CHAPTER-I

INTRODUCTION & SUMMARY

CHAPTER-I

INTRODUCTION AND SUMMARY

Introduction:

This chapter is introductory. In section (1.1) the motivation to the problem is given, in brief, with definition and concept of confidence interval. Section (1.2) contains concept of fixed sample size procedure, its limitations, and the need for sequential procedure. Some preliminary definitions and results that are frequently used in the subsequent chapters of dissertation are also reported. The last section includes the literature survey of fixed-width confidence interval and chapter-wise summary.

1.1. Motivation to the problem:

In the problems of statistical inference, sometimes we are not interested in deriving the point estimator for parameters, or obtaining test of the hypothesis, but like to establish a lower and upper bound for the real valued parameter. For example, if T is time of survival of an equipment, we are interested in obtaining a lower bound for the probability that the equipment will survive beyond some specified time, that is a reliability function. If X denotes the nicotine content of a particular brand of cigeratte, one may interested in obtaining a lower and an upper bound to the average nicotine content. In other words, experimenter is interested in constructing the family of sets that contains true value of parameter with specified (usually high) probability. The problem of this kind is called the problem of confidence estimation or interval estimation. The method of interval estimation consist of determining a subset of parameter space that contains the true value of the parameter with certain specified probability. The theory of confidence intervals was first introduced by Neyman during (1941) and further developed by many people.

Now, in the following we give general definition of confidence interval.

Definition (1.1.1) : Confidence Interval:

Let $\mathbb{F} = \left\{ F_\theta : \theta \in \Theta \right\}$ be a family of distribution functions (d.f.) of a random variable, where Θ is an interval on real line. Consider a subinterval of Θ, say $S(X) = (L(X), U(X))$ where limits depends on observed random variable X. Let $\alpha \in (0,1)$. Then $S(X)$ is called a $(1-\alpha)$ level confidence interval for θ, if for all $\theta \in \Theta$,

$$P_{\theta}\left\{ S(X) \ni \theta \right\} \geq (1-\alpha). \qquad \ldots(1.1.1)$$

The $(1-\alpha)$ on right hand side of (1.1.1) is called <u>confidence level</u>, L(X) and U(X) are called <u>lower</u> and <u>upper</u> <u>confidence</u> <u>limits</u> respectively.

Note that, S(X) is random interval in Θ. The quantity

$$\sup_{\theta\in\Theta} P_{\theta}\left\{ S(X) \ni \theta \right\} \qquad \ldots(1.1.2)$$

is called <u>confidence</u> <u>coefficient</u> of S(X).

Generally the precision of a family of interval estimators is measured in terms of length of confidence interval and their coverage probability. In many situations, we are interested in constructing a confidence interval of specified width and having a specified coverage probability. Such a confidence interval is called as fixed-width confidence interval. Now we define the fixed-width confidence interval.

<u>Definition (1.1.2) : Fixed-width confidence interval:</u>

A family of confidence interval S(X) is called $(1-\alpha)$ level fixed-width confidence interval if for pre-specified d $(d>0)$, S(X) has a fixed length d and

$$P_{\theta}\left\{ S(X) \ni \theta \right\} \geq (1-\alpha), \text{ for all } \theta \in \Theta. \qquad \ldots(1.1.3)$$

Generally θ is a real parameter or some real functional on family $\mathbb{F}$ of distribution functions. The problem of fixed-width confidence interval is that of determining whether such family exists and providing the sampling scheme, if required, which will gaurantee that a proper statistic T(X) which satisfies the definition (1.1.2).

1.2. Fixed sample size procedure its limitations and Concept of Sequential procedure, Some Preliminary definitions and Results:

In Statistical inference problems (point estimation, testing of hypothesis and interval estimation) the decision is usually based on the assumption that sample size 'n', fixed in advance, is available. This type of decision procedure is known as fixed sample size procedure. The fixed sample size procedure doesnot take into account the cost of sampling. Thus, the problem is to determine the minimum sample size, such that the decision should be taken with prefixed desired accuracy.

But in some situations,if we fix sample size in advance, we may not able to achieve the desired goal. To illustrate the same we consider in the following the problem of fixed-width confidence interval for mean μ of normal distribution with variance σ^2, when both μ and σ^2 are unknown.

