, \theta_0)} = \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2}}{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta_0)^2}}$$

$$= \frac{e^{-\frac{1}{2}(x-\theta_1)^2/\sigma^2}}{e^{-\frac{1}{2\sigma^2}(x-\theta_0)^2}}$$

$$= e^{-\frac{1}{2\sigma^2}\left[(x-\theta_1)^2 - (x-\theta_0)^2\right]}$$

$$= e^{-\frac{1}{2\sigma^2}\left[(\theta_1^2 - \theta_0^2) - 2x(\theta_1 - \theta_0)\right]}$$

Since, $\quad M_z(t) = E(e^{tz}) = E[\{e^z\}^t]$

where $\quad z = \log \frac{F(x, \theta_1)}{F(x, \theta_0)} \Rightarrow \frac{F(x, \theta_1)}{F(x, \theta_0)} = e^z$

Therefore,

$$M_z(t) = E\left[\frac{F(x, \theta_1)}{F(x, \theta_0)}\right]^t$$

We have $\quad M(t) = 1$ as $t = h$

Therefore, $\quad M(h) = 1 = E\left[\frac{F(x, \theta_1)}{F(x, \theta_0)}\right]^h$

$$1 = \int_{-\infty}^{\infty} \left\{\frac{F(x, \theta_1)}{F(x, \theta_0)}\right\}^h F(x, \theta)\,dx$$

or $$1 = \int_{-\infty}^{\infty} e^{-\frac{h}{2\sigma^2}\left[\left(\theta_1^2-\theta_0^2\right)-2x\left(\theta_1-\theta_0\right)\right]} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} dx$$

or $$1 = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{h}{2\sigma^2}\left(\theta_1^2-\theta_0^2\right)} \int_{-\infty}^{\infty} e^{\frac{2h}{2\sigma^2}\left(\theta_1-\theta_0\right)x-\frac{1}{2\sigma^2}(x-\theta)^2} dx$$

or $$1 = C_1 \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[x^2+\theta^2-2x\theta-2hx\left(\theta_1-\theta_0\right)\right]} dx$$

NOTES

$$= C_1 \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[\left\{x-\left(\theta+h\theta_1-h\theta_0\right)\right\}\right]^2-\left[\left(\theta+h\theta_1-h\theta_0\right)^2+\theta^2\right]} dx$$

or $$1 = e^{-\frac{h}{2\sigma^2}\left(\theta_1^2-\theta_0^2\right)} \cdot e^{\frac{1}{2\sigma^2}\left[\left(\theta+h\theta_1-h\theta_0\right)^2-\theta^2\right]} \times \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[x-\left(\theta+h\theta_1-h\theta_0\right)\right]^2} dx$$

Put $$\frac{x-\left(\theta+h\theta_1-h\theta_0\right)}{\sigma} = y.$$

$\Rightarrow$ $$dx = \sigma\, dy$$

Therefore, $$1 = C_2 \boxed{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy}$$

$\Rightarrow$ $$C_2 = 1$$

$$e^{-\frac{1}{2\sigma^2}\left[h\left(\theta_1^2-\theta_0^2\right)-\left(\theta+h\theta_1-h\theta_0\right)+\theta^2\right]} = 1.$$

Since, $$e^x = 1 \quad \Rightarrow \quad x = 0$$

$\Rightarrow$ $$\theta_1^2-\theta_0^2-h\theta_1^2-h\theta_0^2-2\theta\theta_1+2\theta\theta_0+2h\theta_0\theta_1 = 0$$

or $$-h\left(\theta_1^2+\theta_0^2-2\theta_1\theta_2\right)-2\theta\left(\theta_1-\theta_0\right)+\left(\theta_1^2-\theta_0^2\right) = 0$$

or $$-h\left(\theta_1-\theta_0\right)^2-2\theta\left(\theta_1-\theta_0\right)+\left(\theta_1-\theta_0\right)\left(\theta_1+\theta_0\right) = 0$$

or $$-h\left(\theta_1-\theta_0\right)-2\theta+\left(\theta_1+\theta_0\right) = 0, \text{ as } \theta_1-\theta_0 \neq 0$$

or $$h\left(\theta_1-\theta_0\right) = \left(\theta_1+\theta_0\right)-2\theta$$

$$\boxed{h = \frac{\theta_1+\theta_0-2\theta}{\theta_1-\theta_0}}$$

or $$\boxed{\theta = \frac{\left(\theta_1+\theta_0\right)-h\left(\theta_1-\theta_0\right)}{2}}$$

