CHAPTER 1

Sequential Analysis and Some Related Works

1.1 Introduction

In Statistics, Sequential analysis is statistical analysis where the sample size is not fixed in advance. Instead data are evaluated as they are collected, and further sampling is stopped in accordance with a pre-defined stopping rule as soon as significant results are observed. Thus a conclusion may sometimes be reached at a much earlier stage than would be possible with more classical hypothesis testing or estimation, at consequently lower financial and/or human cost. Sequential analysis was first developed by Abraham Wald with Jacob Wolfowitz, W. Allen Wallis and Milton Friedman while at Columbia University's Statistical Research Group as a tool for more efficient industrial quality control during World War II. Another early contribution to the method was made by K.J. Arrow with D. Blackwell and M.A. Girshick. A similar approach was independently developed at the same time by Alan Turing, as part of the Banburismus technique used at Bletchley Park, to test hypotheses about whether different messages coded by German Enigma machines should be connected and analysed together. This work remained secret until the early 1980s. Regarding inferential problems of testing, estimation, ranking, selection etc., problems, in the beginning the only purpose of adopting the sequential procedures was to minimize the average sample number for the problems which could have been dealt with the fixed sample size procedures too.

Later on, researchers faced the situations that many inferential problems could not be tackled with the help of the fixed sample size procedures like the ‘given precision problems’ for which only sequential methods provided solutions. Moreover, recently sequential procedures are finding wide application in various statistical analysis concerning Multivariate
Problems, Multihypothesis Problems, Confidence Sets and Intervals, Adaptive Non-Parametric Procedures, Ranking and Selection Problems, Secretary Problems, Decision Theory, Stochastic Approximation, Detection and Change-Point Problems, Quality Control, Clinical Trials, Reliability and Life Testing, to name a few. In the next sections, let us review some facets of Sequential Analysis.

1.2 Sequential and Multi-Stage Procedures for ‘Given Precision Problems’

A sequential procedure in which the number of stages but naturally not the number of observations in every stage is bounded is called a multi-stage procedure. Researchers have developed various sequential and multi-stage procedures to deal with the ‘given precision problems’ which includes

(a) Tests whose power is pre-assigned

(b) Confidence interval(region) estimation with pre-assigned width and coverage probability

(c) Point estimation problems

(d) Ranking and selection problems.

A brief review of these problems and corresponding sequential and multi-stage procedures follows:

1.2.1 Tests with Preassigned Power

Let $X$ be a random variable following normal distribution with unknown mean $\mu$ and unknown variance $\sigma^2$. Student’s problem was to test the hypothesis $H_0$: $\mu = \mu_o$ against the alternative $H_1$: $\mu > \mu_o$. Dantzig(1940) posed the problem of devising a test procedure for this hypothesis, whose power function is independent of $\sigma$. He proved the non-existence of meaningful fixed sample-size procedures for this problem. Stein(1945) proposed a two-
sample test having this property, in which the size of the second-stage sample stage depends on the result of the first-sample, thus making the sample size a random variable. Sequential procedures for such problems have been developed and studied by various authors. Lorden(1983) proposed and studied a class of three-stage tests for the exponential family. Govindarajulu(1985) and Ghosh and Sen(1991) have illustrated various sequential and multi-stage techniques available in the literature. Joshi and Shah(1990) developed a sequential probability ratio test for a simple hypothesis concerning mean of an inverse Gaussian distribution and showed that the test can be extended to test the composite one sided hypothesis. Chaturvedi, Kumar and Kumar(1998, 2000) generalised the problem of sequential testing by introducing a class of distributions representing various probability models and proposing sequential procedures besides studying robustness of the SPRT for various hypothesis.

