PROPERTIES OF SEQUENTIAL TESTS WHEN THE CONDITIONS FOR STANDARD APPROXIMATE FORMULAE ARE NOT SATISFIED

by

David Gerhard Hoel
University of North Carolina

Institute of Statistics Mimeo Series No. 480

June 1966

This research was supported by the Army Research Office - Durham - Grant No. DA-ARO-D-31-124-G670.

DEPARTMENT OF STATISTICS
University of North Carolina
Chapel Hill, N. C

PROPERTIES OF SEQUENTIAL TESTS WHEN THE CONDITIONS FOR STANDARD APPROXIMATE FORMULAE ARE NOT SATISFIED

David Gerhard Hoel

Institute of Statistics
Mimeo Series No. 480

Name	Date

ACKNOWLEDGMENTS

I am deeply indebted to Professor Norman L. Johnson, not only for proposing the topic of this dissertation, but also for his guidance and encouragement throughout its development. It has been an invaluable experience to have had the opportunity to work under his direction.

I also wish to give very special thanks to Professors William J. Hall and Wassily Hoeffding for both their helpful suggestions and their creative teachings.

The computer time needed for the tables in chapters one and five which was provided by the Mathematics Division of the Oak Ridge National Laboratories is greatly appreciated.

Finally, I would like to thank Mrs. Elsie Camp and Miss Martha Jordan of the secretarial staff and especially Mrs. Judy Zeuner for her cheerful typing.

TABLE OF CONTENTS

CHAPTER PAGE

INTRODUCTION

Prior to 1943, sequential analysis as it is known and used today did not exist except for various sequential type procedures each designed with a specific problem in mind. Sequential analysis actually came into its own as a recognized sector of statistical inference with the publication of a series of papers ([32][1/], [33], [34], [35], [36], [37]) by Abraham Wald. It was here that he proposed and determined many of the properties of the now well-known <u>sequential probability ratio test</u> (s.p.r.t.) for discriminating between two simple hypotheses.

As a brief summarization of this procedure, suppose that we have a sequence of random variables $x_1, x_2, \ldots$ such that $x_1, \ldots, x_m$ has a probability density function $f_{\theta,m}(x_1, \ldots, x_m)$[2/], and we wish to discriminate between the two simple hypotheses

$$H_o: \ \theta = \theta_o \quad \text{and} \quad H_1: \theta = \theta_1 \ .$$

Then if A and B are two constants such that $0 < B < 1 < A$ and $f_{i,m} = f_{\theta_i,m}(x_1, \ldots, x_m)$, the s.p.r t requires that we continue sampling as long as

---

1/ The numbers in square brackets refer to the bibliography.

2/ Here and in the following chapters, most of the results are also valid when probability density function and the operation of integration are replaced by probability mass function and summation, respectively.

$$(0.1) \qquad B < \frac{f_{1,m}}{f_{o,m}} < A$$

and stop sampling as soon as one of these inequalities is violated. If sampling is terminated at trial m then H_o is accepted if $\frac{f_{1,m}}{f_{o,m}} \leq B$ and H_1 is accepted if $\frac{f_{1,m}}{f_{o,m}} \geq A$. If we define

$$b = \log B$$

$$a = \log A$$

$$Z_o = o$$

$$Z_i = \log \frac{f_{1,i}}{f_{o,i}} \qquad \text{for } i = 1,2, \ldots$$

$$z_i = Z_i - Z_{i-1} \qquad \text{for } i = 1,2, \ldots$$

then (0.1) may be written as "continue"sampling as long as

$$(0.2) \qquad b < \sum_{i=1}^{m} z_i < a$$

and stop sampling as soon as one of these inequalities is violated. If sampling is terminated at trial m then H_o is accepted if $\sum_{i=1}^{m} z_i \leq b$ and H_1 is accepted if $\sum_{i=1}^{m} z_i \geq a$. As a matter of notation, we shall assume that the symbols a, b, z_i, Z_i refer to the s.p.r.t. given by (0.2).

If we allow the constants a and b in (0.2) to become functions of m (to be denoted as a_m, b_m), we then shall refer to the procedure as a generalized sequential probability ratio test (g.s.p.r.t.). One of the most interesting examples of such a procedure is given by Anderson [1] (see also Weiss [39]).

For any sequential test procedure, the distribution of the sample size, N, and the operating characteristic function, $P_\theta[\text{accept } H_o]$, will, for practical purposes, completely describe the procedures behavior. Because of the difficulties involved in obtaining the distribution of N, the expected value of N, commonly called the average sample number (a.s.n.)along with the operating characteristic function has been used to represent the performance of a sequential test.

This paragraph contains a brief list of some of the properties of an s.p.r.t. all of which may be found in more detail in Wald's book [38]. If we define

$$1 - \alpha = P\left[\text{accept } H_o \mid H_o\right]$$

$$\beta = P\left[\text{accept } H_o \mid H_1\right]$$

then Wald's fundamental inequalities

$$(0.3) \qquad a \leq \log \frac{1-\beta}{\alpha}$$

$$b \geq \log \frac{\beta}{1-\alpha}$$

hold quite generally. Now, for the case when $z_1, z_2, \ldots$ are independent and identically distributed, Wald has given the following well-known approximations

$$\text{a.s.n.} \doteq \frac{b\,P\left[\text{accept } H_o\right] + a\left[1-P\left[\text{accept } H_o\right]\right]}{\mathcal{E}(z_1)} \quad \text{for } \mathcal{E}(z_1) \neq o$$

(0.4)

$$P\left[\text{accept } H_o\right] \doteq \frac{e^{at_o}-1}{e^{at_o}-e^{bt_o}}$$

where t_o is the nonzero solution of $\mathcal{E}[\,e^{tz_1}\,] = 1$. Page [21], and later Kemp [18], has improved these approximations for the case when z_i is normally distributed. More recently Tallis and Vagholkar [29] attempted to give general improvements to (0.4). Unfortunately, however, there are errors present in their paper (e.g. equation (11) on page 77). Continuing with the assumption of independent and identically distributed variables, Wald has also given the following bounds. Define

$$\xi = \sup_r \mathcal{E}\left[z_1 - r \mid z_1 \geq r \right] \quad \text{for a-b} > r > o$$

$$\xi' = \inf_r \mathcal{E}\left[z_1 + r \mid z_1 \leq -r \right] \quad \text{for a-b} > r > o$$

$$\delta = \sup_\rho \rho\, \mathcal{E}\left[e^{z_1 t_o} \mid e^{z_1 t_o} \geq \frac{1}{\rho} \right] \quad \text{for } \rho > o$$

$$\eta = \inf_\rho \rho\, \mathcal{E}\left[e^{z_1 t_o} \mid e^{z_1 t_o} \leq \frac{1}{\rho} \right] \quad \text{for } \rho > 1.$$

Then

$$\frac{(b+\xi')\,P\left[\text{accept } H_o\right] + a\left[1-P\left[\text{accept } H_o\right]\right]}{\mathcal{E}(z_1)} \leq \text{a.s.n} \leq$$

$$\frac{b \; P\left[\text{accept } H_o\right] + (a+\xi)\left[1-P\left[\text{accept } H_o\right]\right]}{\mathcal{E}(z_1)}$$

(0.5)

$$\frac{e^{at_o}-1}{e^{at_o}-\eta e^{bt_o}} \leq P\left[\text{accept } H_o\right] \leq \frac{\delta e^{at_o}-1}{\delta e^{at_o}-e^{bt_o}} \qquad \text{for } t_o > o$$

$$\frac{1-e^{at_o}}{\delta e^{bt_o}-e^{at_o}} \leq P\left[\text{accept } H_o\right] \leq \frac{1-\eta e^{at_o}}{e^{bt_o}-\eta e^{at_o}} \qquad \text{for } t_o < o \,.$$

In the four following chapters we shall be concerned with determining the properties of an s.p.r.t. when Wald's assumptions of independent and identically distributed variables are relaxed. Previous work on this problem has been mainly directed towards the class of s.p.r.t.'s obtainable by the invariance method (see Hall, Wijsman and Ghosh [13]). For some members of this class of tests, termination with probability one, ($P[\, N < \infty] = 1$), and monotonicity of the o.c. function have been established (for more details and references, see pages 586 and 587 of [13]). More generally, Bhate [7] and Bartholomew [6] have given conjectured approximations to the a.s.n. function and the o.c. function respectively. Both of these conjectures will be discussed in chapters two and three.

In chapter one, we begin by considering a particular s.p.r.t. in which the observations are neither independent nor identically distributed. The properties of this test are obtained exactly and thus we are furnished with a model with which the various bounds, approximations and conjectures of the following chapters may be compared.

