SEQUENTIAL METHODS IN ESTIMATION AND PREDICTION

By

NITIS MUKHOPADHYAY

Thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy

CALCUTTA

JUNE, 1975

A_C_K_N_O_W_L_E_D_G_E_M_E_N_T

I thank the Statistics and Mathematics Division of the Indian Statistical Institute for extending all the facilities for my work. I gratefully thank Professor D. Basu for creating a nice atmosphere for research in the Math.-Stat. building of ISI.

I gratefully acknowledge my indebtedness to Dr. Malay Ghosh under whose supervision and constant encouragement this thesis was written. He suggested some interesting topics for my research and guided me through every step. His editing of the manuscript has been of immense value to me. I am most thankful for the benefit of his time and attention I could draw upon so liberally, though he is too busy with visiting appointments in United States.

I express my sincere thanks to Professor J. K. Ghosh. He made several interesting remarks during long discussions at different stages, which were extremely helpful in improving the quality of the thesis. Also I owe my sincere thanks to Professor G. Simons of University of North Carolina, Chapel Hill for going through parts of the manuscript and some helpful comments.

I wish to express my deep gratitude to Dr. Malay Ghosh and Dr. Bimal Kumar Sinha for giving me permission to incorporate here our joint works. These form the materials of Chapters 2, 3, 8 and a part of Chapter 7. I benefited a lot from the discussions and correspondences I had with Dr. Sinha. My heartfelt thanks are due to him.

My friend Shri Anjanananda Thakur took keen interest in my works and assisted me with programming in an excellent way. It is impossible to express how much I owe to him. I would like to thank Professor J. Roy, Head of the Computer Science Unit, ISI for making available machine time of H-400 electronic computer. Thanks are due to other members of Computer Science Unit who were always helpful.

I acknowledge my sincerest debt to all my teachers and colleagues at the Research and Training School who have all been helpful and encouraging to me. I extend my thanks to Shri T. K. Chandra and Shri D. Majumdar for their help at final stages.

I also thank Shri G. M. Das for excellent typing.

Calcutta,
June 1975.

Nitis Mukhopadhyay

ABBREVIATIONS AND NOTATIONS

Abbreviations

rv	random variable
iid	independent and identically distributed
iidrv	independent and identically distributed random variable
p.d	positive definite
a.s.	almost surely
SPRT	sequential probability ratio test
SLLN	strong law of large numbers
df.	distribution function
pdf	probability density function

Notations

$N_p(\mu, \Sigma)$	p variate normal with mean μ and dispersion matrix Σ
$N(\mu, \sigma^2)$	$N_1(\mu, \sigma^2)$
χ^2_p	central chi-square with p degrees of freedom
I_p	Identity matrix of order p
I	Indicator function

TABLE OF CONTENTS

p.t.o

PAGE

p.t.o.

PAGE

CHAPTER 1

INTRODUCTION AND SUMMARY

1.1 A brief review of the literature

Sequential analysis has made great strides in modern statistical development. Starting with the poineering work of the late Abraham Wald, there has been considerable development in the area of sequential hypothesis testing over the past three decades. A good deal of work is available also in sequential estimation. Broadly speaking, there are two basic reasons why sequential methods are suggested. First to reduce the sample size on an average as compared to the corresponding fixed sample size procedure which meets the same error requirements. Second, to solve certain problems which cannot be solved with a predetermined sample size. Our work concentrates on certain estimation and prediction problems which call for sequential methods in the absence of any fixed sample size procedures for their solution.

We start with two basic estimation problems.

(I) For a $N(\mu, \sigma^2)$ population, μ real (unknown), $\sigma > 0$, consider the problem of finding a confidence interval for μ of prescribed length $2d$ (> 0) and preassigned confidence coefficient $1-\alpha$ ($0 < \alpha < 1$).

(II) Consider the same set up as of (I). The problem now is the point estimation of μ based on a random sample $X_1, \cdots, X_n$ from the $N(\mu, \sigma^2)$ population. Suppose the loss incurred in estimating μ pointwise by $\partial(X) = \partial(X_1, \cdots X_n)$ is

(1.1.1) $$L_n = A|\partial(X) - \mu|^s + cn^t$$

where A, s, c, t are all known positive constants. The problem is to minimize the risk $E_\sigma(L_n)$ for all σ.

For (I), in estimating μ by $\bar{X}_n = n^{-1} \sum_1^n X_i$, the coverage probability is given by

$$P_{\mu,\sigma}[|\bar{X}_n - \mu| \leq d] = 2\Phi(n^{1/2} d/\sigma) - 1.$$

If σ is known, the smallest sample size n^* for which the above coverage probability is $\geq 1-\alpha$ for all μ (and σ) is given by n^* = smallest integer $\geq a^2\sigma^2 d^{-2}$, where $a = a_\alpha$ = the upper $100 \frac{\alpha}{2}$ % point of the $N(0,1)$ distribution. If σ is not known, but it is given that $0 < \sigma < \sigma_0$, where σ_0 is specified, one can choose n^* = smallest integer $\geq a^2 \sigma_0^2 d^{-2}$ to meet the requirement.

For (II) in estimating μ pointwise by $\bar{X}_n$, the risk

$$R_n(c, \sigma) = 2s^{-1} k \sigma^s n^{-\frac{1}{2}s} + c \cdot n^t,$$

where $k = \frac{1}{2} \cdot As\, 2^{\frac{1}{2}s} \Gamma(\frac{1}{2}(s+1)) / \Gamma(\frac{1}{2})$, and the loss is given by (1.1.1). If σ is known, the optimal sample size minimizing $R_n(c, \sigma)$ is $n^* = (k\sigma^s/ct)^{2/(s+2t)}$, where in the above, and and in the sequel we conveniently ignore the fact that n^* need not be an integer.

If σ is not known, or it is not also known that $0 < \sigma < \sigma_o$, where σ_o is known, then there does not exist any fixed sample size procedure for which the goal is achieved in (I). To see this intuitively, note that the coverage probability tends to zero as $\sigma \to \infty$, i.e. as the scatter gets wider making the distribution more flat, we cannot bound μ in a prescribed confidence interval around the sample mean with preassigned confidence. This non-existence result was first proved by Dantzig (1940) in the dual problem of testing of hypothesis. The non-existence result is more clearly propounded in Lehmann (1951) for distribution with pdf's of the form $\sigma^{-1} f((x-\mu)/\sigma)$, $\sigma > 0$ and for a more general loss including the ones considered in (I) and (II) as particular cases. A very pertinent question to ask now is whether the objective can be achieved by using a sequential sampling scheme.

Stein (1945, 1949) gave an ingenious technique of a two-stage procedure which achieves the objective in (I). His procedure consists in taking an initial sample $X_1, \dots, X_{n_o}$ of size $n_o (\geq 2)$ from the $N(\mu, \sigma^2)$ population, and then define the stopping time N as

$$N = \max(n_o, [a^2_{n_o-1} s^2_{n_o} d^{-2}] + 1)$$

where $s^2_k = (k-1)^{-1} \sum_1^k (X_i - \bar{X}_k)^2$, $k \geq 2$ and a_{n_o-1} is the upper $100 \frac{\alpha}{2}$ % point of Student's t-distribution with (n_o-1) degrees of freedom, $[x]$ denoting the greatest integer $\leq x$. Unfortunately, the Stein's procedure usually results in a large average sample number (ASN). This is because Stein utilized only the first stage observations in estimating the variance. This hurts, especially when d is small.

We need the following definition for a critical appraisal of the subsequent development.

Def. 1.1 If N is the random sample size, a procedure is 'consistent' for (I) if

$$P[|\bar{X}_N - \mu| \leq d] = 1 - \alpha \tag{1.1.2}$$

and is 'asymptotically efficient' for (I) if

$$\lim_{d \to 0} E(N/C) = 1, \tag{1.1.3}$$

where $C = a^2\sigma^2/d^2$.

With this definition, Stein's two-stage procedure is not 'asymptotically efficient', though it is 'consistent'. So, the next question is the existence of an 'asymptotically efficient' and 'consistent' procedure for (I). The first attempt in this direction is due to Ray (1957) who suggested a sequential procedure. However, inadequate computations led Ray to the misleading conjecture that his procedure was not '<u>asymptotically</u> (as $d \to 0$) <u>consistent</u>', i.e.

$$\lim_{d \to 0} P[|\bar{X}_N - \mu| \leq d] = 1 - \alpha .$$

Chow-Robbins (1965) suggested the following sequential procedure. Let $X_1, X_2, \ldots$ be iidrv's with df F, having mean μ and variance σ^2 $(0 < \sigma < \infty)$. The problem is the same as in (I). Then,

(1.1.4) The stopping time $N \equiv N_d$ is the first integer n $(\geq n_o)$ for which $n \geq a^2 (s_n^2 + n^{-1})/d^2$, where $s_n^2 = (n-1)^{-1} \sum_1^n (X_i - \bar{X}_n)^2$, $n_o (\geq 2)$ is the starting sample size.

Chow and Robbins proved that the procedure was 'asymptotically consistent' and 'asymptotically efficient'. Starr (1966a) carried out extensive numerical computations in the normal case.

Simons (1968) proved an interesting result regarding the consistency of the Chow-Robbins procedure in the normal case. For $n_o \geq 3$, he proved the existence of an integer $k (\geq 0)$ such that

(1.1.5) $P[|\bar{X}_{N+k} - \mu| \leq d] \geq 1 - \alpha$ for all μ, σ and d;

(1.1.6) $$E(N+k) \leq C + n_o + k,$$

where N is given by (1.1.4). In actual practice, however, k is not known, and there is very little idea how large it is going to be.

Gleser (1965, 1966) extended the Chow-Robbins procedure to get a confidence region for regression parameters in a simple linear regression model with given coverage probability, such that the length of the interval cut off on each axis is fixed. For estimating the multivariate normal mean vector with unknown dispersion matrix, the Stein two-stage procedure was extended by Chatterjee (1959), while Chow-Robbins sequential procedure was extended by Srivastava (1967).

For (II), when σ is unknown, the sequential procedure suggested by Robbins (1959)(in the case $s = t = 1$) and Starr (1966b) is as follows:

(1.1.7) The stopping time $N \equiv N_c$ is the first integer $n\ (\geq n_o)$ for which

$$n \geq (k\, s_n^s/ct)^{2/(s+2t)},$$

where $s_n^2 = (n-1)^{-1} \sum_1^n (X_i - \bar{X}_n)^2$, $(n \geq 2)$, $n_o (\geq 2)$ being the starting sample size.

When σ is known, the minimum risk is

$$R(c, \sigma) = R_{n*}(c, \sigma) = c(\tfrac{2}{s} \cdot t + 1)(n^*)^t .$$

Since the event $[N = n]$ and $\bar{X}_n$ are independent for any fixed n,

$$(1.1.8) \qquad \bar{R}(c,\ \sigma) = E(L_N) = \frac{2}{s} k\ \sigma^s\ E(N^{-s/2}) + c.E(N^t).$$

The '<u>risk efficiency</u>' and '<u>regret</u>' are defined respectively by

$$(1.1.9) \qquad \begin{cases} \eta\ (c,\ \sigma) = \bar{R}(c,\ \sigma)/R(c,\ \sigma), \\ w\ (c,\ \sigma) = \bar{R}(c,\ \sigma) - R(c,\ \sigma)\ . \end{cases}$$

Starr (1966b) proved that, for all fixed $0 < \sigma < \infty$,

$$\lim_{c \to 0} \eta(c,\ \sigma) = 1 \quad \text{if} \quad n_o > s^2/(s+2t) + 1$$
$$= 1 + \gamma \quad \text{if} \quad n_o = s^2/(s+2t) + 1$$
$$= \infty \quad \text{if} \quad n_o < s^2/(s+2t) + 1$$

where γ is a known positive constant depending on n_o, s, t. Starr and Woodroofe (1969) proved (with $t = 1$) that,

$$\lim_{n \to 0} w(c,\ \sigma) = 0 \quad \text{if and only if} \quad n_o \geq s + 1,$$

for all fixed $0 < \sigma < \infty$.

A basic difference between the point and confidence interval estimation procedures is that, in deriving limiting properties of Chow-Robbins procedure we really needed Anscombe's (1952) result, while in the procedures of Robbins (1959) and Starr (1966b) independence of $[N = n]$ and $\bar{X}_n$ plays a very important role. A very interesting example of minimum risk point estimation procedure is given in Starr and Woodroofe (1972) where this independence is lost, leading thereby to a more difficult analysis.

A sequential analogue of Behrens-Fisher problem was dealt by Robbins et al (1967) and Srivastava (1970). Sequential point estimation problems for multivariate normal mean vector were done by Rohatgi and O'Neill (1973), when the dispersion and weight matrices are diagonal positive definite (p.d.). Sinha and Mukhopadhyay (1974a, 1974b) have studied the above point estimation problem in the bivariate case when (i) the dispersion matrix is arbitrary p.d., but the weight matrix is diagonal p.d., and (ii) the weight matrix is arbitrary p.d. and the variances are equal. The loss structure involved is squared error plus cost.

1.2 Summary of the results in Chapters 2-8

In all the sequential estimation problems considered in section 1.1, the evaluation of the exact distribution of the stopping time is quite tedious and time-consuming. An algorithm developed by Moyal (see Robbins (1959)) has been used by Robbins (1959) and Starr (1966a) to find the exact distribution of N. The problem gets quite formidable for small d and c. Also there are cases (e.g. Chow-Robbins procedure in the non-normal situations) where the distribution of N cannot be obtained analytically, and Monte-Carlo techniques need be adopted.

With the above considerations in background, we have developed in chapter 2 a theorem regarding asymptotic normality of stopping times. The main idea is to apply the theorem to sequential estimation problems. But interestingly enough, the theorem is quite general to include some examples of sequential hypothesis testing as well. The theorem is applied almost in all the chapters in finding the asymptotic distribution of N. The major tool used is the asymptotic normality of U-statistics, and thus, in particular, of the sample sum, with random indices.

In chapter 3, we have considered the sequential point estimation problem for a bivariate normal mean vector, when both the dispersion (unknown) and weight matrices are arbitrary p.d. The loss structure involved is weighted squared error plus cost. The sequential procedure suggested enjoys the asymptotic (as cost per observation tends to zero) optimality properties from the point of view of 'risk efficiency' and 'regret' as compared to the fixed sample size optimal procedure when the dispersion matrix is known. The procedure also performs very well for moderate sample sizes, as studied by Monte-Carlo methods. The asymptotic normality of the stopping time is also proved.

In chapter 4, a sequential procedure is given to get a point estimator of the parameter $\partial = \mu_1 - \mu_2$, when the underlying populations are $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. The loss structure is squared error plus cost. All the four parameters are assumed to be unknown. It is shown that our procedure is 'asymptotically risk efficient' and the 'regret' is bounded as the cost per observation tends to zero. The cost of not knowing the variances is also studied; we have shown that excess (if any) of the optional sample size over the optimal sample size had σ_1, σ_2 been known is less than a fixed constant. The asymptotic distribution of the stopping time is studied, and, finally the moderate sample size behaviour of the proposed procedure is studied through Monte-Carlo methods using pseudo-random normal deviates.

In chapter 5, the following problems are dealt. Suppose μ_i, σ_i^2 are the mean and variance of the ith normal population, $i = 1,2,3$. Assume that all the parameters are unknown and the variances are not necessarily equal. For given non-zero constants $\lambda_1, \lambda_2, \lambda_3$, the aim is to estimate the parameter $\mu = \lambda_1\mu_1 + \lambda_2\mu_2 + \lambda_3\mu_3$ with prescribed accuracy using sample sizes not necessarily equal. Sequential procedures have been developed to arrive at a fixed-width confidence interval for μ, with a given coverage probability. Also, we have given a sequential procedure to estimate μ pointwise, the loss being squared error and the cost of sampling proportional to sample size. 'Asymptotic consistency' and 'asymptotic efficiency' are proved in the first situation, while 'risk efficiency' and 'regret' have been studied in the latter case. The cost of ignorance of the variances has been given proper attention in both cases. Lastly, we have shown by Monte-Carlo methods that the sequential fixed-width confidence interval procedures behave quite satisfactorily for moderate sample sizes.

In chapter 6, the population of interest is $\sigma^{-1} \exp(-\frac{x-\mu}{\sigma})$, $x > \mu$, ($\sigma > 0$, $-\infty < \mu < \infty$, both unknown). With a general loss structure, a sequential point estimator of μ is suggested. The same is shown to be 'asymptotically risk efficient' under some conditions on the starting sample size (kept fixed). In this connexion, we point out two serious mistakes in Basu (1971). To arrive at a fixed width confidence interval for μ with preassigned coverage probability, we suggest a sequential procedure which is 'asymptotically consistent and efficient'. Moderate sample size performances of our procedures in either problem are studied through Monte-Carlo methods.

In chapter 7, we consider some more results on sequential estimation. Firstly, in the Gauss-Markoff linear estimation with quadratic loss structure, a sequential point estimator for the regression parameters is suggested. This procedure is shown to have asymptotic 'risk efficiency' and 'bounded regret'. The asymptotic normality of the stopping time is also proved. Second, a sequential procedure is proposed in point estimation of θ in $R(0, \theta)$ by the sample maximum, when the loss is absolute error plus cost. The procedure is shown to be asymptotically risk efficient. Moderate sample size performance of this procedure is studied in detail through Monte-Carlo methods and is found quite satisfactory.

In chapter 8, we consider some results on sequential prediction. As in the estimation set up, the problem is two-fold in prediction set up also. First, to provide a point predictor of a future observation Y on the basis of a series of observations $X_1, X_2, \cdots$ which has minimum risk (considering a suitable loss function) for <u>all</u> parameter points and second, to provide a fixed-width prediction interval for Y which has a specified coverage probability for <u>all</u> parameter points. The results of this chapter are:

(1) Considering a location and scale parameter family of symmetrical unimodal densities $f(|\frac{x-\alpha}{\beta}|)$, $-\infty < x < \infty$, $-\infty < \alpha < \infty$, $0 < \beta < \infty$ (both unknown), it has been

proved that there exists no fixed-width prediction interval for Y which has a specified coverage probability (whatever α and β), under any sampling procedure which terminates with probability one. It is also pointed out that the symmetry of f around α is not essential.

(2) The problem of providing an optimum point predictor of Y with squared error as loss and cost of sampling proportional to sample size has been tackled considering truncated exponential, exponential and normal densities. A sequential procedure is given in each case, which is shown to be asymptotically risk efficient in the first case, in the other two cases 'regret' is shown to bounded.

CHAPTER 2

ASYMPTOTIC NORMALITY OF STOPPING TIMES IN SEQUENTIAL ANALYSIS

2.1 Introduction

The need for obtaining the asymptotic distribution of the stopping time in the absence of a suitable analytic method to find its exact distribution has been revealed in 1.2. We may add that the task is equally difficult even for most of the well-known sequential probability ratio tests (SPRT's). Mention may also be made of some of the more recently developed sequential testing procedures by Darling and Robbins (1967a, 1967b, 1968).

Recently Siegmund (1968) and Bhattacharyya and Mallik (B- M)(1973) have succeeded obtaining asymptotic distributions of stopping times in some important cases arising in sequential analysis. Their procedures rely heavily on the asymptotic normality of sample sums with random indices.

In this chapter we have first proved (in section 2.2) a general result regarding asymptotic normality of stopping times. Our formulation is slightly different, but the basic idea is essentially similar to that of B- M. The advantage of the present formulation is that unlike B- M we do not need necessarily express our stopping rule in terms of a sample sum type statistic, and thus get more direct results in many applications. Mention may be made of the much shorter proof of theorem 2 of B-M given in section 2.5.

The theorem is applied in Section 2.3 in deriving the asymptotic normality of stopping times in some sequential point estimation problems. Then, follows, in Section 2.4 its applications in a particular modified SPRT problem. Later, in this section, we consider certain sequential procedure proposed by Robbins (1970) to obtain tests with power one.

The asymptotic normality of U-statistics with random indices is used extensively in Section 2.5. The major tool is available in the work of Sproule (1969). Using this, we have been able to get more applications of our main result including a proof of theorem 2 of B-M under relaxed conditions.

Section 2.6 is devoted to the study of asymptotic distributions of stopping-times associated with fixed length interval estimation of means of U-statistics. Some applications are mentioned.

Finally in Section 2.7, our results are applied to problems connected with general linear regression models.

2.2 The main result

Consider a sequence $\{N_\upsilon, \upsilon \geq 1\}$ of positive integer valued rv's defined as follows:

(2.2.1) N_υ is the smallest integer $n(\geq n_o)$ for which $n \geq \Psi_\upsilon T_n$, where n_o is the starting sample size, $\{\Psi_\upsilon\}$ is a sequence of positive constants $\to \infty$ as $\upsilon \to \infty$, and $T_n (n \geq n_o)$ are statistics such that $P(T_n \leq 0) = 0$ for all $n \geq n_o$. Then, our main theorem can be stated as follows:

<u>Theorem 2.1.</u> <u>For the sequence of stopping times defined in (2.2.1), if</u>

(2.2.2) $$N_\upsilon^{1/2}(T_{N_\upsilon} - a)/b \xrightarrow{\mathcal{L}} N(0, 1) \text{ as } \upsilon \to \infty$$

<u>and</u>

(2.2.3) $$N_\upsilon^{1/2}(T_{N_\upsilon - 1} - a)/b \xrightarrow{\mathcal{L}} N(0,1) \text{ as } \upsilon \to \infty$$

<u>where</u> $a(>0)$ <u>and</u> $b(>0)$ <u>are constants</u>, <u>then</u>,

(2.2.4) $$a^{1/2}(N_\upsilon - a\Psi_\upsilon)/(b\,\Psi_\upsilon^{1/2}) \xrightarrow{\mathcal{L}} N(0,1) \text{ as } \upsilon \to \infty$$

<u>Proof:</u> Let Φ denote the distribution function (d.f.) of the $N(0, 1)$ distribution. Use the basic inequalities,

$$N_\upsilon \geq \Psi_\upsilon T_{N_\upsilon} \quad \text{and} \quad N_\upsilon - 1 < n_o - 1 + \Psi_\upsilon \cdot T_{N_\upsilon - 1}$$

i.e.,

(2.2.5) $$\Psi_\upsilon T_{N_\upsilon} \leq N_\upsilon < n_o + \Psi_\upsilon T_{N_\upsilon - 1}.$$

Now, from (2.2.1) $N_\upsilon \xrightarrow{P} \infty$ as $\upsilon \to \infty$. Hence from (2.2.2) and

(2.2.3), $T_{N_\upsilon} \xrightarrow{P} a$, $T_{N_\upsilon - 1} \xrightarrow{P} a$ as $\upsilon \to \infty$. Dividing now both the sides of (2.2.5) by $a \cdot \psi_\upsilon$ and using the fact that $\psi_\upsilon \to \infty$, as $\upsilon \to \infty$, we get $N_\upsilon/(a\,\psi_\upsilon) \xrightarrow{P} 1$ as $\upsilon \to \infty$.

Thus (2.2.2) and (2.2.3) can be alternatively expressed as

$$(2.2.6) \qquad \begin{cases} (a\,\psi_\upsilon)^{1/2}(T_{N_\upsilon} - a)/b \xrightarrow{\mathcal{L}} N(0,1) \\ (a\,\psi_\upsilon)^{1/2}(T_{N_\upsilon - 1} - a)/b \xrightarrow{\mathcal{L}} N(0,1). \end{cases}$$

Using (2.2.5) and (2.2.6), one gets now,

$$\begin{aligned} \Phi(x) &= \lim_{\upsilon \to \infty} P\left\{ (a\psi_\upsilon)^{1/2}(T_{N_\upsilon} - a)/b \le x \right\} \\ &\le \liminf_{\upsilon \to \infty} P\left\{ a^{1/2}(N_\upsilon - a\psi_\upsilon)/(b\psi_\upsilon^{1/2}) \le x \right\} \\ (2.2.7) \qquad &\le \limsup_{\upsilon \to \infty} P\left\{ a^{1/2}(N_\upsilon - a\psi_\upsilon)/(b\psi_\upsilon^{1/2}) \le x \right\} \\ &\le \lim_{\upsilon \to \infty} P\left\{ (a\psi_\upsilon)^{1/2}(T_{N_\upsilon} - a)/b + n_o \cdot a^{1/2}/(b\psi_\upsilon^{1/2}) \le x \right\} \\ &= \Phi(x). \end{aligned}$$

Hence, the result.

<u>Remark 1.</u> It is immediate to check from the proof that Φ can be replaced by any nondegenerate d.f.

<u>Remark 2.</u> It might appear that (2.2.2) implies (2.2.3), but that is not so. In general, the distribution of T_{N_υ} and $T_{N_\upsilon - 1}$ may be quite different.

Remark 3. The usual sufficient conditions to ensure (2.2.2) and (2.2.3) are the ones due to Anscombe (1952). These are (i) for a sequence $\{n_\upsilon\}$ of positive integers ($n_\upsilon \to \infty$ as $\upsilon \to \infty$), $N_\upsilon / n_\upsilon \xrightarrow{P} 1$ as $\upsilon \to \infty$, (ii) $m^{1/2}(T_m - a)/b \xrightarrow{\mathcal{L}} N(0, 1)$ as $m \to \infty$, and (iii) given $\epsilon > 0$ and $\eta > 0$, there exists $\partial(>0)$ and a positive integer n_0, such that

$$P\left\{ \sup_{|n'_\upsilon - n_\upsilon| \leq \partial n_\upsilon} |T_{n'_\upsilon} - T_{n_\upsilon}| > \epsilon \cdot n_\upsilon^{-1/2} \right\} < \eta$$

for $n_\upsilon \geq n_0$.

2.3 Some sequential point estimation problems

We start with a sequential point estimation problem considered by Starr and Woodroofe (1972).

Example 1. Let $Y_1, Y_2, \ldots$ be iidrv's with p.d.f.

$$(2.3.1) \qquad f_\mu(y) = \mu^{-1} \exp(-y/\mu), \; y > 0, \; \mu > 0.$$

With the end of minimizing risk with sample mean estimators for μ, weighted squared error loss (weight A being known) and cost c per unit, the sequential procedure adopted by Starr-Woodroofe (1972) is as follows:

(2.3.2) The stopping time $N \equiv N_c$ is the first integer $n (\geq n_o$, initial sample size) for which

$$n \geq \psi_c \bar{Y}_n \quad \text{where} \quad \psi_c = (Ac^{-1})^{1/2}, \; \bar{Y}_n = n^{-1} \sum_{1}^{n} Y_i,$$

$(n \geq 1)$.

Note that $\psi_c \to \infty$ as $c \to 0$, $P(\bar{Y}_n \leq 0) = 0$ for $n \geq 1$, and using remark 3 of Section 2.2, (2.2.2) and (2.2.3) are satisfied with $a = b = \mu$ (> 0). Hence from theorem 2.1, we get

$$(2.3.3) \qquad (N_c - \xi_c)/\xi_c^{1/2} \xrightarrow{\mathcal{L}} N(0, 1),$$

where

$$\xi_c = \mu \cdot \psi_c = \mu(A c^{-1})^{1/2} .$$

<u>Example 2.</u> Consider the setup in (II) of Section 1.1. The rule in (1.1.7) differs slightly from Starr's (1966b) rule, where $\delta = 1$. In conformity with B-M, we have considered the asymptotic approach when $c \to 0$, σ fixed, rather than the one (adopted by Starr) when $\sigma \to \infty$, δ fixed

Using the Helmert Transformation

$$Y_2 = (X_1 - X_2)/\sqrt{2}, \ldots, Y_n = (X_1 + \ldots + X_{n-1} - (n-1)X_n)/\sqrt{n(n-1)},$$

one can write $(n-1)\, s_n^2 = \sum_{2}^{n} Y_i^2$ $(n \geq 2)$, where Y_i^2/σ^2 are iid χ_1^2 variables. Note that $P(s_n^2 \leq 0) = 0$ for all $n \geq 2$, and $\psi_c = (kc^{-1} t^{-1})^{2/(s+2t)} \to \infty$ as $c \to 0$.

Using Anscombe's (1962) result, once again, one has

(2.3.4) $$N_c^{1/2}\,(s_{N_c}^2 - \sigma^2)/(\sqrt{2}\sigma^2) \xrightarrow{\mathcal{L}} N(0,\ 1)$$

as $c \to 0$. Using now a theorem of Mann and Wald (see e.g. Rao (1965), P. 319), one gets

(2.3.5) $$N_c^{1/2}\,(s_{N_c}^{2s/(s+2t)} - \sigma^{2s/(s+2t)})/(\sqrt{2}\sigma^2 \cdot \tfrac{s}{s+2t}(\sigma^2)^{\frac{s}{s+2t}-1}) \xrightarrow{\mathcal{L}} N(0,\ 1) \text{ as } c \to 0,$$

and

$$N_c^{1/2}(s_{N_c-1}^{2s/(s+2t)} - \sigma^{2s/(s+2t)})/(\sqrt{2}\sigma^2 \cdot \tfrac{s}{s+2t} \cdot (\sigma^2)^{\frac{s}{s+2t}-1}) \xrightarrow{\mathcal{L}} N(0,\ 1) \text{ as } c \to 0.$$

Thus (2.2.2) and (2.2.3) are satisfied with $a = \sigma^{2s/(s+2t)}$, $b = \sqrt{2}s(s+2t)^{-1}a$. Using theorem 2.1, we are thus led to the result

(2.3.7) $$(N_c - \xi_c)/(d\,\xi_c^{1/2}) \xrightarrow{\mathcal{L}} N(0,\ 1) \text{ as } c \to 0,$$

where $d = \sqrt{2}\,s(s+2t)^{-1}$, $\xi_c = a\eta_c$. B-M (1973) have considered the case $s = 2$, $t = 1$, $A = 1$. In that case $k = 1$, and (2.3.7) reduces to

(2.3.8) $$(N_c - \sigma\,c^{-1/2})/(\tfrac{1}{2}\cdot\sigma\cdot c^{-1/2})^{1/2} \xrightarrow{\mathcal{L}} N(0,\ 1)$$

as $c \to 0$, as result obtained by B-M (1973). Also by putting

$s = 1$, $t = 1$, we can get the result in Robbins' (1959) case.

2.4. Application in sequential hypothesis testing

We first apply theorem 2.1 in a simple modified SPRT problem.