1.2.2 Confidence Interval(Region) Estimation with Prescribed Width and Coverage Probability

Ruben(1950) discussed the situation when it is imperative to keep the parameters of a population, in probability sense, within prescribed limits. We know that corresponding to every test procedure, we can construct an equivalent confidence region(interval). As a consequence of the result of Dantzig(1940), there does not exist any fixed sample size procedure with the help of which one can construct a confidence interval of ‘pre-assigned width and coverage probability’ for a normal mean when the variance is unknown. In order to handle this problem, Stein(1945) developed a two-stage procedure and properties of which were further studied and developed by Ruben(1961). Chow and Robbins (1965) developed a sequential procedure for constructing fixed-width confidence interval for the mean of an arbitrary population. They studied the asymptotic properties of the proposed sequential procedure and proved its ‘asymptotic efficiency’ and asymptotic consistency’, that is, the
asymptotic coverage probability of the sequential procedure meets the target value. Starr (1966) considered the problem of constructing fixed-width confidence interval for the mean of a normal population and proposed a sequential procedure. He proved the ‘asymptotic efficiency and consistency’ of his sequential procedures and derived the formula for obtaining the exact distribution of the stopping time. For the sequential procedure of Starr (1966a), Simons (1968) proved that the ‘cost of ignorance’ of the variance remained bounded and one can achieve the target value of the coverage probability at the cost of some ‘additional’ observations after the stopping time. Singh and Chaturvedi (1988) obtained a lower bound for the initial sample size required to meet the target value of the coverage probability. Hall (1981), through his theoretical and numerical evaluations, showed that if a single stage is appended to Stein’s (1945) two-stage procedure, the resulting three-stage procedure becomes strongly competitive to both the two-stage and purely sequential procedures. Hall’s (1981) three-stage procedure is simple to apply, it is ‘asymptotically efficient’, the ‘cost of ignorance’ of the variance remains bounded and one can achieve the target value of the coverage probability as closely as one pleases at the cost of only a finite number of observations. Later on, for the same reasons behind the use of the three-stage procedure, Hall (1983) developed an ‘accelerated’ sequential procedure, in which the number of stages can be reduced by a predetermined factor at the cost of only a finite number of observations. For further and later works in this direction one can refer to Hamdy and Pallotta (1987), Mukhopadhyay and Mauromoustakos (1987), Chaturvedi (1988), Hamdy (1988), Kuo and Mukhopdhyay (1990), Mukhopdhyay (1990), Mukhopdhyay and Solanky (1991), Chaturvedi, Shukla and Shukla (1992), Hamdy and El-Bassiouni (1992) and Chaturvedi, Tiwari and Pandey (1993), Chaturvedi and Gupta(1996), Ghosh and Mukhopadhyay (1981), Mukhopadhyay and Hamdy (1984), Costanza, Hamdy and Son (1986), Lam (1986), Mukhopadhyay and Abid (1986 a;b), Singh and Chaturvedi (1988),

1.2.3 The Point Estimation Problems

Robbins (1959) considered the problem of the minimum risk point estimation of the mean of a normal population under absolute error loss function and linear cost of sampling. He proved the failure of the fixed sample size procedure for the unknown variance case. In order to deal with the situation, he proposed a sequential procedure, from his numerical computations; he concluded that the difference between the risks of the sequential procedure and that of the optimal fixed sample size procedure was negligible for the values of the variance considered by him.