Now, given different values to θ, we get corresponding h. Hence a graph between θ and L (θ) gives a O.C. function.

In particular $\quad H_0 : \theta = \theta_0 = 1$

$H_1 : \theta = \theta_1 = 2$

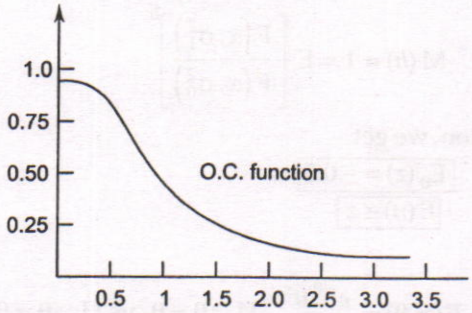
$\alpha = 0.05 = \beta$

Now, $$A = \frac{1-\beta}{\alpha} = 19$$

$$B = \frac{\beta}{1-\alpha} = 1/19$$

$$h = \frac{\theta_1+\theta_2-2\theta}{\theta_1-\theta_0}$$

h	θ	A^h	B^h	$L(\theta)$
−1	2.0	1/19	19	0.05
−2	2.5	1/361	361	0.0028
−3	3.0	1/6859	6859	0.00025
+1	1.0	19	1/19	0.95
+2	0.5	361	1/361	0.997
+3	0	6289	1/6859	0.998



For A.S.N.

$$E(n) = [1 - L(\theta)] \log A + L(\theta) \log B / E_\theta(z)$$

Now $$z = \log \frac{F(x, \theta_1)}{F(x, \theta_0)} = -\frac{1}{2\sigma^2}\left[\left(\theta_1^2 - \theta_0^2\right) - 2x\left(\theta_1 - \theta_0\right)\right]$$

$$E_\theta(z) = -\frac{1}{2\sigma^2}\left[\left(\theta_1^2 - \theta_0^2\right) - 2\theta\left(\theta_1 - \theta_0\right)\right], \quad \text{as} \quad E(x) = \theta$$

Now $$\sigma = 1, \theta_0 = 1, \theta_1 = 2$$
$$\alpha = 0.05 = \beta$$

Thus, giving different values to h, we get θ and calculate to A^h, B^h and $L(\theta)$ and then O.C. function.

Now, A.S.N., we have expected value of n under θ

$$E_\theta(n) = [1 - L(\theta)] \log A + L(\theta) \log B / E_\theta(z)$$

where $$z = \log \frac{F(x, \theta_1)}{F(x, \theta_0)}$$

$$= \log \left(\frac{\theta_1}{\theta_0}\right)^x e^{-(\theta_1 - \theta_0)}$$

$$= x \log\left(\frac{\theta_1}{\theta_0}\right) - (\theta_1 - \theta_0)$$

$\Rightarrow$ $$E_\theta(z) = \theta \log\left(\frac{\theta_1}{\theta_0}\right) - (\theta_1 - \theta_0), \text{ as } E(x) = \theta.$$