Example 3. Let $X_1, X_2, \ldots$ be iid $N(\theta, 1)$, where θ is unknown, $-\infty < \theta < \infty$. Consider the modified SPRT problem for testing $H_0: \theta = -\frac{1}{2}$ against $H_1: \theta = \frac{1}{2}$ with boundaries $-\psi_n$ and ψ_n, ψ_n being a sequence of positive constants $\to \infty$ as $n \to \infty$. The stopping time for this problem is given by:

N is the smallest integer $n (\geq 1)$ for which

$$(2.4.1) \qquad \left| \sum_1^n X_i \right| \geq \psi_n \quad \text{i.e.,} \quad n \geq \psi_n \, |\bar{X}_n|^{-1}.$$

To develop the asymptotic theory here, we modify the stopping time in (2.4.1) as follows:

N is the smallest integer $n (\geq 1)$ for which

$$(2.4.2) \qquad n \geq \phi_\upsilon \, |\bar{X}_n|^{-1}$$

where $\phi_\upsilon = \psi_{n_\upsilon}$ is a sequence of positive constants $\to \infty$ as $\upsilon \to \infty$, and n_υ is a sequence of positive integers $\to \infty$ as $\upsilon \to \infty$. Using strong law of large numbers (SLLN),

$|\bar{X}_n|^{-1} \to |\theta|^{-1}$ a.s. for $\theta \neq 0$ as $n \to \infty$. Then, using a theorem similar to theorem 2.1 (with obvious changes in the proof), Anscombe's (1952) result and the Mann-Wald theorem, we are led to

$$(2.4.3) \qquad |\theta|^{-1/2}(N_\upsilon - |\theta|^{-1}\phi_\upsilon)/(|\theta|^{-2}\phi_\upsilon^{1/2}) \xrightarrow{\mathcal{L}} N(0, 1),$$

as $\upsilon \to \infty$, for $\theta \neq 0$. It is also easy to see how the result can be modified if we want to test $H_0 : \theta = -d$ against $H_1: \theta = d$, $d > 0$, but not necessarily equal to $1/2$.

The above result can be generalized in a particular test with power 1 problem considered by Robbins (1970). This is evidenced in the following example.

Example 4. Consider again a $N(\theta, 1)$ distribution, with θ unknown, $-\infty < \theta < \infty$. We want to test $H^-: \theta < 0$ against $H^+: \theta > 0$ ($\theta = 0$ being excluded). To get a test with power 1 for this problem, Robbins (1970) defines the stopping time N as the first integer n (≥ 1) s.t. $|\sum_1^n X_i| \geq c_n$ and accept H^+ or H^- according as $\sum_1^n X_i \geq c_n$ or $\sum_1^n X_i \leq -c_n$, where c_n is a sequence of positive constants such that $c_n/n \to 0$ as $n \to \infty$. In many important particular cases as considered by Robbins, $c_n \to \infty$ as $n \to \infty$, subject to the above condition. Then, similarly as in example 3, we can obtain the asymptotic distribution of the stopping time.

2.5 Asymptotic normality of U-statistics with random indices and a few more applications.

We use a result of Sproule (1969) on asymptotic normality of U-statistics with random indices and get asymptotic normalit proof for the stopping time in the Chow-Robbins (1965) procedure discussed in (1.1.4). We shall see that the normality assumption of the original observations is not required for the above. We shall also see that the normality assumption is not required in example 2 to get (2.3.5) and (2.3.6).

The first thing is to define U-statistics and to state the result we are going to use. With this end, we proceed as follows:

Let $X_1, X_2, \ldots$ be iidrv's, and let $f(x_1, \ldots, x_r)$ be a symmetric function of r arguments. Then as in Hoeffding (1948), we define U-statistics U_n $(n \geq r)$ by

$$(2.5.1) \qquad U_n = \binom{n}{r}^{-1} \Sigma\, f(X_{\alpha_1}, \ldots, X_{\alpha_r}), \qquad n \geq r$$

where the summation extends over all $1 \leq \alpha_1 < \ldots < \alpha_r \leq n$. $f(x_1, \ldots, x_r)$ is referred to as a 'kernel of the U-statistic with degree r'.

Suppose $\theta = E\, f(X_1, \ldots, X_r)$. Define

$$(2.5.2) \qquad f_1(x) = Ef\,(x, X_2, \ldots, X_r) = E[f(X_1, \ldots, X_r) \mid X_1 = x] .$$

Let $\varsigma_1 = V[f_1(X_1)]$. Then the needed theorem (see Sproule (1969), p. 58) can be stated as follows:

Theorem 2.2. Assume $E[f(X_1, \ldots, X_r)]^2 < \infty$ and $\varsigma_1 > 0$. Let n_υ be an increasing sequence of positive integers tending to ∞ as $\upsilon \to \infty$, and let N_υ be a sequence of proper rv's, taking on positive integral values such that,

$$N_\upsilon / n_\upsilon \xrightarrow{P} 1 \quad \text{as} \quad \upsilon \to \infty.$$

Then with $\sigma^2 = r^2 \cdot \varsigma_1$,

$$N_\upsilon^{1/2} (U_{N_\upsilon} - \Theta)/\sigma \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as} \quad \upsilon \to \infty.$$

The fundamental U-statistic is the sample mean. The sample variance $s_n^2 = (n-1)^{-1} \sum_1^n (X_i - \bar{X}_n)^2 = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} (X_i - X_j)^2/2$ is also an U-statistic.

Theorem 2.2 ensures that for iidrv's with finite fourth moments and for which $\mu_4 \neq \mu_2^2$ (μ_r denoting the r^{th} moment about mean of the original rv's)

$$N_\upsilon^{1/2} (s_{N_\upsilon}^2 - \mu_2)/(\mu_4 - \mu_2^2)^{1/2} \xrightarrow{\mathcal{L}} N(0, 1) \text{ as } \upsilon \to \infty.$$

In particular, for iid normal variables with non-zero and finite variance σ^2, one has

$$N_\upsilon^{1/2} (s_{N_\upsilon}^2 - \sigma^2)/(\sqrt{2}\sigma^2) \xrightarrow{\mathcal{L}} N(0, 1), \text{ as } \upsilon \to \infty,$$

a fact already used in example 2.

Theorems 2.1 and 2.2 are now used in the following problem of fixed length interval estimation for the population mean.

Example 5. We consider the problem (I) of section 1.1, in the Chow-Robbins' set up (1965) which led to the stopping rule in (1.1.4). We introduce a slight change in notation by defining c as the upper $100 \frac{\alpha}{2}$ % pt of the $N(0, 1)$ distribution. The asymptotic (as $d \to 0$) distribution of N_d in this case can be obtained very easily. Take $\psi_d = c^2/d^2 \to \infty$ as $d \to 0$. Put $T_n = s_n^2 + n^{-1}$ so that $P(T_n \leq 0) = 0$. With the assumption that the population fourth moment is finite and $\mu_4 \neq \sigma^4$, theorem 2.2 now leads to (2.2.2) and (2.2.3) with $a = \sigma^2$, $b = (\mu_4 - \sigma^4)^{1/2}$. Theorem 2.1 now yields,

$$\sigma(N_d - (c^2 \sigma^2/d^2))/[(\mu_4 - \sigma^4)^{1/2} . c/d]$$

$$\xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as} \quad d \to 0, \quad \text{i.e.,}$$

(2.5.3) $$(N_d - \xi_d)/((\beta_2 - 1)\,\xi_d)^{1/2} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{as} \quad d \to 0$$

where $\xi_d = c^2\sigma^2/d^2$, $\beta_2 = \mu_4/\sigma^4$. In the particular case of normal distribution, $\beta_2 = 3$, and we get,

(2.5.4) $$(N_d - \xi_d)/(2\xi_d)^{1/2} \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as} \quad d \to 0.$$

Consider again example 2; in the case $s = 2$, $E|\bar{X}_n - \mu|^2 = \sigma^2/n$, even when the X_i's are not normal, so that in this case, (1.1.7) can still be motivated even without the normality assumption. It is not clear whether the optimality properties claimed by Starr (1966b) of his procedure hold even without the normality assumption, but it is true that theorem 2 of B-M (our example 2 with $s = 2$, $t = 1$, $A = 1$) holds true even without the normality assumption if the fourth moment is finite, $\mu_4 \neq \sigma^4$ and s_n is changed to $(s_n + n^{-1})$ in defining the stopping time in (1.1.7). Theorem 2.2 and the Mann-Wald theorem lead to (2.2.2) and (2.2.3) where $T_n = s_n + n^{-1}$ $(n \geq 2)$, $a = \sigma$, $b = \frac{1}{2}\sigma^{-1}(\mu_4 - \sigma^4)^{1/2}$. The change needed in the final conclusion is that, instead of (2.3.8) we would then have,

$$(2.5.5) \quad \sigma^{1/2}(N_c - \sigma c^{-1/2})/[\tfrac{1}{2}\sigma^2(\mu_4 - \sigma^4)^{1/2} c^{-1/4}] \xrightarrow{\mathcal{L}} N(0,1)$$

as $c \to 0$, i.e.,

$$(N_c - \sigma c^{-1/2})/[\tfrac{1}{2}(\beta_2 - 1)^{1/2}(\sigma c^{-1/2})^{1/2}] \xrightarrow{\mathcal{L}} N(0, 1)$$

as $c \to 0$, which particularizes to (2.3.8) for a $N(\mu, \sigma^2)$ population.

In all these examples, a basic difference between our approach and the earlier approach by B-M is that the latter are expressing the sample variance as a sample mean, whereas we are viewing the same as a U-statistic. The former representation is possible for samples from a normal population, but not in general, whereas the U-statistic property is always true. Our result has thus wider applicability than the earlier known results in this direction.

Example 5 opens up another question. We have been able to derive the asymptotic distribution of the stopping time in fixed length interval estimation of the population mean by using the sample mean. (see (2.5.4)). The sample mean being the basic U-statistic, a pertinent question to ask is whether the given result can be generalized to similar problems associated with means of U-statistics. The key to success lies in finding suitable estimates of variances of U-statistics whose asymptotic distributions with random indices can be obtained. This is the problem we are going to tackle in the next section. Fortunately it turns out from the work of Sproule (1969) that variance of U-statistics can be consistently estimated by linear functions of U-statistics. This fact will be exploited in getting the desired asymptotic distributions with random indices.

2.6. Asymptotic normality of stopping times in fixed length interval estimation of means of U-statistics.

We first introduce a few notations in order to define an estimate of the variance of a U-statistic. We have a sequence

$X_1, X_2, \ldots,$ of iidrv's. The basic notations introduced in Section 2.5 in connection with U-statistics remain the same. Further as in Sproule (1969, p. 15) some more notations are introduced.

For each $i = 1, 2, \ldots, n$, define a U-statistic based on $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ by

(2.6.1) $$U_{(i)n} = \binom{n-1}{r}^{-1} \Sigma^{(n-1,\ r)} f(X_{\alpha_1}, \ldots, X_{\alpha_r})$$

where the summation is over all combinations $(\alpha_1, \ldots, \alpha_r)$ formed from $(1, \ldots, i-1, i+1, \ldots, n)$. Define now the statistics.

(2.6.2) $$W_{in} = n\, U_n - (n - r)\, U_{(i)n}, \quad i = 1, 2, \ldots, n$$

and $\bar{W}_n = n^{-1} \sum_1^n W_{in}$. Now, Sen (1960) has proposed the following estimate s^2_{wn} for $r^2 \varrho_1$:

(2.6.3) $$s^2_{Wn} = (n-1)^{-1} \sum_1^n (W_{in} - \bar{W}_n)^2 .$$

This estimate of $nV(U_n) = r^2 \varrho_1 + O(n^{-1})$ (see e.g. Sproule (1969) p. 6) has been used by Sproule (1969) in developing the sequential procedure. His stopping rule for obtaining a fixed length $2d$ (> 0) confidence interval for $\theta = Ef\,(X_1, \ldots, X_r)$ is given by:

(2.6.4) $\quad N_d$ is the smallest integer $n(\geq 2)$ such that $n \geq (u^2/d^2)(s^2_{Wn} + n^{-1})$

where $2\Phi(u) - 1 = 1 - \alpha$, which is the preassigned coverage probability.

Sen (1960) has shown that if $E[f^2(X_1, \ldots, X_r)] < \infty$, then $s^2_{Wn} \xrightarrow{P} r^2 \varrho_1$. Stronger results are available in theorem 3.1 of Sproule (1969, p. 35). Writing now $\Psi_d = u^2/d^2$, $\Psi_d \to \infty$ as $d \to 0$. It remains to verify (2.2.2) and (2.2.3) for applying theorem 2.1. With this end, we need the following representations:

(2.6.5) $$s^2_{Wn} = r^2(U_n^{(1)} - U_n^{(0)}) + \sum_{c=0}^{r} \alpha_n^{(c)} . U_n^{(c)}$$

given by Sproule (1969, (3.18) of p.36); the constants $\alpha_n^{(c)}$ are $O(n^{-1})$ for all $c = 0, 1, \ldots, r$. To define $U_n^{(c)}$ $U_n^{(c)}$ $(c = 0, 1, \ldots, r)$ we proceed as Sproule (1969 p.8). For $c = 0, 1, \ldots, r$, define

(2.6.6) $$q^{(c)}(x_1, \ldots, x_{2r-c})$$
$$= \binom{2r-c}{c}^{-1}\binom{r}{c}^{-1}\Sigma^{(c)} f(X_{\alpha_1}, \ldots, X_{\alpha_r}) . f(X_{\beta_1}, \ldots, X_{\beta_r})$$

where the summation $\Sigma^{(c)}$ is over all combinations $(\alpha_1, \ldots, \alpha_r)$ and $(\beta_1, \ldots, \beta_r)$, each formed from $1, 2, \ldots, 2r-c$ and such that there are exactly c integers in common. Now for each

$c = 0, 1, \ldots, r$, we define

$$(2.6.7) \qquad U_n^{(c)} = \binom{n}{2r-c}^{-1} \Sigma^{(n,2r-c)} q^{(c)}(X_{\alpha_1}, \ldots, X_{\alpha_{2r-c}})$$

where the summation extends over all combinations $\alpha_1, \ldots, \alpha_{2r-c}$ formed from the first n positive integers. Sproule (1969, p. 30) has shown that $V(U_n^{(1)} - U_n^{(0)}) = \partial_1 n^{-1} + O(n^{-2})$ where a simplified expression for ∂_1 is given in p.33. The exact expression for ∂_1 is not important, but we shall give it later for the sake of completeness. Of immediate importance is a proof of asymptotic normality of $s^2_{WN_d}$. Note that $U_n^{(c)}$'s being U-statistics with finite expectations, by Hoeffding's (1961) (see also Berk (1966)) we get $U_n^{(c)} \xrightarrow{a.s.} Eq^{(c)}(X_1, \ldots, X_{2r-c})$ as $n \to \infty$ for $c = 0, 1, \ldots, r$. Further $\alpha_n^{(c)} = O(n^{-1})$ for all $c = 0, 1, \ldots, r$. Hence, $s^2_{Wn} - r^2(U_n^{(1)} - U_n^{(0)}) \xrightarrow{a.s.} 0$ as $n \to \infty$. Sproule ((1969), pp.30-31) has shown that $E(U_n^{(1)} - U_n^{(0)}) = \varrho_1$. Using asymptotic normality of U-statistics (see Hoeffding (1948)),

$$n^{1/2}\left\{ (U_n^{(1)} - U_n^{(0)}) - \varrho_1 \right\} \xrightarrow{\mathcal{L}} N(0, \partial_1)$$

as $n \to \infty$. Hence

$$n^{1/2}(s^2_{Wn} - r^2\varrho_1) \xrightarrow{\mathcal{L}} N(0, \partial_1)$$

as $n \to \infty$. Also, proceeding as in theorem 4.5 of Sproule (1969), it is easy to verify also Anocombe's (1952) uniform continuity in probability condition, getting thereby

$$N_d^{1/2}(s^2_{WN_d} - r^2\varrho_1) \xrightarrow{\mathcal{L}} N(0, \partial_1) \text{ as } d \to 0.$$

Our theorem 2.1 will now lead to

$$(r\varrho_1^{1/2})(N_d - r^2\varrho_1\psi_d^{1/2})/(\partial_1\psi_d^{1/2}) \xrightarrow{\mathcal{L}} N(0,1) \text{ as } d \to 0.$$

To define ∂_1, a few more notations need be introduced: let

$$g_1(x_1) = f_1(x_1) - \theta, \; g_2(x_1,x_2) =$$

$$= E[f(X_1, \ldots, X_r) | X_1 = x_1, X_2 = x_2] - \theta .$$

Then define $g^{(1)}(x_1) = g_1(x_1)$, $g^{(2)}(x_1,x_2) = g_2(x_1,x_2) - g_1(x_1) - g_2(x_2)$, and $g_o(x_1) = E[g^{(1)}(X_2) \cdot g^{(2)}(x_1, X_2)]$. Now from pp. 31-33 of Sproule (1969),

$$\partial_1 = \beta_1 + r(r-1)\beta_2 + 4(r-1)^2\beta_3$$

where

$$\beta_1 = V[\{g^{(1)}(X_1)\}^2]$$

$$\beta_2 = \text{cov}[\{g^{(1)}(X_1)\}^2, \; g_o(X_1)], \; \beta_3 = V[g_o(X_1)].$$

As a simple application of the result, the case of estimation of the population mean given in example 5 can be included. Another application lies in obtaining the asymptotic distribution of the stopping time associated with fixed length interval estimation of the population variance as considered by Sproule (1969). Here finiteness of eighth moment of the X_i's is required.

In this case (see Sproule, pp. 49-51) $r = 2$, $\varrho_1 = \frac{1}{4}(\mu_4 - \mu_2^2)$ and $\partial_1 = \beta_1 + 4\beta_2 + 4\beta_3$, where

$$\beta_1 = 1/16\ (\mu_8 + 8\mu_2^2\mu_4 - 4\mu_2\mu_6 - \mu_4^2 - 4\mu_2^4),$$

$$\beta_2 = -\mu_3(\mu_5 - 2\mu_2\mu_3)/8,\ \beta_3 = \frac{1}{4}\cdot\mu_3^2\mu_2,$$

where μ_k denotes the k^{th} central moment. The variance of the asymptotic distribution of the stopping time is rather involved in general, but considerable simplification can be effected in the normal case where $\beta_2 = \beta_3 = 0$. Other U-statistics examples may also be considered.

2.7. Regression problems

The final application of our result lies in the problem of sequential confidence band estimation of the regression parameter in the general linear model. Consider a sequence $Z_1, Z_2, \ldots$ of independent normal variables with unknown common variance σ^2 and $E(Z_i) = \underline{x}'_{(i)}\,\underline{\beta}$, where $\underline{\beta}$ is $m \times 1$ vector of unknown parameters, $\underline{x}_{(i)}$ is a $m \times 1$ vector of non-stochastic known constants. It is assumed that the matrices $X'_n = (\underline{x}_1, \ldots, \underline{x}_n)$, have full ranks for all $n \geq m$, i.e. rank $X_n = m$ for all $n \geq m$.

Under the above set up, Gleser (1965, 1966) finds a confidence region I for β in R^m such that the length of the interval cut off on the β_i-axis by I is 2d (> 0) ($1 \leq i \leq m$) and $\lim_{d \to 0} P(\beta \in I) = 1 - \alpha$ ($0 < \alpha < 1$).

It may be mentioned that Gleser does not require the normality assumption of the Z_i's, but for us even this special case has some importance of its own. Let $Y_n = (Z_1, \ldots, Z_n)'$ and $\beta_n = (X_n' X_n)^{-1} X_n' Y_n$ the least squares estimator of β based on Y_n under the given linear model. The error sum of squares are denoted by $R_{on}^2 = Y_n' Y_n - Y_n' X_n \beta_n$ ($n \geq m$). One can express (Rao (1965)) R_{on}^2 as $\sum_1^{n-m} U_i$ ($n > m$) where U_i/σ^2's are iid χ_1^2 variables. We assume that $n^{-1} X_n' X_n \to \Sigma$ a p.d. matrix as $n \to \infty$. Gleser's (1965) stopping time is given by

(2.7.1) N_d the smallest integer n ($\geq m+1$) such that $n \geq (u/d^2)(R_{on}^2 / (n-m))$,

where u is the upper 100 $\alpha/2$ % point of the distribution of the weighted sum of m independent χ_1^2 variables, the weights being the latent roots of Σ. As in example 2, we end up with

(2.7.2) $(N_d - \xi_d)/(\sqrt{2}\, \xi_d) \xrightarrow{\mathcal{L}} N(0, 1)$ as $d \to 0$,

where $\xi_d = u\sigma^2/d^2$.

CHAPTER 3

BIVARIATE SEQUENTIAL POINT ESTIMATION

3.1 Introduction

Multivariate extension of the sequential point estimation procedures developed by Robbins (1959), Starr (1966b) and Starr and Woodroofe (1969) are made by Khan (1968), Rohatgi and O'Neill (1973) when the dispersion matrix is diagonal. The assumption of diagonality of the dispersion matrix, besides being of a very particular nature, makes the multivariate problem essentially the same as the univariate problem, and does not call for any new analysis beyond the one required in the univariate case.

Recently Sinha and Mukhopadhyay (1974a, 1974b) have studied the above point estimation problem in the bivariate case when (i) the dispersion matrix is arbitrary positive definite (p.d), but the weight matrix (also p.d.) is diagonal, and (ii) the weight matrix is arbitrary p.d., but the variances are equal. Also Callahan (1969), has studied the problem when the dispersion matrix is arbitrary but the weight matrix is diagonal with both elements equal (see also Gleser (1969)). The loss structure involved was weighted squared error plus cost.

In this chapter, we have considered the bivariate point estimation problem under similar loss structure as Sinha and Mukhopadhyay, when both the dispersion and weight matrices are arbitrary (p.d.). The sequential procedures suggested here enjoys the asymptotic optimality properties from the point of view of 'risk efficiency' and 'regret' (to be defined in section 3.2 in conformity with (1.1.9)) as compared to the fixed sample size optimal procedures when the dispersion matrix is known. These procedures also perform very well for moderate sample sizes.

The sequential procedure is introduced in section 3.2 and some preliminary results are proved in section 3.3. The asymptotic behaviour of the 'regret' and 'risk efficiency' are studied in section 3.4. We shall study the moderate sample size behaviour of the proposed procedure in section 3.5 by Monte-Carlo techniques generating pseudo-random bivariate normal deviates. The asymptotic normality of the stopping time is proved in section 3.6.

3.2. The sequential procedure

Let $\underline{Z}_i = (X_i, Y_i)'$, $i = 1, 2, \ldots$ be a sequence of independent bivariate normal variables with unknown mean vector $\underline{\mu} = (\mu_1, \mu_2)'$ and dispersion matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where $-\infty < \mu_1, \mu_2 < \infty$, $0 < \sigma_1, \sigma_2 < \infty$, $-1 < \rho < 1$.

Having recorded n observations, suppose the loss incurred in estimating $\underline{\mu}$ by $\overline{\underline{Z}}_n = n^{-1} \sum_{i=1}^{n} \underline{Z}_i$ is

(3.2.1) $$L_n = (\overline{\underline{Z}}_n - \underline{\mu})' A (\overline{\underline{Z}}_n - \underline{\mu}) + c\, n ,$$

where A is a known 2×2 p.d. matrix and $c(> 0)$ is the known cost per unit sample, with risk

(3.2.2) $$R_n(c) = E(L_n) = n^{-1} \operatorname{tr} (A\, \Sigma) + c\, n$$

where tr stands for trace. The risk (3.2.2) is minimized for $n = n^*$ where

(3.2.3) $$n^* = [c^{-1} \operatorname{tr} (A\Sigma)]^{1/2}$$

with minimum risk

(3.2.4) $$R(c) = R_{n^*}(c) = 2cn^* .$$

Thus if Σ were known, we could have taken a sample of size $[n^*] + 1$ and estimated $\underline{\mu}$ by the sample mean, where $[p]$ stands for the largest integer not exceeding p. But in the absence of any knowledge of Σ, no fixed sample size procedure will minimize (3.2.2) simultaneously for all Σ - matrices. So the possibility of utilising a sample of random size N is considered, which is shown to achieve the objective of attaining the minimum risk asymptotically (as $c \to 0$).

The following stopping rule, motivated from (3.2.3) is therefore suggested:

(3.2.5) The stopping time $N \equiv N_c$ is the smallest integer $n\ (\geq n_o)$ for which

$$n \geq [c^{-1} \operatorname{tr} (AS_n)]^{1/2},$$

where $(k-1)S_k = \sum_{i=1}^{k} (\underline{Z}_i - \overline{\underline{Z}}_k)(\underline{Z}_i - \overline{\underline{Z}}_k)'$, $k \geq 2$ and $n_o (\geq 2)$ is the starting sample size. When we stop, we estimate $\underline{\mu}$ by $\overline{\underline{Z}}_N$.

Using the Helmert orthogonal transformation

$$U_k = [X_1 + \ldots + X_{k-1} - (k-1)X_k]/(k(k-1))^{1/2}$$

$$V_k = [Y_1 + \ldots + Y_{k-1} - (k-1)Y_k]/(k(k-1))^{1/2}$$

one can write $(k-1)S_k = \sum_{2}^{k} \underline{W}_i \underline{W}_i'$ where $\underline{W}_i' = (U_i, V_i)$, $k = 2, 3, \ldots$.

Note that $\{(U_i, V_i)', i = 2, 3, \ldots\}$ is a sequence of iid $N_2(\underline{0}, \Sigma)$ variables. Now A and Σ being p.d., there exist non-singular matrices B and D such that $B\Sigma B' = I_2$ and $A = D'D$. Hence,

(3.2.6) $$(n-1) \operatorname{tr} (AS_n) = \sum_{2}^{n} \operatorname{tr} (A\underline{W}_i\underline{W}_i') = \sum_{2}^{n} \operatorname{tr} (D'DB^{-1}B\underline{W}_i\underline{W}_i'B'(B')^{-1})$$

$$= \sum_{2}^{n} \operatorname{tr} (Q'Q\ \underline{W}_i^* \underline{W}_i^{*\prime})$$

where $Q = DB^{-1}$, $\underline{W}_i^* = B\underline{W}_i$ $(i = 2, 3, \ldots)$ are iid $N_2(\underline{0}, I_2)$. $Q'Q$ being p.d., there exists an orthogonal matrix P such that $P'(Q'Q)P = \text{diag}(\lambda_1, \lambda_2)$, λ_1 and λ_2 being the eigen values of $Q'Q$, each positive.

Hence from (3.2.6),

(3.2.7) $$(n-1)\ \text{tr}\ (AS_n) = \sum_{2}^{n} \text{tr}\ [\text{Diag}\ (\lambda_1, \lambda_2)\underline{T}_i\underline{T}_i'\ ]$$
$$= \sum_{2}^{n} (\lambda_1\ T_{1i}^2 + \lambda_2 T_{2i}^2), \quad n \geq 2$$

where $\underline{T}_i = P\underline{W}_i^* = (T_{1i}, T_{2i})'$, $i = 2, 3, \ldots$ are iid $N_2(\underline{0}, I_2)$.

Note that,

(3.2.8) $$\lambda_1 + \lambda_2 = \text{tr}\ (Q'Q) = \text{tr}\ (A\Sigma).$$

With these reductions in mind, the rule in (3.2.5) can be alternately stated as:

(3.2.9) The stopping number $N \equiv N_c$ is the smallest integer $n\ (\geq n_o)$ for which

$$cn^2(n-1) \geq \lambda_1 \sum_{i=2}^{n} T_{1i}^2 + \lambda_2 \sum_{i=2}^{n} T_{2i}^2$$

where $n_o(\geq 2)$ is the starting sample size. Analogous to lemma 3 of Robbins (1959), we have the following lemma.

<u>Lemma 3.1</u>. <u>For any fixed</u> $n(\geq 2)$, $\bar{\underline{Z}}_n$ <u>is independent of</u> $S_2, S_3, \ldots, S_n$.

Proof. The proof is an easy consequence of a result of Basu (1955) when we observe that $\bar{Z}_n$ is complete sufficient for μ and $S_2, S_3, \ldots, S_n$ have distributions independent of μ.

Lemma 3.2. For any fixed $n (\geq n_o)$, $I_{[N=n]}$ and L_n (defined in 3.2.1) are independent, I denoting the usual indicator function.

Proof: Proof is trivial if we note that the event $[N=n]$ is described solely through $S_{n_o}, S_{n_o+1}, \ldots, S_n$ while L_n depends only on $\bar{Z}_n$ and lemma 3.1 applies.

Lemma 3.3.

(3.2.10) N is well defined, and if $0 < \sigma_i^2 < \infty$ $(i = 1,2)$, then $P[N < \infty] = 1$.

(3.2.11) N(c) is decreasing in c and $\lim_{c \to 0} (N/n^*) = 1$ a.s.

Proof: First past of (3.2.10) is obvious. Also,

$$P[N = \infty] = \lim_{n \to \infty} P[N > n)$$
$$\leq \lim_{n \to \infty} P(cn^2 < \mathrm{tr}\,(AS_n)).$$

Using (3.2.7) and SLLN, $\mathrm{tr}\,(AS_n) \to \mathrm{tr}\,(A\Sigma)$ a.s. as $n \to \infty$. Also $0 < \mathrm{tr}\,(A\Sigma) < \infty$, since $0 < \sigma_i^2 < \infty$ $(i = 1,2)$. Hence (3.2.10).

To prove (3.2.11), the first part is obvious from the definition of N. It follows immediately that $N_c \to \infty$ a.s. as $c \to 0$. Hence $\mathrm{tr}\,(AS_{N_c}) \to \mathrm{tr}\,(A\Sigma)$ a.s. and $\mathrm{tr}\,(AS_{N_c-1}) \to \mathrm{tr}\,(A\Sigma)$ a.s. as $c \to 0$.

We now use the inequality

(3.2.12) $[c^{-1} \text{ tr } (AS_N)]^{1/2} \leq N \leq n_o + [c^{-1} \text{ tr } (AS_{N-1})]^{1/2}$ a.s.

Dividing both sides of (3.2.12) by $n^* = (c^{-1} \text{ tr } (A\Sigma))^{1/2}$ and making $c \to 0$, we get (3.2.11).

Using lemma 3.2 and (3.2.10), we find that

(3.2.13) $\bar{R}(c) = E(L_N) = c[(n^*)^2 E(N^{-1}) + E(N)]$.