Chapter two is concerned with the sample size of an s.p.r.t. The bounds given in (0.5) for the a.s.n. are extended in a natural way to the case of z_1, z_2, ... independent but not identically distributed. With this assumption on the z_i's, bounds similar to (0.5) are obtained for $\mathcal{E}(N^2)$. We also consider extending the applicability of the lower bounds for the a.s.n of a general sequential test procedure given by Hoeffding [14], [15]. We then conclude by giving a method of general applicability for obtaining upper and lower bounds for $\mathcal{E}(N^m)$.

In chapter three the o.c. function of a s.p.r.t. is studied. We begin by introducing the concept of "consistent conjugate densities" in connection with applying Wald's o.c. formula given in (0.4) to cases in which the z_i's are not independent and identically distributed. Also the bounds for P[accept H_o] given in (0.5) are extended to cases in which the z_i's are independent but not necessarily identically distributed. Next we consider conditions under which P[accept H_o] is increased by replacing each z_i by a stochastically larger variable. Also some discussion is given concerning the effect upon P[accept H_o] after the z_i's have undergone a scale change. Finally we conclude by giving a general method for obtaining upper and lower bounds for P[accept H_o].

In chapter four we apply the results of chapters two and three to the s.p.r.t. given in chapter one. We then conclude by attempting to determine P[accept H_o] for the s.p.r.t. in which $z_1, z_2, \ldots$ have a multivariate normal distribution.

In the fifth and final chapter we briefly discuss a simple sequential test procedure (not an s.p.r.t.) whose properties may be determined exactly.

CHAPTER I

SEQUENTIAL SAMPLE SIZE TEST

1.1 Introduction

In certain "life-testing" problems, observations $x_1, x_2, \ldots, x_r$ are made in time until a predetermined number r have been obtained from a total sample of n. Assuming that the sampling is random, we then have available the first r order statistics from a sample of size n. From this information an inference is made concerning the distribution from which the observations have been taken. In this chapter we will be concerned with the case when the population distribution is known but the sample size is unknown. Johnson [17] has given the properties of the fixed sample size test based upon the likelihood ratio for discriminating between two values of n. He has also given the maximum likelihood estimate of n. In the following section, the properties of the s.p.r.t. discriminating between two values of n will be considered. One reason for being concerned with this problem other than for its own sake, is that although the observations are neither independent nor identically distributed, the properties of this s.p.r.t. can be obtained exactly without resorting to quadrature. This will then provide a model to which the various bounds, approximations and conjectures that are to appear in the following chapters may be applied.

1.2 Properties of the sequential sample size test (s.s.s.t.)

Suppose we have a random sample of size n from an absolutely continuous distribution G, with g denoting the probability density function associated with G. Let $x_1, \ldots, x_n$ represent the n observations after they have been ordered such that $x_1 \leq \cdots \leq x_n$. Suppose that we wish to choose between two possible values of n by means of an s.p.r.t. discriminating between

$$(1.1) \qquad H_o: \; n = n_o \quad \text{and} \quad H_1: \; n = n_1$$

where $n_o < n_1$. Letting $f(x_1, \ldots, x_r | n)$ denote the joint density of $x_1, \ldots, x_r$ we have

$$f(x_1,\ldots,x_k|n) = \frac{n\,!}{(n-k)!}\left[\prod_{i=1}^{k} g(x_i)\right]\left[1 - G(x_k)\right]^{n-k} \text{ for } k=1,\ldots, n$$

Define

$$R(k) = \frac{f(x_1,\ldots,x_k|n_1)}{f(x_1,\ldots,x_k|n_o)} = \frac{n_1!(n_o-k)!}{n_o!(n_1-k)!}\left[1-G(x_k)\right]^{n_1-n_o} \quad \text{for } k=1,\ldots, n_o$$

and

$$R(n_o + 1) = \begin{cases} A & \text{if } n > n_o \\ B & \text{if } n = n_o \end{cases}$$

and then it follows that Wald's s.p.r.t. of (1.1) is to continue sampling if $B < R(k) < A$ and to stop if either of these inequalities is violated, at which time H_o is accepted if $R(k) \leq B$ while

H_1 is accepted if $R(k) \geq A$. If sampling proceeds to the $n_o + 1$ stage and x_{n_o+1} does not exist, then H_o is true. Thus the procedure will be logical as long as we reject H_o if x_{n_o+1} is observed and accept H_o if x_{n_o+1} does not exist. In this respect there is considerable freedom in defining $R(n_o+1)$. It should also be noted that at each stage the sequential procedure depends solely on the most recently observed variable and the stage index.

For purposes of determing the properties of the s.s.s.t., we shall restrict attention to the family of hypotheses with $n_o \leq n \leq n_1$. Suppose we are able to obtain

$$(1.2)\quad p_n(j) = P\left[R(j) \leq B \text{ and } B < R(i) < A \text{ for } i=1,\dots,j-1 \,\middle|\, n\right]$$

$$q_n(j) = P\left[R(j) \geq A \text{ and } B < R(i) < A \text{ for } i=1,\dots,j-1 \,\middle|\, n\right]$$

for $j = 1, \dots, n_o$; then the properties of the test are known, since

$$P[\text{accept } H_o | n] = \begin{cases} \sum_{j=1}^{n_o} p_n(j) & \text{if } n > n_o \\ 1 - \sum_{j=1}^{n_o} q_n(j) & \text{if } n = n_o \end{cases}$$

$$P\left[N = j | n\right] = \begin{cases} p_n(j) + q_n(j) & \text{if } j \leq n_o \\ 1 - \sum_{i=1}^{n_o} p_n(i) - \sum_{i=1}^{n_o} q_n(i) & \text{if } j = n_o + 1 \\ 0 & \text{if } j > n_o + 1 . \end{cases}$$

Define

$$u_k = G(x_k)$$

$$A(k) = \frac{n_1!\ (n_o - k)!}{n_o!\ (n_1 - k)!}$$

$$b_i = \max\left\{ 0,\ 1 - \left[\frac{B}{A(i)} \right]^{\frac{1}{n_1 - n_o}} \right\}$$

$$a_i = \max\left\{ 0,\ 1 - \left[\frac{A}{A(i)} \right]^{\frac{1}{n_1 - n_o}} \right\}$$

and observe that since $A(i+1) \geq A(i)$, we have $1 \geq b_i \geq a_i \geq 0$, $b_{i+1} \geq b_i$ and $a_{i+1} \geq a_i$. Also let

$$B_i(t) = \int_{a_i}^{t} \int_{a_{i-1}}^{b_{i-1}} \cdots \int_{a_1}^{b_1} c(u_1, \ldots, u_i)\ du_1 \ldots du_i$$

where

$$c(u_1, \ldots, u_i) = \begin{cases} 1 & \text{if } o \leq u_1 \leq \cdots \leq u_i \leq 1 \\ o & \text{otherwise} \end{cases}$$

and notice that $B_i(t)$ is a polynomial in t of degree at most i. Now

$$p_n(j) = P\left\{ A(j)[\ 1-u_j]^{n_1-n_o} \leq B \text{ and } B < A(i)[\ 1-u_i]^{n_1-n_o} < A \text{ for } i = 1, \ldots, j-1 | n \right\} =$$

$$P \left\{ b_j \leq u_j \leq 1 \text{ and } a_i < u_i < b_i \text{ for } i = 1,\ \ldots,\ j-1 |\ n \right\} .$$

Since u_k is the $k\underline{th}$ order statistic of a sample of size n from a (0,1) uniform distribution, we have

$$p_n(j) = \frac{n!}{(n-j)!} \int_{b_j}^{1} (1-t)^{n-j} B_{j-1} (\min (b_{j-1}, t)) \, dt$$

$$= \frac{n!}{(n-j+1)!} (1 - b_j)^{n-j+1} B_{j-1} (b_{j-1})$$

and

$$p_n(1) = (1 - b_1)^n .$$

Similarly,

$$q_n(j) = P\left\{a_j \geq u_j \geq a_{j-1} \text{ and } a_i < u_i < b_i \text{ for } i=1, \ldots, j-1 \middle| n\right\}$$

$$= \frac{n!}{(n-j)!} \int_{a_{j-1}}^{a_j} (1-t)^{n-j} B_{j-1} (\min(b_{j-1}, t)) \, dt$$

$$= \begin{cases} \dfrac{n!}{(n-j)!} \displaystyle\int_{a_{j-1}}^{a_j} (1-t)^{n-j} B_{j-1}(t) \, dt & \text{for } b_{j-1} \geq a_j \\[2ex] \dfrac{n!}{(n-j)!} \displaystyle\int_{a_{j-1}}^{b_{j-1}} (1-t)^{n-j} B_{j-1}(t) \, dt + \dfrac{n!}{(n-j+1)!} \Big[(1-b_{j-1})^{n-j+1} & \\[2ex] \qquad - (1-a_j)^{n-j+1} \Big] B_{j-1}(b_{j-1}) & \text{for } b_{j-1} \leq a_j . \end{cases}$$