Later on, Starr (1966b) considered this estimation problem under a family of loss functions and a cost function of the general form, he considered an optimality for the sequential minimum risk point estimation procedure as its ‘asymptotic risk-efficiency’. According to Starr (1966b), “if the ratio between the risks of the sequential procedure to that of the optimal fixed sample size procedure converges to unity, the proposed sequential procedure is asymptotically risk-efficient”. He obtained a condition on the initial sample size ensuring this property. Similar results under different probabilistic models have been obtained by various authors. For a brief review, one may refer to Wang (1973, 1980), Mukhopadhyay (1974, 1976), Ghosh and Mukhopadhyay (1975, 1979), Ghosh, Sinha and Mukhopadhyay (1976), Chow and Yu (1981), Hayre (1983), Martinsek (1983b, 1987), Mukhopadhyay, Hamdy, Ghosh and Wackerley (1983), Chaturvedi (1985, 1986a;b, 1987), Mukhopadhyay and Hilton (1986), Mukhopadhyay and Ekwo (1987), Sriram
Starr and Woodroofe (1969) introduced another measure of the optimality of a sequential point estimation procedure as its ‘regret’, which they defined as the difference between the risk of the sequential procedure and that of the optimal fixed sample size procedure. A sequential procedure is ‘optimal’ if its ‘regret’ is asymptotically bounded. For the sequential procedure of Starr (1966b), taking the cost function to be linear, Starr and Woodroofe (1969) determined a necessary and sufficient condition on the starting sample size ensuring this property. Woodroofe (1977) introduced the concept of the ‘second-order approximations’ in the area of sequential estimation. He obtained these approximations for the sequential procedures for the minimum risk point estimation and the fixed-width confidence interval estimation of a normal mean assuming variance to be unknown. Improving the results of Starr (1966a,b), he studied the asymptotic behavior of the average sample number, ‘regret’ and the associated coverage probability with the corresponding sequential procedures up to second-degree remainder terms. For other important results concerning different point estimations, one may refer to Starr and Woodroofe (1972), Rohatgi and Rastogi (1973), Ghosh, Sinha and Mukhopadhyay (1976), Nagao and Takada (1980), Chow and Martinsek (1982), Mukhopadhyay (1982b), Chaturvedi (1986b,1987b),Mukhopadhyay and Hilton (1986), Aras (1987) and Mukhopadhyay and Ekwo (1987a;b). In order to have a brief review of the literature on two-stage, three-stage and ‘accelerated’ sequential point estimation procedures, one may refer to O’Neill and Rohatgi (1973), Mukhopadhyay, Hamdy, Ghosh and Wackerley (1983), Mukhopadhyay (1985, 1987a, 1990,1992), Mukhopadhyay and Hilton (1986), Hamdy and Pallotta (1987), Mukhopadhyay, Hamdy, Al-Mahmeed and Costanza (1987), Mukhopadhyay and
Al-Mahmeed and Hamdy (1989), Kubokawa (1989a;b), Kuo and Mukhopadhyay (1990), Son 
and Hamdy (1990), Mukhopadhyay and Solanky (1991), Mukhopadhyay and Abid (1992), 

1.2.4 The Ranking and Selection Problems

Bechhofer (1954) considered the problem of selecting the largest mean of several 
normal populations. For the case when the populations have known variances, they proposed 
a fixed sample size procedure to achieve the desired probability of correct selection. Later on, 
Bechhofer, Dunnet and Sobel (1954) proved the non-existence of the fixed sample size procedure to deal with the problem when the populations have a common unknown variance. 
In order to handle the situation, they developed a two-stage procedure. Robbins, Sobel and Starr (1968) proposed and studied a purely sequential procedure. Dudewicz (1971) showed 
the failure of the fixed sample size procedure for the purpose of selecting the largest mean of 
several normal populations with a given probability of correct selection, when they have 
unknown and unequal variances. Dudewicz and Dalal (1975) proposed a two-stage procedure 
for handling the situation. Rinott (1978) provided a modification over the two-stage procedure 
of Dudewicz and Dalal (1975). Later on work on sequential and multistage methods for 
ranking and selection procedures can be reviewed by refereeing to some significant 
contributions by Blumenthal (1975), Desu, Narula and Villareal (1977), 
1.3 Sequential and Multi-Stage Procedures for other Problems

Researchers have been successfully developing sequential and multi-stage techniques for various other problems, a brief introduction of some of which follows ahead.

1.3.1 Multivariate Problems

The particular area of Multivariate problems using sequential analysis has been enriched by valuable contributions of many researchers over the years. Sequential estimation of the difference between two multivariate normal means, and simultaneous sequential estimation of both the mean vector and the dispersion matrix of a multinormal distribution can be seen in Mukhopadhyay(1979), Chou and Hwang(1986), and Nickerson(1987). Sequential inference procedures of Stein’s for a class of multivariate regression problems have been developed by Chatterjee(1963). Other important contributions include Ghosh, Sinha and Mukhopadhyay(1976), Khan(1968) and Finster(1983).