Regarding the usefulness of the sequential procedure in (3.2.5), we define the 'regret' as

(3.2.14) $$w(c) = \bar{R}(c) - R(c)$$

and 'risk efficiency' as

(3.2.15) $$\eta(c) = \bar{R}(c)/R(c)$$

Now, our main results can be stated as follows.

Theorem 3.1 As $c \to 0$,

(3.2.16) $w(c) = O(c)$,

(3.2.17) $\eta(c) \to 1$.

The proof is deferred to section 3.4. In the meantime, in section 3.3, we develop some desirable properties of the rule given in (3.2.5) in the form of several lemmas and theorems The property (3.2.17) is referred to as asymptotic risk efficiency.

3.3. Some properties of N

Lemma 3.4 $E(N) \le n^* + n_o + 1$

Proof: By looking at the rule in (3.2.5), one gets,

$$(N-1)^2 \le (n_o-1)^2 + c^{-1}\,\mathrm{tr}\,(AS_{N-1})$$

2/
$$\le (n_o-1)^2 + c^{-1}(N-2)^{-1}\left[\lambda_1 \sum_2^N T_{1i}^2 + \lambda_2 \sum_2^N T_{2i}^2\right]$$

$$= (n_o-1)^2 + c^{-1}(N-1)(N-2)^{-1}\left\{\lambda_1 (N-1)^{-1} \sum_2^N T_{1i}^2 + \lambda_2 (N-1)^{-1} \sum_2^N T_{2i}^2\right\}$$

Thus
$$(N-1)(N-2) \le (n_o-1)^2(N-2)(N-1)^{-1} + c^{-1}\left[(N-1)^{-1}\sum_2^N (\lambda_1 T_{1i}^2 + \lambda_2 T_{2i}^2)\right]$$

so that we obtain

$$(N-2)^2 \le (n_o-1)^2 + c^{-1}\left[(N-1)^{-1} \sum_2^N (\lambda_1 T_{1i}^2 + \lambda_2 T_{2i}^2)\right], \tag{3.3.1}$$

Thus using a theorem of Robbins (see Starr-Woodroofe (1968)), we get

$$(E(N) - 2)^2 \le (n_o-1)^2 + c^{-1}(\lambda_1 + \lambda_2)$$
$$= (n_o-1)^2 + n^{*2},$$

which leads to

(3.3.2) $$E(N) \leq 2 + (n_o-1) + n^* = n^* + n_o + 1.$$

Corrolary 3.1 $$\lim_{c \to 0} E(N/n^*) = 1.$$

Proof is trivial, since $\liminf_{c \to 0} E(N/n^*) \geq E(\liminf_{c \to 0} N/n^*)$

(by Fatou's lemma)

$= 1$ by (3.2.11)

and $\limsup_{c \to 0} E(N/n^*) \leq 1$ by lemma 3.4

Lemma 3.5 $P(N = n_o) = 0_e(c^{n_o-1})$ as $c \to 0$.

Proof: Let $Q_j^{(n-1)} = \sum_{i=2}^{n} T_{ji}^2$ $(j = 1,2, \quad n \geq 2)$,

Then $Q_1^{(n-1)}$, $Q_2^{(n-1)}$ are independent χ^2_{n-1} variables. With the notation $\lambda_{01} = \min(\lambda_1, \lambda_2)$, $\lambda_{02} = \max(\lambda_1, \lambda_2)$, so that $0 < \lambda_{01} \leq \lambda_{02} < \infty$, we get

(3.3.3) $$P(N = n_o) \leq P(\lambda_{01} \chi^2_{2(n_o-1)} < c\, n_0^2 (n_0-1))$$

$$= \int_0^{\frac{c\, n_0^2 (n_0-1)}{2\lambda_{01}}} e^{-x} x^{n_0-2} / \Gamma(n_0-1)\, dx$$

$$\leq [(2\lambda_{01})^{-1} c\, n_0^2 (n_0-1)]^{n_0-1} / \Gamma(n_0)$$

$$= 0(c^{n_0-1})$$

$$(3.3.4)\quad P(N = n_0) \geq P(\lambda_{02}\chi^2_{2(n_0-1)} < c\, n_0^2(n_0-1))$$

$$= \int_0^{\frac{c\, n_0^2\,(n_0-1)}{2\lambda_{02}}} e^{-x}\, x^{n_0-2} / \Gamma(n_0-1)\; dx$$

$$\geq e^{-\frac{c\, n_0^2\,(n_0-1)}{2\lambda_{02}}} [(2\lambda_{02})^{-1}\, c\, n_0^2(n_0-1)]^{n_0-1} / \Gamma n_0$$

$$= O(c^{n_0-1}), \quad \text{as} \quad c \to 0.$$

This completes the proof of lemma 3.5.

Lemma 3.6. For any fixed $0 < \theta < 1$,

$$P(N \leq \theta\, n^*) = O(c^{n_0/2}) \quad \underline{\text{as}} \quad c \to 0.$$

Proof: With $Q_j^{(\cdot)}$'s same as in the proof of lemma 3.5, note that

$$(3.3.5)\quad P(N \leq \theta n^*) = P(N = n_0) + \sum_{n=n_0+1}^{[\theta n^*]} P(N = n)$$

$$\leq P(N = n_0) + \sum_{n=n_0+1}^{[\theta n^*]} P(\lambda_1 Q_1^{(n-1)} + \lambda_2 Q_2^{(n-1)} < cn^2(n-1))$$

$$\leq P(N = n_0) + \sum_{n=n_0+1}^{[\theta n^*]} \inf_{h>0} \Big\{ E[e^{hcn_0^2(n_0-1)}],$$

$$[e^{-h(\lambda_1 Q_1^{(n-1)} + \lambda_2 Q_2^{(n-1)})}] \Big\}$$

$$= P(N = n_0) + \sum_{n=n_0+1}^{[\theta n^*]} \inf_{h>0} \Big\{ e^{hcn_0^2(n_0-1)}$$

$$(1+2\lambda_1 h)^{-\frac{n-1}{2}} (1+2\lambda_2 h)^{-\frac{n-1}{2}} \Big\}$$

$$\leq P(N = n_0) + \sum_{n=n_0+1}^{[\theta n^*]} \inf_{h>0} \left\{ e^{hcn_0^2(n_0-1)} [1+2(\lambda_1+\lambda_2)h]^{-\frac{n-1}{2}} \right\}$$

$$= P(N = n_0) + \sum_{n=n_0+1}^{[\theta n^*]} e^{h_0 cn_0^2(n_0-1)} [1+2(\lambda_1+\lambda_2)h_0]^{-\frac{n-1}{2}}$$

where $h_0 = \frac{1}{2(\lambda_1+\lambda_2)}(\frac{\lambda_1+\lambda_2}{cn^2} - 1) > 0$ for $n \leq [\theta n^*]$

since then $(\lambda_1+\lambda_2)/cn^2 \geq (\lambda_1+\lambda_2)/c\theta^2 n^{*2} = \theta^{-2} > 1.$

Thus, from (3.3.5),

$$(3.3.6) \quad P(N \leq \theta n^*) \leq P(N = n_0) + \sum_{n=n_0+1}^{[\theta n^*]} [e^{1-\frac{cn^2}{\lambda_1+\lambda_2}} \frac{cn^2}{\lambda_1+\lambda_2}]^{\frac{n-1}{2}}$$

$$= 0_e(c^{n_0-1}) + c^{\frac{n_0}{2}} \sum_{n=n_0+1}^{[\theta n^*]} (\frac{n^2}{\lambda_1+\lambda_2} e^{1-\frac{cn^2}{\lambda_1+\lambda_2}})^{n_0/2}$$

$$[e^{1-\frac{cn^2}{\lambda_1+\lambda_2}} (\frac{cn^2}{\lambda_1+\lambda_2})]^{\frac{n-n_0-1}{2}}$$

Now, for $n \leq [\theta n^*]$, $cn^2/(\lambda_1+\lambda_2) \leq \theta^2$. Noting that $xe^{1-x} \uparrow$ in x for $0 < x \leq 1$, we have from (3.3.6),

$$(3.3.7) \quad P(N \leq \theta n^*) \leq c^{n_0/2} \sum_{n=n_0+1}^{[\theta n^*]} \epsilon^{n-n_0-1} \left\{ \frac{n^2}{\lambda_1+\lambda_2} e^{1-\frac{cn^2}{\lambda_1+\lambda_2}} \right\}^{\frac{n_0}{2}}$$

$$+ 0_e(c^{n_0-1})$$

$$< c^{n_0/2} e^{n_0/2} \sum_{n=n_0+1}^{[\theta n^*]} (n^2/\lambda_1+\lambda_2)^{n_0/2} \epsilon^{n-n_0-1}$$

$$+ 0_e(c^{n_0-1}),$$

where $\epsilon = (\theta^2 e^{1-\theta^2})^{1/2} < 1$. Using the ratio rule of convergence,

$$P(N \leq \theta n^*) \leq O_e(c^{n_0-1}) + O(c^{n_0/2}) = O(c^{n_0/2}) \quad \text{as} \quad c \to 0.$$

<u>Lemma 3.7</u>. <u>If</u> $n \geq \theta n^*$ $(\theta > 1)$, <u>then as</u> $c \to 0$,

$$P(N > n) = O(\eta^{n-1}), \quad 0 < \eta < 1.$$

<u>Proof:</u> $P(N > n) = P(\lambda_1 Q_1^{(n-1)} + \lambda_2 Q_2^{(n-1)} > cn^2(n-1))$

$$\leq \inf_{0<h<1/2(\lambda_1+\lambda_2)} \left\{ e^{-hcn^2(n-1)} (1-2\lambda_1 h)^{-\frac{n-1}{2}} (1-2\lambda_2 h)^{-\frac{n-1}{2}} \right\}$$

$$\leq \inf_{0<h<1/2(\lambda_1+\lambda_2)} [e^{-hcn^2(n-1)} \{1-2(\lambda_1+\lambda_2)h\}^{-\frac{n-1}{2}}]$$

$$= e^{-h_0 cn^2(n-1)} [1-2(\lambda_1+\lambda_2)h_0]^{-\frac{n-1}{2}}.$$

where $h_0 = \frac{1}{2(\lambda_1+\lambda_2)}(1 - \frac{\lambda_1+\lambda_2}{cn^2})$; $0 < h_0 < \frac{1}{2(\lambda_1+\lambda_2)}$ for $n \geq \theta n^*$,

since then $cn^2/(\lambda_1+\lambda_2) \geq \theta^2 > 1$.

Hence,

(3.3.8) $P(N > n) \leq [e^{1-\frac{cn^2}{\lambda_1+\lambda_2}} \cdot \frac{cn^2}{\lambda_1+\lambda_2}]^{\frac{n-1}{2}}$, and noting that

$e^{1-x} x \downarrow$ in x for $x > 1$, one gets

(3.3.9) $P(N > n) \leq [e^{1-\theta^2} \theta^2]^{\frac{n-1}{2}} = \eta^{n-1}$

with $0 < \eta = (\theta^2 e^{1-\theta^2})^{1/2} < 1$.

In the following lemmas I stands for the usual indicator function, and $\epsilon(<1)$ is some positive number.

Lemma 3.8 $E[\frac{(N-n^*)^2}{N} I_{[|N-n^*| \le \epsilon \sqrt{n^*}]}] = O(1)$, as $c \to 0$.

Proof: $E[\frac{(N-n^*)^2}{N} I_{[|N-n^*| \le \epsilon \sqrt{n^*}]}]$

$$\le \frac{\epsilon^2 n^*}{(n^* - \epsilon \sqrt{n^*})} P(|N - n^*| \le \epsilon \sqrt{n^*})$$

$$\le \epsilon^2/[1 - \epsilon(n^*)^{-1/2}] = O(1) \text{ as } c \to 0.$$

Lemma 3.9: $E[\frac{(N-n^*)^2}{N} I_{[\epsilon \sqrt{n^*} < N - n^* \le \epsilon n^*]}] = O(1)$ as $c \to 0$.

Proof: $E[\frac{(N-n^*)^2}{N} I_{[\epsilon \sqrt{n^*} < N-n^* \le \epsilon n^*]}]$

$$\le \frac{n^*}{(n^* + \epsilon \sqrt{n^*})} E[\frac{(N-n^*)^2}{n^*} I_{[\epsilon < \frac{(N-n^*)}{\sqrt{n^*}} \le \epsilon \sqrt{n^*}]}]$$

$$\le E[X^2 I_{[\epsilon < X \le \epsilon \sqrt{n^*}]}], \text{ where } X = \frac{N-n^*}{\sqrt{n^*}}.$$

Suppose $F(x)$ is the d.f. of X. Then,

$$(3.3.10) \quad E[X^2 I_{[\epsilon < X \le \epsilon \sqrt{n^*}]}] = -\int_{\epsilon}^{\epsilon\sqrt{n^*}} x^2 d(1 - F(x))$$

$$\le \epsilon^2(1 - F(\epsilon)) + 2\int_{\epsilon}^{\epsilon\sqrt{n^*}} x(1 - F(x))dx.$$

Now, $1 - F(x) = P\{N > n^* + x\sqrt{n^*}\} = P(N > t)$

where $t = t(x) = [n^* + x\sqrt{n^*}]$. Thus, we have,

$$P(N > t) \leq P(\lambda_1 Q_1^{(t-1)} + \lambda_2 Q_2^{(t-1)} > ct^2(t-1)$$

$$= P[\lambda_1 Q_1^{(t-1)} + \lambda_2 Q_2^{(t-1)} - (\lambda_1 + \lambda_2)(t-1) > (t-1)(ct^2 - \lambda_1 - \lambda_2)].$$

Now $ct^2 - \lambda_1 - \lambda_2 \geq c(n^* + x\sqrt{n^*} - 1)^2 - \lambda_1 - \lambda_2$

$$= c(n^{*2} + 2xn^{*3/2} + x^2 n^* - 2n^* - 2x\sqrt{n^*} + 1) - cn^{*2}$$

$$\geq 2cxn^{*3/2} - 2cn^* - 2cx\sqrt{n^*}$$

$> k\,x\,c^{1/4}$, for small c, where k is some positive constant. Hence

$$P(N > t) \leq P[\lambda_1 \{Q_1^{(t-1)} - (t-1)\} + \lambda_2 \{Q_2^{(t-1)} - (t-1)\} > k\,x\,c^{1/4}(t-1)].$$

Note that for small c and positive x, $kx(t-1)c^{1/4}$ is positive. Thus

$$P[N > t] \leq \frac{E[\lambda_1 \{Q_1^{(t-1)} - (t-1)\} + \lambda_2 \{Q_2^{(t-1)} - (t-1)\}]^4}{k^4 x^4 c(t-1)^4}$$

$$= \frac{(\lambda_1^4 + \lambda_2^4)[12(t-1)^2 + 48(t-1)] + 24\lambda_1^2\lambda_2^2(t-1)^2}{k^4 x^4 c(t-1)^4}$$

$$\leq \frac{k_1 (t-1)^2}{x^4 c(t-1)^4} \leq k_2/x^4, \quad \text{for small } c,$$

noting the fact that $t = O(c^{-1/2})$, where k's are positive constants independent of c. Thus from (3.3.10), one gets,

$$E[X^2 I_{[\epsilon < X \leq \epsilon\sqrt{n^*}]}] \leq \epsilon^2(1 - F(\epsilon)) + 2k_2 \int_{\epsilon}^{\epsilon\sqrt{n^*}} x^{-3}\,dx$$

$$= \epsilon^2(1 - F(\epsilon)) + k_2\,\epsilon^{-2}(1 - 1/n^*)$$

$$= o(1) \quad \text{as} \quad c \to 0.$$

Lemma 3.10. $E[\frac{(N-n^*)^2}{N} I_{[-\epsilon n^* < N - n^* < -\epsilon\sqrt{n^*}]}] = o(1)$

as $c \to 0$.

Proof: $E[\frac{(N-n^*)^2}{N} I_{[-\epsilon n^* < N - n^* < -\epsilon\sqrt{n^*}]}]$

$$\leq \frac{n^*}{(n^* - \epsilon\sqrt{n^*})} E[\frac{(N-n^*)^2}{n^*} I_{[-\epsilon\sqrt{n^*} < \frac{N-n^*}{\sqrt{n^*}} < -\epsilon]}]$$

$$= (1 + o(1))\, E[X^2 I_{[-\epsilon\sqrt{n^*} < X < -\epsilon]}], \text{ for}$$

small c, where $X = (N - n^*)/\sqrt{n^*}$.

But

(3.3.11) $E[X^2 I_{[-\epsilon\sqrt{n^*} < X < -\epsilon]}]$

$$= \int_{-\epsilon\sqrt{n^*}}^{-\epsilon} x^2\,dF(x), \quad \text{where } F(x) = P(X \leq x)$$

$$= \epsilon^2 F(-\epsilon) - 2\int_{-\epsilon\sqrt{n^*}}^{-\epsilon} x\,F(x)\,dx$$

$$= \epsilon^2 F(-\epsilon) + 2 \int_{\epsilon}^{\epsilon\sqrt{n^*}} x F(-x)dx .$$

Now $F(-x) = P(N \leq n^* - x \sqrt{n^*})$

$$= P[N \leq (1-\epsilon)n^*] + P[(1-\epsilon)n^* < N \leq n^* - x\sqrt{n^*}]$$

$$\leq O(c^{n_o/2}) + P((1-\epsilon)n^* < N \leq n^* - x\sqrt{n^*}),$$

for small c, by lemma 3.6.

Let $m_1 = [(1-\epsilon)n^*]$, $m_2 = [n^* - x\sqrt{n^*}]$.

Thus,

$$(3.3.12) \quad P[m_1 + 1 \leq N \leq m_2] \leq P[\bigcup_{n=m_1+1}^{m_2} \{\lambda_1 Q_1^{(n-1)} + \lambda_2 Q_2^{(n-1)} < cn^2(n-1)\}]$$

$$= P[\bigcup_{n=m_1+1}^{m_2} \{\lambda_1(Q_1^{(n-1)} - (n-1)) + \lambda_2(Q_2^{(n-1)} - (n-1)) < (n-1)(cn^2 - \lambda_1 - \lambda_2)\}].$$

Noting that for $n \leq n^* - x\sqrt{n^*}$

$$cn^2 - \lambda_1 - \lambda_2 \leq c(n^* - x\sqrt{n^*})^2 - c\, n^{*2}$$

$$= -2cx(n^*)^{3/2} + cx^2 n^*$$

$< -k\, c^{1/4} x$, for small c, where k is some positive constant independent of c.

Thus (3.3.12) is less than or equal to,

$$(3.3.13) \quad P[\min_{m_1+1 \leq n \leq m_2} \{\lambda_1(Q_1^{(n-1)} - (n-1)) + \lambda_2(Q_2^{(n-1)} - (n-1))\} < -(m_1 - 1)k\, c^{1/4} x].$$

Also $\left\{\lambda_1 Q_1^{(n-1)} + \lambda_2 Q_2^{(n-1)} - (\lambda_1+\lambda_2)(n-1) = Q^{(n-1)};\ n \geq 2\right\}$ is a stationary martingale sequence. Using Kolmogorov's inequality for martingales (see p. 399, Loeve (1968)), (3.3.13) cannot exceed

$$E[Q^{(m_2-1)} - Q^{(m_1)}]^4 \Big/ [(m_1-1)^4 k^4 c\, x^4]$$

$$= E[Q^{(m_2-m_1-1)}]^4 \Big/ [(m_1-1)^4 k^4\, c\, x^4]$$

$$\leq k_1 \frac{(m_2-m_1-1)^2}{(m_1-1)^4 c\, x^4} \leq \frac{k_2}{(n^*)^2 c\, x^4} \leq \frac{k_3}{x^4}$$

where k's are all positive constants.

Hence, from (3.3.11) we get,

$$E[X^2 I_{[-\epsilon\sqrt{n^*} < X < -\epsilon]}]$$

$$\leq \epsilon^2 F(-\epsilon) + O(c^{n_o/2}) \int_{\epsilon}^{\epsilon\sqrt{n^*}} x\,dx + k_3 \int_{\epsilon}^{\epsilon\sqrt{n^*}} x^{-3}\,dx$$

$$\leq \epsilon^2 F(-\epsilon) + O(c^{\frac{n_o-1}{2}}) + k_3/2\epsilon^2$$

$$\leq O(1) \text{ as } c \to 0 \text{ since } n_o \geq 2.$$

3.4. Proof of Theorem 3.1

Asymptotic Risk Efficiency

Risk efficiency, $\eta(c) = \frac{1}{2}[n^* E(N^{-1}) + E(N/n^*)]$.

To prove (3.2.17), in view of Corrolary 3.1, it is sufficient to prove that

(3.4.1) $$n^*E(N^{-1}) \to 1 \quad \text{as} \quad c \to 0$$

By Fatou's lemma and (3.2.11) we get

(3.4.2) $$\liminf_{n \to 0} \; n^*E(N^{-1}) \geq 1 \, .$$

Now, for ϵ in (0, 1), let $\alpha = [(1-\epsilon)n^*]$

Then $n^*E(N^{-1}) \leq (n^*/n_o)P(N \leq \alpha) + (n^*/(\alpha+1))\,P(N > \alpha)$ and thus by lemma 3.6 and (3.2.11),

$$\limsup_{c \to 0} \; n^*E(N^{-1}) \leq \limsup_{c \to 0} \; O(c^{\frac{n_o-1}{2}}) + (1-\epsilon)^{-1} \text{ since } n_o \geq 2.$$

Now, ϵ being arbitrary, $\limsup_{c \to 0} \; n^*E(N^{-1}) \leq 1$, and so the procedure given in (3.2.5) is asymptotically risk efficient.

<u>Asymptotic behaviour of regret</u>

We first note that the regret

$$w(c) = c\,E[(N-n^*)^2/N].$$

So, to prove (3.2.16) it is sufficient to show

(3.4.3) $$E[(N-n^*)^2/N] = O(1) \quad \text{as} \quad c \to 0.$$

For ϵ in (0, 1), let $\alpha = [(1-\epsilon)n^*]$, $t_1 = [n^* - \epsilon\sqrt{n^*}]$, $t_2 = [n^* + \epsilon\sqrt{n^*}]$, $\beta = [(1+\epsilon)n^*]$ so that

$$E[(N-n^*)^2/N] = E[\frac{(N-n^*)^2}{N} \{I_{[n\leq \alpha]} + I_{[\alpha < N < t_1]}$$

$$+ I_{[t_1 \leq N \leq t_2]} + I_{[t_2 < N < \beta]} + I_{[N \geq \beta]}\}]$$

$$= O(c^{\frac{n_o - 2}{2}}) + O(1) = O(1) \cdot \text{ if } n_o \geq 2,$$

by using lemmas 3.6, 3.7, 3.8, 3.9 and 3.10. This completes the proof of the main theorem stated in section 3.2.

A remark: The results we have derived here can be extended immediately to the $p(\geq 3)$ - variate normal case. There $tr\,(AS_n)$ = weighted average of p-independent χ^2_{n-1} variables where the sum of the weights = $tr\,(A\,\Sigma)$. However, in view of inadmissibility of the sample mean for the population mean (see Stein (1956), Stein-James (1961)) with respect to squared error loss we are rather reluctant to propose the sample mean as the point estimator of the population mean in the $p\,(\geq 3)$-variate case.

3.5 Moderate sample behaviour of the stopping time

In this section, we present the results of a few Monte-Carlo experiments with pseudo random bivariate normal deviates using the stopping rule given in (3.2.5). We fix

$a_{11} = 2,\quad a_{12} = 1,\quad a_{22} = 3,\quad \sigma_1 = \sigma_2 = 1,$

with $\rho = .01(.02).09$. We compute c from the relation $c = (a_{11} + 2a_{12}\rho + a_{22})/n^{*2}$ where n^* takes the values

$$5(5)25,\ 40,\ 50(25)100,\ 150,\ 200\ .$$

For each entry, we estimate $E(N)$, $E(1/N)$ by repeating the experiment 100 times in H-400 electronic computer. Also we keep $n_o = 2$ fixed. Computations are presented in tables I-V on pages 55-57.

3.6. Asymptotic normality of the stopping time

In this section we prove the asymptotic normality of the stopping time N_c as $c \to 0$ by using Theorem 2.1. With this end observe from (3.2.5) that the sequential procedure adopted by us can be written as follows:

(3.6.1) The stopping time N_c is the first integer $n(\geq n_o \geq 2)$ for which $n \geq \psi_c T_n$ where $\psi_c = c^{-1/2}$, $T_n = (\text{tr}(AS_n))^{1/2}$. Note that $\psi_c \to \infty$ as $c \to 0$, $P(T_n \leq 0) = 0$ for all $n \geq 2$. Remembering from (3.2.7) the representation of $\text{tr}(AS_n)$ as the mean of iidrv's with finite mean and variance, and using Anscombe's (1952) result, we immediately have as $c \to 0$

$$(3.6.2) \qquad N_c^{1/2}(T_{N_c}^2 - (\lambda_1 + \lambda_2)) \Big/ \sqrt{2}\,(\lambda_1^2 + \lambda_2^2)^{1/2} \xrightarrow{\mathcal{L}} N(0,\ 1)$$

and

$$(3.6.3) \quad N_c^{\frac{1}{2}} \left(T^2_{N_c-1} - (\lambda_1 + \lambda_2)\right) \Big/ \sqrt{2}\, (\lambda_1^2 + \lambda_2^2)^{\frac{1}{2}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Using now a theorem of Mann and Wald (Rao (1965), p. 319) one gets as $c \to 0$

$$(3.6.4) \quad N_c^{\frac{1}{2}} \left(T_{N_c} - \sqrt{\lambda_1 + \lambda_2}\right) \Big/ \left[(\lambda_1^2 + \lambda_2^2) / 2(\lambda_1 + \lambda_2)\right]^{\frac{1}{2}} \xrightarrow{\mathcal{L}} N(0, 1)$$

and

$$(3.6.5) \quad N_c^{\frac{1}{2}} \left(T_{N_c-1} - \sqrt{\lambda_1 + \lambda_2}\right) \Big/ \left[(\lambda_1^2 + \lambda_2^2) / 2\, (\lambda_1 + \lambda_2)\right]^{\frac{1}{2}} \xrightarrow{\mathcal{L}} N(0,1).$$

It is now easy to observe that the conditions (2.2.2) and (2.2.3) of theorem 2.1 are satisfied with $a = \sqrt{\lambda_1 + \lambda_2} > 0$ and $b = [(\lambda_1^2 + \lambda_2^2) / 2(\lambda_1 + \lambda_2)]^{1/2} > 0$.

Applying the theorem 2.1, we then have

$$(3.6.6) \quad (N_c - \xi_c) \Big/ (d\xi_c^{\frac{1}{2}}) \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as} \quad c \to 0$$

where $d = (\lambda_1^2 + \lambda_2^2)^{\frac{1}{2}} / \sqrt{2}\, (\lambda_1 + \lambda_2)$, $\xi_c = a\Psi_c = c^{-\frac{1}{2}} (\lambda_1 + \lambda_2)^{\frac{1}{2}}$.

Table I: ASN, Risk Efficiency and Regret when $\rho = 0.1$.

n^*	$c \times 10^2$	$E(N)$	$E(1/N)$	$\eta(c)$	$w(c)$
5	20.8000	4.93	.2209	1.0454	.09451
10	5.2000	9.63	.1133	1.0483	.05024
15	2.3111	14.69	.0731	1.0383	.02656
20	1.3000	19.64	.0544	1.0353	.01836
25	0.8320	25.49	.0397	1.0060	.00249
40	0.3250	39.94	.0253	1.0047	.00122
50	0.2080	49.92	.0201	1.0033	.00069
75	0.0924	56.19	.0100	0.7521	- .03437
100	0.0520	99.87	.0100	1.0013	.00014
150	0.0231	151.39	.0066	1.0010	.00006
200	0.0130	150.59	.0037	0.7506	- .01297

Table II: ASN, Risk Efficiency and Regret when $\rho = 0.3$.

n^*	$c \times 10^2$	$E(N)$	$E(1/N)$	$\eta(c)$	$w(c)$
5	22.4000	4.92	.2205	1.0431	.09667
10	5.6000	9.11	.1275	1.0929	.10408
15	2.4889	14.79	.7269	1.0381	.02848
20	1.4000	19.64	.0520	1.0112	.00627
25	0.8960	25.02	.0407	1.0098	.00438
40	0.3500	40.12	.0251	1.0041	.00116
50	0.2240	50.18	.0200	1.0034	.00077
75	0.0995	75.01	.0133	1.0021	.00032
100	0.0560	99.99	.0100	1.0019	.00022
150	0.0248	149.40	.0067	1.0010	.00007
200	0.0140	149.97	.0037	0.7507	- .01396

Table III: ASN, Risk Efficiency and Regret when ρ = .05.

n*	c x 10^2	E(N)	E(1/N)	η(c)	w(c)
5	24.0000	4.76	.2338	1.0606	.14545
10	6.0000	9.49	.1207	1.0780	.09369
15	2.6667	15.06	.0715	1.0383	.03068
20	1.5000	19.89	.0516	1.0130	.00781
25	0.9600	25.41	.0400	1.0081	.00389
40	0.3750	39.80	.0254	1.0057	.00170
50	0.2400	50.27	.0200	1.0040	.00096
75	0.1067	75.26	.0133	1.0027	.00044
100	0.0600	74.56	.0076	0.7515	-.02982
150	0.0267	149.84	.0067	1.0014	.00011
200	0.0150	199.66	.0050	1.0008	.00005

Table IV: ASN, Risk Efficiency and Regret when ρ = 0.7

n*	c x 10^2	E(N)	E(1/N)	η(c)	w(c)
5	25.6100	4.67	.2392	1.0650	.16657
10	6.4000	9.49	.1209	1.0789	.10104
15	2.8444	14.68	.0735	1.0404	.03449
20	1.6000	19.13	.0543	1.0216	.01381
25	1.0240	24.81	.0416	1.0157	.00802
40	0.4000	40.53	.0249	1.0061	.00195
50	0.2560	50.00	.0202	1.0046	.00118
75	0.1138	55.57	.0102	0.7521	-.04230
100	0.0640	98.64	.0101	1.0027	.00034
150	0.0284	149.79	.0067	1.0017	.00014
200	0.0160	199.14	.0050	1.0010	.00007

Table V: ASN, Risk Efficiency and Regret when $\rho = 0.9$

n*	c x 10^2	E(N)	E(1/N)	$\eta(c)$	w(c)
5	27.2000	4.56	.2445	1.0673	.18310
10	6.8000	9.29	.1266	1.0976	.13269
15	3.0222	14.38	.0798	1.0779	.07064
20	1.7000	19.50	.0584	1.0717	.04875
25	1.0880	25.45	.0422	1.0367	.01999
40	0.4250	39.50	.0257	1.0089	.00306
50	0.2720	49.32	.0204	1.0046	.00125
75	0.1209	73.76	.0137	1.0046	.00084
100	0.0680	99.55	.0100	1.0024	.00032
150	0.0302	149.92	.0067	1.0021	.00019
200	0.0170	199.50	.0050	1.0011	.00007

<u>Remarks</u>: (1) Average sample size is quite near the values of n* in the range of c considered for computations.