Example(1.2.1): Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu,\sigma^2)$, where n is fixed in advance. Then $(1-\alpha)$ level confidence interval for μ is given by

$$\left(\bar{X}_n - S/(n)^{1/2}\, t_{n-1,\alpha/2} \,,\ \bar{X}_n + S/(n)^{1/2}\, t_{n-1,\alpha/2} \right) \quad \ldots(1.2.1)$$

where $S^2=(n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ and $t_{n-1,\alpha/2}$ is $100(\alpha/2)\%$ point of Student's t-distribution with (n-1) degrees of freedom.

The length of this confidence interval is $2S/(n)^{1/2}\, t_{n-1,\alpha/2}$ which is a random variable. In this case, we have no control on the width of the confidence interval. If σ^2 is known, a $(1-\alpha)$ level confidence interval for μ is given by

$$\left(\bar{X}_n - \sigma/(n)^{1/2}\, Z_{\alpha/2} \,,\ \bar{X}_n + \sigma/(n)^{1/2}\, Z_{\alpha/2} \right)\ldots(1.2.2)$$

This confidence interval has length $2\sigma/(n)^{1/2}\, Z_{\alpha/2}$. This length will be atmost 2d, where (d>0) fixed in advance, provided we choose the sample size 'n' such that

$$d \geq \sigma(n)^{-1/2} Z_{\alpha/2}.$$

That is,

$$n \geq \frac{\sigma^2 Z^2_{\alpha/2}}{d^2} . \qquad \ldots(1.2.3)$$

Thus, if we take a random sample of size $n = \left[\frac{\sigma^2 Z^2_{\alpha/2}}{d^2} \right] + 1$ then the confidence interval has width 2d and confidence coefficient atleast $(1-\alpha)$, where $[x]$ stands for the largest integer smaller than x. However, if σ^2 is unknown we cannot find this sample size 'n'. Thus, in fixed sample size procedure it is not always possible to minimise the number of observations that are required to arrive at the decision that is optimum in some sense.

An alternative procedure suggests itself why not take the observations sequentially, that is, one at a time and use information provided by observations to date to determine whether the further observation is necessary or not. In such case sample size is random variable. The procedure of taking a decision by above method is called as <u>sequential decision procedure</u>.

Now we consider some basic notions of sequential decision procedure. Given an infinite sequence of random variables, say, $X_1, X_2, \ldots$, the statistician faces the problem of providing a set of rules that tells experimenter when to stop the sampling; once a sampling is terminated after taking, say n observations, the decision problem is treated as a fixed sample size problem.

Let Θ be a parameter space and $\mathbb{D}$ is the set of decisions available, that is decision space, to the statistician. We assume that the random variables $X_1, X_2, \ldots$ that are observed sequentially are independent and identically distributed (i.i.d.) and let $f(x;\theta)$; $\theta \in \Theta$, be the common probability density function(p.d.f.) of X_i, $i=1,2,\ldots$.

Definition(1.2.1):Components of sequential decision procedure:

A sequential decision procedure has two components.

(a)Sampling plan or Stopping rule:

One component of sequential decision procedure is called sampling plan or stopping rule. The statistician first specifies whether decision in $\mathbb{D}$ should be chosen without any observation or whether atleast one observation should be taken. If atleast one observation is taken, the statistician specifies for every possible set of observed values $X_1=x_1, \ldots, X_n=x_n$ $(n\geq 1)$ whether sampling should be stopped and a decision in $\mathbb{D}$ is chosen without further observation or whether another observation X_{n+1} should be taken.

(b) Decision rule:

The second component of the sequential decision procedure is called decision rule. If no observation is taken, the statistician specifies decision $d_0 \in \mathbb{D}$ that is to be chosen. If atleast one observation is taken, the statistician specifies decision $d_n(x_1, x_2, \ldots, x_n) \in \mathbb{D}$ that is to be chosen for each possible set of observed values $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ after which sampling might be terminated.

In the following we give some more definitions, useful in due course of discussion.

Definition(1.2.1):Stopping Region:

Let $R_n \subset \mathbb{R}_n$, n=1,2,..., be a sequence of Boral-measurable sets such that sampling is terminated after observing $X_1 = x_1, \ldots, X_n = x_n$ if $(x_1, \ldots, x_n) \in R_n$. If $(x_1, \ldots, x_n) \notin R_n$ another observation X_{n+1} is taken. The set $\{R_n, n=1,2,\ldots\}$ are called stopping regions. Here $\mathbb{R}_n$ is n-dimensional euclidean space.