Now since $B_i(t)$ is a polynomial, the above expressions can easily be used to obtain $p_n(j)$ and $q_n(j)$. Although there does not appear to be a simple expression for $B_i(t)$, it can be obtained recursively as follows:

$$B_1(t) = t - a_1$$

$$B_{i+1}(t) = (t - a_{i+1})\, B_i(b_i) \qquad \text{if } b_i \leq a_{i+1}$$

$$B_{i+1}(t) = \begin{cases} \displaystyle\int_{a_{i+1}}^{t} B_i(u)\,du & \text{if } b_i \geq a_{i+1} \text{ and } a_{i+1} \leq t \leq b_i \\[2ex] \displaystyle\int_{a_{i+1}}^{b_i} B_i(u)\,du + (t-b_i)\,B_i(b_i) & \text{if } b_i \geq a_{i+1} \text{ and } b_i \leq t\,. \end{cases}$$

For computing purposes the following algorithm simplifies the determination of $B_i(t)$. Let $B(k,i,j)$ be the coefficient of t^{j-1} in $B_i(t)$ for $b_{k-1} \leq t \leq b_k$. Then, since $B_1(t) = t-a_1$, we have $B(1,1,1)=-a_1$ and $B(1,1,2) = 1$. Now,

(a) if $b_i \leq a_{i+1}$ then

$$B(k,i+1,\cdot) = 0 \qquad \text{for } k \leq i$$

$$B(i+1,i+1,1) = -\,a_{i+1} \sum_{p=1}^{i+1} B(i,i,p)\, b_i^{\,p-1}$$

$$B(i+1,i+1,2) = \sum_{p=1}^{i+1} B(i,i,p)\, b_i^{\,p-1}$$

$$B(i+1,i+1,j) = 0 \qquad \text{for } j > 2$$

(b) if $b_i \geq a_{i+1}$ and $k \leq i$ then

$$B(k,i+1,j) = \frac{1}{j-1} B(k,i,j-1) \quad \text{for } j > 1$$

$$B(k,i+1,1) = \sum_{n=1}^{k-i_o-1} \sum_{p=1}^{i+1} \frac{1}{p} B(i_o+n,i,p)(b_{i_o+n}^p - b_{i_o+n-1}^p)$$

$$+ \sum_{p=1}^{i+1} \frac{1}{p} B(i_o,i,p)(b_{i_o}^p - a_{i+1}^p) - \sum_{p=1}^{i+1} \frac{1}{p} B(k,i,p)\, b_{k-1}^p$$

where i_o is defined as the smallest j such that $b_j \geq a_{i+1}$.

(c) if $b_i \geq a_{i+1}$ and $k = i+1$ then

$$B(i+1,i+1,j) = o \quad \text{for } j > 2$$

$$B(i+1,i+1,1) = \sum_{p=1}^{i+2} B(i,i+1,p)\, b_i^{p-1} - \sum_{p=1}^{i+1} B(i,i,p)\, b_i^p$$

$$B(i+1,i+1,2) = \sum_{p=1}^{i+1} B(i,i,p)\, b_i^{p-1} \quad .$$

The following comparison was made with the fixed sample results to be found in table 2 on page 62 of Johnson [17]. The values of A and B were varied so that the error probabilities agree with those found by Johnson.

TABLE 1.1

r	n_o	n_1	$\alpha = \beta$	$\mathcal{E}_{n_o}(N)$	$\mathcal{E}_{n_1}(N)$
5	10	15	.294	3.23	3.73
5	10	20	.185	2.98	3.81
5	10	25	.123	2.78	3.83
5	15	20	.360	3.32	3.68
5	15	25	.266	3.13	3.75
5	20	15	.394	3.38	3.65
10	15	20	.256	6.34	7.33
10	15	25	.134	5.79	7.43
10	20	25	.326	6.35	7.08
15	20	25	.238	9.82	11.18

r is the fixed sample size and α, β are the type I, II errors, respectively.

The s.s.s.t. has the interesting feature that although the sampling rule depends upon the distribution G, the properties of the test do not. Additional tables of these properties are given in section 1.3. One additional feature of the test which is necessary for future reference will be given before proceeding. That is, if we define $u_o = 0$, then

$$z_k = \log R(k) - \log R(k-1) =$$

$$= \log \frac{n_1-k+1}{n_o-k+1} + (n_1 - n_o)\left[\log(1-u_k) - \log(1-u_{k-1})\right]$$

$$\text{for } k = 1, \ldots, n_o \; .$$

If we set

$$v_i = \frac{1 - u_i}{1 - u_{i-1}} \qquad \text{for } i = 1, \ldots, n \quad ,$$

then the joint density of $(v_1, \ldots, v_n)$ is

$$n\,! \prod_{i=1}^{n} v_i^{n-i} \qquad \text{for } o \leq v_i \leq 1 \text{ and } i = 1, \ldots, n \quad .$$

Hence, $v_1, \ldots, v_n$ are independent and the density of v_i is

$$(n - i + 1)\, v_i^{n-i} \qquad \text{for } o \leq v_i \leq 1 \quad .$$

Since z_k may be written as

$$\log \frac{n_1 - i + 1}{n_o - i + 1} + (n_1 - n_o) \log v_i \quad ,$$

we have $z_1, \ldots, z_{n_o}$ are independent and the distribution function F_i of z_i is

$$F_i(z_i) = \begin{cases} \exp\,[\, c_i(z_i - a_i)] & \text{for } z_i \leq a_i \\ 1 & \text{for } z_i > a_i \end{cases}$$

where

$$c_i = \frac{n-i+1}{n_1 - n_o} \text{ and } a_i = \log \frac{n_1 - i + 1}{n_o - i + 1} \,.$$

Therefore, by redefining $R(n_o+1)$, we have a sequential test with $z_1, z_2, \ldots$ independent while $x_1, x_2, \ldots$ are dependent.

1.3 Data

In this section, the o.c. function and a.s.n. are given in tablesfor various pairs of n_o and n_1. All tables are for $\alpha = \beta = .05$ and the values of A and B which produce these errors are given.

TABLE 1.2

$n_o=5$ $n_1=10$ B=.1898 A=7.881

n	o.c	a.s.n.	n	o.c.	a.s.n.
5	.950	3.763	8	.175	4.915
6	.526	4.387	9	.094	4.918
7	.313	4.762	10	.050	4.839

TABLE 1.3

$n_o=5$ $n_1=15$ B=.2150 A=9.4216

n	o.c.	a.s.n	n	o.c.	a.s.n.
5	.950	2.643	11	.197	4.057
6	.795	3.077	12	.140	4.041
7	.647	3.453	13	.100	3.995
8	.502	3.738	14	.071	3.931
9	.375	3.924	15	.050	3.858
10	.274	4.024			

TABLE 1.4

$n_o=5$ $n_1=20$ B=.2574 A=10.0055

n	o.c.	a.s.n.	n	o.c.	a.s.n.
5	.950	2.110	13	.253	3.455
6	.875	2.398	14	.201	3.456
7	.786	2.673	15	.160	3.437
8	.684	2.916	16	.127	3.404
9	.579	3.114	17	.100	3.363
10	.481	3.264	18	.079	3.315
11	.392	3.367	19	.063	3.265
12	.316	3.428	20	.050	3.214

TABLE 1.5

$n_o=10$ $n_1=15$ B=.14918 A=9.304

n	o.c.	a.s.n.	n	o.c.	a.s.n.
10	.950	7.569	13	.188	9.305
11	.568	8.489	14	.098	9.286
12	.342	9.070	15	.050	9.121

TABLE 1.6

$n_o=10$ $n_1=20$ B =.14845 A=11.518

n	o.c.	a.s.n.	n	o.c.	a.s.n.
10	.950	5.126	16	.218	7.532
11	.837	5.813	17	.152	7.487
12	.704	6.442	18	.105	7.376
13	.358	6.946	19	.073	7.226
14	.421	7.291	20	.050	7.056
15	.307	7.478			

TABLE 1.7

$n_o=10$ $n_1=25$ B=.16529 A=11.735

n	o.c.	a.s.n.	n	o.c.	a.s.n.
10	.950	3.829	18	.286	6.209
11	.901	4.286	19	.225	6.209
12	.831	4.741	20	.176	6.166
13	.742	5.165	21	.137	6.092
14	.643	5.531	22	.107	5.996
15	.541	5.820	23	.083	5.888
16	.445	6.026	24	.064	5.774
17	.359	6.152	25	.050	5.658