(2) Negative regrets at places are, we believe, due to sampling fluctuations.

(3) The performance of the rule R, on the whole, is very satisfactory and it can be recommended for use in practice.

CHAPTER 4

SEQUENTIAL ESTIMATION OF THE DIFFERENCE OF TWO NORMAL MEANS

4.1 Introduction. A sequential fixed-width confidence interval procedure for estimating the difference of the means of two normal populations (with unknown and unequal variances) was studied by Robbins et al (1967) and Srivastava (1970). In this chapter we shall investigate the possibility of utilising sequential procedures to get hold of a point-estimator of the difference of two normal means.

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two independent sequences of rv's, the X's iid $N(\mu_1, \sigma_1^2)$, the Y's iid $N(\mu_2, \sigma_2^2)$ with $-\infty < \mu_1, \mu_2 < \infty$, $0 < \sigma_1, \sigma_2 < \infty$, μ_1, μ_2 being unknown. We wish to find an estimator for the parameter $\mu = \mu_1 - \mu_2$. Taking samples of sizes r and s from X's and Y's respectively, suppose the loss incurred in estimating μ by $W = \bar{X}_r - \bar{Y}_s$ is

(4.1.1) $$L_{r,s} = A(W - \mu)^2 + c(r + s),$$

$\bar{X}_r$ and $\bar{Y}_s$ being the respective sample means, A and c are known positive constants, c being the cost per observation.

The risk is

$$(4.1.2) \qquad \upsilon_{r,s}(c) = A(\sigma_1^2/r + \sigma_2^2/s) + c(r+s).$$

Throughout we assume σ_1, σ_2 to be fixed, while c tends to zero. For known σ_1, σ_2 and fixed c, the pair (r^*, s^*) for which the risk (4.1.2) is a minimum, is given by

$$(4.1.3) \qquad r^* = b\sigma_1 , \quad s^* = b\sigma_2$$

with $b^2 = Ac^{-1}$.

For this pair

$$(4.1.4) \qquad r^*/s^* = \sigma_1/\sigma_2$$

and the total sample size is

$$(4.1.5) \qquad n^* = r^* + s^* = b(\sigma_1 + \sigma_2),$$

the corresponding minimum risk being

$$(4.1.6) \qquad \upsilon(c) = \upsilon_{r^*,s^*}(c) = 2cn^*.$$

But when we are ignorant about σ_1, σ_2, which is the case in most applications, no fixed sample size procedure will minimize (4.1.2) simultaneously for all $0 < \sigma_1, \sigma_2 < \infty$. However, we shall propose a sequential procedure determining r and s as rv's as follows. Define for $i, j \geq 2$

$$u_i^2 = (i-1)^{-1} \sum_{k=1}^{i} (X_k - \bar{X}_i)^2, \quad v_j^2 = (j-1)^{-1} \sum_{k=1}^{j} (Y_k - \bar{Y}_j)^2,$$

these being the usual estimators of σ_1^2 and σ_2^2 for which $u_i \to \sigma_1$ a.s., $v_j \to \sigma_2$ a.s. as i and $j \to \infty$. We take n_o (≥ 3) observations on X and Y each to start with. Then if at any stage we have taken i observations on X and j observations on Y with $n = i + j$ ($\geq 2n_o$), we take the next observation on X or Y according as

(4.1.7) $\quad i/j \leq u_i/v_j \quad$ or $\quad i/j > u_i/v_j$.

The motivation seems to be clear when one looks at (4.1.4). We now propose a stopping rule which is motivated from (4.1.3) and is as follows:

(4.1.8) The stopping time $N \equiv N_c$ is the first integer $n(\geq 2n_o)$ such that if $R = r$ observations on X and $S = s$ observations on Y have been taken, with $r + s = n$,

$$r \geq bu_r \quad \text{and} \quad s \geq bu_s;$$

$N = R + S$.

In section 4.2, the cost of ignorance of σ_1, σ_2 is studied, and we are able to show that $E(N) - n^*$ is less than some finite constant for all $0 < c, \sigma_1, \sigma_2 < \infty$.

We use the notation L_n instead of $L_{r,s}$. Using independence of $(\bar{X}_i, \bar{Y}_j ; i \geq 1, j \geq 1)$ and $(u_i^2, v_j^2 ; i \geq 2, j \geq 2)$, one may observe that L_n is independent of $I_{[N=n]}$ for all $n \geq 2n_o$, where I stands for the usual indicator function. Hence

$$(4.1.9) \quad \bar{\upsilon}(c) = E(L_N) = AE(\sigma_1^2 R^{-1} + \sigma_2^2 S^{-1}) + cE(N) .$$

As possible measures of usefulness of our procedure in (4.1.8), we define for each c, the 'risk efficiency'

$$(4.1.10) \quad \eta(c) = \bar{\upsilon}(c)/\upsilon(c) = \tfrac{1}{2}[b(\sigma_1 + \sigma_2)^{-1} E(\sigma_1^2 R^{-1} + \sigma_2^2 S^{-1}) + E(N/n^*)]$$

and the 'regret'

$$(4.1.11) \quad w(c) = \bar{\upsilon}(c) - \upsilon(c) = cE[(R - r^*)^2/R + (S - s^*)^2/S].$$

Regarding $\eta(c)$ and $w(c)$, we prove the following theorems.

Theorem 4.1. For fixed σ_1, σ_2, $\lim_{c \to 0} \eta(c) = 1$.

Theorem 4.2. For fixed σ_1, σ_2, $\lim_{c \to 0} w(c) = O(c)$.

The proofs of these theorems are deferred to section 4.4. In section 4.3 we describe a few properties of the rule given in (4.1.8). These results being asymptotic in nature, in section 4.5 we have presented results of a Monte-Carlo investigation for moderate sample size behaviour of the sequential procedure using pseudo-random normal deviates.

Finally, asymptotic distribution of the stopping time N has been studied in section 4.6. With this end, we prove the following theorem.

<u>Theorem 4.3</u> $(N - n^*)(2/n^*)^{\frac{1}{2}} \xrightarrow{\mathcal{L}} N(0,1)$ <u>as</u> $c \to 0$.

4.2. The cost of ignorance of σ_1 and σ_2.

In this section we prove the following result.

<u>Lemma 4.1</u> $E(N) \leq n^* + 2n_o$, <u>for all</u> $0 < c, \sigma_1, \sigma_2 < \infty$.

<u>Proof:</u> Suppose that $R > n_o$ and that just before the Rth observation on X there were (R-1 observations on X and) j observations on Y. Then

$$(4.2.1) \qquad R - 1 \leq bu_{R-1}$$

because otherwise since $j \geq bv_j$, we would have stopped at the (R-1, j)th stage. It follows from (4.2.1) and the definition of u_{R-1} that for $R > n_o$,

(4.2.2) $$(R-1)^2(R-2) \leq b^2 \sum_1^{R-1} (X_i - \bar{X}_{R-1})^2$$
$$\leq b^2 \sum_1^{R} (X_i - \bar{X}_R)^2$$
$$\leq b^2 \sum_1^{R} (X_i - \mu_1)^2 .$$

Now,

$$(R-1)^2(R-2) - (R-n_o)^2 \cdot R$$
$$= R[(R-1)^2 - (R-n_o)^2] - 2(R-1)^2$$
$$= (n_o-1)\, R\, (2R- n_o - 1) - 2(R-1)^2$$
$$\geq (n_o - 1)(R- 1)^2 - 2(R- 1)^2 \geq 0 \qquad \text{if} \quad n_o \geq 3$$

Thus for $R > n_o$, from (4.2.2) we obtain

(4.2.3) $$b^2 \sum_1^{R} (X_i - \mu_1)^2 \geq R(R - n_o)^2$$

which trivially holds even if $R = n_o$.

Using Wiener's dominated ergodic theorem, from (4.2.1) we conclude that $E(R) < \infty$. Using convexity of the function $(R- n_o)^2 \cdot R$ (since $R \geq n_o$ a.s.), Jensen's inequality and Wald's lemma, we get

(4.2.4) $$b^2\sigma_1^2\, E(R) \geq E(R)(E(R) - n_o)^2$$

which gives

(4.2.5) $$E(R) \leq r^* + n_o .$$

We can similarly obtain

(4.2.6) $$E(S) \leq s^* + n_o \; .$$

Combining (4.2.5) and (4.2.6) we get the lemma 4.1.

4.3. Some properties of the rule (4.1.8)

Lemma 4.2. $$\lim_{c \to 0} E(N/n^*) = 1$$

Proof: From lemma 4.1, we have

$$E(N/n^*) \leq 1 + (2n_o/n^*)$$

which leads to

(4.3.1) $$\limsup_{c \to 0} E(N/n^*) \leq 1 \; .$$

Clearly, from (4.1.8), $\liminf_{c \to 0} (N/n^*) \geq 1$ a.s. using the a.s. convergence of u_r and v_s to σ_1 and σ_2 respectively as $r \to \infty$, $s \to \infty$. Also $R \to \infty$ a.s., $S \to \infty$ a.s. as $c \to 0$. Thus, by Fatou's lemma,

(4.3.2) $$\liminf_{c \to 0} E(N/n^*) \geq E(\liminf_{c \to 0} N/n^*) \geq 1 .$$

Combining (4.3.1) and (4.3.2) we get the lemma 4.2.

It may be noted that from (4.1.8), (4.2.1) (and a relation similar to (4.2.1) with S) we can conclude

$$\lim_{c \to 0} (R/r^*) = \lim_{c \to 0} E(S/s^*) = \lim_{c \to 0} E(N/n^*) = 1 \quad \text{a.s.} \tag{4.3.3}$$

Lemma 4.3. For fixed θ in $(0,1)$, $P(R \leq \theta r^*) = O(c^{\frac{n_o - 1}{2}})$ as $c \to 0$.

Proof. Let $\alpha = [\theta r^*]$, $p(r, \sigma_i; c) = \dfrac{cr^2(r-1)}{A\sigma_i^2}$, $i = 1,2$; $r \geq n_o$.

$$P(R \leq \theta r^*) = P(R = n_o) + \sum_{r=n_o+1}^{\alpha} P(R = r) \tag{4.3.4}$$

$$\leq P(u_{n_o} \leq (Ac^{-1})^{1/2} n_o) + \sum_{r=n_o+1}^{\alpha} P(R = r)$$

$$\leq P(\chi^2_{n_o - 1} \leq p(n_o, \sigma_1; c)) + \sum_{r=n_o+1}^{\alpha} P(\chi^2_{r-1} \leq p(r, \sigma_1; c)).$$

Now, the 2nd term in (4.3.4) cannot exceed

$$\sum_{r=n_o+1}^{\alpha} \inf_{h>0} E[\exp\left\{\frac{hcr^2(r-1)}{A\sigma_1^2} - h\chi^2_{r-1}\right\}]$$

$$= \sum_{r=n_o+1}^{\alpha} \inf_{h>0} [(1+2h)^{-\frac{r-1}{2}} \exp\left\{\frac{hcr^2(r-1)}{A\sigma_1^2}\right\}]$$

$$= \sum_{r=n_o+1}^{\alpha} (1+2h_o)^{-\frac{r-1}{2}} \exp\left\{\frac{h_o cr^2(r-1)}{A\sigma_1^2}\right\}$$

where $h_o = \frac{1}{2}\left(\dfrac{A\sigma_1^2}{cr^2} - 1\right)$.

For $r \leq \alpha$,

$$cr^2 \le c\alpha^2 \le c\theta^2 r^{*2} = A\sigma_1^2\,\theta^2 < A\sigma_1^2, \text{ so that } h_o > 0.$$

Thus,

$$P(n_o + 1 \le R \le \theta r^*) \le \sum_{r=n_o+1}^{\alpha} \left[\left\{ \exp\left(1 - \frac{cr^2}{A\sigma_1^2}\right) \right\} \cdot \frac{cr^2}{A\sigma_1^2} \right]^{\frac{r-1}{2}}.$$

We note that for $0 < x < 1$, $xe^{1-x} \uparrow$ in x, so that

$$e^{1-\frac{cr^2}{A\sigma_1^2}} \frac{cr^2}{A\sigma_1^2} \le e^{1-\theta^2}\theta^2 = \epsilon(\text{say}) < 1,$$

using the fact that $e^{x-1} \ge x$ for real x, equality if and only if $x = 1$.

Hence,

$$P(n_o+1 \le R \le \theta r^*) \le c^{\frac{n_o}{2}} \sum_{r=n_o+1}^{\alpha} \epsilon^{\frac{r-n_o-1}{2}} \left(\frac{r^2}{A\sigma_1^2}\right)^{\frac{n_o}{2}} \left(e^{1-\frac{cr^2}{A\sigma_1^2}}\right)^{\frac{n_o}{2}}$$

$$\le c^{\frac{n_o}{2}} e^{\frac{n_o}{2}} \sum_{r=n_o+1}^{\alpha} \epsilon^{\frac{r-n_o-1}{2}} \left(\frac{r^2}{A\sigma_1^2}\right)^{\frac{n_o}{2}}$$

$$= 0_e(c^{\frac{n_o}{2}}), \quad \text{as} \quad c \to 0,$$

using the ratio test of convergence of series of positive terms.

Getting back to (4.3.4),

$$P(R \le \theta r^*) \le 0_e(c^{\frac{n_o-1}{2}}) + 0_e(c^{\frac{n_o}{2}}) \quad \text{as} \quad c \to 0$$

$$= 0_e(c^{\frac{n_o-1}{2}}) \qquad \text{as} \quad c \to 0$$

which leads to lemma 4.3.

<u>Lemma 4.4</u>. $\lim_{c \to 0} b(\sigma_1 + \sigma_2)^{-1} E(\sigma_1^2 \bar{R}^1 + \sigma_2^2 S^{-1}) = 1.$

Lemma 4.4 has an equivalent variant form in

<u>Lemma 4.5</u> $\lim_{c \to 0} (\sigma_1 + \sigma_2)^{-1} [\sigma_1 E(r^*/R) + \sigma_2 E(s^*/S)] = 1.$

To prove lemma 4.5, it suffices to prove the following:

<u>Lemma 4.6.</u> $\lim_{c \to 0} E(r^*/R) = 1$ <u>and</u> $\lim_{c \to 0} E(s^*/S) = 1.$

<u>Proof:</u> To prove the first part, using Fatou's lemma and (4.3.3),

(4.3.5) $$\liminf_{c \to 0} E(r^*/R) \geq E(\liminf_{c \to 0} r^*/R) = 1.$$

Now, for arbitrary ϵ in (0, 1), let $\theta = 1 - \epsilon$. Hence with $\alpha = [\theta r^*]$,

(4.3.6) $$E(R^{-1}) = \sum_{r=n_0}^{\alpha} r^{-1} P(R = r) + \sum_{r > \alpha} r^{-1} P(R = r)$$

$$\leq n_0^{-1} P(R \leq \theta r^*) + (\alpha + 1)^{-1} P(R > \alpha).$$

Thus, using lemma 4.3,

$$r^* E(R^{-1}) \leq 0_e(c^{\frac{n_0 - 2}{2}}) + (1 - \epsilon)^{-1} P(R > \alpha)$$

so that

$$\limsup_{c \to 0} r^* E(R^{-1}) \leq 1 + \partial, \qquad 0 < \partial = \partial(\epsilon) < 1,$$

since $n_o \geq 3$. ϵ being arbitrary, combining (4.3.5) and (4.3.7), the first part of the lemma follows. The other part follows in the same way.

4.4. Proofs of Theorem 4.1 and Theorem 4.2

Asymptotic Risk Efficiency. The proof of theorem 4.1 is trivial in view of lemmas 4.2 and 4.6.

Asymptotic Behaviour of Regret

In view of (4.1.11), for proving theorem 4.2, it is sufficient to prove the following lemma.

Lemma 4.7. For the rule in (4.1.8),

(4.4.1) $E[(R-r^*)^2/R] = O(1)$ as $c \to 0$ if and only if $n_o \geq 3$,

(4.4.2) $E[(S-s^*)^2/S] = O(1)$ as $c \to 0$ if and only if $n_o \geq 3$.

Proof: We shall prove only (4.4.1). (4.4.2) will follow in a similar way. First note that

$$E[(R-r^*)^2/R] \geq \frac{(n_o-r^*)^2}{n_o} \cdot P(R = n_o)$$

$$= O_e(c^{-1}) \cdot O_e(c^{\frac{n_o-1}{2}})$$

$$= O_e(c^{\frac{n_o-3}{2}}) \quad \text{as} \quad c \to 0$$

which proves the only if part.

For the if part, let us write $\alpha = [(1-\epsilon)r^*]$, $t_1 = [r^* - \epsilon(r^*)^{\frac{1}{2}}]$, $t_2 = [r^* + \epsilon(r^*)^{1/2}]$, $\beta = [(1+\epsilon)r^*]$, where $[u]$ denotes the largest integer $\leq u$, ϵ being arbitrary with $0 < \epsilon < 1$.

Then,

$$(4.4.3) \quad E[(R-r^*)^2/R] = E[\frac{(R-r^*)^2}{R}\{I_{[n_0 \leq R \leq \alpha]} + I_{[\alpha < R < t_1]} + I_{[t_1 \leq R \leq t_2]} + I_{[R > t_2]}\}] .$$

In view of lemma 4.3, the proof of if part of (4.4.1) will be complete if we prove the following lemmas.

<u>Lemma 4.8</u> $E(R^{-1}(R-r^*)^2 I_{[t_1 \leq R \leq t_2]}] = O(1)$ <u>as</u> $c \to 0$

<u>Proof:</u> $$E[R^{-1}(R-r^*)^2 I_{[t_1 \leq R \leq t_2]}] \leq \frac{\epsilon^2 r^*}{r^* - \epsilon(r^*)^{1/2} - 1} P(t_1 \leq R \leq t_2)$$

$$\leq \epsilon^2(1 - \epsilon(r^*)^{-1/2} - (r^*)^{-1})^{-1} = O(1) \quad \text{as} \quad c \to 0.$$

<u>Lemma 4.9.</u> $E[R^{-1}(R-r^*)^2 I_{[R > t_2]}] = O(1)$ <u>as</u> $c \to 0$.

<u>Proof:</u> $$E[\frac{(R-r^*)^2}{R} I_{[R > t_2]}] \leq E[\frac{(R-r^*)^2}{r^*} I_{[R > t_2]}]$$

$$= E[X^2 I_{[X > \epsilon]}], \text{ where } X = (R-r^*)/(r^*)^{1/2}$$

$$= -\int_\epsilon^\infty x^2 d(1-F(x)), \quad \text{where } F(x) = P(X \leq x)$$

$$\leq \epsilon^2(1-F(\epsilon)) + 2\int_\epsilon^\infty x(1-F(x))dx.$$

Now, $1-F(x) = P(R > r^* + x(r^*)^{\frac{1}{2}}) \leq P(R > t)$, where $t = t(x) = [r^* + x(r^*)^{1/2}]$. Use now the inequality,

$$P(R > t) \leq P(u_t^2 > cA^{-1} t^2) = P\left\{\chi^2_{t-1} - (t-1) > (t-1)(t^2(r^*)^{-2} - 1)\right\}.$$

But $t^2(r^*)^{-2} - 1 \geq (r^*)^{-2}(r^* + x(r^*)^{1/2} - 1)^2 - 1$

$$\geq 2x(r^*)^{-\frac{1}{2}} - 2(r^*)^{-1} - 2x(r^*)^{-\frac{3}{2}}$$

$$\geq Kxc^{1/4}, \quad \text{for small } c,$$

where K is a generic symbol for a positive constant independent of c. The, Markov's inequality gives

$$P(R > t) \leq KE(\chi^2_{t-1} - (t-1))^4 / \{(t-1)^4 x^4 c\}$$

$$\leq K(t-1)^{-2} x^{-4} c^{-1} \leq K/x^4, \quad \text{for small } c.$$

Hence,

$$\int_\epsilon^\infty x^2 d F(x) \leq \epsilon^2(1-F(\epsilon)) + K \int_\epsilon^\infty x^{-3}\, dx \quad \text{as } c \to 0$$

$$= \epsilon^2(1-F(\epsilon)) + K\epsilon^{-2}.$$

This completes the proof of lemma 4.9.

<u>Lemma 4.10.</u> $E[R^{-1}(R-r^*)^2 I_{[\alpha < R < t_1]}] = O(1)$ <u>as</u> $c \to 0$.

<u>Proof:</u> $E[R^{-1}(R-r^*)^2 I_{[\alpha < R < t_1]}] \leq (1-\epsilon)^{-1} E[X^2 I_{[-\epsilon(r^*)^{\frac{1}{2}} < X < -\epsilon]}]$

where $X = (R-r^*)/(r^*)^{\frac{1}{2}}$. Now, writing $F(x) = P(X \leq x)$, we get,

(4.4.4)

$$E[X^2 I_{[-\epsilon(r^*)^{\frac{1}{2}} < X < -\epsilon]}] = \int_{-\epsilon(r^*)^{1/2}}^{-\epsilon} x^2 dF(x) \leq \epsilon^2 F(\epsilon)$$

$$-2 \int_{-\epsilon(r^*)^{1/2}}^{-\epsilon} xF(x)dx$$

$$= \epsilon^2 F(\epsilon) - 2 \int_{\epsilon(r)^{1/2}}^{\epsilon} (-x)\ F(-x)(-dx) = \epsilon^2 F(\epsilon) + \int_{\epsilon}^{\epsilon(r^*)^{1/2}} xF(-x)dx.$$

Also,

$$F(-x) = P(R \leq r^* - x(r^*)^{\frac{1}{2}}) = P(n_o \leq R \leq \alpha) + P(\alpha < R \leq r^* - -x(r^*)^{\frac{1}{2}})$$

$$= O_e\ (c^{\frac{1}{2}(n_o-1)}) + P(\alpha < R \leq r^* - x(r^*)^{\frac{1}{2}}).$$

Now,

(4.4.5) $$P(\alpha < R \leq r^* - x(r^*)^{\frac{1}{2}}) = P(\alpha < R \leq u),\ u = [r^* - x(r^*)]$$

$$\leq P(u_n^2 < cA^{-1} n^2 \text{ for at least one } n \text{ such that } \alpha < n \leq u)$$

$$= P(\chi^2_{n-1} < cn^2(n-1)/A\sigma_1^2 \text{ for at least one } n \text{ such that } \alpha < n \leq u).$$

But, for $\alpha < n \leq u$, $(cn^2/A\sigma_1^2) - 1 \leq (cn^2/A\sigma^2) - 1$

$$\leq c(r^* - x(r^*)^{\frac{1}{2}})^2/A\sigma_1^2 - 1 = -2cx\ (r^*)^{\frac{3}{2}}(A\sigma_1^2)^{-1} + cx^2\ r^*(A\sigma_1^2)^{-1}$$

$\leq -Kx\, c^{1/4}$ for small c, where K is a generic possitive constant.

Let us write $u_n^2 = \sum_1^{n-1} U_i$ where U_i's are iid χ_1^2 variables.

Then, from (4.4.5) we get

(4.4.6) $P(\alpha < R \leq r^* - x(r^*)^{\frac{1}{2}}) \leq P\left\{ \inf_{\alpha < n \leq u} \sum_{1}^{n-1} (U_i - 1) < -K\, x . c^{\frac{1}{4}} . \alpha \right\}$.

Using Kolmogorov's inequality, right hand side of (4.4.6) cannot exceed

(4.4.7) $K . E\left[\sum_{1}^{u-\alpha-1} (U_i - 1) \right]^4 / x^4 . c . \alpha^4$

$\leq K . (u - \alpha - 1)^2 / x^4 . c . r^{*4} \leq K / x^4$, for small c.

Hence, from (4.4.4) and (4.4.7),

$$E[X^2 \, I_{[-\epsilon (r^*)^{1/2} < X < -\epsilon]}]$$

$$= \epsilon^2 F(\epsilon) + 2 \int_{\epsilon}^{\epsilon (r^*)^{1/2}} x[O_e(c^{\frac{1}{2}(n_o - 1)}) + K \bar{x}^4] dx$$

$$\leq \epsilon^2 F(\epsilon) + K [\, c^{(n_o - 2)/2} + \epsilon^{-2}]$$

which proves the lemma 4.10.

This completes the proof of theorem 4.2.

4.5. Moderate sample size behaviour of the rule in (4.1.8).

Here we present the results of an experiment using pseudo-random normal deviates carried out in H-400 electronic computer. We write (4.1.5) as

(4.5.1) $n^* = \partial_1 (1 + \partial_2)$

where $\partial_1 = Ac^{-1}\sigma_2$, $\partial_2 = \sigma_1/\sigma_2$. We have considered the cases for

$$\partial_2 = 1,\ 1/2,\ 1/4,\ 1/8,\ 1/16$$

with

$$n^* = 10,\ 15,\ 20,\ 25,\ 40,\ 50,\ 60,\ 70,\ 100$$

$$n_o = 5, \text{ the starting sample size.}$$

To estimate $E(N)$, $E(R^{-1})$, $E(S^{-1})$ we repeated the experiment 200 times for each each entry using the stopping rule (4.1.8) and inturn we estimated the 'risk efficiency' η and 'regret' w (look at (4.1.10) and (4.1.11)). The tables 4.1 - 4.5 show the results. Computations are presented in tables 4.1 - 4.5 on pages 79-81.

4.6. Asymptotic normality of the stopping time N.

Here we follow a similar technique as the one discussed in chapter 2, to get the asymptotic normality of the stopping time N for the rule in (4.1.8)

Suppose W_i's are iid with mean μ_{11} and variance σ_{11}^2, Z_j's are iid with mean μ_{22} and variance σ_{22}^2, W_i's are independent of Z_j's, $(i = 1, \ldots, m_\upsilon,\ j = 1, \ldots, n_\upsilon;\ \upsilon \geq 1)$. Letting $m_\upsilon = m$, $n_\upsilon = n$, $T \equiv T_\upsilon = m_\upsilon + n_\upsilon$, assume $T \to \infty$ as $\upsilon \to \infty$ with

$$m/T \to \lambda \qquad \text{as} \quad \upsilon \to \infty$$

where λ is in (0, 1), and we have

(4.6.1) $$T^{\frac{1}{2}} (\bar{W}_m - \mu_{11}, \ \bar{Z}_n - \mu_{22}) \xrightarrow{\mathcal{L}} N_2 (\underline{0}, \Sigma)$$

where $$\Sigma = \begin{pmatrix} \sigma_{11}/\lambda & 0 \\ 0 & \sigma_{22}/(1-\lambda) \end{pmatrix}$$

<u>Lemma 4.7.</u> <u>Suppose</u> $R = R_\upsilon$ <u>and</u> $S = S_\upsilon$ <u>are proper</u> rv's <u>such that</u> $R/m \to 1$ <u>a.s.</u>, $S/n \to 1$ <u>a.s.</u> <u>and</u> $N/T \to$ <u>a.s.</u> as $\upsilon \to \infty$ (<u>where</u> $N = R + S$). <u>Then</u>

$$N^{\frac{1}{2}} (\bar{W}_R - \mu_{11}, \ \bar{Z}_S - \mu_{22}) \xrightarrow{\mathcal{L}} N_2(\underline{0}, \Sigma)$$

<u>as</u> $\upsilon \to \infty$.

<u>Proof</u>: Since $N/T \to 1$ a.s. as $\upsilon \to \infty$, it suffices to prove the following:

(4.6.2) $$T^{\frac{1}{2}} (\bar{W}_R - \mu_{11}, \ \bar{Z}_S - \mu_{22}) \xrightarrow{\mathcal{L}} N_2 (\underline{0}, \Sigma) \text{ as } \upsilon \to \infty.$$

Fix ϵ_1, ϵ_2, two arbitrary positive numbers.

(4.6.3) $$P[\sqrt{T} (\bar{W}_R - \mu_{11}) \le x, \ \sqrt{T} (\bar{Z}_S - \mu_{22}) \le y]$$

$$\le P[\sqrt{T} (\bar{W}_R - \mu_{11}) \le x, \ \sqrt{T} (\bar{Z}_S - \mu_{22}) \le y, \ \sqrt{T}|\bar{W}_R - \bar{W}_m| \le \epsilon_1, \ \sqrt{T}|\bar{Z}_S - \bar{Z}_n| \le \epsilon_2]$$

$$+ P[\sqrt{T}|\bar{W}_R - \bar{W}_m| > \epsilon_1] + P[\sqrt{T}|\bar{Z}_S - \bar{Z}_n| > \epsilon_2].$$

in view of R/m -> 1 a.s., S/n -> 1 a.s., the last two terms in the r.h.s. of (4.6.3) can be made arbitrarily small for large υ, since sample means follow the uniform continuity in probability condition of Anscombe (1952). Thus,

$$(4.6.4) \quad P[\sqrt{T}(\bar{W}_R - \mu_{11}) \leq x, \ \sqrt{T}(\bar{Z}_S - \mu_{22}) \leq y]$$
$$\leq P[\sqrt{T}(\bar{W}_m - \mu_{11}) \leq x + \epsilon_1, \ \sqrt{T}(\bar{Z}_n - \mu_{22}) \leq y + \epsilon_2] + \eta_1 + \eta_2$$
$$\leq G(x, y) + \eta, \quad \text{for large } \upsilon,$$

where η, η_1, η_2 are positive quantities, and $G(x,y)$ is the d.f. of $N_2(\underline{0}, \Sigma)$, after utilising (4.6.1). Similarly one can show that

$$(4.6.5) \quad P[\sqrt{T}(\bar{W}_R - \mu_{11}) \leq x, \ \sqrt{T}(\bar{Z}_S - \mu_{22}) \leq y]$$
$$\geq G(x, y) - \eta,$$

for large υ, $\eta > 0$. This completes the proof of lemma 4.7.