Definition(1.2.2):Stopping Rule:

With every sequential stopping rule we associate a stopping random variable N, which takes on values 1, 2, ... N, is the (random) total number of observations taken before sampling is terminated.

Let [N = n] denotes the event that sampling is stopped after observing n values $x_1, x_2, \ldots, x_n$ and not before. Thus

$[N = 1] = R_1$ and

$[N = n] = [(x_1, x_2, \ldots, x_n) \in \mathbb{R}_n :$ sampling is stopped after observing x_n and not before $]$

$$= (R_1 \cup R_2 \cup \ldots \cup R_{n-1})' \cap R_n$$

$$= R_1' \cap R_2' \cap \ldots \cap R_{n-1}' \cap R_n .$$

Note that $[N = n]$, and the event $[N \leq n] = \bigcup_{k=1}^{n} [N = k]$ depends only on observations $X_1, X_2, \ldots, X_n$ and not on $X_{n+1}, X_{n+2}, \ldots \ldots$

Definition(1.2.3):Closed sequential procedure:

A sequential procedure for which sampling eventually terminates with probability one is called closed sequential procedure. That is,

$$P_{\theta}[N < \infty] = 1$$

$$\text{or } P_{\theta}[N = \infty] = 1 - P_{\theta}[N < \infty] = 0.$$

Definition(1.2.4):Sequential procedure:

It is assumed that, we may record as many observations as we like. Suppose that n-observations are recorded and let $T_n = T(X_1, X_2, \ldots, X_n)$ be a statistic. Let d $(d>0)$ be a given positive number and Let

$$I_n = \left[T_n - d , T_n + d \right] \qquad \ldots(1.2.4)$$

A sequential procedure S, for constructing a fixed-width confidence interval, is a pair (I_N, N), where $N=N(d)$ is stopping rule and when $N=n$,the interval I_N is to be used as fixed-width confidence interval.

In following, we introduce the definitions of asymptotic consistency, asymptotic efficiency, asymptotic relative efficiency for fixed-width sequential confidence intervals. These definitions are due to Chow and Robbins(1965).

Assume that $F \in \mathbb{F}$, the class of d.f. having a finite second moment and Let $\mu(F)=E_F(X)$ and $\sigma^2(F) = Var_F(X)$. Then

$$\lim_{n \to \infty} P\left\{ \frac{n^{1/2}(\bar{X}_n - \mu(F))}{\sigma(F)} \le x \right\} = \Phi(x), \text{ all } F \in \mathbb{F}, \quad \ldots(1.2.5)$$

where $\bar{X}_n$ is the sample mean based on $X_1, X_2, \ldots, X_n$ and $\Phi(.)$ is d.f. of standard normal variate.

Definition (1.2.5) : Exact Consistency:

Given a preassigned number α, $\alpha \in (0,1)$, we have to construct a confidence interval for any parametric function $g(F)$ such that,

$$P_F\left\{ g(F) \in I_n \right\} \ge (1-\alpha). \quad \ldots(1.2.6)$$

The property (1.2.6) is reffered as consistency or exact consistency.

Definition(1.2.6):Asymptotic consistency:

A fixed-width confidence interval

$$I_N = \left[T_N - d \, , \, T_N + d \right]$$

based on stopping random variable N is said to be asymptotically consistent if,

$$\lim_{d \to 0} P_F \left\{ g(F) \in I_N \right\} \geq (1-\alpha), \text{ for all } F \in \mathbb{F},$$

for some α, $\alpha \in (0,1)$.

Definition(1.2.7):Asymptotic efficiency:

An asymptotically consistent sequential procedure S, for constructing confidence interval for $\mu(F)$, when $\sigma^2(F)$ is unknown, is said to be asymptotically efficient if,

$$\lim_{d \to 0} \frac{E_F\left[N_{(d)} \right]}{n_0\left[d, \sigma^2(F) \right]} = 1, \text{ for all } F \in \mathbb{F},$$

where $n_0(d,\sigma^2(F)) = \dfrac{a^2\sigma^2(F)}{d^2}$ and $\sigma^2(F)$ is known and $a=\Phi^{-1}(\alpha/2)$.