TABLE 1.8

$n_o=15$ $n_1=20$ B=.13533 A=9.9515

n	o.c.	a.s.n.	n	o.c.	a.s.n.
15	.950	11.779	18	.194	13.894
16	.584	12.891	19	.100	13.855
17	.354	13.606	20	.050	13.618

TABLE 1.9

$n_o=15$ $n_1=25$ B=.12595 A=12.4578

n	o.c.	a.s.n.	n	o.c.	a.s.n.
15	.950	8.097	21	.229	11.303
16	.852	8.990	22	.159	11.229
17	.727	9.819	23	.109	11.053
18	.583	10.497	24	.074	10.806
19	.443	10.970	25	.050	10.496
20	.324	11.230			

TABLE 1.10

$n_o=20$ $n_1=25$ B=.12835 A=10.3237

n	o.c.	a.s.n.	n	o.c.	a.s.n.
20	.950	16.210	23	.198	18.618
21	.593	17.461	24	.101	18.686
22	.361	18.274	25	.050	19.002

CHAPTER II

SAMPLE SIZE

2.1 Introduction

In this chapter we will consider extending the usual bounds for $\mathcal{E}(N)$ of a sequential test procedure for discriminating between two simple hypotheses to problems in which the assumption of $z_1, z_2, \ldots$ being independent and identically distributed has been relaxed. Also we shall look at bounds for the higher moments of N. In section 2.2, there will be some discussion concerning a conjectured approximation to $\mathcal{E}(N)$. Before proceeding, however, we shall need three forms of Wald's equation which have been proved in the literature. They are given as follows:

Lemma 2.1 Let $x_1, x_2, \ldots$ be a sequence of random variables, $\mu_1, \mu_2, \ldots$ a sequence of constants and N a stopping variable[1] depending upon $x_1, x_2, \ldots$. If the conditions

(a) $|\mu_i| \leq C < \infty$

(b) $\mathcal{E}|x_i| < \infty$

(c) $\mathcal{E}(x_{i+1} - \mu_{i+1} | x_1, \ldots, x_i) = 0$ and $\mathcal{E}(x_1 - \mu_1) = 0$

(d) $\mathcal{E}(N) < \infty$

(e) $\mathcal{E}\left(\sum_{j=1}^{N} |x_j| \right) < \infty$

[1] A stopping variable is defined to be a random variable (N) with positive integer values such that the event N=n depends only on $x_1, \ldots, x_n$.

are satisfied for all i, then

$$\mathcal{E}\left(\sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{\infty} \mu_i \, P\,[\,N \geq i\,] \;.$$

Proof: By lemma 2 of Chow, Robbins and Teicher [10], we have

$\mathcal{E}\left(\sum_{i=1}^{N} (x_i - \mu_i)\right) = o$. Therefore,

$$\mathcal{E}\left(\sum_{i=1}^{N} x_i\right) = \mathcal{E}\left(\sum_{i=1}^{N} \mu_i\right) = \sum_{i=1}^{\infty} \mu_i \, P\,[N \geq i\,] \;.$$

<u>Corollary 2.1.1</u> Let $x_1, x_2, \ldots$ be a sequence of independent random variables and N a stopping variable. Then, if

(a) $\mathcal{E}\,|x_i| \leq C < \infty$

(b) $\mathcal{E}(N) < \infty$

are satisfied for all i, it follows that

$$\mathcal{E}\left(\sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{\infty} \mathcal{E}(x_i) \, P\,[\,N \geq i\,] \;.$$

Proof: The result follows from either lemma 2.1 or the proof of theorem 1 by Johnson [16].

<u>Lemma 2.2</u> Let $x_1, x_2, \ldots$ be a sequence of independent random variables and N a stopping variable. If the following conditions,

(a) $\mathcal{E}\, x_i = o$

(b) $\mathcal{E}\,(x_i^2) \leq C < \infty$

(c) $\mathcal{E}\,(N) < \infty$

are satisfied for all i, then

$$\mathcal{E}\left(\sum_{i=1}^{N} x_i\right) = o$$

and

$$\mathcal{E}\left(\sum_{i=1}^{N} x_i\right)^2 = \sum_{i=1}^{\infty} \mathcal{E}\,(x_i^2)\; P\,[\,N \geq i\,] \;.$$

Proof: This result is a special case of theorem 2 by Chow, Robbins and Teicher [10].

2.2 A conjectured approximation to the a.s.n.

Bhate [7] has conjectured that for an s.p.r.t., $\mathcal{E}(N)$ may be obtained approximately as the solution of

$$(2.1) \qquad \mathcal{E}(Z_m) = L\,b + (1 - L)\,a$$

for m, where L is the probability of accepting H_o. For the case where Z_n is a sum of independent and identically distributed random variables, (2.1) is simply Wald's formula for approximating $\mathcal{E}(N)$. In the more general situation it appears to be a natural extension of Wald's formula.

In some cases of interest, for example, the sequential t test, the calculation of the left hand side of (2.1) is quite laborious. So, Ray [22] has proposed a further conjecture. He suggests that the left hand side of (2.1) be replaced by $\log \ell\,[\,\mathcal{E}(t_m)]$ where $\log \ell\,(t_N) = Z_N$.

By way of illustration, both approximations will be applied to the s.s.s.t. of chapter 1. From section 1.2 we have for $n_1 \geq n \geq n_o \geq k \geq 1$

$$(2.2) \qquad \mathcal{E}_n\,[Z_k] = \log A(k) + (n_1 - n_o)\,\mathcal{E}_n\,[\log\,(1-u_k)\,]$$

Since u_k is the $k\underline{th}$ order statistic of a sample of size n from a $[0,1]$ uniform distribution, we have

$$(2.3) \qquad \mathcal{E}_n\,[\log\,(1-u_k)\,] = \int_o^1 \log\,(1-u)\,\frac{n\,!}{(k-1)!\,(n-k)!}\,u^{k-1}(1-u)^{n-k}\,du.$$

Define

$$I\,(k) = -\int_o^1 (1-x)^{n-k}\,x^{k-1}\,\log\,(1-x)\,dx$$

and by integrating by parts we obtain for $k > 1$

$$I(k) = \frac{(k-1)!\,(n-k)\,!}{n\,!\,(n-k+1)} + \frac{k-1}{n-k+1}\,I(k-1)$$

$$= c(k) + d(k)\,I(k-1),$$

let us say.

Now since

$$I(1) = -\int_o^1 (1-x)^{n-1}\,\log(1-x)dx = \int_o^\infty xe^{-nx}\,dx = \frac{1}{n^2} = c(1)$$

we have for $k \geq 1$

$$I(k) = c(k) + \sum_{i=1}^{k-1} c(i) \prod_{j=i+1}^{k} d(j)$$

$$= \frac{(k-1)!\ (n-k)!}{n\ !\ (n-k+1)} + \sum_{i=1}^{k-1} \frac{(k-1)!\ (n-k)!}{n\ !\ (n-i+1)}$$

$$= \frac{(k-1)!\ (n-k)!}{n\ !} \sum_{i=1}^{k} \frac{1}{n-i+1} \quad .$$

Therefore, by (2.2) and (2.3), $\mathcal{E}_n(Z_k)$ is equal to

$$\text{(2.4)} \qquad \log A(k) + (n_1 - n_o) \sum_{i=1}^{k} \frac{1}{n-i+1} \quad .$$

For Ray's approximation we replace $\mathcal{E}_n[Z_k]$ by

$$\log A(k) + (n_1 - n_o) \log (1 - \mathcal{E}_n(u_k)) ,$$

which is equal to

$$\text{(2.5)} \qquad \log A(k) + (n_1 - n_o) \log (1 - \frac{k}{n+1}) .$$

Now by equating (2.4) and (2.5) to the previously determined values of L b + (1 - L) a and solving for k, we arrive at the conjectured values of $\mathcal{E}_n(N)$.

TABLE 2.1

$n_o = 10$ $n_1 = 25$ $\alpha = \beta = .05$

n	$\mathcal{E}_n(N)$	Bhate	Ray
10	3.83	2.41	2.67
11	4.29	2.69*	3.02*
12	4.74	2.83*	3.23*
13	5.17	2.65*	3.13*
14	5.53	1.71*	2.19*
15	5.82	10.17	9.72
16	6.03	9.43	8.05
17	6.15	8.01	7.06
18	6.21	7.20	6.55
19	6.21	6.72	6.26
20	6.17	6.36	6.04
21	6.09	6.09	5.82
22	6.00	5.86	5.63
23	5.89	5.65	5.46
24	5.77	5.46	5.30
25	5.66	5.29	5.15

Note: For those values which are followed by an asterisk, there are two solutions; the one which is given and another which falls between 10 and 11.