Now, with X, Y, u_i, v_j, R, S the same as in sections 4.1-4.4, we have the following corollaries.

<u>Corollary 4.1</u> $N^{\frac{1}{2}}(u_R^2 - \sigma_1^2, \ v_S^2 - \sigma_2^2) \xrightarrow{\mathcal{L}} N_2(\underline{0}, \Sigma)$

<u>as</u> c -> 0, <u>where</u> $\Sigma = \begin{pmatrix} 2\sigma_1^4/\lambda & 0 \\ 0 & 2\sigma_2^4/(1-\lambda) \end{pmatrix}$.

Proof: The proof follows from lemma 4.7 if we express u_R^2/σ_1^2 and v_S^2/σ_2^2 as means of iid χ_1^2 variables.

Corollary 4.2. $N^{1/2}(u_R - \sigma_1, v_S - \sigma_2) \xrightarrow{\mathcal{L}} N_2(\underline{0}, \Sigma_0)$

as $c \to 0$, where $\Sigma_0 = \begin{pmatrix} \sigma_1^2/2\lambda & 0 \\ 0 & \sigma_2^2/2(1-\lambda) \end{pmatrix}$.

Its proof is trivial, once we use Corollary 4.1 and invoke Mann-Wald's theorem (Rao (1965), p. 319). Since $N/n^* \to 1$ a.s., the above corollary can be rephrased as

$$(4.6.6) \qquad n^{*\frac{1}{2}}(u_R - \sigma_1, v_S - \sigma_2) \xrightarrow{\mathcal{L}} N_2(\underline{0}, \Sigma_0)$$

as $c \to 0$, Σ_0 being same as in corollary 4.2. Note Note that $\lambda = \sigma_1/(\sigma_1 + \sigma_2)$, so that

$$\tfrac{1}{2}\left\{\sigma_1^2/\lambda + \sigma_2^2/(1-\lambda)\right\} = \tfrac{1}{2}(n^*/b)^2$$

where $b = (Ac^{-1})^{\frac{1}{2}}$.

Proof of Theorem 4.3. We have

$$(4.6.7) \qquad R \geq bu_R \text{ a.s. and } S \geq bv_S \text{ a.s.}$$

So,

$$P\left\{ n^{*\frac{1}{2}} (u_R + v_S - \sigma_1 - \sigma_2) \Big/ 2^{\frac{1}{2}} (n^*/b) \le x \right\}$$

$$= P\left\{ b(n^*/2)^{-\frac{1}{2}} (u_R + v_S - \sigma_1 - \sigma_2) \le x \right\}$$

$$\ge P\left\{ b(n^*/2)^{-\frac{1}{2}} \left(\frac{R+S}{b} - \sigma_1 - \sigma_2\right) \le x \right\}, \quad \text{by (4.6.7)}$$

$$= P\left\{ (n^*/2)^{-\frac{1}{2}} (N - n^*) \le x \right\} .$$

Now using corollary 4.2,

(4.6.8) $$\limsup_{c \to 0} P\left\{ (n^*/2)^{-\frac{1}{2}} (N - n^*) \le x \right\} \le \Phi(x)$$

where Φ stands for d.f. of $N(0, 1)$.

We recall (4.2.1) as

(4.6.9) $$R - n_o \le b u_{R-1} \quad \text{a.s.} \quad \text{and} \quad S - n_o \le b v_{S-1} \quad \text{a.s.}$$

so that

$$P\left\{ n^{*\frac{1}{2}} (u_{R-1} + v_{S-1} - \sigma_1 - \sigma_2) \Big/ 2^{-\frac{1}{2}} (n^*/b) \le x \right\}$$

$$= P\left\{ b(n^*/2)^{-\frac{1}{2}} (u_{R-1} + v_{S-1} - \sigma_1 - \sigma_2) \le x \right\}$$

$$\le P\left\{ (N - n^* - 2n_o)(n^*/2)^{-\frac{1}{2}} \le x \right\} \quad \text{by (4.6.9)}$$

and hence

$$\liminf_{c \to 0} P\left\{ (n^*/2)^{-\frac{1}{2}} (N - n^*) \leq x \right\} \geq \Phi(x) . \tag{4.6.10}$$

Combining (4.6.8) and (4.6.10), we have

$$(2/n^*)^{\frac{1}{2}} (N - n^*) \xrightarrow{\mathcal{L}} N(0, 1)$$

as $c \to 0$, which is the Theorem 4.3, as stated in section 4.1.

Table 4.1: $\partial_2 = 1$

∂_1	n*	E(N)	η	W
5.0	10	11.195	1.0162	0.3248
7.5	15	14.560	1.0346	1.0380
10.0	20	18.625	1.0548	2.1945
12.5	25	23.775	1.0420	2.1030
20.0	40	38.035	1.0344	2.7521
25.0	50	48.740	1.0180	1.8091
30.0	60	58.540	1.0168	2.0225
35.0	70	67.875	1.0171	2.4075
50.0	100	98.710	1.0068	1.3635

Table 4.2: $\partial_2 = 1/2$

∂_1	n*	E(N)	η	W
6.66	10	11.825	1.0480	0.9606
10.00	15	15.245	1.0401	1.2036
13.33	20	19.575	1.0391	1.5644
16.66	25	23.925	1.0417	2.0870
26.66	40	38.340	1.0360	2.8813
33.33	50	47.950	1.0281	2.8105
40.00	60	58.085	1.0183	2.1980
46.66	70	68.090	1.0171	2.4054
66.66	100	98.255	1.0077	1.5495

Table 4.3: $\partial_2 = 1/4$

∂_1	n*	E(N)	η	w
8.0	10	12.635	1.1181	2.3628
12.0	15	16.655	1.0612	1.8361
16.0	20	20.030	1.0650	2.6008
20.0	25	24.675	1.0247	1.2373
32.0	40	38.825	1.0148	1.1186
40.0	50	48.590	1.0169	1.6960
48.0	60	58.670	1.0153	1.8415
56.0	70	68.885	1.0139	1.9485
80.0	100	97.995	1.0119	2.3918

Table 4.4: $\partial_2 = 1/8$

∂_1	n*	E(N)	η	w
8.88	10	13.465	1.1935	3.8708
13.33	15	17.970	1.1224	3.6721
17.77	20	22.005	1.0675	2.7024
22.22	25	26.040	1.0591	2.9580
35.54	40	40.005	1.0105	0.8425
44.44	50	50.025	1.0096	0.9523
53.33	60	58.105	1.0093	1.1218
62.22	70	69.490	1.0079	1.1086
88.88	100	98.550	1.0074	1.4963

Table 4.5 : $\partial_2 = 1/16$

∂_1	n*	E(N)	η	w
9.41	10	13.835	1.2428	4.8573
14.11	15	18.685	1.1536	4.6095
18.82	20	23.140	1.1034	4.1370
23.52	25	27.370	1.0820	4.1004
37.64	40	41.740	1.0507	2.4631
47.05	50	51.405	1.0469	4.6906
56.46	60	60.800	1.0102	1.2249
65.88	70	69.595	1.0075	1.0582
94.10	100	99.275	1.0044	0.8807

Remarks

1) In the range of σ_1, σ_2 considered, the average sample size is generally very near the optimal sample size.

2) The regret is always positive, as expected, meaning thereby that the risk of the procedure (4.1.8) is larger than the minimum risk; however, the boundedness of the regret is substantiated.

3) The risk efficiencies are quite near unity in all the cases.

CHAPTER 5

SEQUENTIAL ESTIMATION OF A LINEAR FUNCTION OF MEANS OF THREE NORMAL POPULATIONS

5.1 Introduction. In chapter 4 we investigated a sequential procedure to get hold of a point estimator of the difference of two means, the underlying populations being normal with unequal variances. It is easy to extend those results and the results of Robbins et al (1967) for estimating a linear function of two normal means. In this chapter, we have considered sequential estimation problems for a linear function of means of three normal populations (having unequal variances), sample sizes being not necessarily equal. Here the sampling scheme is quite complicated having though some inherrent symmetry.

5.2 Notations and preliminaries

Let X_{ij}, $j = 1,2,\ldots$ be a sequence of independent and normally distributed rv's having mean μ_i and variance σ_i^2, $i = 1, 2, 3$, all the parameters being unknown. Also we assume that the three populations are independent. For given non-zero constants λ_1, λ_2 and λ_3 we wish to estimate the linear compound of the means, $\mu = \lambda_1\mu_1 + \lambda_2\mu_2 + \lambda_3\mu_3$. Now, there is no loss of generality if we assume $\lambda_1 = \lambda_2 = \lambda_3 = 1$. Having r observations on X_{1j}, s observations on X_{2j} and t observations on X_{3j}, let

$$(5.2.1)\quad \begin{cases} \bar{X}_{1r} = r^{-1} \sum_{j=1}^{r} X_{1j}\ ;\quad \bar{X}_{2s} = s^{-1} \sum_{j=1}^{s} X_{2j},\quad \bar{X}_{3t} = t^{-1} \sum_{j=1}^{t} X_{3j} \\ u_r^2 = (r-1)^{-1} \sum_{j=1}^{r} (X_{1j} - \bar{X}_{1r})^2,\quad v_s^2 = (s-1)^{-1} \sum_{j=1}^{s} (X_{2j} - \bar{X}_{2s})^2, \\ w_t^2 = (t-1)^{-1} \sum_{j=1}^{t} (X_{3j} - \bar{X}_{3t})^2 \end{cases}$$

and we propose the estimator

$$(5.2.2)\qquad W = \bar{X}_{1r} + \bar{X}_{2s} + \bar{X}_{3t}$$

for μ.

First, we want to find a confidence interval I of width 2d having coverage probability $\geq 1 - \alpha$ for μ, where $0 < d < \infty$, $0 < \alpha < 1$ are preassigned. We propose the interval

$$(5.2.3)\qquad I = [W - d,\ \ W + d]$$

of width 2d, centered at W. In sections 5.3 and 5.4, sampling scheme and sequential procedures have been suggested to attain the objective approximately. The procedures are shown to have 'asymptotic consistency' and 'asymptotic efficiency' in the sense of (1.1.2) and (1.1.3). The cost of not knowing the σ_i's is also studied.

In section 5.5, our attention is confined to estimate μ pointwise with a view to minimizing the risk (loss being squared error plus cost). Suppose the loss incurred in estimating μ by W is

$$(5.2.4) \qquad L_{r,s,t} = A(W - \mu)^2 + c(r + s + t),$$

A, c are known positive constants, c being the cost of a single unit. Here again, we have studied a sequential procedure to achieve the objective approximately. The procedure is shown to be 'asymptotically risk efficient' (as c -> 0). The 'regret' is also shown to be bounded (as c -> 0). The cost of not knowing the σ_i's (i = 1, 2, 3) is also investigated.

Moderate sample size behaviour of sequential confidence interval procedures (given in section 5.4) is studied by Monte-Carlo methods, and is found to be quite satisfactory.

5.3. Fixed-width confidence interval estimation for μ

(5.2.3) yields,

$$(5.3.1) \qquad P(\mu \in I) = 2\,\Phi\left[d\left(\frac{\sigma_1^2}{r} + \frac{\sigma_2^2}{s} + \frac{\sigma_3^2}{t}\right)^{-1/2}\right] - 1$$

where Φ denotes the N(0, 1) distribution function. Let a and b be constants given by

$$2\Phi(a) - 1 = 1 - \alpha, \quad b = (a/d)^2.$$

In order that $P(\mu \in I) \geq 1 - \alpha$, r, s, t need satisfy the inequality

$$(5.3.2) \qquad \frac{\sigma_1^2}{r} + \frac{\sigma_2^2}{s} + \frac{\sigma_3^2}{t} \leq \frac{1}{b}.$$

Regarding r, s, t as continuous variables, the idea is to find the triplet (r*, s*, t*) which satisfies (5.3.2) and for which n = r + s + t is a minimum, and it is quite easy to obtain

$$(5.3.3) \qquad \begin{cases} r^* = b\,\sigma_1\,(\sigma_1 + \sigma_2 + \sigma_3) \\ s^* = b\,\sigma_2\,(\sigma_1 + \sigma_2 + \sigma_3) \\ t^* = b\,\sigma_3\,(\sigma_1 + \sigma_2 + \sigma_3) \end{cases}$$

For this triplet

$$(5.3.4) \qquad r^*/s^* = \sigma_1/\sigma_2, \quad s^*/t^* = \sigma_2/\sigma_3, \quad r^*/t^* = \sigma_1/\sigma_3$$

and the total sample size is

$$(5.3.5) \qquad n^* = r^* + s^* + t^* = b\,(\sigma_1 + \sigma_2 + \sigma_3)^2.$$

When σ_i's (i = 1, 2, 3) are unknown, we cannot find with fixed sample size procedure any interval I of preassigned length 2d for μ with preassigned coverage probability $1-\alpha$. However, we shall propose sequential procedures determining r, s, and t as r v 's where the goal can be achieved asymptotically. The usual estimates of σ_i's given in (5.2.1) have the desirable property that

$$u_i \to \sigma_1 \text{ a.s.}, \quad v_j \to \sigma_2 \text{ a.s.}, \quad w_k \to \sigma_3 \text{ a.s.}$$
$$\text{as } i, j, k \to \infty .$$

Now we give the scheme for sampling at any stage and the stopping rules.

5.4 Sampling scheme and stopping rules.

We take n_o (≥ 2) observations on the three populations to start with. Then if at any stage we have taken r observations on X_{1j}, s observations on X_{2j} and t observations on X_{3j}, we take the next observation

(A) on X_{1j} if $r/s \leq u_r/v_s$, $r/t \leq u_r/w_t$;

(B) on X_{2j} if $r/s > u_r/v_s$, $s/t \leq v_s/w_t$;

(C) on X_{3j} if $s/t > v_s/w_t$, $r/t > u_r/w_t$.

The motivations seem to be clear when one looks at (5.3.4). Also one may refer to Robbins et al (1967) or chapter 4 of our work. We now propose four more or less equivalent stopping rules, easily motivated from (5.3.2) - (5.3.5).

R_1 : The stopping time $N \equiv N(d)$ is the smallest positive integer $n \geq 3\,n_o$ such that if $R = r$ observations on X_{1j}, $S = s$ observations on X_{2j} and $T = t$ observations on X_{3j} have been taken with $r + s + t = n$

(5.4.1) $$n \geq b\,(u_r + v_s + w_t)^2$$

R_2 : The same, with (5.4.1) replaced by

(5.4.2) $$\frac{u_r^2}{r} + \frac{v_s^2}{s} + \frac{w_t^2}{t} \leq \frac{1}{b}$$

R_3 : The same, with (5.4.1) replaced by

(5.4.3) $$\begin{cases} r \geq bu_r(u_r + v_s + w_t),\; s \geq bv_s(u_r + v_s + w_t), \\ t \geq bw_t(u_r + v_s + w_t) \end{cases}$$

R_4 : The same, with (5.4.1) replaced by

(5.4.4) $$r^2 \geq bnu_r^2\,,\quad s^2 \geq bnv_s^2,\quad t^2 \geq bnw_t^2\,.$$

In each case the confidence interval for μ is taken to be I as in (5.2.3), the random sample sizes being R, S, T, the total sample size $N = R + S + T$. Suppose $N \equiv N(d)$ is the general notation for the stopping time. Main results regarding

N in relation to (5.4.1)-(5.4.4), are stated in the following theorem. Note that if N_i denotes the sample size required under rule R_i $(i = 1,\cdots,4)$, then $N_1 \leq N_2 \leq N_3 \leq N_4$ a.s.

Theorem 5.1. Assume $0 < \sigma_i < \infty$ $(i = 1,2,3)$ are fixed. Then for any of the stopping rules $R_1 - R_4$ with $n^* = b(\sigma_1 + \sigma_2 + \sigma_3)^2$ (see (5.3.5)),

(5.4.5) $$\lim_{d \to 0} N/n^* = 1 \quad \text{a.s.}$$

(5.4.6) $$\lim_{d \to 0} E(N)/n^* = 1$$

(5.4.7) $$\lim_{d \to 0} P(\mu \in I) = 1 - \alpha$$

Properties (5.4.6) and (5.4.7) are referred to as the 'asymptotic consistency' and 'asymptotic efficiency' of the proposed procedure. To prove the theorem, we need some results which we prove in the following lemma.

Lemma 5.1. For all $d > 0$, under the assumption $0 < \sigma_i < \infty$ fixed $(i = 1,2,3)$, for any of the stopping rules $R_1 - R_4$,

(5.4.8) $$E(N) \leq n^* + 6n_o.$$

Proof. First note that it suffices to prove the lemma for the rule R_4. Suppose that $r > n_o$ and that just before the rth observation on X_{1j}, there were ($r-1$ observations on X_{1j}) s observations on X_{2j} and t observations on X_{3j}. Then according to (A) of our sampling scheme,

(5.4.9) $(r-1)/s \leq u_{r-1}/v_s$ and $(r-1)/t \leq u_{r-1}/w_t$.

Then

(5.4.10) $$(r-1)^2 \leq b.(n-1)u^2_{r-1},$$

for otherwise, if $(r-1)^2 > b(n-1)u^2_{r-1}$, then, from (5.4.9),

$$s^2 \geq b(n-1)v^2_s \text{ and } t^2 \geq b(n-1)w^2_t ,$$

and sampling would have stopped at $(r-1, s, t)$th stage for R_4. Also, then, from (5.4.10),

(5.4.11) $$(r-1)^2(r-2) \leq b.n \sum_{j=1}^{r-1} (X_{1j} - \bar{X}_{1\overline{r-1}})^2$$
$$\leq b\, n \sum_{j=1}^{r} (X_{1j} - \bar{X}_{1r})^2 .$$

This leads to the inequality,

(5.4.12) $$bU^2_R \geq (R-1)^2(R-2).N^{-1} I_{[R>n_o]} + (R-n_o)^2.R\, I_{[R=n_o]}$$
$$\geq (R-n_o)^2 RN^{-1} , \text{ where } U^2_r = (r-1)u^2_r . (r \geq 2).$$

If we assume $E(R) < \infty$, using convexity of $(R-n_o)^2 R/N$ and Jensen's inequality, we get from (5.4.12),

$$(ER - n_o)^2 \, E(R)/E(N) \leq b\sigma_1^2 \, E(R)$$

and hence

(5.4.13) $$(ER - n_o)^2 \leq b\sigma_1^2 \, E(N) .$$

Similarly,

$$(ES - n_o)^2 \leq b\sigma_2^2 \, E(N) \qquad \text{if} \quad E(S) < \infty ,$$

$$(ET - n_o)^2 \leq b\sigma_3^2 \, E(N) \qquad \text{if} \quad E(T) < \infty .$$

Hence, since $N - 3n_o = (R - n_o) + (S - n_o) + (T - n_o)$, we have

$$(EN - 3n_o)^2 \leq b(\sigma_1 + \sigma_2 + \sigma_3)^2 \, E(N)$$

which implies

$$(EN)^2 - 6n_o \, E(N) \leq n^* \, E(N)$$

leading thereby to

(5.4.14) $$E(N) \leq n^* + 6n_o .$$

Also then, for the rule R_4,

(5.4.15) $$\begin{cases} E(R) \leq r^* + n_o + p\sigma_1 \\ E(S) \leq s^* + n_o + p\sigma_2 \\ E(T) \leq t^* + n_o + p\sigma_3 \end{cases}$$

with $p^2 = 6bn_o$, using the fact that $(a+b)^{\frac{1}{2}} \leq a^{\frac{1}{2}} + b^{\frac{1}{2}}$ where a and b are nonnegative. If we do not assume $E(R) < \infty$, $E(S) < \infty$, $E(T) < \infty$, define $R_k = \min (R, k)$, $S_k = \min (S, k)$, $T_k = \min (T, k)$ and get bounds for $E(R_k)$, $E(S_k)$, $E(T_k)$ and

$E(N_k)$, $N_k = R_k + S_k + T_k$. Using Monotone Convergence theorem, we then get the result.

This completes the proof of lemma 5.1

Proof of theorem 5.1. For proving theorem 5.1, we need one more lemma, which is an immediate extension of the lemma in Robbins et al (1967).

Lemma 5.2. Given constants $c_{i,j}$, $d_{j,k}$ $(i,j,k = 1,2,3,\ldots)$ such that $0 < c_{i,j} \to C > 0$, $0 < d_{j,k} \to D > 0$ as $i,j,k \to \infty$ and any integer $n_o \geq 1$, define $i(3n_o) = j(3n_o) = k(3n_o) = n_o$ and for $n > 3n_o$ let

(A_1) $i(n+1) = i(n) + 1$, $j(n+1) = j(n)$, $k(n+1) = k(n)$ if

$$i(n)/j(n) \leq c_{i(n),j(n)}, \quad i(n)/k(n) \leq [c_{i(n),j(n)} \cdot d_{j(n),k(n)}]$$

(B_1) $i(n+1) = i(n)$, $j(n+1) = j(n) + 1$, $k(n+1) = k(n)$ if

$$i(n)/j(n) > c_{i(n),j(n)}, \quad j(n)/k(n) \leq d_{j(n),k(n)}$$

(C_1) $i(n+1) = i(n)$, $j(n+1) = j(n)$, $k(n+1) = k(n) + 1$ if

$$j(n)/k(n) > d_{j(n),k(n)}, \quad i(n)/k(n) > [c_{i(n),j(n)} \cdot d_{j(n),k(n)}].$$

Then $i(n)/j(n) \to C$ and $j(n)/k(n) \to D$ as $n \to \infty$.

Proof of lemma 5.2.

Clearly $i(n)$, $j(n)$ and $k(n) \to \infty$ as $n \to \infty$. Call an integer $n \geq 3n_o$ of class I, class II or class III according as the conditions of (A_1), (B_1) or (C_1) are satisfied for n. Then for all sufficiently large n, there exist largest integers n_1, n_2, n_3 respectively of class I, II and III with $3n_o \leq n_1, n_2, n_3 < n$. Also $n_1, n_2, n_3 \to \infty$ as $n \to \infty$.

Now, $$i(n_1)/j(n_1) \leq c_{i(n_1),\, j(n_1)}$$

and $i(n_1 + 1) = i(n_1 + 2) = \dots = i(n)$

$$j(n) \geq j(n_1)$$

so that $$i(n)/j(n) \leq i(n)/j(n_1)$$
$$= i(n_1 + 1)/j(n_1) = \frac{i(n_1)}{j(n_1)} + \frac{1}{j(n_1)}$$
$$\leq c_{i(n_1),j(n_1)} + \frac{1}{j(n_1)}$$

which implies $\limsup_{n \to \infty} i(n)/j(n) \leq C$.

Also $$j(n) = j(n_2 + 1) = j(n_2) + 1, \quad i(n) \geq i(n_2)$$

so that $$i(n)/j(n) \geq i(n_2)/j(n_2) + 1$$
$$= \frac{i(n_2)}{j(n_2)} [1 - \frac{1}{j(n_2) + 1}]$$
$$\geq c_{i(n_2),\, j(n_2)} [1 - \frac{1}{j(n_2) + 1}]$$

which implies $\liminf_{n \to \infty} i(n)/j(n) \geq C$, so that $\lim_{n \to \infty} i(n)/j(n) = C$.

Similarly the other part follows. Thus, proof of lemma 5.2 is complete.

In our case we put $c_{i,j} = \frac{u_i}{v_j}$, $d_{j,k} = \frac{v_j}{w_k}$, and obtain the results $i(N)/j(N) \to \frac{\sigma_1}{\sigma_2}$ a.s. and $j(N)/k(N) \to \frac{\sigma_2}{\sigma_3}$ a.s. and $j(N)/k(N) \to \frac{\sigma_2}{\sigma_3}$ a.s. as $N \to \infty$.

For any of our stopping rules, since $N \to \infty$ a.s. as $d \to 0$, it follows that

$$(5.4.25) \qquad \frac{R}{S} \to \frac{\sigma_1}{\sigma_2} \text{ a.s. and } \frac{S}{T} \to \frac{\sigma_2}{\sigma_3} \text{ a.s. as } d \to 0.$$

If N_i denotes the total samples sizes required by the rule R_i $(i = 1,\cdots,4)$, note the fact that $N_1 \le N_2 \le N_3 \le N_4$ a.s. So it is sufficient to prove the theorem 5.1 for the rule R_4. (5.4.5) follows easily from (5.4.4) together with the fact that $R \to \infty$ a.s., $S \to \infty$ a.s. and $T \to \infty$ a.s. as $d \to 0$.

By Fatou's lemma and (5.4.5) it follows that $\liminf_{d \to 0} E(N/n^*) \ge 1$. From (5.4.8), one gets $\limsup_{d \to 0} E(N/n^*) \le 1$, which gives (5.4.6). To prove (5.4.7), note that for any integer n, the two events $[\mu \in I]$ and $[N = n]$ are stochastically independent, since $(\bar{X}_{1r}, \bar{X}_{2s}, \bar{X}_{3t})$ is independent of (u_r, v_s, w_t). So,

$$(5.4.26) \quad P[\mu \in I] = 2E\Phi\left[d\left(\frac{\sigma_1^2}{R} + \frac{\sigma_2^2}{S} + \frac{\sigma_3^2}{T}\right)^{-\frac{1}{2}}\right] - 1 .$$

Again, from the definition of the stopping rule R_4, it follows that

$$\lim_{d \to 0} b\left(\frac{\sigma_1^2}{R} + \frac{\sigma_2^2}{S} + \frac{\sigma_3^2}{T}\right) = 1 \quad \text{a.s.}$$

Using now the dominated convergence theorem

~~$\lim_{d \to 0} P(\mu \in I) = 2\Phi(a) - 1 - 1 - \alpha$~~

and the fact that $b = (a/d)^2$, we get

$$\lim_{d \to 0} P(\mu \in I) = 2\Phi(a) - 1 = 1 - \alpha .$$

This completes the proof theorem 5.1.

5.5 Point estimation of μ with minimum risk

In this section we shall investigate the possibility of utilising a sequential procedure to get hold of a point estimator of μ ($= \mu_1 + \mu_2 + \mu_3$) as in (5.2.2), with the loss structure (5.2.4), having obtained samples of sizes r, s and t from the three populations, objective being to minimize the risk. In what follows, we consider σ_1, σ_2 and σ_3 to be kept fixed . Risk is

$$(5.5.1) \quad v_{r,s,t}(c) = A(\sigma_1^2/r + \sigma_2^2/s + \sigma_3^2/t) + c(r+s+t).$$

In case $\sigma_1, \sigma_2, \sigma_3$ were known, the triplet (r^*, s^*, t^*) for which the risk (5.5.1) is a minimum is obtained as

(5.5.2) $$r^* = b\sigma_1 \, , \quad s^* = b\sigma_2 \, , \quad t^* = b\sigma_3$$

with $b^2 = Ac^{-1}$.

For this triplet, as in (5.3.4),

(5.2.3) $$r^*/s^* = \sigma_1/\sigma_2, \quad s^*/t^* = \sigma_2/\sigma_3 \, , \quad r^*/t^* = \sigma_1/\sigma_3 \; ;$$

and the total sample size is

(5.5.4) $$n^* = r^* + s^* + t^* = b(\sigma_1 + \sigma_2 + \sigma_3),$$

the minimum risk being

(5.5.5) $$\upsilon(c) = \upsilon_{r^*, s^*, t^*}(c) = 2cn^* \, .$$

But in ignorance of σ_i's $(i = 1,2,3)$, no fixed sample size procedure will minimize (5.5.1) simultaneously for all $0 < \sigma_1, \sigma_2, \sigma_3 < \infty$. However, we shall propose a sequential procedure determining r, s and t as rv's as follows. Define u_r, v_s and w_t as in (5.2.1). We begin with $\underline{n_o (\geq 3)}$ observations on X_{ij}, $i = 1,2,3$. Then at any stage, the sampling scheme is same as given in the section 5.4. We now give a stopping rule motivated from (5.5.2).

R^* : The stopping time $N \equiv N_c$ is the first integer $n(\geq 3n_o)$ such that if $R = r$ observations on X_{1j}, $S = s$ observations on X_{2j} and $T = t$ observations on X_{3j} have been taken with $r + s + t = n$,

(5.5.6) $$r \geq bu_r, \quad s \geq bv_s, \quad t \geq bw_t;$$
$$N = R + S + T$$

For notational convenience we shall write L_n instead of $L_{r,s,t}$, as in (5.2.4). Observing that L_n and $I_{[N=n]}$ are independent for all n $(\geq 3n_o)$, one finds that

(5.5.7) $$\bar{\upsilon}(c) = E(L_N) = AE\,[\sigma_1^2/R + \sigma_2^2/S + \sigma_3^2/T] + cE(N).$$

The 'risk efficiency'

(5.5.8) $$\eta(c) = \bar{\upsilon}(c)/\upsilon(c)$$
$$= \tfrac{1}{2}[b(\sigma_1 + \sigma_2 + \sigma_3)^{-1} E(\sigma_1^2/R + \sigma_2^2/S + \sigma_3^2/T) + E(N/n^*)],$$

and the 'regret'

(5.5.9) $$W(c) = \bar{\upsilon}(c) - \upsilon(c)$$
$$= c\left\{E(r^{*2}/R - r^* + R - r^*) + E(s^{*2}/S - s^* + S - s^*) + E(t^{*2}/T - t^* + T - t^*)\right\}$$
$$= cE[(R - r^*)^2/R + (S - s^*)^2/S + (T - t^*)^2/T].$$

By using the same techniques as in proving lemma 4.2.1, we shall get

$$(5.5.10)\quad \begin{cases} E(N) \leq n^* + 3n_o, & E(R) \leq r^* + n_o \\ E(S) \leq s^* + n_o, & E(T) \leq t^* + n_o \end{cases}$$

which study the cost of ignorance of the variances. Regarding the stopping time N in (5.5.6), $\eta(c)$ and $w(c)$, we have the following result.

Theorem 5.2. For all fixed $0 < \sigma_1, \sigma_2, \sigma_3 < \infty$,

$$(5.5.11)\quad \lim_{c \to 0} N/n^* = 1 \text{ a.s.}, \quad \lim_{c \to 0} E(N/n^*) = 1,$$

$$(5.5.12)\quad \lim_{c \to 0} \eta(c) = 1,$$

$$(5.5.13)\quad \lim_{c \to 0} w(c) = O(c).$$

We omit the proof as all the results follow similarly as in theorem 4.1 and theorem 4.2.