Definition(1.2.8):Asymptotic relative efficiency:

Let S_1 and S_2 ~~are~~ be two closed sequential procedures, then asymptotic relative efficiency(ARE) of S_1 with respect to S_2 is defined as

$$e(S_1,S_2) = \lim_{d\to 0}\left[\frac{E_\theta(N_{S1})}{E_\theta(N_{S2})}\right]$$

and procedure S_1 is asymptotically more efficient than that of S_2 if $e(S_1,S_2) \le 1$. If equality holds then both procedures are equally efficient.

In asymptotic theory of sequential procedures we have to obtain asymptotic distribution of the randomly indexed random variables. Anscombe(1952) has introduced the conditions of uniform continuity in probability. These conditions are stated as follows,

Let $\{Y_n\}$ be an infinite sequence of random variables and suppose that there exist a real number θ, a sequence of positive numbers $\{w_n\}$ and a d.f. F(x), such that the following conditions are satisfied.

(C-I)Convergence in law of Y_n:

For any continuity point x of F(x),

$$P\left\{ Y_n - \theta \le xw_n \right\} \xrightarrow{L} F(x), \text{ as } n \to \infty.$$

(C-II)Uniform continuity in probability of Y_n:

Given any $\varepsilon>0$ and $\eta>0$, there exists a large ν and small positive c such that for any $n>\nu$,

$P\left\{ |Y_{n'} - Y_n| < \varepsilon w_n \text{ simultaneously for all integers } n' \text{ such that } |n'-n| < cn \right\} > 1-\eta.$

Theorem(1.2.1): (Anscombe)

Let $\{n_t\}$ be increasing sequence of integers tending to ∞ and let $\{N_{(t)}\}$ be a sequence of positive integer valued proper random variables such that $\frac{N_{(t)}}{n_t} \to 1$ in probability as $n \to \infty$ then, if the sequence of random variables Y_n satisfies the conditions (C-I) and (C-II),

$$P\left\{ Y_{N(t)} - \theta < x w_{n_t} \right\} \xrightarrow{L} F(x) \text{ as } t \to \infty.$$

Theorem(1.2.2):

Let $\{X_n, Y_n\}$, $n=1,2,\ldots$, be a sequence of pairs of random variables. Let $X_n \xrightarrow{L} X$ and $Y_n \xrightarrow{P} C$, where C be a constant. Then

(a) $X_n \pm Y_n \xrightarrow{L} X \pm C$.

Cramer Rao reference

(b) $X_n Y_n \xrightarrow{L} CX$, if $C \neq 0$ and $X_n Y_n \xrightarrow{P} 0$, if $C = 0$.

(c) $\dfrac{X_n}{Y_n} \xrightarrow{L} \dfrac{X}{C}$, if $C \neq 0$.

Theorem(1.2.3):(Lebesgue Dominated Convergence)

Let $\{X_n\}$ be a sequence of random variables. If $|X_n| \leq Y$

a.s., Y integrable, then if

$$X_n \xrightarrow{P} X \Rightarrow E(X_n) \longrightarrow E(X).$$

Theorem(1.2.4):(Wald)

If $\{X_n\}$ be a sequence of i.i.d. random variables distributed like X, satisfying $E|X|<\infty$. For any sequential stopping rule yielding $E(N)<\infty$,

$$E\left\{\sum_{i=1}^{n} X_i\right\} = E(X)E(N).$$

1.3 Literature survey on fixed width confidence interval and Chapterwise summary:

The problem of construction of fixed-width confidence interval is first considered in case of normal distribution.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables from normal distribution with mean μ and variance σ^2, that is $N(\mu, \sigma^2)$. Given a preassigned number α, $\alpha\in(0,1)$, we have to construct a confidence interval for μ such that,

$$P_\theta\left\{ I_n \ni \mu \right\} = P_\theta\left\{ |T_n - \theta| \le d \right\} \ge (1-\alpha). \ldots(1.3.1)$$

Dantzig(1940) proves that, for fixed sample size n, no fixed-width confidence interval of type (1.2.4) satisfying (1.3.1) can be constructed for mean μ. That is no fixed sample size procedure exists.

Stein(1945) proposed a two-stage procedure to construct a fixed-width confidence interval of type (1.2.4) satisfying (1.3.1). The procedure due to Stein can be described in brief as follows.