TABLE 2.2

$\alpha = \beta = .05$

		$n=n_o$		$n=n_1$	
n_o	n_1	$\mathcal{E}_n(N)$	Bhate	$\mathcal{E}_n(N)$	Bhate
5	10	3.76	3.04	4.84	4.68
5	15	2.64	1.39	3.86	3.44
5	20	2.11	0.77	3.21	2.77
10	15	7.57	7.32	9.12	9.22
10	20	5.13	4.06	7.06	6.89
10	25	3.83	2.41	5.66	5.29
15	20	11.78	11.97	13.62	13.94
15	25	8.10	7.37	10.50	10.67
20	25	16.21	16.75	19.00	18.76

Table 2.1 is typical of the cases which were considered. From the values it appears that the approximation cannot be relied on for all n. Siskind [25] considered both of these approximations in connection with the two-sided sequential t test. He found the approximation to be adequate except midway between his hypothesized values.

Bhate's conjecture consists of two approximations. The first is the neglect of overshoot which is usually considered reasonable. In general, however, when $\mathcal{E}(N)$ is small, the overshoot might be quite large. From table 2.2, we observe that the approximation is good for the larger values of $\mathcal{E}(N)$ but poor for several of the small values. Thus, for these small values, we might be justified in explaining away part of the inaccuracy of Bhate's approximation by his neglect of overshoot.

The second approximation in Bhate's conjecture consists of setting $\mathcal{E}(N) = m$ where m is the solution of $\mathcal{E}(Z_m) = \mathcal{E}(Z_N)$. Suppose we restrict our attention to the case when $z_1, z_2, \ldots$ are independent and the conditions of corollary 2.1.1 are satisfied (this is, in effect, the case for the s.s.s.t., since $z_1, \ldots, z_{n_o}$ are independent). Thus, by corollary 2.1, Bhate's second approximation consists of finding an m such that

$$(2.6) \qquad \sum_{i=1}^{m} \mathcal{E}(z_i) \leq \sum_{i=1}^{\infty} \mathcal{E}(z_i)\, P\,[N \geq i\,] \leq \sum_{i=1}^{m+1} \mathcal{E}(z_i) \quad ,$$

then approximating $\mathcal{E}(N)$ by a value, which is found by interpolation, between m and m+1. Hopefully it will then be the case that

$$(2.7) \qquad m \leq \sum_{i=1}^{\infty} P\,[\,N \geq i\,] \leq m + 1 \quad .$$

If (2.7) holds approximately, then we may conclude that the second approximation is reasonable. Now in the s.s.s.t. with $n=n_o=5$, $n_1 = 10$ and $\alpha = \beta = .05$, we have

$$\sum_{i=1}^{\infty} P\,[\,N \geq i\,] = 3.76 \quad .$$

Thus by (2.7), $m = 3$. Now, $m = 3$ implies, by (2.6), that

$$-1.432 \leq \sum_{i=1}^{\infty} \mathcal{E}(z_i)\, P[N \geq i] \leq -2.679;$$

but

$$\sum_{i=1}^{\infty} \mathcal{E}(z_i)\, P[N \geq i] = .1612\, \mathcal{E}(z_6) - 3.114$$

where

$$\mathcal{E}(z_6) \leq o \quad .$$

Therefore in general, (2.6) does not necessarily imply (2.7).

It can also be shown for an s.p.r.t. that $\mathcal{E}_{H_o}(z_i) \leq 0$ and $\mathcal{E}_{H_1}(z_i) \geq 0$, but in general it is not necessarily true that $\mathcal{E}(z_i) \geq o$ for all i or $\mathcal{E}(z_i) \leq o$ for all i. Thus $\sum_{i=1}^{n} \mathcal{E}(z_i)$ need not be either increasing or decreasing in n. Hence there is no assurance that a unique m which satisfies (2.6) exists (this is illustrated in table 2.1 by $n = 11, \ldots, 14$).

If $u \geq \mathcal{E}(z_i) \geq \ell > o$, it then follows from (2.6) that

$$m \ell \leq \sum_{i=1}^{m} \mathcal{E}(z_i) \leq \sum_{i=1}^{\infty} \mathcal{E}(z_i)\ P[N \geq i] \leq \sum_{i=1}^{m+1} \mathcal{E}(z_i) \leq (m+1)\ u$$

and

$$\ell \sum_{i=1}^{\infty} P[N \geq i] \leq \sum_{i=1}^{\infty} \mathcal{E}(z_i)\ P[N \geq i] \leq u \sum_{i=1}^{\infty} P[N \geq i] \quad .$$

Hence we have

$$\frac{\ell}{u}\ m \leq \sum_{i=1}^{\infty} P\ [N \geq i\ ] \leq \frac{u}{\ell}\ (m+1)$$

and a similar expression for $o > u \geq \mathcal{E}(z_i) \geq \ell$. Therefore we may conclude that the second approximation in Bhate's conjecture will be reasonable, in the case of independence, if there is little relative variation among the values of $\mathcal{E}(z_i)$ and they are bounded away from zero. In general, however, it is clear that Bhate's conjecture may often lead to fallacious results when applied to cases other than H_o or H_1.

2.3 Wald's bounds for $\mathcal{E}(N)$

Suppose $z_1, z_2, \ldots$ are independent, with $\mathcal{E}(|z_i|) \leq C < \infty$ and $\mathcal{E}(N) < \infty$. It then follows by Wald's argument (see Mallows [20]) that

$$(2.8) \qquad a(1-L) + (b+\xi')\ L \leq \mathcal{E}(Z_N) \leq (a+\xi)\ (1-L) + bL$$

where

$$L = P\,[\text{ accept } H_o]$$

$$\xi = \sup_i \; \sup_r \; \mathcal{E}[z_i - r \mid z_i \geq r] \quad \text{for } a-b > r > o$$

$$\xi' = \inf_i \; \inf_r \; \mathcal{E}[z_i + r \mid z_i \leq -r] \quad \text{for } a-b > r > o .$$

By corollary 2.1.1 we have

$$\ell\, \mathcal{E}(N) \leq \mathcal{E}(Z_N) \leq u\, \mathcal{E}(N) \tag{2.9}$$

where

$$\ell \leq \mathcal{E}(z_i) \leq u \qquad \text{for all } i.$$

Now combining (2.8) and (2.9) we arrive at the following extension of Wald's bounds given in the introduction:

$$\begin{aligned}
\mathcal{E}(N) &\geq \frac{1}{u}\left[a(1-L) + (b+\xi')L\right] && \text{for } u > o \\
\mathcal{E}(N) &\geq \frac{1}{\ell}\left[(a+\xi)\ (1-L) + b\ L\right] && \text{for } \ell < o \\
\mathcal{E}(N) &\leq \frac{1}{u}\left[\ a(1-L) + (b+\xi')\ L\right] && \text{for } u < o \\
\mathcal{E}(N) &\leq \frac{1}{\ell}\left[(a+\xi)\ (1-L) + b\ L\right] && \text{for } \ell > o \ .
\end{aligned} \tag{2.10}$$

As long as $\ell > o$ or $u < 0$, we will have both an upper and lower bound for $\mathcal{E}(N)$ from (2.10). However, in some instances the lower bound may be trivial. One could also consider using approximative bounds by neglecting overshoot through setting $\xi = \xi' = o$ in (2.10).

In equation (2.9), one will typically set $\ell = \inf_i \mathcal{E}(z_i)$ and $u = \sup_i \mathcal{E}(z_i)$. But if there exists an n such that $P[N \geq n] = o$ and $\mathcal{E}(z_i)$ is increasing or decreasing in i, (2.9) may be improved as follows:

By induction we can show

$$\left(\sum_{i=1}^{m} c_i\right)\left(\sum_{i=1}^{m} d_i\right) - m\sum_{i=1}^{m} c_i d_i = \sum_{m\geq i>j\geq 1} (c_i - c_j)(d_j - d_i) \quad .$$

So, if c_i is increasing and d_i is decreasing, then

$$\frac{1}{m}\left(\sum_{i=1}^{m} c_i\right)\left(\sum_{i=1}^{m} d_i\right) \geq \sum_{i=1}^{m} c_i d_i \quad .$$

Therefore, if $P\,[N \geq n\,] = o$ and if $\mathcal{E}(z_i)$ is increasing, it would imply that

$$\mathcal{E}(z_N) \leq \left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{E}(z_i)\right)\ \mathcal{E}(N) \quad .$$

Similarly, if $\mathcal{E}(z_i)$ is decreasing then

$$\mathcal{E}(z_N) \geq \left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{E}(z_i)\right)\ \mathcal{E}(N) \quad .$$

2.4 Hoeffding's lower bound for $\mathcal{E}(N)$

In two papers, Hoeffding ([14], [15]) has derived lower bounds for the expected sample size of a sequential test procedure. In each case the assumption of independent and identically distributed observations has been made. This assumption may be relaxed somewhat by the use of the versions of Wald's equation given in section 2.1.