1.3.2 Multihypothesis Problems

Most of the theory in hypothesis testing is concerned with tests between two hypotheses but there are also multihypothesis problems where it is natural to consider three or more hypothesis. One can have a machine set correctly, too high, or too low. One can have one type of yarn stronger, weaker, or equal to another type. Wetherill and Glazebrook(1986) and Liteanu and Rice(1980) have discussed some sequential tests for these problems. Tests for multihypothesis testing have also been provided by Lorden(1976) and Whitehead and Brunier(1990).
1.3.3 Sequential Nonparametrics

The past 30 years have witnessed a phenomenal growth of the literature on sequential nonparametric methods. Flexibility of the model with respect to the form of the underlying probability distributions, robustness (against departure from model-based assumptions, error contamination, and/or outliers), and asymptotic considerations dominate the scenario in sequential nonparametric. Now, sequential tests based on robust and nonparametric statistics are well regarded as good competitors of the conventional sequential probability ratio tests which are based on certain parametric i.e., specific distribution models. Significant contributions in this area have come from Ghosh and Sen(1971, 1977), who developed sequential confidence intervals based on rank tests and sequential rank test for regression. Sen(1985, 1987, 1990) and Sen and Ghosh(1974, 1981) have added to the development of sequential inference in nonparametric setup.

1.3.4 Sequential Design

Most of the work in sequential designs can be broadly classified into two closely related areas: n-armed bandit problems and allocation rules in clinical trials. Besides there have been some interesting theoretical studies of the general sequential design problems, for example, by Lalley and Lorden(1986) as initiated by Chernoff(1959) and continued in Kiefer and Sacks(1963). Some work on sequential designs has been done in problems of ranking and selection by Bather(1983,1985), Bechohofer(1985), Lai and Robbins(1984) to name a few. Moreover two excellent books on related bandit problem by Berry and Fristedt(1985) and Gittins(1989) are also available for understanding and review.
1.3.5 Detection and Change –Point problems

The change-point problem has attracted much attention during the last 30 years, and the number of papers on this and related problems is growing. Most of the papers in the literature deal with the fixed-sample problem, and relate to the atmost-one change(AMOC) model. Sequential methods are especially important in problems of statistical control of processes with stochastic input, early warning of changes in the distributions, and tracking of processes. A review of sequential methods and comprehensive bibliography can be found in Zacks(1983). Procedures of sequential detection of changes in distribution laws are of special importance for statistical process control. One of the pioneering works in this area is Shewart(1931), who proposed the famous “3-sigma control charts” which are still being widely used in industries all over the world for controlling manufacturing processes. Recently more efficient procedures namely cumulative sum procedures (CUSUM) and Bayes procedures can be seen in the literature to detect and observe change-points. Related sequential methods can be classified into three categories: Bayes sequential stopping rules, cumulative sum procedures (CUSUM), and tracking methods. Shiryayev(1963a, b), Zacks and Brazily(1981), Pollak(1985) and others have contributed a lot to the Bayes sequential stopping rules. CUSUM procedures have been developed by Lorden(1971), Zacks(1981) and Pollak(1985). Chernoff and Zacks(1964) and Gordon and Smith(1988) have done good work in tracking methods.

1.3.6 Sequential Methods in Reliability and Life Testing

Since the pioneering work of Wald (1947), considerable work on sequential procedures in the area of reliability and life testing has been done. A good number of text books and monographs contain materials on sequential analysis in reliability or survival