(II) $X \sim N(\theta, \sigma^2)$, θ is known

$$H_0 : \sigma = \sigma_0 \text{ vs } H_1 : \sigma = \sigma_1$$

$$F(x, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2}$$

Now, $$\frac{F(x, \sigma_1^2)}{F(x, \sigma_0^2)} = \frac{1}{\sigma_1\sqrt{2\pi}} \times \frac{e^{-\frac{1}{2\sigma_1^2}(x-\theta_1)^2}}{e^{-\frac{1}{2\sigma_0^2}(x-\theta_0)^2}} \times \sigma_0\sqrt{2\pi}$$

$$= \frac{\sigma_0}{\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\theta_0)^2} \times e^{+\frac{1}{2\sigma_0^2}(x-\theta_1)^2}$$

$$= \frac{\sigma_0}{\sigma_1} e^{\frac{1}{2\sigma^2}\left[\frac{(x-\theta_0)^2}{\sigma_0^2} - \frac{(x-\theta_1)^2}{\sigma_1^2}\right]}$$

$$= \left(\frac{\sigma_0}{\sigma_1}\right) e^{-\frac{1}{2}(x-\theta)^2\left[\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right]}$$

NOTES

Now, $\quad M(h) = 1 = E\left[\frac{F(x, \sigma_1^2)}{F(x, \sigma_0^2)}\right]^h$

After simplification, we get

$$\boxed{E_\theta(z) = -0.5}$$

$$\boxed{E(n) \cong z}$$

(III) $X \sim P(\theta)$

$$F(x, \theta) = \frac{e^{-\theta}\theta^x}{\lfloor x}, H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta = \theta_1$$

$$\frac{F(x, \theta_1)}{F(x, \theta_0)} = \frac{e^{-\theta_1}\theta_1^x}{\lfloor x \times e^{-\theta_0}\theta_0^x} \times \lfloor x$$

$$= e^{\theta_0 - \theta_1} \frac{\theta_1^x}{\theta_0^x} = \left(\frac{\theta_1}{\theta_0}\right)^x e^{\theta_0 - \theta_1}$$

Now $\quad M(h) = 1 = E\left[\frac{F(x, \theta_1)}{F(x, \theta_0)}\right]^h$

$$= \sum_{x=0}^{\infty}\left[\frac{F(x, \theta_1)}{F(x, \theta_0)}\right]^h F(x, \theta)$$

or

$$1 = \sum_{x=0}^{\infty}\left(\frac{\theta_1}{\theta_0}\right)^{hx} e^{-h(\theta_1 - \theta_0)} \frac{e^{-\theta}\theta^x}{\lfloor x}$$

$$1 = e^{-h(\theta_1 - \theta_0) - \theta} \sum_{x=0}^{\infty}\left[\left(\frac{\theta_1}{\theta_0}\right)^h \theta\right]^x \Big/ \lfloor x$$

$$1 = e^{-h(\theta_1 - \theta_0) - \theta} \sum_{x=0}^{\infty} \frac{m^x}{\lfloor x}, \text{ where } \quad m = \left(\frac{\theta_1}{\theta_0}\right)^h \cdot \theta$$

$$1 = e^{-h(\theta_1 - \theta_0) - \theta} \cdot e^m$$

$$1 = e^{-h(\theta_1 - \theta_0) - \theta} \cdot e^{\left(\frac{\theta_1}{\theta_0}\right)^h \theta}$$

$$1 = e^{\theta\left(\frac{\theta_1}{\theta_0}\right)^h - h(\theta_1 - \theta_0) - \theta}$$

$$1 = e^{\theta\left[\left(\frac{\theta_1}{\theta_0}\right)^h - 1\right] - h(\theta_1 - \theta_0)}$$

Since $\quad e^x = 1 \;\Rightarrow\; x = 0.$

Now, we get

$$\theta\left[\left(\frac{\theta_1}{\theta_0}\right)^h - 1\right] - h(\theta_1 - \theta_0) = 0$$

$$\Rightarrow \qquad \theta = \frac{h(\theta_1 - \theta_0)}{\left[\left(\frac{\theta_1}{\theta_0}\right)^h - 1\right]}.$$