Let $x_1, x_2, \ldots$ be a sequence of random variables with $x_1, x_2, \ldots, x_n$ having a joint probability density function $f_{\theta,n}(\underline{x}_n)$ where θ belongs to some interval Ω. Let S be a sequential test for deciding between H_o: $\theta \in \omega_o$, and H_1: $\theta \in \omega_1$, such that

$$P_\theta\ (S \text{ accepts } H_1) \leq \alpha \text{ if } \quad \theta \in \omega_o \subset \Omega$$

$$P_\theta\ (S \text{ accepts } H_o) \leq \beta \text{ if } \quad \theta \in \omega_1 \subset \Omega$$

where ω_o and ω_1 are disjoint, $\alpha > o$ and $\beta > o$. Assume that $P_\theta(S \text{ accepts } H_o) + P_\theta(S \text{ accepts } H_1) = 1$ for $\theta \in \Omega$ and $\alpha + \beta \leq 1$. Define N to be the sample size of the test S. Also define the following variables

$$z_{j1} = \log \frac{f_{\theta,1}(x_1)}{f_{\theta_j,1}(x_1)} \qquad j = o,1$$

$$z_{ji} = \log \frac{f_{\theta,i}(\underline{x}_i)}{f_{\theta_j,i}(\underline{x}_i)} - \log \frac{f_{\theta,i-1}(\underline{x}_{i-1})}{f_{\theta_j,i-1}(\underline{x}_{i-1})} \quad j = o,1$$

and

$$Z_{jn} = \sum_{i=1}^{n} z_{ji} \qquad j = o,1 \ .$$

Now Wald has demonstrated that

$$\mathcal{E}_\theta\left[\log \frac{f_{\theta,N}(\underline{x}_N)}{f_{\theta',N}(\underline{x}_N)}\right] \geq L(\theta) \log \frac{L(\theta)}{L(\theta')} + (1-L(\theta)) \log \frac{1-L(\theta)}{1-L(\theta')}$$

where $L(\theta) = P_\theta$ [S accepts H_o]. Therefore, we may apply Hoeffding's [14] result to obtain

$$(2.11) \qquad c\, \mathcal{E}_\theta\left[Z_{oN}\right] + (1-c)\, \mathcal{E}_\theta\left[Z_{1N}\right] \geq -\log\left[\alpha^c(1-\beta)^{1-c} + (1-\alpha)^c\, \beta^{1-c}\right]$$

for $o < c < 1$, $\theta_o \in \omega_o$ and $\theta_1 \in \omega_1$.

Next we assume that the sequence of random variables $z_{j1}, z_{j2}, \ldots$ satisfies the conditions of corollary 2.1.1, or,more generally, lemma 2.1. Then by defining $\mu_j = \sup_i \mathcal{E}_\theta(z_{ji})$ it follows that

$$(2.12) \qquad \mathcal{E}_\theta\left[Z_{jN}\right] \leq \mu_j\, \mathcal{E}_\theta(N) \qquad j = o,1.$$

Since $\mathcal{E}_\theta(z_{ji}) \geq o$, we may apply (2.11) and (2.12) to obtain

$$(2.13) \qquad \mathcal{E}_\theta(N) \geq \sup_{o<c<1} \frac{-\log\{\alpha^c(1-\beta)^{1-c} + (1-\alpha)^c\, \beta^{1-c}\}}{c\, e_o + (1-c)\, e_1}$$

where $e_i = \inf_{\theta_i \in \omega_i} \mu_i$. Hoeffding [14] has also shown that

$$(2.14) \qquad \mathcal{E}_\theta(N) \geq \frac{\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha) \log \frac{1-\alpha}{\beta}}{e_1} \qquad \text{for} \quad \theta \in \omega_o$$

and

$$(2.15) \qquad \mathcal{E}_{\theta}(N) \geq \frac{\beta \log \frac{\beta}{1-\alpha} + (1-\beta) \log \frac{1-\beta}{\alpha}}{e_o} \quad \text{for } \theta \in \omega_1 .$$

For Hoeffding's [15] second bound, let $x_1, x_2, \ldots$ be a sequence of independent random variables and let the probability density function of x_i be f_i (with respect to a σ-finite measure μ). Consider a nonrandomized test, S, such that the probability of making a wrong decision is less than or equal to α_i, when $f_j = f_{ij}$ for each j ($i = 1,2$). Also let N denote the number of observations required by S. Assume that $\mathcal{E}_o(N) < \infty$ (the zero subscript denotes $f_j = f_{oj}$ for all j) and suppose that $\alpha_1 + \alpha_2 < 1$. Now following Hoeffding's notation and proof, define

$$\zeta_j^i = \int f_{oj} \log \frac{f_{oj}}{f_{ij}} \, d\mu \qquad i = 1,2$$

$$\zeta_j = \max \left[\zeta_j^1, \zeta_j^2 \right]$$

$$\tau_j^2 = \int \left(\log \left(\frac{f_{2j}}{f_{1j}}\right) - \zeta_j^1 + \zeta_j^2 \right)^2 f_{oj} \, d\mu$$

$$Y_j = \log \left(\frac{f_{2j}}{f_{1j}} \right) - \zeta_j^1 + \zeta_j^2$$

and assume that $\tau^2 = \sup_j \tau_j^2$ and $\zeta = \sup_j \zeta_j$ are finite. From Hoeffding's proof we have

$$\alpha_1 + \alpha_2 \geq \exp\left\{ -\mathcal{E}_o\left[\max (Z_{1N}, Z_{2N})\right] -\zeta\, \mathcal{E}_o(N)\right\}$$

where

$$Z_{in} = \sum_{j=1}^{n} \left(\log \frac{f_{oj}}{f_{ij}} - \zeta_j^i\right) \quad i = 1,2 \quad .$$

Since ζ is finite we have, by corollary 2.1.1, $\mathcal{E}_o(Z_{iN}) = o$. Since τ^2 is finite we have by lemma 2.2

$$\mathcal{E}_o \left(\sum_{i=1}^{N} Y_i\right)^2 = \sum_{i=1}^{\infty} \mathcal{E}(Y_i^2)\, P\,[\, N \geq i\,] \quad .$$

Therefore

$$\mathcal{E}_o\left[\max (Z_{1N}, Z_{2N})\right] = \frac{1}{2}\, \mathcal{E}_o\, (Z_{1N} + Z_{2N}) +$$

$$\frac{1}{2}\, \mathcal{E}_o(|\, Z_{1N} - Z_{2N}|) = \frac{1}{2}\, \mathcal{E}_o(|\, Z_{1N} - Z_{2N}|)$$

$$= \frac{1}{2}\, \mathcal{E}_o\left(\left|\sum_{i=1}^{N} Y_i\right|\right) \leq \frac{1}{2}\sqrt{\mathcal{E}_o\left(\sum_{i=1}^{N} Y_i\right)^2} =$$

$$\frac{1}{2}\sqrt{\sum_{j=1}^{\infty} \tau_j^2\, P_o[N \geq j]} \leq \frac{1}{2}\,\tau\sqrt{\sum_{j=1}^{\infty} P_o\,[N \geq j]} = \frac{1}{2}\,\tau\,\sqrt{\mathcal{E}_o(N)} \quad .$$

Thus we arrive at Hoeffding's lower bound,

$$\text{(2.16)} \qquad \mathcal{E}_o(N) \geq \frac{1}{\zeta^2}\left\{\left[\left(\frac{\tau}{4}\right)^2 - \zeta \log (\alpha_1 + \alpha_2)\right]^{\frac{1}{2}} - \frac{\tau}{4}\right\}^2 \quad .$$

2.5 Wald type bounds for $\mathcal{E}(N^2)$

Part of the attractiveness of a s.p.r.t. is that often $\mathcal{E}(N)$ is **smaller** than the sample size required for an equivalent fixed sample size procedure. However, if Var(N) is large compared with $\mathcal{E}(N)$, much of this attractiveness is lost making it desirable to resort to truncated procedures. Therefore, before proceeding with an s.p.r.t. procedure, one should consider the value of $\mathcal{E}(N^2)$.

We shall now obtain bounds on $\mathcal{E}[N^2]$ by employing the same technique which Wald used for obtaining his bounds on $\mathcal{E}(N)$.