5.6. Moderate sample size behaviour of the stopping rules R_1-R_3

We consider here numerical results regarding fixed width confidence interval estimation procedures for a linear function of means of k normal populations dealt with in sections 5.3 and 5.4 $(k = 2, 3)$.

Case I: $k = 2$

Scheme: We take $n_0 (\geq 3)$ observations on X_{1j} and X_{2j} to start with. Then if at any stage we have taken r observations on X_{1j}, s observations on X_{2j}, we take the next observation on X_{1j} or X_{2j} according as (with positive λ's)

(5.6.1) $r/s \leq \lambda_1 u_r / \lambda_2 v_s$ or $r/s > \lambda_1 u_r / \lambda_2 v_s$.

Analogous to those in (5.4.1) - (5.4.3), we have here three stopping rules, R_i (i = 1,2,3).

(5.6.2) $n^* = b(\lambda_1\sigma_1 + \lambda_2\sigma_2)^2 = \{a\, \partial_1(\partial_2 + 1)\}^2$

where $\partial_1 = \lambda_2\sigma_2/d$, $\partial_2 = \lambda_1\sigma_1/\lambda_2\sigma_2$.

We fix the coverage probability at 0.95, i.e. $a = 1.96$. We take

$$n^* = 10(5)25,\ 40,\ 50,\ 70,\ 100(50)200,$$

$$\partial_2 = 1,\ \tfrac{1}{2},\ \tfrac{1}{4},\ \tfrac{1}{16},\ 2,\ 4,$$

and $\partial = \sigma_1/\sigma_2 = 0.5$ (kept fixed).

This covers the values of λ_1 through:

$$\cdot 111,\ \cdot 333,\ \cdot 500,\ \cdot 667,\ \cdot 800,\ \cdot 889$$

and $\lambda_2 = 1 - \lambda_1$.

E_i and P_i stand for the expected sample size and coverage probability respectively for the rule R_i $(i = 1,2,3)$, and for each entry these are estimated through 200 repetitions of an experiment using pseudo-random normal deviates in H - 400 electronic computer. We take $n_o = 5$ in all the cases. Computations are presented in tables 5.1 - 5.6 on pages 101 - 103.

<u>Case II:</u> $k = 3$

We consider the scheme and stopping rules $R_1 - R_3$ of section 5.4, replacing u_r, v_s, w_t respectively by $\lambda_1 u_r$, $\lambda_2 v_s$ and $\lambda_3 w_t$ for positive λ's.

We take

$$\beta_1 = \sigma_1/\sigma_2 = 0.5, \qquad \beta_2 = \sigma_2/\sigma_3 = 1.5,$$

$$n^* = \left\{ a\partial_1 (\partial_2 + \partial_3 + 1)\right\}^2 ,$$

where $\partial_1 = \lambda_2\sigma_2/d$, $\partial_2 = \lambda_1\sigma_1/\lambda_2\sigma_2$, $\partial_3 = \lambda_3\sigma_3/\lambda_2\sigma_2$.

We fix the coverage probability at 0.95, i.e. $a = 1.96$.

We take

$$n^* = 10(5)25,\ 40,\ 50,\ 70,\ 100(50)200,$$

i) $\partial_1 = 1.25$, $\quad \partial_2 = 0.25$

ii) $\partial_1 = 2.00$, $\quad \partial_2 = 1.00$

that is, we consider λ-values as

i) $\lambda_1 = .6818$, $\quad \lambda_2 = .2727$, $\quad \lambda_3 = .0455$

ii) $\lambda_1 = .7059$, $\quad \lambda_2 = .1765$, $\quad \lambda_3 = .1176$.

We then proceed as in Case I, with $n_o = 5$, the notations E_i, P_i having the same interpretations. We present the results in the tables 5.7 and 5.8 on page 104.

Remarks on computations:

(1) The objective was to get a confidence coefficient as 0.95 and in most of the cases we are very near that target. Also we note that generally, the expected sample sizes are less than the optimal sample sizes. It is interesting to note that the coverage target cannot be attained unless $E(N)$ exceeds n^*, which follows as a consequence of Jensen's inequality applied to the formula in (5.4.26). To estimate the coverage probabilities, instead of taking possibly a stronger approach based on (5.4.26), we take the frequency approach since this really seems to be the situation in applications. The rule R_3 is performing best among R_i's $(i = 1,2,3)$ with regard to achieved coverage probability, the reason being that R_3 is using more samples to stop than others. However, R_4 is expected to perform in a still better way.

(2) The numerical results presented in tables 4.1 - 4.5 can be looked upon as moderate sample size behaviours of the rule (4.1.8) in the two population case with $\lambda_1 = \lambda_2 = 1$.

Table 5.1: $\partial_2 = 1/16$, $\lambda_1 = .111$

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	1.518	11.420	.905	13.505	.940	13.760	.940
15	1.860	13.765	.845	16.695	.890	17.420	.905
20	2.147	17.385	.885	20.720	.950	21.310	.955
25	2.401	21.280	.835	24.870	.900	25.975	.905
40	3.307	33.020	.885	37.015	.885	38.310	.895
50	3.395	45.390	.940	48.400	.935	49.765	.940
70	4.018	66.080	.915	67.660	.930	69.065	.925
100	4.802	97.765	.925	98.900	.925	99.475	.935
150	5.881	147.280	.955	147.435	.955	148.075	.950
200	6.791	197.605	.940	198.555	.945	198.990	.945

Table 5.2: $\partial_2 = 1/4$ $\lambda_1 = .333$

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	1.291	11.380	.935	12.300	.945	13.015	.950
15	1.580	13.660	.920	14.880	.945	15.685	.965
20	1.825	17.515	.895	18.335	.905	19.350	.920
25	2.041	21.865	.890	22.665	.890	23.640	.915
40	2.581	36.750	.910	37.010	.910	37.655	.905
50	2.886	44.495	.910	44.790	.910	45.535	.915
70	3.415	63.340	.885	63.380	.885	64.295	.910
100	4.081	93.440	.950	93.455	.950	94.170	.955
150	4.999	148.000	.955	148.040	.955	148.640	.950
200	5.772	196.285	.940	196.285	.940	196.955	.930

Table 5.3: $\partial_2 = 1/2$, $\lambda_1 = .500$

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	1.076	11.300	.945	11.525	.950	11.895	.955
15	1.317	14.365	.915	14.650	.940	15.170	.935
20	1.521	17.310	.945	17.430	.945	17.970	.935
25	1.700	20.585	.875	20.760	.880	21.595	.900
40	2.151	36.945	.910	36.990	.910	37.775	.925
50	2.405	44.430	.915	44.900	.910	45.260	.900
70	2.846	65.690	.925	65.760	.930	66.700	.930
100	3.401	94.585	.920	94.586	.920	95.155	.915
150	4.166	146.555	.965	146.555	.965	147.130	.965
200	4.810	198.700	.940	198.705	.940	199.450	.935

Table 5.4: $\partial_2 = 1$, $\lambda_1 = .667$

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	.807	11.215	.960	11.285	.960	11.550	.960
15	.988	14.365	.945	14.505	.945	14.950	935
20	1.141	17.305	.920	17.405	.920	18.000	.930
25	1.276	21.350	.920	21.465	.920	22.170	.930
40	1.613	34.230	.910	34.270	.910	34.950	910
50	1.804	44.380	.900	44.410	.895	45.365	.905
70	2.134	63.415	.910	63.435	.910	64.405	.905
100	2.551	94.450	.950	94.450	.950	95.100	.950
150	3.124	147.195	.940	147.195	.940	147.945	.935
200	3.608	196.345	.955	196.345	.955	196.990	.955

Table 5.5: $\partial_2 = 2$, $\lambda_1 = .800$

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	.538	11.470	.970	11.755	.970	12.155	.965
15	.659	13.790	.920	14.000	.925	14.680	.960
20	.761	16.710	.910	16.905	.925	17.565	.915
25	.850	21.765	.910	22.000	.920	22.535	.915
40	1.075	34.320	.925	34.495	.920	35.310	.920
50	1.202	45.865	.900	46.010	.905	46.640	.910
70	1.423	62.270	.900	62.420	.905	63.245	.915
100	1.701	95.925	.940	95.930	.935	96.625	.935
150	2.083	144.715	.930	144.725	.930	145.315	.930
200	2.405	197.480	.935	197.480	.935	197.995	.935

Table 5.6: $\partial_2 = 4$, $\lambda_1 = .889$

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	.323	11.165	.915	11.895	.920	12.375	.925
15	.395	13.530	.870	14.615	.885	15.440	.915
20	.456	11.570	.910	18.690	.930	19.640	940
25	.510	21.175	.895	21.770	.900	22.825	.910
40	.645	36.710	.930	36.905	.930	37.690	.945
50	.721	46.030	.935	46.220	.935	47.185	.925
70	.854	65.135	.935	65.185	.935	65.845	.935
100	1.020	95.550	.910	95.640	.915	96.100	.915
150	1.249	147.865	.935	147.875	.935	148.415	.930
200	1.443	197.375	.920	197.390	.920	198.020	.920

Table 5.7: λ_1 = .6818, λ_2 = .2727, λ_3 = .0455,
∂_2 = 1.25, ∂_3 = 0.25

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	0.6453	15.205	.975	15.815	.980	16.320	.975
15	0.7904	16.260	.935	17.340	.930	18.295	.935
20	0.9127	18.545	.930	20.010	.940	21.155	.945
25	1.0204	21.655	.915	23.185	.915	24.780	.935
40	1.2907	33.180	.885	34.425	.885	35.870	.885
50	1.4431	42.715	.885	43.515	.885	45.580	.920
70	1.7075	62.945	.905	63.465	.905	65.095	.920
100	2.0408	94.365	.935	94.460	.935	96.010	.945
150	2.4995	146.350	.965	146.350	.965	148.040	.965
200	2.8861	194.620	.920	194.670	.920	196.240	.925

Table 5.8: λ_1 = .7059, λ_2 = .1765, λ_3 = 1176,
∂_2 = 2.00, ∂_3 = 1.00

n^*	∂_1	E_1	P_1	E_2	P_2	E_3	P_3
10	0.4033	15.180	.975	15.395	.975	15.835	.975
15	0.4940	16.635	.930	17.065	.935	18.065	.950
20	0.5704	18.805	.895	19.360	.915	20.320	.930
25	0.6377	22.085	.915	22.655	.920	23.955	.965
40	0.8067	33.530	.905	33.930	.910	35.550	.930
50	0.9019	40.530	.885	40.895	.880	42.195	.895
70	1.0672	61.065	.910	61.170	.910	63.220	.915
100	1.2755	93.135	.925	93.170	.925	94.735	.920
150	1.5622	145.425	.955	145.435	.955	147.065	.965
200	1.8038	194.635	.915	194.650	.915	196.145	.925

CHAPTER 6

SEQUENTIAL ESTIMATION OF LOCATION PARAMETER IN EXPONENTIAL DISTRIBUTIONS

6.1. Introduction. Life testing model is usually taken to be governed by the distribution

$$(6.1.1) \qquad f(x; \mu, \sigma) = \sigma^{-1} \exp\left(-\frac{x-\mu}{\sigma}\right), \quad x > \mu$$
$$= 0 \qquad , \quad x \leq \mu$$

where $-\infty < \mu < \infty$, $0 < \sigma < \infty$ are two unknown parameters. Consider a sequence $X_1, X_2, \ldots$ of iidrv's with the density in (6.1.1). Here our object is to estimate μ, the minimum life. In section 6.2 we derive a sequential procedure to estimate μ pointwise with minimum risk, which is shown to be 'asymptotically risk efficient'. Basu (1971) considered the same problem with a loss structure which is a particular case of ours. Also, we point out two mistakes in his paper;

i) the proof of his theorem 3 is incorrect;

ii) applicability of his sequential procedure in practice is not substantiated by his computations, since he used an algorithm of J.E.Moyal (given in Robbins (1959)) to compute the distribution of the stopping time, which is not applicable in this context.

In section 6.3 we give some proofs of some properties of the stopping time. Since there is no algorithm available to compute the exact distribution of the random sample size, we study the moderate sample size behaviour of our procedure in section 6.4 by Monte-Carlo methods using pseudo-random exponential deviates (i.e. density is $f(x; 0, 1)$). In section 6.5, a sequential fixed width confidence interval procedure is developed for estimating μ. The same is shown to be 'asymptotically consistent' and 'asymptotically efficient' in the analogous sense of (1.1.2) and (1.1.3). For this problem also, moderate sample size behaviour of our procedure is studied and presented in section 6.6.

6.2. Point estimation of μ

Suppose $X_{n(1)} = \min(X_1, \ldots, X_n)$. Suppose A, s, t, and c are all known positive constants. Suppose the loss incurred in estimating μ by $X_{n(1)}$ from a sample of size n is

(6.2.1) $$L_n = A(X_{n(1)} - \mu)^s + c.n^t$$

with risk

(6.2.2) $$v_n(c) = E(L_n) = K\sigma^s/sn^s + cn^t$$

where $K = As^2\Gamma s$, $s > 0$.

Basu (1971) considered the loss structure L_n with $c = t = 1$.

The risk (6.2.2) is minimum for $n = n^*$ where

(6.2.3) $$n^* = (K\sigma^s / ct)^{1/(t+s)}$$

and the minimum risk

(6.2.4) $$\upsilon(c) = \upsilon_{n^*}(c) = c(1 + t/s)(n^*)^t.$$

Then, as in Robbins (1959), Starr (1966b), Basu (1971), we consider a sequential sampling scheme:

Let $\sigma_n = (n-1)^{-1} \sum_{i=1}^{n} (X_i - X_{n(1)}), \quad n \geq 2.$

(6.2.5) The stopping number $N \equiv N_c$ is the first integer $n\ (\geq n_o)$ such that

$$n \geq (K\, \sigma_n^s / ct)^{1/(t+s)},$$

$n_o (\geq 2)$ being the starting sample size. When we stop, we estimate μ by $X_{N(1)}$. Basu's (1971) stopping rule GR is same as (6.2.5) with $c = t = 1$, $s = p$. Some simple properties of N are expressed in the following lemma.

Lemma 6.1 The stopping time N has the following properties:

(6.2.6) N is well defined, nonincreasing as a function of c and $E(N) < \infty$ for all $c > 0$.

(6.2.7) $\lim_{c \to 0} N = \infty$ a.s., $\lim_{c \to 0} E(N) = \infty$

(6.2.8) $\lim_{c \to 0} (N/n^*) = 1$ a.s.

The proof is immediate from lemma 1 of Chow-Robbins (1965), noting the fact that $\sigma_n \to \sigma$ a.s. as $n \to \infty$ and $2(n-1)\sigma_n/\sigma \sim \chi^2_{2(n-1)}$.

Using Basu's (1955) theorem, we can prove that L_n and the $I_{[N=n]}$ are stochastically independent for any $n(\geq n_o)$. Using this and lemma 6.1, one finds that

(6.2.9) $\bar{\upsilon}(c) = E(L_N) = (K\sigma^s/s)\, E(N^{-s}) + cE(N^t).$

The 'risk efficiency' $\eta(c)$ and 'regret' $w(c)$ for the rule in (6.2.5) turns out to be

(6.2.10) $$\begin{cases} \eta(c) = (1+t/s)^{-1}[(t/s)(n^*)^s E(N^{-s}) + E(N/n^*)^t], \\ w(c) = \bar{\upsilon}(c) - \upsilon(c). \end{cases}$$

We shall give a condition on n_o so that $\eta(c) \to 1$ as $c \to 0$. In this respect we have the following result. However, there is no mathematical result proved for $w(c)$

<u>Theorem 6.1.</u> $\lim_{c \to 0} \eta(c)$

$$= 1 \qquad \text{if } n_o > 1 + s^2/(t+s)$$

$$= 1 + \gamma \qquad \text{if } n_o = 1 + s^2/(t+s)$$

$$= \infty \qquad \text{if } n_o < 1 + s^2/(t+s)$$

<u>where γ is a known positive constant.</u>

For its proof, we need some properties of N, which can be proved in the same lines as Starr's (1966b). However, our proofs in section 6.3 are new and seem to be simpler than the existing ones.

6.3. Some properties of the rule in (6.2.5)

<u>Lemma 6.2.</u> <u>For</u> $0 < \sigma < \infty$ <u>fixed, for all fixed</u> $w > 0$,

$$\lim_{c \to 0} E(N/n^*)^w = 1.$$

<u>Proof.</u> Using Fatou's lemma and (6.2.8), we get

$$\text{(6.3.1)} \qquad \liminf_{c \to 0} E(N/n^*)^w \geq E[\liminf_{c \to 0} (N/n^*)^w] = 1.$$

Also,

$$\text{(6.3.2)} \qquad E(N/n^*)^w = \sum_{n_o \leq n \leq \beta} (n/n^*)^w P(N = n) + (n^*)^{-w} T(\beta)$$

where $T(\beta) = \sum_{n \geq \beta+1} n^w P(N = n)$, $\beta = [(1+\epsilon)^{w^{-1}} n^*] + 1$, $0 < \epsilon < 1$.

Define $V_n = 2(n-1)\sigma_n/\sigma$,

$$p(n, \sigma; c) = (ct/K)^{1/s} \frac{2(n-1)n^{\frac{t+s}{s}}}{\sigma}$$

$$= c^{1/s}.p(n, \sigma), \text{ say}$$

Then

(6.3.3)
$$T(\beta) = \sum_{n \geq \beta} (n+1)^w P(N = n+1)$$

$$\leq \sum_{n \geq \beta} (n+1)^w P(V_n > p(n, \sigma; c))$$

$$\leq \sum_{n \geq \beta} (n+1)^w \inf_{0<h<1/2} [\exp\{-h.p(n,\sigma;c)\}. E\{\exp(h.V_n)\}]$$

$$= \sum_{n \geq \beta} (n+1)^w \inf_{0<h<1/2} [\exp\{-h.p(n,\sigma;c)\}. (1-2h)^{-(n-1)}]$$

Writing $\tau = (t+s)/sw$, for $n \geq \beta$, $p(n,\sigma; c) \geq a(n)$ where $a(n) = 2(n-1)(1+\epsilon)^{\tau}$.

Thus, from (6.3.3) we get

(6.3.4)
$$T(\beta) \leq \sum_{n \geq \beta} (n+1)^w \inf_{0<h<1/2} [\exp\{-h.a(n)\}. (1-2h)^{-(n-1)}]$$

$$= \sum_{n \geq \beta} (n+1)^w \exp\{-h_o.a(n)\}.(1-2h_o)^{-(n-1)}$$

where h_o is given by the relation

$$1 - 2h_o = 2(n-1)/a(n) = (1+\epsilon)^{-\tau};$$

so $0 < h_o < \frac{1}{2}$.

Hence, for $c < c(\epsilon)$, (6.3.4) reduces to

$$T(\beta) \leq \sum_{n \geq \beta} (n+1)^w . (1+\epsilon)^{\tau(n-1)} \exp\left[-(n-1)\left\{(1+\epsilon)^{\tau}-1\right\}\right]$$

$$= \sum_{n \geq \beta} (n+1)^w . \rho^{n-1},$$

where $\rho = (1+\epsilon)^{\tau} \exp\left\{1-(1+\epsilon)^{\tau}\right\} < 1$. Hence, using the ratio rule of convergence $T(\beta) < L$, a constant independent of c. The lemma now follows from (6.3.2).

<u>Lemma 6.3.</u>

$$P(N = n_o) = O_e\left(c^{\frac{n_o-1}{s}}\right) \text{ as } c \to 0 .$$

<u>Proof</u>. $N(N = n_o) = P(V_{n_o} \leq p(n_o, \sigma; c))$

$$\geq \frac{1}{2^{n_o-1}(n_o-1)!} e^{-p(n_o,\sigma;c)} \left\{p(n_o,\sigma; c)\right\}^{n_o-1}$$

$$\geq K_1 c^{\frac{n_o-1}{s}}$$ as $c \to 0$, for some positive constant K_1. Also,

$$P(N = n_o) = P(V_{n_o} \leq p(n_o, \sigma; c))$$

$$\leq \frac{1}{2^{n_o-1}(n_o-1)!}\left\{p(n_o, \sigma; c)\right\}^{n_o-1}$$

$$\leq K_2 c^{\frac{n_o-1}{s}} \qquad \text{as} \quad c \to 0,$$

for some positive constant K_2.

This completes the proof of lemma 6.3.

Lemma 6.4. For fixed θ in $(0, 1)$, $P(N \leq \theta n^*) = O_e(c^{\frac{n_o-1}{s}})$ as $c \to 0$.

Proof. Let $\alpha = [\theta n^*]$.

(6.3.6) $\quad P(n_o \leq N \leq \theta n^*)$

$$= \sum_{n=n_o}^{\alpha} P(N = n)$$

$$\leq \sum_{n=n_o}^{\alpha} P(V_n \leq p(n, \sigma; c))$$

$$\leq \sum_{n=n_o}^{\alpha} \inf_{h>0} E[\exp\left\{h\, p(n, \sigma; c)\right\}(1+2h)^{-(n-1)}]$$

$$= \sum_{n=n_o}^{\alpha} \exp\left\{h_o \cdot p(n, \sigma; c)\right\}(1+2h_o)^{-(n-1)}$$

where $h_o = \frac{1}{2}\left[\frac{2(n-1)}{p(n, \sigma; c)} - 1\right]$.

For $n \leq \alpha$, $\frac{2(n-1)}{p(n, \sigma; c)} \geq \bar{\theta}^{\frac{t+s}{s}} > 1$, so $h_o > 0$.

From (6.3.6) we get

(6.3.7) $P(n_o \leq N \leq \theta n^*)$

$$\leq \sum_{n=n_o}^{\alpha} [\exp\left\{1-(\frac{ct}{K})^{\frac{1}{s}} \cdot \frac{n}{\sigma}^{\frac{t+s}{s}}\right\} \cdot (\frac{ct}{K})^{\frac{1}{s}} \cdot \frac{n}{\sigma}^{\frac{t+s}{s}}]^{(n-1)}.$$

Note that for $0 < x < 1$, $xe^{1-x} \uparrow$ in x, so that for $n \leq \alpha$,

$$\exp\left\{1 - (\frac{ct}{K})^{\frac{1}{s}} \cdot \frac{n}{\sigma}^{\frac{t+s}{s}}\right\} (\frac{ct}{K})^{\frac{1}{s}} \cdot \frac{n}{\sigma}^{\frac{t+s}{s}}$$

$$\leq \exp(1-\theta^{\frac{t+s}{s}}) \cdot \theta^{\frac{t+s}{s}} = \epsilon \text{ (say)} < 1,$$

using $e^{x-1} \geq x$ for real x, equality iff $x = 1$.

Hence, (6.3.7) reduces to

(6.3.8) $P(n_o \leq N \leq \theta n^*)$

$$\leq c^{\frac{1}{s}(n_o-1)} \sum_{n=n_o}^{\alpha} \epsilon^{n-n_o} [(t/K)^{\frac{1}{s}} \cdot \frac{n}{\sigma}^{\frac{t+s}{s}} \cdot$$

$$\exp\left\{1 - (\frac{ct}{K})^{\frac{1}{s}} \cdot \frac{n}{\sigma}^{\frac{t+s}{s}}\right\}]^{n_o-1}$$

$$\leq c^{\frac{(n_o-1)}{s}} e^{n_o-1} \sum_{n=n_o}^{\alpha} \epsilon^{n-n_o} \left\{(\frac{t}{K})^{\frac{1}{s}} \cdot \frac{n}{\sigma}^{\frac{t+s}{s}}\right\}^{n_o-1}$$

$$= O_e(c^{\frac{n_o-1}{s}}),$$

by the ratio test of convergence, for small c. Also, from lemma 6.3, $P(N \le [\theta n^*]) \ge P(N = n_o) = 0_e(e^{(n_o-1)/s})$ as $c \to 0$. Thus, using lemma 6.3,

$$P(N \le \theta n^*) = 0_e(c^{\frac{n_o-1}{s}}) \quad \text{as} \quad c \to 0$$

which proves lemma 6.4. It follows also from the lemma that

$$P(n_o + 1 \le N \le \theta n^*) = 0\,(c^{n_o/s}).$$

The following lemma states conditions on the starting sample size n_o under which

$$EN^{-w} \sim (n^*)^{-w} \qquad \text{as} \quad c \to 0,$$

for fixed $w > 0$.

We write $b(w) = (K\sigma^s/t)^{w/(t+s)}$, $d(n_o, w) = n_o^{-w}/2^{n_o-1}(n_o-1)!$,

$$\partial(n_o, w; \sigma) = d(n_o, w).b(w).\{p(n_o, \sigma)\}^{n_o-1}$$

and note that $\partial(n_o, w, \sigma)$ is independent of σ when $n_o' = 1 + sw/(t+s)$. Let ∂ be this value.

<u>Lemma 6.5.</u> <u>For fixed</u> $w > 0$

$$\begin{aligned} \lim_{c \to 0} (n^*)^w E(N^{-w}) &= 1 && \text{\underline{if}} \quad n_o > 1 + sw/(t+s) \\ &= 1 + \partial && \text{\underline{if}} \quad n_o = 1 + sw/(t+s) \\ &= \infty && \text{\underline{if}} \quad n_o < 1 + sw/(t+s) \end{aligned}$$

Proof. We shall proceed in the same lines as Starr's (1966b). Fix ϵ in (0, 1); let $\alpha = (1-\epsilon)^{w^{-1}} \cdot n^*$, $\beta = (1+\epsilon)^{w^{-1}} \cdot n^*$.

Write

$$B_1 = n_o^{-w} P(N = n_o),$$

$$B_2 = \sum_{n_o+1 \leq n \leq \alpha} n^{-w} P(N = n),$$

$$B_3 = \alpha^{-w} P(N \geq \alpha) .$$

Observe that

$$(6.3.9) \qquad (n^*)^w \left\{ p(n_o, \sigma; c) \right\}^{n_o - 1} = b(w) . \left\{ p(n_o, \sigma) \right\}^{n_o - 1} . c^{\left(\frac{n_o - 1}{s} - \frac{w}{t+s} \right)}$$

so that, lemma 6.5 follows if we prove that

$$(6.3.10) \qquad \lim_{c \to 0} (n^*)^w E(N^{-w}) = 1 + d(n_o, w) \lim_{c \to 0} \left\{ (n^*)^w . p^{n_o - 1}(n_o \sigma; c) \right\} .$$

By using same types of computations as in lemma 6.3, we have

$$(n^*)^w E(N^{-w}) \geq d(n_o, w) . \left\{ (n^*)^w . p^{n_o - 1}(n_o, \sigma; c) \right\} . \exp \left\{ -p(n_o, \sigma; c)/2 \right\} + (1+\epsilon)^{-1} P(n_o < N \leq \beta).$$

But $\lim_{c \to 0} p(n, \sigma; c) = 0$ for any fixed n. Therefore,

$$\liminf_{c \to 0} (n^*)^w E(N^{-w}) \geq d(n_o, w) \lim_{c \to 0} \left\{ (n^*)^w p^{n_o - 1}(n_o, \sigma; c) \right\} + 1 - \theta .$$

where $0 < \theta = \theta(\epsilon) < 1$. Thus,

(6.3.12) $$\liminf_{c \to 0} (n^*)^w E(N^{-w}) \geq 1 - \theta \quad \text{if } n_o > 1 + sw/(t+s)$$
$$\geq 1 + \partial - \theta \quad \text{if } n_o = 1 + sw/(t+s)$$
$$= \infty \quad \text{if } n_o < 1 + sw/(t+s).$$

Now,

(6.3.13) $$E(N^{-w}) = \sum_{n=n_o}^{\infty} n^{-w} P(N = n)$$
$$< B_1 + B_2 + B_3$$

(6.3.14) $$B_1 = n_o^{-w} P(V_{n_o} \leq p(n_o, \sigma; c))$$
$$= d(n_o, w) \left\{ p(n_o, \sigma; c) \right\}^{n_o - 1}, \text{ from lemma (6.3).}$$

Also,

(6.3.15) $B_2 \leq n_o^{-1} \; 0_e(c^{\frac{n_o}{s}})$, by the remark following lemma 6.4.

From (6.3.13) - (6.3.15), we have

(6.3.16) $$\limsup_{c \to 0} (n^*)^w E(N^{-w}) \leq d(n_o, w) \lim_{c \to 0} \left\{ (n^*)^w p^{n_o - 1}(n_o, \sigma; c) \right\} + \limsup_{c \to 0} \left\{ (n^*)^w B_2 \right\} + 1 - \epsilon.$$

We note a fact that from (6.3.9),

$$\lim_{c \to 0} (n^*)^w p^{n_o - 1}(n_o, \sigma; c) < \infty \text{ implies}$$

$$\lim_{c \to 0} (n^*)^w c^{n_o/s} = 0.$$

Now utilising (6.3.12), (6.3.14) and (6.3.15), we get from (6.3.16)

$$\text{(6.3.17)} \quad \limsup_{c \to 0} (n^*)^w E(N^{-w}) \begin{cases} = \infty & \text{if } n_o < 1 + sw/(t+s) \\ \leq 1 + \partial - \epsilon & \text{if } n_o = 1 + sw/(t+s) \\ \leq 1 - \epsilon & \text{if } n_o > 1 + sw/(t+s). \end{cases}$$

ϵ being arbitrary, the lemma follows from (6.3.12) and (6.3.17). Basu's (1971) theorem 3 is same as the lemma 6.5 with $t = 1$. His proof is not correct. The mistake occurs in his (25).

<u>Proof of Theorem 6.1</u>. In view of lemma 6.2 and lemma 6.5, the proof of theorem 6.1 is straightforward, where

$$\gamma = (\tfrac{t}{s}).(1 + \tfrac{t}{s})^{-1}.\partial = t\,(t+s)^{-1}\,\partial.$$

<u>Remarks</u>. Though the results of our theorem 6.1 (and corollary in Basu (1971))are purely asymptotic in nature, Basu (1971) used an algorithm (available in Robbins (1959)) to evaluate the exact distribution of random sample size (given in (6.2.5)), risk efficiency etc. and studied the behaviour of the procedure

for moderate sample sizes. But we want to stress the point that the basic algorithm used by Basu (1971) is not at all applicable to this case, so that his computations do not really show any importance or usefulness of this sequential sampling rule in practice. The reason is that

$$2\sigma_n = (n-1)^{-1} \sum_{i=2}^{n} Y_{n,i} ,$$

where $\sigma^{-1} Y_{n,i}$'s are iid χ_2^2 variables and the rv's in the sum goes on changing with n. For that reason, to study the behaviour of the rule in case of moderate sample sizes, we take resort to Monte-Carlo methods using pseudo-random exponential deviates (i.e. density is $f(x;\ 0,\ 1)$). Though we have no theoretical results regarding bounds for the regret $w(c)$, even for small c, our computations really show that the regrets are very small even for moderate sample sizes. We present our results in the following section.