NOTES

$$1 = \int_{-\infty}^{\infty}\left[\frac{F(x,\sigma_1^2)}{F(x,\sigma_0^2)}\right]^h F(x,\sigma^2)\,dx$$

$$1 = \int_{-\infty}^{\infty}\left(\frac{\sigma_0}{\sigma_1}\right)^h\left[e^{-\frac{1}{2}(x-\theta)^2\times\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)}\right]^h dx \times \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$

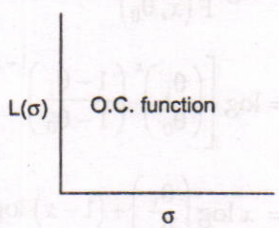
$$1 = \left(\frac{\sigma_0}{\sigma_1}\right)^h \cdot \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2\left[\frac{1}{\sigma^2}+\frac{h}{\sigma_1^2}-\frac{h}{\sigma_0^2}\right]}dx$$

$$1 = \left(\frac{\sigma_0}{\sigma_1}\right)^h \cdot \frac{1}{\sigma} \times \frac{\sigma^*}{\sigma^*\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^{*2}}(x-\theta)^2}\,dx$$

where $$\sigma^{*^2} = \frac{1}{\frac{1}{\sigma^2}+\frac{h}{\sigma_1^2}-\frac{h}{\sigma_0^2}}$$

$$\Rightarrow \qquad 1 = \left(\frac{\sigma_0}{\sigma_1}\right)^h \cdot \frac{1}{\sigma} \cdot \frac{1}{\sqrt{\frac{1}{\sigma^2}+\frac{h}{\sigma_1^2}-\frac{h}{\sigma_0^2}}}$$

For different choice of h, we get σ and hence L (σ) and then O.C. function.



For A.S.N., we have expected value

$$E_\sigma(n) = \frac{[1 - L(\sigma)]\log A + L(\sigma)\log B}{E_\sigma(z)}$$

where $$z = \log\frac{F(x,\sigma_1^2)}{F(x,\sigma_0^2)}$$

$$= \log\left[\frac{\sigma_0}{\sigma_1} - \frac{(x-\theta)^2}{2}\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)\right]$$

$$\Rightarrow \qquad E_\sigma(z) = \log\frac{\sigma_0}{\sigma_1} - \frac{\sigma^2}{2}\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right) \text{ as } E(x-\theta)^2 = \sigma^2.$$

(IV) $X \sim B(1, \theta)$.

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta = \theta_1$$

We have

$$F(x, \theta) = \theta^x (1-\theta)^{1-x}$$

and

NOTES

$$\frac{F(x,\theta_1)}{F(x,\theta_0)} = \frac{\theta_1^x (1-\theta_1)^{1-x}}{\theta_0^x (1-\theta_0)^{1-x}}$$

$$= \left(\frac{\theta_1}{\theta_0}\right)^x \left(\frac{1-\theta_1}{1-\theta_0}\right)^{(1-x)}$$

Now,
$$M(h) = 1 = E\left[\frac{F(x,\theta_1)}{F(x,\theta_0)}\right]^h$$

$$= \sum_{x=0}^{1} \left(\frac{\theta_1}{\theta_0}\right)^{hx} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(1-x)} \theta^x (1-\theta)^{1-x}$$

$$1 = \left(\frac{1-\theta_1}{1-\theta_0}\right)^h (1-\theta) + \left(\frac{\theta_1}{\theta_0}\right)^h \theta$$

$$\Rightarrow \qquad \theta = 1 - \left(\frac{1-\theta}{1-\theta}\right)^h \Big/ \left(\frac{\theta_1}{\theta_0}\right)^h - \left\{\frac{(1-\theta_1)}{(1-\theta_0)}\right\}^h.$$

Thus, given different value to h, we get θ and hence $L(\theta)$ and then O.C. function.