1.3.7 Sequential Methods in Clinical Trials

Clinical research comprises all experimental studies of medical interventions in which human subjects are involved. It is the involvement of human subjects which sets this form of investigation apart from any other scientific, endeavour, and introduces uniquely stringent ethical constraints. In particular, in the design of clinical studies it is desirable that no person should receive a treatment which is already known to be inferior to an available alternative. There is a limit to the extent to which subjects can be protected against a treatment which is beginning to register inferior results while still maintaining scientific objectivity. Human subjects are consenting volunteers. Some are healthy volunteers being paid for their participation; others are patients hoping for a cure of their diseases which is not yet available outside clinical trials. A study which withheld a treatment on the first suspicion of its inferiority would never progress to a definitive conclusion. However, sequential methods offer an opportunity to stop a trial as soon as the evidence for or against a treatment has reached the conventional levels of strength required in scientific work. A lot of contribution is being made to the design and analysis of clinical research studies by the concepts and methods of sequential analysis. The vast majority of studies have been concerned with comparative phase III trials (comparison of experimental package against competitors for a limited and well defined group of patients) and a few have concerned phase II (involves patients suffering from the target disease, and combines further investigation on safety and dosage with the first consideration of therapeutic efficacy). Formal schemes of interim analysis and sequential methods are on rise. Useful review of sequential methods for clinical

1.3.7 Contents of the Thesis

The following is the brief introduction to the chapters that follow in this research output.

Chapter 2- Sequential Procedures for the Simultaneous Estimation of the Parameters of Several Populations

In this Chapter is considered a generalized problem of simultaneous estimation of parameters of several populations $f(X_i; \theta_i, \psi_i)$, where $\theta_i$ and $\psi_i$ are the unknown parameters under a family of general loss function and a linear cost function. The problem of fixed sample size procedure to deal with such problem is established and classes of “purely sequential”, “three stage” and “accelerated” sequential procedure to tackle the problems are proposed. Positive and negative moments for the stopping times are obtained and they are used for deriving the asymptotic expression for the “regret” associated with the classes of “purely sequential”, “three stage” and “accelerated” sequential estimation procedures. Finally, illustrations of simultaneous estimation problems are provided which can be dealt with the help of the proposed classes.

Chapter 3- Estimation Problems Related to Exponential Distribution and their Sequential Solutions

In Chapter 3 is considered three estimation problems related to the exponential distribution, which are the problem of minimum risk point estimation of its scale parameter under the squared-error loss function, the problem of construction of fixed width confidence interval for scale parameter and the problem of constructing fixed-ratio width confidence interval for the reliability function concerning the exponential distribution. The failure of the
fixed sample size procedures to deal with these estimation problems is established. Sequential procedure is proposed in order to estimate scale parameter point wise and obtained are the associated second-order approximations. Also proposed is an ‘improved’ estimator and have shown its dominance over the UMVUE. A sequential fixed-width confidence interval for scale parameter is also developed and has proved its asymptotic efficiency and consistency. Finally, a sequential procedure is proposed to construct a fixed – ratio width confidence interval for the reliability function and have shown it to be asymptotically efficient and consistent.

Chapter 4- Multi-stage Estimation of the Common Mean of Several Normal Populations Having Unequal and Unknown Variances

In this chapter, is considered the problem of estimating the common mean of several normal populations having unequal and unknown (possibly) variances. We consider both, the fixed-width confidence interval estimation and point (minimum risk, as well as, the bounded risk) estimation. We consider the estimation problems and establish the failure of the fixed sample size procedures to deal with them. Various multi-stage procedures like two-stage; three-stage, accelerated sequential and purely sequential procedures are developed for the fixed-width confidence interval estimation problem for the common mean of several normal distributions. Also, multi-stage procedure is proposed for the bounded risk point estimation problem for the common mean of several normal distributions. Finally, multi-stage procedures are proposed for the minimum risk point estimation problem. The second order approximations are obtained for the stopping times and various properties are derived to study the nature of the procedures.
Chapter 5- Sequential Point and Region Estimation of the Difference of Two Multinormal Means

In this Chapter the problem of estimating the difference of means of two multinormal distributions is considered. In the first problem of point estimation of the difference of means, sequential procedure is developed and proved are its asymptotic properties, particularly, second order approximations for the regret has been derived. For the second problem of region(interval) estimation of the difference of means another sequential procedure is developed and proved are its associated asymptotic properties besides second order approximations of the consistency of the procedure.