To begin, assume that $z_1, z_2, \ldots$ are independent, $\mathcal{E}(N) < \infty$ and $\mathcal{E}(z_i - \mu_i)^2 = \sigma_i^2 \leq C < \infty$ where $\mathcal{E}(z_i) = \mu_i$ for all i. Now we must restrict our attention to two cases, $\ell > o$ and $u < o$ where $\ell \leq \mu_i \leq u$ for all i. By these assumptions and lemma 2.2, we have

$$(2.17) \qquad \mathcal{E}\left(\sum_{i=1}^{N} \sigma_i^2\right) = \mathcal{E}\left[\sum_{i=1}^{N} (z_i - \mu_i)\right]^2 = \mathcal{E}\left(\sum_{i=1}^{N} z_i\right)^2$$

$$- 2\,\mathcal{E}\left[\left(\sum_{i=1}^{N} z_i\right)\left(\sum_{i=1}^{N} \mu_i\right)\right] + \mathcal{E}\left[\sum_{i=1}^{N} \mu_i\right]^2 .$$

We now define:

$$\xi = \sup_i \sup_r \mathcal{E}\left[z_i - r \mid z_i \geq r\right] \quad \text{for } a-b > r > o$$

$$\xi' = \inf_i \inf_r \mathcal{E}\left[z_i + r \mid z_i \leq -r\right] \quad \text{for } a-b > r > o$$

$$\zeta = \sup_i \sup_r \mathcal{E}\left[(z_i - r)^2 \mid z_i \geq r\right] \quad \text{for } a-b > r > o$$

$$\zeta' = \sup_i \sup_r \mathcal{E}\left[(z_i + r)^2 \mid z_i \leq -r \right] \quad \text{for } a - b > r > o$$

$$L_n = P\left[\sum_{i=1}^{N} z_i \leq b \mid N = n \right]$$

$$L = P\left[\sum_{i=1}^{N} z_i \leq b \right]$$

and by Wald's method we have

$$b + \xi' \leq \mathcal{E}\left[\sum_{i=1}^{N} z_i \mid N = n, \sum_{i=1}^{N} z_i \leq b \right] \leq b$$

$$a \leq \mathcal{E}\left[\sum_{i=1}^{N} z_i \mid N = n, \sum_{i=1}^{N} z_i \geq a \right] \leq a + \xi \quad .$$

Therefore,

$$(1-L_n)(a-b-\xi') + b + \xi' \leq \mathcal{E}\left[Z_N \mid N=n \right] \leq L_n\ (b-a-\xi) + a + \xi$$

since

$$\mathcal{E}\left[Z_N \mid N=n \right] = L_n\ \mathcal{E}\left[Z_N \mid N=n, Z_N \leq b \right] + (1-L_n)\ \mathcal{E}\left[Z_N \mid N=n, Z_N \geq a \right] \quad .$$

Now

$$\mathcal{E}\left[(Z_N)\left(\sum_{i=1}^{N} \mu_i \right) \right] = \sum_{n-1}^{\infty} P\left[N=n \right] (\mu_1 + \ldots + \mu_n)\ \mathcal{E}\left[Z_N \mid N=n \right]$$

and thus

$$(b+\xi')\,\mathcal{E}\left[\sum_{i=1}^{N}\mu_i\right] + (a-b-\xi')\sum_{n=1}^{\infty}P\left[N=n\right](\mu_1+\dots+\mu_n)(1-L_n) \leq$$

$$\mathcal{E}\left[(Z_N)\left(\sum_{i=1}^{N}\mu_i\right)\right] \leq (a+\xi)\,\mathcal{E}\left[\sum_{i=1}^{N}\mu_i\right] + (b-a-\xi)\sum_{n=1}^{\infty}P\left[N=n\right](\mu_1+\dots\mu_n)L_n$$

which can be written as

$$(b+\xi')\,\mathcal{E}\left(\sum_{i=1}^{N}\mu_i\right) + (a-b-\xi')(1-L)\,\mathcal{E}\left[\sum_{i=1}^{N}\mu_i \,\middle|\, \text{reject } H_o\right] \leq$$

$$\mathcal{E}\left[Z_N\sum_{i=1}^{N}\mu_i\right] \leq (a+\xi)\,\mathcal{E}\left[\sum_{i=1}^{N}\mu_i\right] + (b-a-\xi)L\,\mathcal{E}\left[\sum_{i=1}^{N}\mu_i \,\middle|\, \text{accept } H_o\right].$$

If $\ell > o$, then

$$(2.18)\quad (b+\xi')u\,\mathcal{E}(N) + (a-b-\xi')\ell\,(1-L)\,\mathcal{E}\left[N \,\middle|\, \text{reject } H_o\right] \leq$$

$$\mathcal{E}\left[Z_N\sum_{i=1}^{N}\mu_i\right] \leq (a+\xi)\,u\,\mathcal{E}(N) + (b-a-\xi)\ell\,L\,\mathcal{E}(N \mid \text{accept } H_o).$$

If bounds for $\mathcal{E}(N \mid \text{reject } H_o)$ and $\mathcal{E}(N \mid \text{accept } H_o)$ are not available, we may replace (2.18) by

$$(2.19)\quad (b+\xi')\,u\,\mathcal{E}(N) + (a-b-\xi')\,\ell\,(1-L)$$

$$\leq \mathcal{E}\left[Z_N\sum_{i=1}^{N}\mu_i\right] \leq (a+\xi)\,u\,\mathcal{E}(N) + (b-a-\xi)\,\ell L.$$

Corresponding to (2.18) and (2.19), we have for the case $u < o$

$$(a+\xi)\ell\, \mathcal{E}(N) + (b-a-\xi)\, u\, L\, \mathcal{E}(N\,|\text{accept } H_o) \leq \mathcal{E}\left[Z_N \sum_{i=1}^{N} \mu_i \right]$$

$$\leq (b+\xi')\ell\, \mathcal{E}(N) + (a-b-\xi')\, u\, (1-L)\, \mathcal{E}(N\,|\text{ reject } H_o)$$

and

$$(a+\xi)\ell\, \mathcal{E}(N) + (b-a-\xi)uL \leq \mathcal{E}\left[Z_N \sum_{i=1}^{N} \mu_i \right] \leq$$

$$(b+\xi')\ell\, \mathcal{E}(N) + (a-b-\xi')\, u\, (1-L) \quad .$$

Again by an argument analogous to that used by Wald [38] we have

$$(2.20) \quad L\, b^2 + (1-L)\, a^2 \leq \mathcal{E}\left[Z_N^{\,2}\right] \leq L\, (b^2 + 2b\,\xi' + \zeta')$$

$$+ (1-L)\, (a^2 + 2\, a\xi + \zeta).$$

By defining $s = \inf_i \sigma_i^2$, $t = \sup_i \sigma_i^2$ and combining (2.17), (2.19) and (2.20) we arrive at the following inequalities for $\mathcal{E}(N^2)$:

For $\ell > o$,

$$(2.21) \qquad \mathcal{E}(N^2) \geq \frac{1}{u^2} \Big[s\, \mathcal{E}(N) - L(b^2 + 2b\,\xi' + \zeta')$$

$$- (1-L)\, (a^2 + 2a\xi + \zeta) + 2(b+\xi')\, u\, \mathcal{E}(N) + 2(a-b-\xi')\ell(1-L) \Big]$$

$$\mathcal{E}(N^2) \leq \frac{1}{\ell^2} \left[t\, \mathcal{E}(N) - L\, b^2 - (1-L)\, a^2 + 2(a+\xi)\, u\, \mathcal{E}(N) + 2(b-a-\xi)\, \ell\, L \right],$$

and similarly for $u < o$

$$(2.22) \qquad \mathcal{E}(N^2) \geq \frac{1}{\ell^2} \Big[s\, \mathcal{E}(N) - L(b^2 + 2b\,\xi' + \zeta')$$

$$- (1-L)(a^2+2a\xi + \zeta) + 2(a+\xi)\ell\, \mathcal{E}(N) + 2\,(b-a-\xi)\, u\, L \Big]$$

$$\mathcal{E}(N^2) \leq \frac{1}{u^2}\Big[t\, \mathcal{E}(N) - L\, b^2 - (1-L)\, a^2 + 2(b+\xi')\ell\, \mathcal{E}(N)$$

$$+ 2(a-b-\xi')\, u\, (1-L)\Big] .$$

In the application of the above bounds, $\mathcal{E}(N)$ and L will be replaced by their appropriate bounds. If possible (2.18) should be used in place of (2.19). It should further be mentioned that the lower bound for $\mathcal{E}(N^2)$ will often be trivial. However, the maximum of the lower bound given above and the square of the best available lower bound for $\mathcal{E}(N)$ can be used as a lower bound for $\mathcal{E}(N^2)$.