For A.S.N., we have $E_\theta(n) = \dfrac{[1-L(\theta)]\log A + L(\theta)\log B}{E_\theta(z)}$

where,
$$Z = \log \frac{F(x,\theta_1)}{F(x,\theta_0)}$$

$$= \log\left[\left(\frac{\theta_1}{\theta_0}\right)^x \left(\frac{1-\theta_1}{1-\theta_0}\right)^{1-x}\right]$$

$$= x \log\left(\frac{\theta_1}{\theta_0}\right) + (1-x)\log\left(\frac{1-\theta_1}{1-\theta_0}\right)$$

$$E_\theta(z) = \theta \log\left(\frac{\theta_1}{\theta_0}\right) + (1-\theta)\log\left(\frac{1-\theta_1}{1-\theta_2}\right) \quad \text{as} \quad E(x) = \theta.$$

Example 5.5. *Give the S.P.R.T. for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 (> \theta_0)$ in sampling from a normal density:*

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

where σ is known. Also obtain its O.C. function and A.S.N. function.

Solution.
$$\frac{f(x_i,\theta_1)}{f(x_i,\theta_0)} = \exp\left[-\frac{1}{2\sigma^2}\left((x_i-\theta_1)^2 - (x_i-\theta_0)^2\right)\right]$$

$$= \exp\left[-\frac{1}{2\sigma^2}\{(\theta_0-\theta_1)(2x_i-\theta_0-\theta_1)\}\right] \quad \ldots(5.26)$$

$$\therefore \quad z_i = \log\frac{f(x_i,\theta_1)}{f(x_i,\theta_0)} = \frac{\theta_1-\theta_0}{\sigma^2}\left(x_i-\frac{\theta_0+\theta_1}{2}\right) \quad \ldots(5.27)$$

$$\Rightarrow \quad \log\lambda_m = \sum_{i=1}^{m} z_i = \frac{\theta_1-\theta_0}{\sigma^2}\left[\sum_i x_i - \frac{m(\theta_0+\theta_1)}{2}\right]$$

Hence the S.P.R.T. for $H_0:\theta=\theta_0$ against $H_1:\theta=\theta_1$ is given by (5.2)

(*i*) Reject H_0 if

$$\frac{\theta_1-\theta_0}{\sigma^2}\left[\sum x_i - \frac{m(\theta_0+\theta_1)}{2}\right] \geq \log\left(\frac{1-\beta}{\alpha}\right)$$

$$\Rightarrow \quad \sum_{i=1}^{m} x_i \geq \frac{\sigma^2}{\theta_1-\theta_0}\log\left(\frac{1-\beta}{\alpha}\right)+\frac{m(\theta_0+\theta_1)}{2}; \; (\theta_1>\theta_0)$$

(*ii*) Accept H_0 if

$$\frac{\theta_1-\theta_0}{\sigma^2}\left[\sum x_i - \frac{m(\theta_0+\theta_1)}{2}\right] \leq \log\left(\frac{\beta}{1-\alpha}\right)$$

$$\Rightarrow \quad \sum_{i=1}^{m} x_i \leq \frac{\sigma^2}{\theta_1-\theta_0}\log\left(\frac{\beta}{1-\alpha}\right)+\frac{m(\theta_0+\theta_1)}{2}; \; (\theta_1>\theta_0) \text{ and}$$

(*iii*) Continue taking additional observations as long as

$$\log\left(\frac{\beta}{1-\alpha}\right) < \frac{\theta_1-\theta_0}{\sigma^2}\left[\sum x_i - \frac{m(\theta_0+\theta_1)}{2}\right] < \log\left(\frac{1-\beta}{\alpha}\right)$$

$$\Rightarrow \quad \frac{\sigma^2}{\theta_1-\theta_0}\log\left(\frac{\beta}{1-\alpha}\right)+\frac{m(\theta_0+\theta_1)}{2} < \sum x_i < \frac{\sigma^2}{\theta_1-\theta_0}\log\left(\frac{1-\beta}{\alpha}\right) + \frac{m(\theta_0+\theta_1)}{2}, \; \theta_1>\theta_0$$