6.4. Moderate sample behaviour of the rule in (6.2.5)

We consider two cases

(i) $s = 2$, (ii) $s = 1$

while $A = 2$, $\sigma = 1$, $t = 1$ in either case.

In either case, n^* runs through

5(5)25, 40, 50, 70, 100(50)200.

We take $n_0 = 3$ in all the cases. The values of c are obtained from (6.3.3), viz

$$c = K/(n^*)^{s+1} .$$

To estimate $E(N)$, $E(\bar{N}^s)$ we repeat the experiment using the rule given in (6.2.5), for 100 times in H- 400 electronic computer using pseudo-random exponential deviates (i.e. density is $f(x;\ 0,\ 1)$ as in (6.1.1)). We present our results in the tables 6.1 - 6.2 on pages 126-127 .

6.5. Fixed-width confidence interval for μ

In this section our objective is to find a confidence interval of prescribed width $d(>0)$ and prescribed coverage probability $1 - \alpha$ $(0 < \alpha < 1)$ for μ. If σ were known, this could be achieved as follows. Given a sample $X_1, \ldots, X_n$ of size n, we take the estimator of μ as $X_{n(1)}$ (as in section 6.2). We propose the confidence interval $I =$ $I = [X_{n(1)} - d,\ X_{n(1)}]$ for μ. In order that the confidence coefficient is $1-\alpha$, we require

$$1 - \alpha \leq P(X_{n(1)} - d \leq \mu \leq X_{n(1)})$$
$$= 1 - \exp(-nd/\sigma)$$

so that the sample size n is the smallest positive integer exceeding $d^{-1} \sigma \log (1/\alpha) = C$, say. When σ is unknown, the goal cannot be achieved with a fixed sample size procedure. To meet the end, the following sequential procedure is proposed:

(6.5.1) The stopping time $N \equiv N_d$ is the first integer $n (\geq n_o)$ for which $n \geq d^{-1} \sigma_n \log (1/\alpha)$ where $\sigma_n = (n-1)^{-1} \sum_{i=1}^{n} (X_i - X_{n(1)})$, $n_o (\geq 2)$ is the starting sample size. The confidence interval for μ is then taken to be $I = [X_{N(1)} - d, X_{N(1)}]$.

The following theorem is then proved.

Theorem 6.2. The stopping time N defined by (6.5.1) satisfies the following properties:

(6.5.2) N is well defined, non-increasing as a function of d, and $E(N) < \infty$ for all $d > 0$.

(6.5.3) $\lim_{d \to 0} N = \infty$ a.s., $\lim_{d \to 0} E(N) = \infty$.

(6.5.4) $\lim_{d \to 0} (N/C) =$ a.s.

(6.5.5) $\lim_{d \to 0} P(\mu \in I) = 1 - \alpha$

(6.5.6) $E(N) \leq C + n_o + 1$

(6.5.7) $\lim_{d \to 0} E(N/C) = 1$.

Th ... 5.5.7) and (6.5.5) are referred to as 'asympto ... mptotic consistency' of the proposed sequential procedure.

Proof of

The first proof of (6.5.2) is obvious from the definition of N in (6.5.1). To prove the second part, use the inequality

(6.5.8)
$$E(N) - 2 = \sum_{n=2}^{\infty} P(N > n)$$
$$\leq \sum_{n=2}^{\infty} P[n < d^{-1} \sigma_n \log (1/\alpha)].$$

Noting that $2(n-1)\sigma_n \sim \sigma\chi^2_{2(n-1)}$, one gets from (6.5.1),

$$E(N) - 2 \leq \sum_{n=2}^{\infty} \inf_{0<h<1/2} \left\{ \exp \left[-\frac{2hd(n-1)n}{\sigma \log (1/\alpha)}\right](1-2h)^{-(n-1)} \right\}$$
$$\leq \sum_{n=2}^{\infty} 2^{n-1} \exp \left[-\frac{dn(n-1)}{2\sigma \log(1/\alpha)}\right], \text{ (taking } h = \tfrac{1}{4})$$
$$< \infty.$$

(6.5.3) is immediate from the definition of N and the monotone convergence theorem.

To prove (6.5.4), first note that

$$\sigma_n \to \sigma \quad \text{a.s.} \quad \text{as } n \to \infty.$$

Then we use the inequality (follows from the definition of N given in (6.5.1), (6.5.9) $$d^{-1}\sigma_N \log(1/\alpha) \le N \le n_o + d^{-1}\sigma_{N-1}\log(1/\alpha)$$

Multiplying both sides of (6.5.9) by $d/\log(1/\alpha)$ and using previous remarks the result is proved.

It follows from (6.5.4) that

(6.5.10) $$\exp(-Nd/\sigma) \to \alpha \quad \text{a.s.} \quad \text{as } d \to 0.$$

Using (6.5.2) and noting that the $I_{[N=n]}$ and $X_{n(1)}$ are independent for any fixed $n(\ge 2)$, we have

$$P(\mu \in I) = 1 - E[\exp(-Nd/\sigma)]$$

$$\to 1 - \alpha \quad \text{as} \quad d \to 0$$

from (6.5.10) and the dominated convergence theorem which gives (6.5.5).

To prove (6.5.6) use the inequality

$$N-1 \le d^{-1} a(N-2)^{-1} \sum_1^{N-1} (X_i - X_{\overline{N-1}(1)}) + (n_o - 1) I_{[N=n_o]}$$

where $a = \log(1/\alpha)$. That is,

$$(N-1)(N-2) \le d^{-1} a \sum_1^{N-1} (X_i - X_{\overline{N-1}(1)}) + (n_o-1)(N-2) I_{[N=n_o]}$$

$$\le d^{-1} a \sum_1^{N} (X_i - X_{N(1)}) + N(n_o - 2)$$

$$\le d^{-1} a \sum_1^{N} (X_i - \mu) + N(n_o - 2)$$

Hence, taking expectation and using Wald's first equation (since $E(N) < \infty$ by (6.5.2)),

$$(6.5.11) \qquad E\left\{(N-1)(N-2)\right\} \leq (C + n_o - 2)\, E(N).$$

But $E\left\{(N-1)(N-2)\right\} > E(N^2 - 3N) \geq (EN)^2 - 3E(N)$.

One gets now from (6.5.11), $E(N) \leq C + n_o + 1$, which is (6 5 6). (6.5.7) now follows from (6.5.6), (6.5.4) and Fatou's lemma.

In the following section 6.6, we present the results of Monte-Carlo experiments using pseudo-random exponential deviates about the behaviour of the rule given in (6.5.1) for moderate optimal sample sizes C.

6.6. Moderate sample size behaviour of the sequential procedure in (6.5.1).

In what follows we fix $\alpha = 0.05$, $\sigma = 1$. We consider the following values for d:

1, 1/2, 1/3, 1/5, 1/8, 1/10, 1/14, 1/20, 1/25, 1/30, 1/40, 1/45, 1/55, 1/70, 1/85 .

To estimate $E(N)$ and coverage frequency for each C, we repeat the experiment 100 times using pseudo-random exponential deviates in H- 400 electronic computer. We present the results of our experiment in the following table. E and P denote the estimated average sample size and coverage probability.

For each entry, we take $n_o = 3$ observations to start with.

Results: Table 6.3: Average Sample Size and Coverage Probability : $\alpha = 0.05$

d	C	E	P
1.0000	2.9957	3.77	.97
0.5000	5.9915	5.60	.92
0.3333	8.9872	8.47	.88
0.2500	11.9829	11.00	.93
0.2000	14.9787	13.68	.90
0.1250	23.9659	22.64	.93
0.1000	29.9573	28.98	.90
0.0714	41.9403	41.75	.96
0.0400	74.8933	74.17	.94
0.0333	89.8720	90.34	.94
0.0250	119.8293	119.90	.97
0.0222	134.8080	134.01	.92
0.0182	164.7653	164.46	.95
0.0143	209.7013	210.79	.96
0.0117	254.6372	256.18	.94

Remarks:

1. We note that the achieved coverage probability is quite near the target. Also the average sample sizes are very near the optimal sample sizes C. Generally, the performances of the rule in (6.5.1) for moderate C-values are very encouraging.

<u>Remarks</u> : (contd.)

(2) One may wonder about the use of n_o (≥ 2) in the stopping rule in (6.5.1). Following remarks in Starr (1966b), we can show that

$$\lim_{d \to 0} E(dN/a\sigma)^{-w} = 1 \quad \text{if } n_o > w+1$$
$$= 1 + \partial^* \quad \text{if } n_o = w+1$$
$$= \infty \quad \text{if } n_o < w+1$$

for fixed $w > 0$, for some known positive ∂^*.

Results: Table 6.1: Average Sample Size, Risk Efficiency Regret For Moderate Sample Sizes: Case (i).

n^*	c	$E(N)$	$E(N^{-2})$	$\upsilon(c)$	$\bar{\upsilon}(c) \times 10^5$	$\eta(c)$	$w(c) \times 10^5$
5	.064000	5.28	.04816	.48000	53056.0	1.10533	5055.77
10	.008000	9.48	.01866	.12000	15049.0	1.23405	3048.58
15	.002370	14.77	.00797	.05333	6890.0	1.23452	1357.43
20	.001000	19.76	.00393	.03000	3548.0	1.18269	548.07
25	.000512	24.95	.00170	.01920	1959.0	1.02032	39.40
40	.000125	39.99	.00065	.00750	759.0	1.01215	9.12
50	.000064	49.59	.00042	.00480	486.0	1.01193	5.73
70	.000023	69.72	.00021	.00245	246.0	1.00665	1.63
100	.000008	100.26	.00010	.00120	120.6	1.00466	0.56
150	.000002	150.35	.00005	.00053	53.5	1.00274	0.15
200	.000001	201.55	.00002	.00029	30.0	1.00256	0.07

Table: Average Sample Size, Risk Efficiency, Regret For Moderate Sample Sizes: Case (ii)

n^*	c	$E(N)$	$E(N^{-1})$	$\upsilon(c)$	$\bar{\upsilon}(c) \times 10^5$	$\eta(c)$	$w(c) \times 10^5$
5	.080000	5.21	.20596	.80000	82872.9	1.03591	2872.85
10	.020000	9.77	.11089	.40000	41719.8	1.04299	1719.78
15	.008889	15.05	.07057	.26667	27492.6	1.03097	825.97
20	.005000	20.26	.05007	.20000	20144.2	1.00721	144.23
25	.003200	25.15	.04026	.16000	16099.8	1.00624	99.83
40	.001250	40.17	.02508	.10000	10036.7	1.00367	36.73
50	.000800	49.93	.02013	.08000	8019.5	1.00244	19.80
70	.000408	70.03	.01433	.05714	5724.0	1.00187	10.71
100	.000200	100.16	.01000	.04000	4004.7	1.00118	4.73
150	.000089	150.42	.00666	.02667	2669.0	1.00089	2.37
200	.000050	201.34	.00497	.02000	2001.5	1.00073	1.47

Remarks:

1) In both the cases, $E(N)$ is very close to n^* for moderate n^* also. The starting sample size n_o being 3, $\eta(c)$ is close to unity in case (i) and case (ii) for moderate n^* as well.

2) We note that $w(c)$ is very near to zero for n^* greater than or equal to 20. It would be nice if it could be shown that $w(c) \to 0$ as $c \to 0$, which is really not too much to expect.

3) These numerical results suggest the use of the stopping rule in practice.

CHAPTER 7

SOME MORE RESULTS ON SEQUENTIAL POINT ESTIMATION

7.1 Introduction. In this chapter we consider two sequential point estimation problems, viz. (1) estimation of regression parameters in Gauss-Markoff set up, and (2) estimation of θ in $R(0, \theta)$ population (i.e., uniform distribution on $(0, \theta)$). We consider problem (1) in sections 7.2-7.3. Sections 7.4-7.6 are devoted to problem (2). It may be remarked that Gleser (1965, 1966) investigated the problem of fixed size bounds for regression parameters with Gauss-Markoff set up.

7.2. Sequential Estimation Of Regression Parameters In Gauss-Markoff Set Up

Consider a sequence $Z_1, Z_2, \ldots$ of independent and normally distributed rv's such that

$$Z_i = \underline{x}'_{(i)} \underline{\beta} + \epsilon_i \qquad (i = 1, 2, \ldots) \tag{7.2.1}$$

where $\underline{\beta}$ is a $m \times 1$ vector of unknown parameters, $\underline{x}_{(i)}$ is a $m \times 1$ vector of non-stochastic known constants with ϵ_i distributed as $N(0, \sigma^2)$, cov $(\epsilon_i, \epsilon_j) = 0$ for all

$i, j\ (i \neq j),\ \sigma(> 0)$ being unknown. As a convention, for any $r \times s$ matrix A, A' and $R(A)$ mean respectively the transpose and rank of A. We start with a sample of size $n_o (\geq m+2)$ making sure that $R(X_{n_o}) = m$ where $X_n' = (\underline{x}_{(1)}, \underline{x}_{(2)}, \ldots, \underline{x}_{(n)})$. Let $\underline{Y}_n' = (Z_1, Z_2, \ldots, Z_n)$ for any $n (\geq n_o)$.

It is well known (see e.g. Rao (1965)) that a least square estimator of $\underline{\beta}$ with model (7.2.1) on the basis of a sample of size n is

(7.2.2) $$\underline{\beta}_n = (X_n' X_n)^{-1} X_n' \underline{Y}_n$$

with dispersion matrix

(7.2.3) $$V(\underline{\beta}_n) = \sigma^2 (X_n' X_n)^{-1}.$$

Suppose the loss incurred in estimating $\underline{\beta}$ by $\underline{\beta}_n$ from a sample of fixed size n is

(7.2.4) $$L_n = n^{-1}(\underline{\beta}_n - \underline{\beta})' (X_n' X_n)(\underline{\beta}_n - \underline{\beta}) + cn$$

with risk

(7.2.5) $$\begin{aligned} \upsilon_n(c) &= E_\sigma(L_n) \\ &= E_\sigma \left\{ n^{-1} \operatorname{tr} (\underline{\beta}_n - \underline{\beta})'(X_n' X_n)(\underline{\beta}_n - \underline{\beta}) \right\} + cn \\ &= n^{-1}\sigma^2 \operatorname{tr} (I_m) + cn \\ &= m\sigma^2/n + cn \end{aligned}$$

where tr A means trace of the matrix A and I_m stands for the identity matrix of order $m \times m$. If σ were known, the problem of finding the value of n, say n*, for which the risk (7.2.5) is minimum is given by

(7.2.6) $$n^* = m^{\frac{1}{2}} c^{-\frac{1}{2}} \sigma$$

and the minimum risk

(7.2.7) $$\upsilon(c) = \upsilon_{n^*}(c) = 2cn^*.$$

But, in ignorance of σ, no fixed sample size procedure will minimize (7.2.5) simultaneously for all $0 < \sigma < \infty$. So the possibility of utilising a sample of random size N determined by the following sequential rule is considered:

(7.2.8) The stopping time $N \equiv N_c$ is the first integer $n \geq n_o$ such that

$$n \geq [m \, R^2_{on} \, (n - m)^{-1}/c]^{\frac{1}{2}}$$

where $R^2_{on} = \underline{Y}'_n \underline{Y}_n - \underline{Y}'_n X_n \underline{\beta}_n$ is the error sum of squares, starting sample size is $n_o (\geq m+2)$. When we stop, $\underline{\beta}$ is estimated by $\underline{\beta}_N$.

The rule (7.2.8) can be rephrased as:

(7.2.9) The stopping time $N \equiv N_c$ is the first integer n $(\geq n_o)$ such that

$$V_n \leq p(n, \sigma; c)$$

where $V_n = R^2_{on}/\sigma^2$, $p(n, \sigma; c) = n^2 c\,(n-m)/m\sigma^2$.

We now state the following lemma.

Lemma 7.1. For any fixed $n(\geq n_o)$, β_n is independent of the vector $(V_{n_o}, V_{n_o+1}, \ldots, V_n)$.

Proof. For any integer p in $[n_o, n]$,

(7.2.10) $$R^2_{op} = \underline{Y}'_p\,[I_p - X_p(X'_p X_p)^{-1} X'_p]\underline{Y}_p ,$$

I_p being the identity matrix of order p.

Thus,

$$R^2_{op} = \underline{Y}'_p \left(\sum_{i=1}^{p-m} \xi_i \xi'_i \right) \underline{Y}_p = \sum_{i=1}^{p-m} (\underline{\xi}'_i \underline{Y}_p)^2$$

where $\underline{\xi}'_i$s are orthonormal eigen vectors of the idempotent matrix $A = I_p - X_p(X'_p X_p)^{-1} X'_p$, associated eigen values being thereby all unity $(i = 1, \ldots, p-m)$. Using the symbol $\underline{O}$ for the null vector, irrespective of dimension, we can write from (7.2.10)

(7.2.11) $$R^2_{op} = \sum_{i=1}^{p-m} (\underline{\varrho}'_i \underline{Y}_n)^2$$

where $\underline{\varrho}'_i = (\underline{\xi}'_i \,\vdots\, \underline{O}')$ is $1 \times n$ vector. Let $\left(\begin{matrix} X_p \\ \cdots \\ U_{n-p} \end{matrix} \right)$ be the corresponding partition of X_n. From (7.2.2), $\beta_n = B\underline{Y}_n$ where $B = (X_n' X_n)^{-1} X_n'$. A sufficient condition for $B\underline{Y}_n$ and $\underline{\rho}_i' \underline{Y}_n$ to be distributed independently is $B\underline{\rho}_i = \underline{0}$ (see Rao (1965), p. 170). Now, for verifying this sufficient condition, note that $\underline{\xi}_i$ is a vector in the column space of A and $X_p' A = 0$, the null matrix. This gives $X_p' \underline{\xi}_i = \underline{0}$ which implies $X_n' \underline{\rho}_i = \underline{0}$. Hence $B\underline{\rho}_i = \underline{0}$, and it completes the proof of lemma 7.1.

Using this lemma one can say that $I_{[N=n]}$ and L_n are independent for all $n \geq n_o$, and one gets

(7.2.12)
$$\bar{\upsilon}(c) = E_\sigma(L_N) = m\sigma^2 E(N^{-1}) + cE(N)$$

and

(7.2.13)
$$\begin{cases} \eta(c) = \bar{\upsilon}(c)/\upsilon(c) = \frac{1}{2}[n^*E(N^{-1}) + E(N/n^*)], \\ w(c) = \bar{\upsilon}(c) - \upsilon(c) = c\cdot E[(N-n^*)^2/N]. \end{cases}$$

Regarding efficiencies of our procedure in (7.2.8), we have the following theorem.

__Theorem 7.1.__ $\lim_{c \to 0} \eta(c) = 1$ __and__ $\lim_{c \to 0} w(c) = O(c)$.

__Proof of theorem 7.1.__ The first part can be proved in the same lines as the proof of theorem 6.1. The other part can be proved in the same lines as the proof of (4.4.1). We omit the details; the main point to note is that R^2_{on} can be written as

$$(7.2.14) \qquad R^2_{on} = \sum_{i=1}^{n-m} U_i$$

for $n > m$, where U_i/σ^2 are iid χ^2_1 variables which follows from (7.2.11).

7.3 Asymptotic distribution of N defined in (7.2.8)

The problem of finding the asymptotic distribution of N in this case is the same as of example 2 in chapter 2 with $s = 2$, $t = 1$ and $\Psi_c = (m/c)^{1/2}$; $T_n = R^2_{on}/(n-m)$, $n > m$. We are thus led to

$$(7.3.1) \qquad (N-n^*)\left(\frac{2}{n^*}\right)^{\frac{1}{2}} \xrightarrow{\mathcal{L}} N(0,\ 1)$$

as $c \to 0$.

7.4. Sequential Point Estimation Of The Parameter Of A Rectangular Distribution.

Let $X_1, X_2, \ldots$ be a sequence of iid rv's with pdf

$$(7.4.1) \qquad f_\theta(x) = \theta^{-1}, \quad 0 < x < \theta,$$

$\theta \in (0, \infty)$ and is unknown. Given a random sample $X_1, \ldots, X_n$ of fixed size n, suppose the loss incurred in estimating θ by $T_n = \max(X_1, \ldots, X_n)$ is

(7.4.2) $$L_n = A|T_n - \theta| + cn,$$

where $A(> 0)$ is the known weight and $c(> 0)$ is the known cost per unit sample. The risk

(7.4.3) $$v_n(c) = E_\theta L_n = A\theta/(n+1) + cn$$

is minimized for $n = n^*$ where

(7.4.4) n^* is the smallest positive integer $\geq (A\theta/c)^{\frac{1}{2}} - 1$,

the corresponding minimum risk being

(7.4.5) $$v(c) = v_{n^*}(c) = A\theta/(n^*+1) + cn^* .$$

It may be noted that

(7.4.6) $$v(c) \sim 2(dc)^{\frac{1}{2}} \quad \text{as} \quad c \to 0,$$

where $d = A\theta$. However, θ is unknown, and so no fixed sample size will minimize (7.4.3) simultaneously for all θ. We consider the possibility of utilising a sample of random size, and propose the following stopping rule in accordance with (7.4.4).

(7.4.7) The stopping time $N \equiv N_c$ is the first integer $n \geq n_o$ (≥ 1) for which

$$n \geq (AT_n/c)^{\frac{1}{2}} - 1 .$$

Now estimate θ by T_N. The corresponding risk is then given by

$$\bar{\upsilon}(c) = EL_N = AE\,[\Theta - T_N] + cE(N) .$$

Following Starr (1966b) and Starr-Woodroofe (1969) we define the 'risk-efficiency' and 'regret' of our procedure respectively as

(7.4.8) $\eta(c) = \bar{\upsilon}(c)/\upsilon(c)$; $w(c) = \bar{\upsilon}(c) - \upsilon(c)$.

The main result towards these is

<u>THEOREM 7.2</u> $\lim_{c \to 0} \eta(c) = 1$.

The interpretation of the result is that asymptotically (as the cost component tends to zero), the sequential procedure is as much risk efficient as the corresponding procedure where θ is pretended to be known, and a fixed sample size is used.

We postpone the proof of theorem 7.2 to the next section 7.5. Certain lemmas pertaining to the behaviour of the stopping time are also proved there. These lemmas seem to be of independent interest.

We have not been able to say much regarding the asymptotic (as $c \to 0$) behaviour of the regret except that in view of our main theorem and (7.4.6), $w(c) = O(c^{1/2})$. We conjecture that this order can be improved, but do not have any analytic results at the moment to support this

Moderate sample behaviour of the 'risk efficiency' and 'regret' are studied in section 7.6, by using Monte-Carlo techniques, and the performance of the procedure seems to be quite satisfactory even for moderate c (i.e. moderate n^*).

7.5. The main results. The following lemma gives some of the basic properties of the stopping time N (defined in (7.4.7)).

Lemma 7.2. For the stopping rule defined in (7.4.7),

i) N is well defined, N is $\downarrow$ in c;

ii) $N \leq n^* + n_o$ with probability (wp) 1;

iii) $N/n^* \to 1$ a.s. as $c \to 0$;

iv) $E(N/n^*) \to 1$ as $c \to 0$;

v) For any $\emptyset \in (0, 1)$, $P(N \leq [\emptyset n^*]-1) = O_e(c^{n_o})$ as $c \to 0$

where $[u]$ denotes the integer part of u.

Proof of lemma 7.2. i) is obvious from the definition of N. To prove ii) observe the inequality (with I as usual indicator function)

$$(7.5.1)\quad N \le (AT_N/c)^{\frac{1}{2}} I_{[N>n_o]} + n_o I_{[N=n_o]} \le (d/c)^{\frac{1}{2}} I_{[N>n_o]} + n_o I_{[N=n_o]} \ (\text{wp } 1)$$

$$\le (n^*+1) I_{[N>n_o]} + n_o I_{[N=n_o]} \le n^* + n_o .$$

To prove this inequality, we have used the definitions of n^* and N in (7.4.4) and (7.4.7) respectively.

Since $n^* \to \infty$ as $c \to 0$ (in fact $n^* \sim (d/c)^{\frac{1}{2}}$ as $c \to 0$), it follows from (ii) that $\limsup_{c\to 0} (N/n^*) \le 1$ a.s.. Also, from the definition of N, $N \ge (AT_N/c)^{1/2} - 1$. Now, $T_m \to \theta$ a.s. as $m \to \infty$. So $T_N \to \theta$, a.s. as $c \to 0$, since $N \to \infty$ a.s. as $c \to 0$ (from (i)). Thus,

$$\liminf_{c\to 0} (N/n^*) \ge \liminf_{c\to 0} (AT_N/c)^{\frac{1}{2}} (n^*)^{-1} + \liminf_{c\to 0} (-n^{*-1}) = 1.$$

This proves (iii).

Using (iii) and Fatou's lemma, $\liminf_{c\to 0} E(N/n^*) \ge 1$. Also, from (ii) $\limsup_{c\to 0} E(N/n^*) \le 1$. This proves (iv).

To prove (v), first note that

$$(7.5.2)\quad P(N = n_o) = P(n_o + 1 \ge (AT_{n_o}/c)^{\frac{1}{2}}) = P(T_{n_o} \le c(n_o+1)^2/A)$$
$$= (c(n_o+1)^2/d)^{n_o} = 0_e(c^{n_o})$$

Also, for small c, $P(N \leq [\emptyset\, n^*]-1) \geq P(N = n_o) = O_e(c^{n_o})$.

Further,

$$P(N \leq [\emptyset\, n^*]-1) = \sum_{n=n_o}^{[\emptyset n^*]-1} P(N=n) \leq \sum_{n=n_o}^{[\emptyset n^*]-1} P(T_n \leq c(n+1)^2/A)$$

$$= \sum_{n=n_o}^{[\emptyset n^*]-1} (c(n+1)^2/d)^n = \sum_{n=n_o}^{[\emptyset n^*]-1} (c(n+1)^2/d)^{n_o} \cdot$$

$$(c(n+1)^2/d)^{n-u_o}$$

But for $n \leq [\emptyset n^*]-1$, $c(n+1)^2/d \leq c(\emptyset n^*)^2/d \leq (c/d)\emptyset^2(d/c)$

$$= \emptyset^2 \quad (0 < \emptyset^2 < 1) \ .$$

So, $P(N \leq [\emptyset n^*]-1) \leq (c/d)^{n_o} \sum_{n=n_o}^{[\emptyset n^*]-1} (n+1)^{2n_o} \emptyset^{2(n-n_o)}$

$$\leq (c/d)^{n_o} \sum_{n=n_o}^{\infty} (n+1)^{2n_o} \emptyset^{2(n-n_o)} = O_e(c^{n_o}),$$

using the ratio rule of convergence. This proves (V). It completes the proof of lemma 7.2.

<u>Proof of the theorem 7.2.</u>

With the aid of the lemma 7.2, we are now able to prove the theorem 7.2. First, note that in view of (7.4.6) and $n^* \sim (d/c)^{1/2}$ as $c \to 0$, it suffices to show that

(7.5.3) $E(N/n^*) \to 1$ as $c \to 0$; $n^* E(1-\theta^{-1} T_N) \to 1$ as $c \to 0$.

But, $E(1-\theta^{-1} T_N) \geq E(1-\theta^{-1} T_{n^*+n_o})$, using (ii) of the lemma 7.2, and the fact that T_m is ↑ in m.

So,

$$\liminf_{c \to 0} n^* E(1-\theta^{-1} T_N) \geq \liminf_{c \to 0} n^*\left(1-\frac{n^*+n_o}{n^*+n_o+1}\right)$$

(7.5.4) $$= \lim_{c \to 0} [1+(n_o+1)/n^*]^{-1} = 1 .$$

Fix $\emptyset$ in (0, 1). Then,

$$n^*E(1-\theta^{-1} T_N) = n^*E[(1-\theta^{-1} T_N)\{ I_{[N \leq [\emptyset n^*]-1]} + I_{[N \geq [\emptyset n^*]]}\}],$$

I denoting the usual indicator function.

Thus,

$$n^*E(1-\theta^{-1} T_N) \leq n^*P(N \leq [\emptyset n^*]-1) + n^*(1-\theta^{-1} T_{[\emptyset n^*]})$$

(7.5.5) $$= 0_e\left(c^{\frac{2n_o-1}{2}}\right) + n^*\left(1-\frac{[\emptyset n^*]}{[\emptyset n^*]+1}\right)$$

$$= 0_e\left(c^{\frac{2n_o-1}{2}}\right) + \emptyset^{-1} .$$

Hence,

(7.5.6) $$\limsup_{c \to 0} n^*E(1-\theta^{-1} T_N) \leq \emptyset^{-1} ,$$

for $n_o \geq 1$. $\emptyset$ being arbitrary, we get, (7.5.7) $$\limsup_{c \to 0} n^* E(1 - \theta^{-1} T_N) \leq 1.$$

The theorem 7.2 follows now from (7.5.4), (7.5.7) and (iv) of lemma 7.2.

Remarks: One essential distinctive feature of our problem as compared to the usual sequential point estimation problem is the lack of independence of L_n and $I_{[N=n]}$ for all $n \geq n_o$. Thus, unlike the usual cases (see e.g. Robbins (1959), Starr (1966b)) the risk $\bar{v}(c)$ corresponding to the sequential rule cannot be expressed in term of moments of the stopping time. But the analysis in our case becomes simple in view of the boundedness of T_m by $\theta(w_p 1)$ for $m \geq 1$ and the fact that T_m is $\updownarrow$ in m.

7.6. Behaviour of the procedure (7.4.7) for moderate c.

For moderate sample behaviour of the rule in (7.4.7), we consider the following values of n^*:

5, 7, 9, 10, 25, 40, 50, 75, 90, 100

150, 195, 200, 225, 250

We take n_o = 3, 5, 7, 9. For estimating $E(N), E(Y_{N(N)})$ for each entry, we repeat the experiment 200 times in H-400 electronic computer using pseudo-random numbers. We present the results of our Monte-Carlo experiments in tables 7.1-7.4. We provide standard errors (S.E.) of our estimates of $Y_{N(N)}$, $\eta(c)$ and $w(c)$.