At a first stage a random sample of size n, $(n \geq 2)$, is recorded and let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$, $S_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$. Compute S_n^2 based on the first sample and define the sample size 'N' by

$$N = \max\left\{ n, \left[\frac{a_{n-1}^2 S_n^2}{d^2} \right] + 1 \right\}, \qquad \ldots(1.3.2)$$

where a_{n-1} is $100(\alpha/2)\%$ point of Student's t-distribution with (n-1) degrees of freedom, and propose an interval,

$$I_N = \left\{ \bar{X}_N - d \,,\, \bar{X}_N + d \right\}, \qquad \ldots(1.3.3)$$

for μ. The interval (1.3.3) satisfies (1.3.1). Note that, in above procedure the sample size at second stage is random variable. When $N-n = 0$ the sampling is terminated after the first stage.

The problem of construction of fixed-width confidence interval, for mean of the normal distribution, has been also studied by Ray(1957), Starr(1966) etc. The analogous problem for variance of normal distribution has been studied by Graybill and Connel(1964) by using two stage sampling. However, in some cases, no sampling scheme exists with predetermined number of stages, that can give fixed-width confidence interval. In such cases, the purely sequential procedures are required. The results of this kind were published by Blum and Rosenblatt(1966), Farrel(1966) and others. Chow and Robbins(1965) proposed a method to construct fixed-width confidence interval for mean of population with unknown variance. Khan(1969) proposed a general method to construct fixed-width confidence interval for parameter of distribution, when some nuisance parameter is present and distribution satisfies the regularity conditions. Sen(1981) and Sproule(1985) generalized the Chow-Robbins procedure to sequential confidence intervals based on U-Statistic.

This dissertation deals with construction of fixed-width confidence interval for the various models. It is divided into four chapters. In the following, we summerize in brief the contents of the different chapters.

In chapter-II, we study various purely sequential methods to construct a fixed-width confidence interval. The section (2.1) is introductory gives details of the various methods that are discussed in the subsequent sections, in brief. In section (2.2) the purely sequential method to construct fixed-width confidence interval for the population mean, when sample comes from population with unknown variance,is dicussed. The asymptotic properties of the method are also discussed. This method is due to Chow and Robbins(1965). In section (2.3) the general method to construct a fixed-width confidence interval for parameter of distribution (not necessarily mean of the population) is considered. The asymptotic properties of this method are also discussed. The method is illustrated with normal distribution and exponential distribution. This method is due to Khan(1969). In section (2.4), we have extended a method due to Khan(1969) to construct a confidence interval for parametric function $g(\theta)$, a continuous differentiable function of θ. The method is illustrated by constructing a fixed-width confidence interval for reliability function, when sample is drawn from exponential distribution with mean θ. In last section we report simulation results for the exponential model based on proposed method in section (2.4).

In chapter-III, we study two-stage sequential procedures to construct fixed-width confidence interval. In section (3.2), we discuss two-stagw procedure to construct a fixed-width confidence interval for mean of distribution satisfying some assumptions. Also we discuss the asymptotic properties of this method. The procedure is illustrated for normal distribution, negative exponential distribution and for multivariate normal distribution. In section (3.3), we study the modified two-stage procedure, which is asymptotically efficient. The asymptotic properties of these modified two-stage procedures are also given. In section (3.4), we review the problem of estimating the parameters of inverse Gaussian distribution in terms of controlling the risk function corrosponding to a suitable zero-one loss function. In last section, some properties of two-stage procedure to construct a fixed width confidence interval along the lines of Birnbaum and Healy(1960), are reviewed. The procedures discussed in this chapter are due to Mukhopadhyay(1982).

In chapter-IV, we study the non-parametric method to construct a fixed-width confidence interval. Section (4.2) is devoted to general method to construct a non-parametric fixed-width confidence interval, in brief. The general method is illustrated by constructing a fixed-width confidence interval for reliability function, when underlying distribution is completely unknown. In section (4.3), simulation results are reported, for the model described in the section (4.2), when F has exponential distribution with mean θ. The results obtained are compared with the results of parametric method, given in section (2.5). Some comments, on the comparison of these method are reported. In section (4.4), the problem of construction of fixed-width confidece interval for correlation coefficient, when observations are drawn from the bivariate nornal distribution, is reviewed. The asymptotic properties of the proposed method are also studied. This method is proposed by Tahir(1992). We comment on the method and suggest possible improvement over the method.

The dissertation is concluded with list of references duly acknowledged and appendix on algorithms and computer programs.