O.C. Function. First of all we shall determine $h = h(\theta) \neq 0$ from (5.23) *i.e.*, from

$$\int_{-\infty}^{\infty}\left[\frac{f(x,\theta_1)}{f(x,\theta_0)}\right]^h f(x,\theta)\,dx = 1$$

$$\Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right]\cdot\left[\exp\left\{-\frac{1}{2\sigma^2}(\theta_1-\theta_0)\times(-2x+\theta_0+\theta_1)\right\}\right]^h dx = 1,$$

[on using (5.26)]

$$\Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\sigma^2}\left\{x^2-2x\left((\theta_1-\theta_0)h+\theta\right)+\theta^2+\left(\theta_1^2-\theta_0^2\right)h\right\}\right]dx = 1$$

If we take

$$\left.\begin{aligned}\lambda &= (\theta_1-\theta_0)\,h+\theta\\ \lambda^2 &= \left(\theta_1^2-\theta_0^2\right)h+\theta^2\end{aligned}\right\} \quad \ldots(5.28)$$

then L.H.S. becomes: $\dfrac{1}{\sigma\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\sigma^2}(x-\lambda)^2\right]dx,$

which being the total area under normal probability curve with mean λ and variance σ^2 is always unity, as desired. Thus $h = h(\theta)$ is the solution of (5.28) and is given by

$$\left(\theta_1^2 - \theta_0^2\right) h + \theta = \{(\theta_1 - \theta_0)h + \theta\}^2 \quad \Rightarrow \quad \left(\theta_1^2 - \theta_0^2\right) h = (\theta_1 - \theta_0)^2 h^2 + 2\theta(\theta_1 - \theta_0)\, h$$

Since $h = h(\theta) \neq 0$ and $\theta_1 \neq \theta_0$, on dividing throughout by $(\theta_1 - \theta_0)\, h$, we get

$$(\theta_1 + \theta_0) = (\theta_1 - \theta_0)\, h + 2\theta \;\Rightarrow\; h(\theta) = \frac{\theta_1 + \theta_0 - 2\theta}{\theta_1 - \theta_0}$$

NOTES

Substituting for $h(\theta)$ in (5.22), we get the required expression for the O.C. function.

A.S.N. function. We have

$$Z = \log \frac{f(x, \theta_1)}{f(x, \theta_0)} = \frac{\theta_1 - \theta_0}{\sigma^2}\left(x - \frac{\theta_0 + \theta_1}{2}\right), \text{[From (5.27)]}$$

$$\therefore \qquad E(Z) = \frac{\theta_1 - \theta_0}{2\sigma^2}(2E(x) - \theta_0 - \theta_1) = \frac{\theta_1 - \theta_0}{2\sigma^2}(2\theta - \theta_0 - \theta_1)$$

Substituting in (5.24), we get the required A.S.N. function.

Example 5.6. *Let X have the distribution*

$$f(x, \theta) = \theta^x (1 - \theta)^{1-x};\ x = 0, 1;\ 0 < \theta < 1$$

For testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, construct S.P.R.T. and obtain its A.S.N. and O.C. Function.

Solution. We have

$$\lambda_m = \frac{L(x_1, x_2, \ldots, x_m | H_1)}{L(x_1, x_2, \ldots, x_m | H_0)}$$

$$= \left\{\theta_1^{\sum_{i=1}^m x_i} (1 - \theta_1)^{m - \sum_{i=1}^m x_i}\right\} \div \left(\theta_0^{\sum x_i} (1 - \theta_0)^{m - \sum x_i}\right)$$

$$= \left(\frac{\theta_1}{\theta_0}\right)^{\sum_{i=1}^m x_i} \cdot \left(\frac{1 - \theta_1}{1 - \theta_0}\right)^{m - \sum_{i=1}^m x_i}$$

$$\log \lambda_m = \sum x_i \log(\theta_1/\theta_0) + \left(m - \sum x_i\right) \log\left(\frac{1 - \theta_1}{1 - \theta_0}\right)$$

$$= \sum_{i=1}^m x_i \log\left[\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)}\right] + m \log\left(\frac{1 - \theta_1}{1 - \theta_0}\right)$$