Table 7.1: ASN, Risk Efficiency and Regret when $n_o = 3$.

n^*	$E(N)$	$Y_N(N)$ Mean	$Y_N(N)$ S.E.	$\eta(c)$ Mean	$\eta(c)$ S.E.	$w(c)$ Mean	$w(c)$ S.E.
5	4.755	.807	.170	1.064	.526	0.706	5.790
7	6.680	.848	.156	1.093	.624	-1.407	9.366
9	8.760	.896	.119	1.004	.598	0.084	11.359
10	9.720	.895	.106	1.067	.583	1.413	12.250
25	24.820	.959	.040	1.025	.525	1.276	26.757
40	39.810	.975	.025	1.001	.519	0.096	42.012
50	49.840	.979	.020	1.030	.526	3.118	53.167
75	74.880	.987	.011	0.969	.415	- 4.738	62.731
90	89.845	.989	.010	1.013	.484	2.370	87.681
100	99.830	.990	.009	0.998	.469	- 3.796	94.239
150	149.810	.993	.007	1.027	.549	8.087	165.274
195	194.895	.995	.004	0.946	.475	-21.229	185.812
200	199.905	.995	.004	0.924	.407	-30.272	163.189
225	224.845	.996	.005	0.959	.548	-18.688	247.096
250	249.825	.996	.004	1.020	.530	10.309	265.753

Table 7.2: ASN, Risk Efficiency and Regret when $n_o = 5$.

n^*	E(N)	$Y_N(N)$ Mean	S.E.	$\eta(c)$ Mean	S.E.	w(c) Mean	S.E.
5	6.000	.863	.135	0.995	.444	- 0.057	4.883
7	6.865	.880	.115	0.969	.475	- 0.451	7.126
9	8.730	.878	.129	1.099	.650	1.881	12.346
10	9.770	.896	.104	1.065	.578	1.370	12.129
25	24.765	.956	.044	1.060	.571	3.087	29.144
40	39.855	.976	.023	0.995	.477	- 0.388	38.646
50	49.890	.981	.016	0.988	.422	- 1.154	42.688
75	74.855	.987	.011	0.970	.444	- 4.477	67.091
90	89.840	.988	.010	1.015	.500	2.730	90.613
100	99.805	.989	.010	1.032	.512	6.452	102.949
150	149.875	.994	.006	0.959	.496	-12.171	149.409
195	194.855	.995	.005	1.031	.479	12.406	137.461
200	199.835	.995	.006	1.023	.559	9.248	224.371
225	224.825	.995	.005	1.018	.556	8.294	250.712
250	249.850	.996	.004	0.959	.523	-20.052	262.244

Table 7.3: ASN, Risk Efficiency and Regret when $n_o = 7$

n^*	$E(N)$	$Y_N(N)$		$\eta(c)$		$w(c)$	
		Mean	S.E.	Mean	S.E.	Mean	S.E.
5	8.000	.893	.099	1.077	.325	0.845	3.585
7	8.000	.883	.112	1.033	.480	0.504	7.207
9	8.820	.889	.097	1.049	.492	0.943	9.359
10	9.770	.907	.088	0.999	.489	- 0.016	10.288
25	24.805	.958	.042	1.033	.544	1.729	27.748
40	39.830	.974	.026	1.021	.539	1.741	43.708
50	49.845	.981	.017	0.970	.434	- 2.957	43.874
75	74.825	.985	.014	1.046	.543	7.036	81.934
90	89.840	.988	.010	1.019	.457	3.460	82.791
100	99.835	.990	.010	0.991	.519	- 1.828	104.348
150	149.880	.994	.005	0.941	.407	-17.645	127.685
195	194.860	.995	.005	0.960	.486	-15.471	190.179
200	199.835	.995	.005	0.995	.505	- 1.971	202.446
225	224.845	.996	.003	0.961	.435	-17.502	196.271
250	249.840	.996	.003	1.016	.467	8.168	234.343

Table 7.4: ASN, Risk Efficiency and Regret When $n_o = 10$

n^*	E(N)	$\gamma_N(N)$ Mean	$\gamma_N(N)$ S.E.	$\eta(c)$ Mean	$\eta(c)$ S.E.	w(c) Mean	w(c) S.E.
5	11.000	.917	.080	1.272	.262	2.991	2.889
7	11.000	.908	.085	1.126	.362	1.889	5.433
9	11.000	.910	.084	1.051	.440	0.977	8.370
10	11.000	.921	.069	0.976	.402	- 0.493	8.455
25	24.800	.959	.039	1.025	.505	1.275	25.745
40	39.855	.978	.021	0.939	.431	- 4.917	34.972
50	49.820	.980	.018	0.999	.466	- 0.023	47.116
75	74.870	.987	.011	0.994	.434	- 0.877	65.558
90	89.815	.988	.011	1.031	.539	5.627	97.705
100	99.780	.989	.011	1.031	.557	6.327	111.901
150	149.875	.994	.006	0.952	.464	-14.418	139.764
195	194.845	.995	.005	0.997	.490	- 1.279	191.650
200	199.845	.995	.004	1.026	.462	10.083	185.461
225	224.885	.996	.004	0.948	.458	-23.561	206.587
250	249.850	.995	.004	1.007	.522	3.842	261.387

Remarks:

The moderate sample behaviour of the rule in (7.4.7) is **quite** satisfactory, and we recommend its use in practice.

CHAPTER 8

SOME RESULTS ON SEQUENTIAL PREDICTION

8.1. Introduction. So far we have concentrated on sequential estimation - either point estimation or fixed-width interval estimation. In this chapter we consider the problem of predicting a future observation Y from a series of observations $X_1, \ldots, X_n$. We assume, as usual, that $X_1, \ldots, X_n$, Y are iidrv's following a certain distribution which depends on the parameter $\underline{\theta}$. As in estimation set up, the problem is two-fold in the prediction set up also. First, to provide a point predictor of Y on the basis of $X_1, \ldots, X_n$ which has minimum risk (considering a suitable loss function) for all parameter points $\underline{\theta}$ and second, to provide a fixed-width prediction interval for Y which has a specified coverage probability for all parameter points $\underline{\theta}$.

The results of this chapter are as follows.

1) In section 8.2, considering a location and scale parameter family of symmetrical unimodal densities $f(\frac{x-\alpha}{\beta})$, $-\infty < x < \infty$, $-\infty < \alpha < \infty$, $0 < \beta < \infty$ (β unknown), it has been proved that there exists no fixed-width prediction interval for Y which has a specified coverage probability, under

any sampling procedure which terminates with probability one. This is an example of a very simple nature on the non-existence of a sequential procedure for some statistical problem. So far, in the literature, only two results of this kind, although in different contexts, are known to us [Bahadur-Savage (1956), Blum-Rosenblatt (1967)].

2) In section 8.3, the problem of providing an optimum predictor of Y has been tackled, considering three families of densities, viz., truncated exponential, exponential and normal. A sequential procedure is given in each case to get a point predictor of Y, the loss being squared error and cost of sampling proportional to sample size. In the first case, the sequential procedure is shown to have asymptotic risk efficiency. While in the other two cases it is shown by referring to relevant papers [Starr (1966b), Starr-Woodroofe (1969, 1972)] that the procedures are such that the 'regrets' are bounded.

8.2. Non-existence of fixed-width prediction interval with prescribed coverage probability

To fix ideas we first consider the simple case of a $N(\mu, \sigma^2)$ distribution, where the parameters may or may not be known. If $X_1, \ldots, X_n$ denote a random sample from the above population and we want to predict Y by $[\bar{X}_n - d, \bar{X}_n + d]$, where $\bar{X}_n = n^{-1} \Sigma X_i$, $(n \geq 1)$, then,

$$P[\,|\bar{X}_n - Y| \leq d] = 2\Phi[(n/n+1)^{\frac{1}{2}} \cdot \frac{d}{\sigma}] - 1 \leq 2\Phi(d/\sigma) - 1$$
$$< \gamma \text{ (preassigned)}$$

if $d/\sigma <$ the upper $100\,(1+\gamma)/2\,\%$ point of the $N(0, 1)$ distribution. The simple reason is that if Y has a non-degenerate distribution and the scatter is wide enough, it is not possible to put Y in a prediction interval of preassigned length and prescribed coverage probability. This simple idea is the basis of subsequent development in this section. The basic difference with the problem of estimation of μ by a confidence interval $[\bar{X}_n - d, \bar{X}_n + d]$ with prescribed coverage probability $1-\alpha$ is that in the second case μ can be viewed as a rv degenerate at a point.

Let $X_1, X_2, \ldots$ be a sequence of i i d r v, each with the unimodal density (with respect to Lebesgue measure) $f(\frac{x-\alpha}{\beta})$, $-\infty < \alpha < \infty$, $\beta > 0$, $-\infty < x < \infty$, β unknown, which is symmetric about the modal value $x = \alpha$, Let Y be independent of the X-sequence, having the same distribution as that of the X's. We prove the following theorem.

Theorem 8.1. Under the above assumptions, there does not exist a sampling procedure that terminates with probability one for all α, β, under which a prediction interval of length $2d$ for Y is attainable which has coverage probability γ when

$d > 0$ and $0 < \gamma < 1$ are preassigned.

Proof: The proof is by negation. Suppose there exists a sampling plan which terminates with probability one and under which the interval predictor

(8.2.1) $$I_d(Z_N) = [h(Z_N) - d,\ h(Z_N) + d]$$

is a prediction interval for Y with coverage probability γ for all $-\infty < \alpha < \infty$, $\beta > 0$ where $Z_N = (X_1, \dots, X_N)$, $h(\cdot)$ is a measurable function and N is the sample size required by the sampling plan. Then the following holds

(8.2.2) $$P_{\alpha,\beta}[Y \in I_d(Z_N)] \geq \gamma \quad \text{for all } -\infty < \alpha < \infty, \quad \beta > 0.$$

We shall show that (8.2.2) cannot hold for all $\beta > 0$. This is because (8.2.2) can be rewritten as

(8.2.3) $$\gamma \leq P_{\alpha,\beta}[h(Z_N) - d \leq Y \leq h(Z_N) + d]$$

$$= P_{\alpha,\beta}\left[\frac{h(Z_N) - \alpha}{\beta} - \frac{d}{\beta} \leq \frac{Y - \alpha}{\beta} \leq \frac{h(Z_N) - \alpha}{\beta} + \frac{d}{\beta}\right]$$

$$= E\left[P_{0,1}\left\{\frac{h(Z_N) - \alpha}{\beta} - \frac{d}{\beta} \leq Y^* \leq \frac{h(Z_N) - \alpha}{\beta} + \frac{d}{\beta} \,\middle|\, Z_N\right\}\right]$$

$$\text{for all } -\infty < \alpha < \infty, \quad \beta > 0,$$

where $Y^* = (Y - \alpha)/\beta$, the inner probability has to be evaluated regarding Y^* as random (and Z_N as fixed) and the outer expectation is with respect to Z_N. Since Y is independent of Z_N, so is Y^* which has a completely known distribution, symmetric about the modal value of the distribution, the origin.

Now, the inner probability

$$\leq P_{0,1}\left\{|Y^*| \leq \frac{d}{\beta} \mid Z_N\right\} = P_{0,1}\left\{|Y^*| \leq \frac{d}{\beta}\right\}.$$

Let ϵ (>0) be such that $\gamma/2 = P[\,|Y^*| \leq \epsilon]$. Choose $\beta > \frac{d}{\epsilon}$, so that rhs of (8.2.3) < $\gamma/2$, a contradiction! This proves that an interval predictor for Y satisfying (8.2.2) does not exist. This completes the proof of Theorem 8.1. Now we add a few remarks.

Remarks

1. The normal, the double exponential and the cauchy family of densities, among others, satisfy the assumptions in the theorem 8.1 and hence in neither case does there exist a fixed-width prediction interval for a future observation which has a specified coverage probability.

2. The above theorem remains valid even if α is known. Also, the restriction to symmetrical densities is not at all necessary. If common density is unimodal, it follows that Y^* has a completely known distribution having a unique mode at some point y_0, say. Let a (may be $-\infty$) be the lowest end point of the range of Y^*. Let $\epsilon_1, \epsilon_2 > 0$ be such that the interval $[\max(y_0 - \epsilon_2, a),\ y_0 + \epsilon_1]$ is the 'optimum' modal interval of

probability content $\gamma/2$ i.e., the interval around y_o of smallest length and probability content $\gamma/2$. It then follows that if $2d/\beta$ is less than ϵ_1 i.e., $\beta > 2d/\epsilon_1$, then the inner probability in (8.2.3) is always less than $\gamma/2$, independent of the quantity $(h(Z_N) - \alpha)/\beta$.

8.3. Minimum risk point predictor

The problem of providing an 'optimum' point predictor of Y has been tackled in this section. We consider a sequence $X_1, X_2, \ldots$ of independent r v 's distributed identically with finite second moment. While we wish to predict a future observation Y (distributed identically as X's and independent of X-sequence) through $\bar{X}_n$ having recorded $(X_1, X_2, \ldots, X_n)$, suppose the loss incurred is given by

$$(8.3.1) \qquad L_n = A(\bar{X}_n - Y)^2 + cn$$

where A is a known positive constant and the cost of sampling is proportional. Our object is to predict Y such that the risk, viz. $E(L_n)$ is minimized for all parameter points. Though fixed width prediction interval procedure does not exist, we shall show that minimum risk point predictor for Y can be obtained in some cases with desirable properties of asymptotic nature. We consider three populations viz. truncated exponential, exponential and normal in three sections.

8.4. Truncated exponential family

The X's have the density

$$f(x;\ \mu,\ \sigma) = \sigma^{-1} \exp(\frac{x-\mu}{\sigma}), \quad x \geq \mu$$

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

$-\infty < \mu < \infty$, $\sigma > 0$, both the parameters unknown. With (8.3.1), the risk

(8.4.1) $$\upsilon_n(c) = A\sigma^2(1 + \frac{1}{n}) + cn$$

which is minimum for

$$n = n^* = \sqrt{A}\,\sigma\, c^{-\frac{1}{2}},$$

the corresponding minimum risk being $\upsilon(c) = \upsilon_{n^*}(c) = cn^*(n^*+2)$. Thus, in ignorance of σ, no fixed sample size procedure will minimize (8.4.1) simultaneously for all μ, σ. So the possibility of utilising a sample of random size N is considered which is shown to achieve the objective asymptotically. We propose the following stopping rule:

(8.4.2) The stopping time $N \equiv N_c$ is the 1st integer $n\ (\geq n_o)$ such that

$$n \geq \sqrt{A}\,\sigma_n\, c^{-\frac{1}{2}}$$

where $X_{n(1)} \leq \cdots \leq X_{n(n)}$ are the ordered $X_1, \ldots, X_n$,

$\sigma_n = (n-1)^{-1} \sum_{i=1}^{n} (X_i - X_{n(1)})$, $n_o (\geq 3)$ is the starting sample size. When we stop, we predict Y by $\bar{X}_N$.

This rule can be rephrased as:

(8.4.3) The stopping time N is the first integer $n(\geq n_o)$ such that

$$\sigma_n^* \leq n/n^*$$

where $\sigma_n^* = (n-1)^{-1} \sum_{i=1}^{n} (X_i - X_{n(1)})$ and X_i's follow the law of distribution $f(x; 0, 1)$. Thus we get

(8.4.4) $$E(L_N) = c[n^{*2} E(\frac{1}{N^2}) + n^{*2} E\{(1-\frac{1}{N})(\sigma_N^*-1)\}^2 + E(N) + n^{*2}].$$

Define the risk efficiency

(8.4.5) $$\eta(c) = E(L_N)/\,\upsilon(c).$$

Now, about the usefulness of the sequential procedure in (8.4.2), we have the following theorem.

<u>Theorem 8.2.</u>

(8.4.6) N <u>is well defined, non-increasing as a function of</u> c <u>and</u> $N \to \infty$ <u>a.s. as</u> $c \to 0$.

(8.4.7) $$\lim_{c \to 0} N/n^* = 1 \quad \text{a.s.}$$

(8.4.8) $E(N) \leq n^* + n_o + 1$ for all μ, σ

(8.4.9) $E(N^2) \leq (n^* + n_o + 1)^2$ for all μ, σ.

Proof: Proof of (8.4.6) is routine. To prove (8.4.7) observe the basic inequality

(8.4.10) $$n^* \sigma_N/\sigma \leq N \leq n_o I_{[N=n_o]} + n^*\sigma_{N-1}/\sigma + 1$$

Now $\sigma_m \to \sigma$ a.s. as $m \to \infty$. Also, $N \to \infty$ a.s. as $c \to 0$ from (8.4.6). Hence, $\sigma_N \to \sigma$ a.s., $\sigma_{N-1} \to \sigma$ a.s. as $c \to 0$. Dividing all sides of (8.4.10) by n^* and making $c \to 0$, we get the result. To prove (8.4.9) and (8.4.10), first note that from (8.4.10).

(8.4.11) $$N(N-2) \leq n_o(N-2)I_{[N=n_o]} + n^* \sum_{1}^{N-1} (X_i - X_{N-1(1)})/\sigma + N-2$$
$$\leq (n_o - 2)N + n^* \sum_{1}^{N} (X_i - \mu)/\sigma + N .$$

Assume for the moment $E(N) < \infty$. Then using Schwarz's inequality and Wald's equation, one gets from (8.4.11) that

(8.4.12) $$(EN)^2 - 2EN \leq (n_o - 2)E(N) + (n^*+1)\,E(N)$$

Hence $E(N) \leq (n^* + n_o + 1)$ which proves (8.4.8). Also, from (8.4.11),

(8.4.13) $$E(N^2) \leq (n^* + n_o + 1)E(N) \leq (n^* + n_o + 1)^2.$$

If we do not assume $EN < \infty$, define $N_m = \min(N, m)$ and get as before $EN_m \leq n^* + n_o + 1$, $EN_m^2 \leq (n^* + n_o + 1)$. But $N_m \uparrow N$ a.s. as $m \to \infty$. Using Monotone convergence theorem, we get the result.

We state the following theorem about 'risk efficiency'.

<u>Theorem 8.3.</u> $\lim_{c \to 0} \eta(c) = 1.$

<u>Proof</u>. First note that

$$\eta(c) \geq cn^{*2}/(cn^*(n^* + 2)) = \left(1 + \frac{2}{n^*}\right)^{-1}.$$

Hence

(8.4.14) $$\liminf_{c \to 0} \eta(c) \geq 1 .$$

Also, $\eta(c) \leq \left(1 + \frac{2}{n^*}\right)^{-1} [1 + (n^*)^{-2} E(N) + E(\sigma_N^* - 1)^2 + E(N^{-2})]$

Now, $E(N^{-2}) \to 0$ as $c \to 0$. Hence, it suffices to prove the following lemma:

<u>Lemma 8.1.</u>

(8.4.15) $$\lim_{c \to 0} E(N/n^*) = \lim_{c \to 0} E(N/n^*)^2 = 1$$

(8.4.16) $$\lim_{c \to 0} E(\sigma_N^*) = \lim_{c \to 0} E(\sigma_N^*)^2 = 1$$

<u>Proof.</u> Using Fatou's lemma, (8.4.8) and (8.4.9) we get (8.4.15).

Using (8.4.15), the inequality $\sigma_N \leq \sigma.(N/n^*)$, and the fact $\sigma_N \to \sigma$ a.s. as $c \to 0$, along with Fatou's lemma, we get (8.4.16).

Remarks:

1) Suppose $w(c) = E(L_N) - \upsilon(c)$, which is called the 'regret'. It is not known yet whether $\lim_{c \to 0} w(c) \leq O(1)$, which we strongly feel to be true.

2) Though for any fixed positive integer n, the events $[N = n]$ and L_n are dependent, the proof of asymptotic risk efficiency is not at all tedious, that is mainly because in this particular context we did not require the rate of convergence of $E(\sigma_N^* - 1)^2 \to 0$ as $n^* \to \infty$.

8.5. Exponential family

The X's have the density

$$f(x) = \frac{1}{\lambda} \exp(-x/\lambda), \qquad x > 0$$

where $\lambda > 0$ is unknown.

With (8.3.1), the risk

$$\upsilon_n(c) = A\lambda^2 \left(1 + \frac{1}{n}\right) + cn$$

which is minimum for

$$n = n^* = \sqrt{\frac{A}{c}}\, \lambda$$

As in section 8.4, we propose the following sequential sampling scheme which has some desirable asymptotic properties.

(8.5.1) The stopping time N is the 1st integer $n(\geq n_o)$ such that

$$n \geq \sqrt{\frac{A}{c}}\,\bar{X}_n$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$, $n_o(\geq 2)$ is the starting sample size. When we stop, we predict Y by $\bar{X}_N$.

We write L_n in a different but equivalent form as follows.

$$L_n = A(\bar{X}_n - \lambda)^2 + A(Y-\lambda)^2 - 2A(\bar{X}_n - \lambda)(Y-\lambda) + cn.$$

Thus,

(8.5.2) $$E(L_N) = AE(\bar{X}_N - \lambda)^2 + A\lambda^2 + cE(N).$$

Also, the minimum risk

(8.5.3) $$\upsilon(c) = \upsilon_{n*}(c) = (n*^2 + 2n*)\,c.$$

Thus the regret is

$$w(c) = E(L_N) - \upsilon(c)$$
$$= c[AE(\bar{X}_N - \lambda)^2 - n* + E(N) - n*]$$

which is bounded for small c by the theorem of Starr-Woodroofe (1972). Also the asymptotic risk efficiency follows from their results.

Remarks:

1) Here also, for any fixed n, the two events $[N = n]$ and L_n are dependent. But for the regret part, the rates of convergences which we need are done in Starr-Woodroofe (1972) with reference to a sequential point estimation problem.

2) However, the asymptotic risk efficiency of the rule in (8.5.1) can be very easily proved in the same way as that of (8.4.2).

8.6. Normal family

The X's have the density

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp[-(x-\mu)^2/2\sigma^2], \quad -\infty < x < \infty$$

where $-\infty < \mu < \infty$, $\infty > \sigma > 0$ are both unknown.

With (8.3.1), the risk

$$v_n(c) = A\sigma^2 (1 + \tfrac{1}{n}) + cn$$

which is minimum for

$$n = n^* = \sqrt{\frac{A}{c}}\,\sigma\ .$$

As in sections 8.4 and 8.5, we propose the following sequential sampling scheme:

(8.6.1) The stopping time N is the 1st integer $n\ (\geq n_o)$ such that

$$n \geq \sqrt{A}\,.\,c^{-\frac{1}{2}}\, s_n$$

where $s_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$, $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$, $n_o (\geq 3)$, is the starting sample size. When we stop, we predict Y by $\bar{X}_N$.

Now, L_n can be written as

$$L_n = A(\bar{X}_n - \mu)^2 + A(Y - \mu)^2 - 2A(\bar{X}_n - \mu)(Y - \mu) + cn$$

so that

$$E(L_N) = c[n^{*2}\, E(N^{-1}) + n^{*2} + E(N)],$$

since for any fixed integer n, the events $[N = n]$ and L_n are independent.

Also, the minimum risk

$$\upsilon(c) = \upsilon_{n^*}(c) = (n^{*2} + 2n^*)c.$$

From Starr (1966b) it follows that $\lim_{c \to 0} E(N/n^*) = 1$ which gives asymptotic risk efficiency at a stroke. For the regret, viz.,

$$w(c) = E(L_N) - \upsilon(c) = c.E[(N - n^*)^2/N],$$

we apply Starr-Woodroofe's (1969) results to get that $w(c)$ is bounded for small c if and only if $n_o \geq 3$

R E F E R E N C E S

1] Anscombe, F. J. (1952): Large sample theory of sequential estimation. Proc. Comb. Phil. Soc., 48, 600-607.

2] Anscombe, F. J. (1953): Sequential estimation. J.R.S.S. Ser. B, 15, 1-21.

3] Bahadur, R. R. and Savage, L. J. (1956): The non-existence of certain statistical procedures in non-parametric problems. Ann. Math. Statist., 27, 1115 - 1122.

4] Basu, A. P. (1971): On a sequential rule for estimating the location parameter of an exponential distribution. Naval Research Logistics Quarterly, 18, 329-337.

5] Basu, D. (1955): On statistics independent of a complete sufficient statistic. Sankhya, 15, 377-380.

6] Berk, R. H. (1966): Limiting behaviour of posterior distributions when the model is incorrect. Ann. Math. Statist., 37, 51-58.

7] Bhattacharyya, P. K. and Mallik, A. (1973): Asymptotic normality of the stopping times of some sequential procedures. Ann. Statist., 1, 1203-1211.

8] Blum, J. R. and Rosenblatt, J. (1967): On partial a-priori information in statistical information. Ann. Math. Statist., 38, 1671-1678.

9] Callahan, J. (1969): On some topics in sequential multiparameter estimation. Ph. D. thesis, Johns Hopkins University.

10] Chatterjee, S. K. (1959a): On an extension of Stein's two-sample procedure to the multinormal problem. Cal. Statist. Assoc. Bull., 8, 121-148.

11] Chatterjee, S. K. (1959b): Some further results on the multinormal extension of Stein's two-sample procedure. Cal. Statist. Assoc. Bull., 9, 20-28.

12] Chow, Y. S. and Robbins, H. (1965): On the asymptotic theory of fixed width sequential confidence intervals for the mean. Ann. Math. Statist., 36, 457-462.

13] Dantzig, R. B. (1940): On the non-existence of tests of Student's hypothesis having power functions independent of σ. Ann. Math. Statist., 11, 186-192.

14] Darling, D. A. and Robbins, H. (1967a): Iterated logarithm inequalities. Proc. Nat. Acad. Sci. USA., 57, 1188-1192.

15] Darling, D. A. and Robbins, H. (1967b): Inequalities for the sequence of sample means. Proc. Nat. Acad. Sci. USA., 57, 1577-1580.

16] Darling, D. A. and Robbins, H. (1968): Some further remarks on inequalities for sample sums. Proc. Nat. Acad. Sci., USA., 60, 1175-1182.

17] Gleser, L. J. (1965): On the asymptotic theory of fixed-size sequential confidence bands for linear regression parameters. Ann. Math. Statist., 36, 463-467. (Correction: (1966), Ann. Math. Statist., 37, 1053-1055.)

18] Gleser, L. J. (1969): On sequential estimation procedures. Private communication.

19] Hoeffding, W. (1948): A class of statistics with asymptotical normal distribution. Ann. Math. Statist., 19 293-325.

20] Hoeffding, W. (1961): The strong law of large numbers for U-statistics. Institute of Statistics Mimeo Series No.302, University of North Carolina, Chapel Hill.

21] Khan, R. A. (1968): Sequential estimation of the mean vector of a multivariate normal distribution Sankhya, 30, 331-334.

22] Lehmann, E. (1951): Mimeographed notes on the theory of estimation. University of California Press.

23] Loeve, M. (1963): Probability Theory. Van Nostrand, Princeton

24] Rao, C. R. (1965): Linear Statistical Inference and Its Applications. John Wiley and Sons, New York.

25] Ray, W. D. (1957): Sequential confidence intervals for the mean of a normal distribution with unknown variance. J.R.S.S. Ser. B, 19, 133-143.

26] Robbins, H. (1959): Sequential estimation of the mean of a normal population. Probability and Statistics (H. Cramer volume). Almqvist and Wiksell, Uppsala, 235-245.

27] Robbins, H., Simons, G. and Starr, N. (1967): A sequential analogue of the Behrens-Fisher problem. Ann. Math. Statist., 38, 1384-1391.

28] Robbins, H. (1970): Statistical methods related to the law of iterated logarithm. Ann. Math. Statist., 41 1397-1409.

29] Rohatgi, V. K. and O'Neill, R. T. (1973): On sequential estimation of the mean of a multinormal population. Ann. of Inst. of Statist. Math., 25, 321-325.

30] Sen, P. K. (1960): On some convergence properties of U-statistics. Cal. Statist. Assoc. Bull., 10, 1-18.

31] Siegmund, D. (1968): On the asymptotic normality of one-sided stopping times. Ann. Math. Statist., 39, 1493-1497.

32] Simons, G. (1968): On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. Ann. Math. Statist., 39, 1946-1952.

33] Sinha, B. K. and Mukhopadhyay, N. (1974a): Sequential estimation of a bivariate normal mean vector. Submitted for publication.

34] Sinha, B. K. and Mukhopadhyay, N. (1974b): Sequential estimation of a bivariate normal mean vector-II. Submitted for publication.

35] Sproule, R. N. (1969): A sequential fixed-width confidence interval for the mean of a U-statistic. Ph.D. thesis, University of North Carolina, Chapel Hill.

36] Srivastava, M. S. (1967): On fixed width confidence bounds for regression parameters and mean vector. J.R.S.S. Ser. B, 29, 132-140.

37] Srivastava, M. S. (1970): On a sequential analogue of the Behrens-Fisher problem. J.R.S.S. Ser. B, 32, 144-148.

38] Starr, N. (1966a): The performance of a sequential procedure for the fixed-width interval estimation of the mean. Ann. Math. Statist., 37, 36-50.

39] Starr, N. (1966b): On the asymptotic efficiency of a sequential procedure for estimating the mean. Ann. Math. Statist., 37, 1173-1185.

40] Starr, N. and Woodroofe, M. (1968): Remarks on a stopping time. Proc. Nat. Acad. Sci., USA., 61, 1215-1218.

41] Starr, N. and Woodroofe, M. (1969): Remarks on sequential point estimation. Proc. Nat. Accd. Sci. USA., 63, 285-288.

42] Starr, N. and Woodroofe, M. (1972): Further remarks on sequential estimation: The exponential case. Ann. Math. Statist., 43, 1147-1154.

43] Stein, C. (1945): A two sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist., 16, 243-258.

44] Stein, C. (1949): Some problems in sequential estimation (abstract). Econometrica, 17, 77-78.

45] Stein, C. (1956): Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berk. Symp. Vol. 1, 197-206. Univ. Calif. Press.

46] Stein, C. and James, W. (1961): Estimation with quadratic loss. Proc. Fourth Berk. Symp. Vol. 1, 361-379. Univ. Calif. Press.