Hence SPRT for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is given by (5.2)

(*i*) Accept H_0 if $\quad \log \lambda_m \leq \log\left(\frac{\beta}{1 - \alpha}\right) = b$, (say)

i.e., if

$$\sum_{i=1}^m x_i \leq \frac{b - m \log\left[(1 - \theta_1)/(1 - \theta_0)\right]}{\log\left[\theta_1(1 - \theta_0)/\theta_0(1 - \theta_1)\right]} = a_m, \text{ (say).}$$

(*ii*) Reject H_0 (Accept H_1) if $\log \lambda_m \geq \log\left(\frac{1 - \beta}{\alpha}\right) = a$, (say)

i.e., if

$$\sum_{i=1}^m x_i \geq \frac{a - m \log\left[(1 - \theta_1)/(1 - \theta_0)\right]}{\log\left[\theta_1(1 - \theta_0)/\theta_0(1 - \theta_1)\right]} = r_m, \text{ (say).}$$

(iii) Continue sampling if

$$b < \log \lambda_m < a \quad \Rightarrow \quad a_m < \Sigma x_i < r_m$$

O.C. Function. O.C. function is given by:

$$L(\theta) = [A^{h(\theta)} - 1]/[A^{h(\theta)} - B^{h(\theta)}] \quad \text{[using (5.22)]} \quad \ldots(5.29)$$

where for each value of θ, $h(\theta) \neq 0$ is to be determined such that

NOTES

$$E\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} = 1 \quad \text{[using (5.23)]}$$

$$\Rightarrow \quad \sum_{x=0}^{1}\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} f(x, \theta) = 1$$

$$\Rightarrow \quad \sum_{x=0}^{1}\left[\left(\frac{\theta_1}{\theta_0}\right)^x \left(\frac{1-\theta_1}{1-\theta_0}\right)^{1-x}\right]^{h(\theta)} \theta(1-\theta)^{1-x} = 1$$

$$\Rightarrow \quad \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(\theta)} \cdot (1-\theta) + \left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} \cdot \theta = 1 \quad \ldots(5.30)$$

The solution of this equation for $h = h(\theta)$ is very tedious. From practical point of view, instead of solving (5.30) for h we regard h as a parameter and solve it for θ; thus giving

$$\theta\left[\left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} - \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(\theta)}\right] = 1 - \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(\theta)}$$

$$\Rightarrow \quad \theta = \frac{1 - \left[(1-\theta_1)/(1-\theta_0)\right]^{h(\theta)}}{(\theta_1/\theta_0)^{h(\theta)} - \left[(1-\theta_1)/(1-\theta_0)\right]^{h(\theta)}} = \theta(h), \text{(say)}. \quad \ldots(5.31)$$

Using (5.29), we have:

$$L(\theta) = \frac{\left[(1-\beta)/\alpha\right]^h - 1}{\left[(1-\beta)/\alpha\right]^h - \left[\beta/(1-\alpha)\right]^h} = L(\theta, h), \text{(say)}. \quad \ldots(5.32)$$

Various points on the O.C. curve are obtained by assigning arbitrary values to 'h' and computing the corresponding values of θ and L(θ) from (5.31) and (5.32) respectively.

A.S.N. Function

$$Z = \log\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]; \; A = \frac{1-\beta}{\alpha}, \; B = \frac{\beta}{1-\alpha}$$

$$\therefore \quad E(Z) = \sum_{x=0}^{1} \log\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right] \cdot f(x, \theta)$$

$$= \sum_{x=0}^{1} \log\left[\left(\frac{\theta_1}{\theta_0}\right)^x \left(\frac{1-\theta_1}{1-\theta_0}\right)^{1-x}\right] \cdot \theta^x (1-\theta)^{1-x}$$

$$= (1-\theta)\log\left(\frac{1-\theta_1}{1-\theta_0}\right) + \theta \cdot \log\left(\frac{\theta_1}{\theta_0}\right)$$

$$= \theta \log\left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right] + \log\left(\frac{1-\theta_1}{1-\theta_0}\right) \quad \ldots(5.33)$$

A.S.N. is given by

$$E(n) = \frac{L(\theta)\log B + [1 - L(\theta)] \cdot \log A}{E(Z)} \quad \ldots(5.34)$$

Substituting the values of E(Z) and $L(\theta)$ from (5.33) and (5.32) in (5.34) we get the A.S.N. function.

NOTES

Remark 5.1. *If h assumes negative values i.e., if instead of h we take – h where h > 0, then*

$$L(\theta, -h) = \frac{A^{-h} - 1}{A^{-h} - B^{-h}} = \left(\frac{1 - A^h}{B^h - A^h}\right) B^h = \left(\frac{A^h - 1}{A^h - B^h}\right) \cdot B^h$$

$$\Rightarrow \quad L(\theta, -h) = B^h . L(\theta, h) \quad \ldots(5.35)$$

and

$$\theta(-h) = \frac{[(1-\theta_1)/(1-\theta_0)]^h - 1}{[(1-\theta_1)/(1-\theta_0)]^h - (\theta_1/\theta_0)^h} \left(\frac{\theta_1}{\theta_0}\right)^h \text{ [From (5.31)]}$$

$$= \theta(h) \cdot \left(\frac{\theta_1}{\theta_0}\right)^h \quad \ldots(5.36)$$

Formulae (5.35) and (5.36) are very convenient to use for obtaining the points on O.C. curve for arbitrary negative values of h.

5.6 SUMMARY

- The best known procedure in sequential testing is the *Sequential Probability Ratio Test* (SPRT) developed by A. Wald.
- The sample size n in sequential testing is a random variable which can be determined in terms of the true density function $f(x, \theta)$.
- The O.C. function $L(\theta)$ is defined as the probability of accepting $H_0 : \theta = \theta_0$ when θ is the true value of the parameter.

5.7 GLOSSARY

Null Hypothesis : Null hypothesis is a hypothesis which is the to tested for the positive rejection under the assumption that it is true.

Normal distribution : A random variable X is said to follow normal distribution with parameter mean (μ) and Variance (σ^2) if its p.d.f. is given by

$$f(x\,;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\lambda}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty \le x \le \infty, -\infty \le \mu \le \infty \text{ and } \sigma > 0.$$

Poisson distribution has only one parameter.

Binomial distribution. It is a discrete probability distribution which has only two parameters. The p.m.f. of binomial distribution is

$$P(X = x) = {}^nC_x\, p^x q^{n-x}, x = 0, 1, 2, \ldots, n$$

5.8 REVIEW QUESTIONS

1. Define sequential analysis. Explain how the sequential test procedure differs from the Neyman-Pearson test procedure.

2. Describe sequential probability ratio test (SPRT). Drive inequalities involving A, B, α and β where symbols have their usual meanings.
3. Define SPRT and explain its properties.
4. Define the OC function and ASN function in sequential analysis.
5. Let x have the distribution

$$f(x, \theta) = \theta^x (1 - \theta)^{1-x} ; x = 0, 1 ; 0 < \theta < 1$$

For testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, construct SPRT and obtain its ASN and OC functions.

6. Set up SPRT for testing the variance of a normal distribution with known mean.

5.9 FURTHER READINGS

Gupta, S.C. and Kapoor, V.K., fundamentals of mathematical statistics, sultan chand & Sons, New Delhi.

Agarwal, B.L., Basic statistics, New Age International Publishers, New Delhi.

Bradley, J.V., Distribution free Statistical Tests, Prentice Hall, Englewood cliffs, 1968.

Fraser, D.A.S., Nonparametric Methods in Statistics, John Wiley, New York, 1957.

Wilcoxon, F., "Probability Tables for Individual Comparisons by Ranking Methods". Biometrics, 1, 80–83, 1945.