As an illustration, consider the s.p.r.t. for the mean of a normal distribution with known variance. More specifically let

$$H_o:\ x_i \sim N(0,1) \text{ and } H_1:\ x_i \sim N(1,1).$$

Then under H_o, $z_i \sim N(-\frac{1}{2}, 1)$. Employing Wald's approximation of a and b for $\alpha = \beta = .05$ we have $-b = a = 2.944$. For this Wald has also obtained formulae for ξ and ξ', which in this case are computed to be 1.64 and -1.0, respectively. Then by (2.22)

$$\mathcal{E}(N^2) \leq 4\Big[\mathcal{E}(N) - (2.944)^2 + 3.944\, \mathcal{E}(N) - 6.888(1-L)\Big] .$$

For this test Baker [5] has obtained experimentally $\mathcal{E}(N) \doteq 7.0$ and $L \doteq .9646$. Inserting these values we find that

$$\mathcal{E}(N^2) \leq 102.79 \text{ and s.d. } (N) \leq 7.33 \quad ,$$

while the corresponding Monte Carolo values Baker arrived at are $\mathcal{E}(N^2) \doteq 70.5$ and s.d. $(N) \doteq 4.64$. In this problem the lower bound for $\mathcal{E}(N^2)$ turns out to be trivial, but using $\mathcal{E}(N) \doteq 7.0$ we have $\mathcal{E}(N^2) \geq 49$.

2.6 General bounds for $\mathcal{E}(N^k)$

Consider any nonrandomized sequential rule such that after the ith observation x_i, if $Z_i(x_1, \ldots x_i) \in C_i$ then another observation is taken, otherwise sampling is discontinued. Assume that N, the sample size, is greater than or equal to one. Let E_i denote the event $Z_i \in C_i$, then

$$P\left[N > n\right] = P\left[\bigcap_{i=1}^{n} E_i\right] = 1-P\left[\bigcup_{i=1}^{n} \bar{E}_i\right] .$$

Now define

$$p_i = P\left[\bar{E}_i\right], \quad p_{ij} = P\left[\bar{E}_i \bar{E}_j\right], \ldots$$

$$s_1 = \sum_{i=1}^{n} p_i \quad , \quad s_2 = \sum_{n\geq i>j\geq 1} p_{ij}, \ldots \qquad .$$

It then follows that

$$P\left[\bigcup_{i=1}^{n} \bar{E}_i\right] = s_1 - s_2 + s_3 - \ldots \pm s_n .$$

Now, by Bonferroni's inequality,

$$s_1 - s_2 + \ldots - s_k \leq P\left[\bigcup_{i=1}^{n} \bar{E}_i\right] \leq s_1 - s_2 + \ldots + s_\ell \quad ,$$

where ℓ is odd and k is even. It is also clear that $P[E_n] \geq P[N > n]$. Now define

$$a_{nk} = \min\left[P[E_n],\ 1-s_1 + s_2 - \dots + s_k\right] ,$$

$$b_{n\ell} = \max\left[\ o,\ \ 1-s_1 + s_2 - \dots - s_\ell\right]$$

and

$$a_{no} = P\left[E_n\ \right] .$$

Hence, since $P\ [\bigcup_{i=1}^{n} \bar{E}_i] = P[\ N \leq n]$, we have bounds on the distribution function of N. (Bhate [8] has also considered using a_{no} as an upper bound for $P[N \leq n\ ]$ when $x_1, x_2, \dots$ are independent.) Since $\mathcal{E}(N^m) = \sum_{i=1}^{\infty} [\ (i+1)^m - i^m]\ P[N > i\ ] + 1$ and

$$a_{nk} \geq P\left[\ N > n\ \right] \geq b_{n\ell}$$

it then follows that

$$(2.23)\quad 1 + \sum_{i=1}^{\infty} \left[(i+1)^m - i^m\right] a_{i\,k_i} \geq \mathcal{E}(N^m) \geq 1 + \sum_{i=1}^{\infty}\left[(i+1)^m - i^m\right]\ b_{i\ell_i}$$

where k_i is even and ℓ_i is odd. In particular,

$$1 + \sum_{i=1}^{\infty} a_{ik_i} \geq \mathcal{E}(N) \geq\ 1 + \ \sum_{i=1}^{\infty} b_{i\ell_i}$$

and

$$1+\sum_{i=1}^{\infty}(2i+1)\, a_{ik_i} \geq \mathcal{E}(N^2) \geq 1+\sum_{i=1}^{\infty}(2i+1)\, b_{i\ell_i} \quad .$$

In many cases, s_i for $i > 1$ will be quite difficult to obtain. So for $k_i = o$ and $\ell_i = 1$ we have

$$1+\sum_{i=1}^{\infty} P\left[Z_i \epsilon\, C_i\right] \geq \mathcal{E}(N) \geq 1+\sum_{i=1}^{\infty} \max\left\{o, 1-\sum_{j=1}^{i} P\left[Z_j \notin C_j\right]\right\}$$

(2.24)

$$1+\sum_{i=1}^{\infty}(2i+1)P\left[Z_i \;\epsilon\; C_i\right] \geq \mathcal{E}(N^2) \geq 1+\sum_{i=1}^{\infty}(2i+1)\left[\max\left[o, 1-\sum_{j=1}^{i} P\left[Z_j \notin C_j\right]\right]\right].$$

If we let n_o denote the smallest value of n such that

$$1-\sum_{j=1}^{n+1} P\left[Z_j \notin C_j\right] \leq 0,$$

it then follows that

$$1+\sum_{i=1}^{\infty} P\left[Z_i \epsilon\, C_i\right] \geq \mathcal{E}(N) \geq -\frac{1}{2}(n_o+1)(n_o-2)+\sum_{i=1}^{n_o}(n_o+1-i)P\left[Z_i \epsilon\, C_i\right]$$

$$1+\sum_{i=1}^{\infty}(2i+1)P\left[Z_i \epsilon\, C_i\right] \geq \mathcal{E}(N^2) \geq 1+\sum_{i=1}^{n_o}(2i+1)\left[1-i+\sum_{j=1}^{i} P\left[Z_j \epsilon C_j\right]\right]$$

and generally

$$1+\sum_{i=1}^{\infty}\left[(i+1)^m-i^m\right] P\left[Z_i \epsilon C_i\right] \geq \mathcal{E}(N^m) \geq 1+\sum_{i=1}^{\infty}\left[(i+1)^m-i^m\right]\left[1-i+\right.$$

$$\sum_{j=1}^{i} P\left[Z_j \epsilon\ C_j\right] \Big] \ .$$

It should be remarked from a pratical point of view that in applications it will be necessary to determine a value of i, j let us say, and a sufficiently small constant K such that

$$\sum_{i=1}^{\infty} \left[(i+1)^m - i^m\right] a_{ik_i} \leq \sum_{i=1}^{j} \left[(i+1)^m - i^m\right] a_{ik_i} + K \ .$$

By way of illustration, consider the normal test of means given at the end of the previous section. Let $z_i \sim N(\theta,1)$ and Φ denote the distribution function of a standard normal variate. Then

$$P\left[Z_m \epsilon\ C_m\right] = P\left[b < Z_m < a\right] = \Phi\left(\frac{a}{\sqrt{m}} - \theta\sqrt{m}\right) - \Phi\left(\frac{b}{\sqrt{m}} - \theta\sqrt{m}\right).$$

If we assume $\theta = 0$, then

$$P\left[Z_m \epsilon\ C_m\right] = \Phi\left(\frac{a}{\sqrt{m}}\right) - \Phi\left(\frac{b}{\sqrt{m}}\right) \geq$$

$$\Phi\left(\frac{c}{\sqrt{m}}\right) - \Phi\left(\frac{-c}{\sqrt{m}}\right) \qquad \text{where } c = \min\,(a,-b) \ .$$

Now it can be shown (see [40]) that for $x \geq o$

$$\text{(2.25)} \qquad \Phi(x) - \Phi(-x) \geq \left(1 - e^{-\frac{x^2}{2}}\right)^{\frac{1}{2}} \ .$$

Thus

$$\sum_{m=n}^{\infty} P\left[Z_m \in C_m\right] \geq \sum_{m=n}^{\infty} \left(1 - e^{-\frac{c^2}{2m}}\right)^{\frac{1}{2}} \geq$$

$$\int_n^{\infty} \left(1 - e^{-\frac{c^2}{2m}}\right)^{\frac{1}{2}} d\, m \geq \int_n^{\infty} \left(1 - e^{-\frac{c^2}{2m}} \right) d\, m \geq$$

$$\int_n^{\infty} \frac{c^2}{c^2+2m} \quad d\, m = \infty \quad .$$

Therefore we cannot apply the upper bound (2.24) at the midpoint between the two hypotheses.

Next, assume $\theta \neq o$ and noting symmetry, restrict attention to $\theta > o$. Let n be an integer such that $\frac{a}{\sqrt{n}} - \theta\sqrt{n} < o$. Then

$$\sum_{m=n+1}^{\infty} P\left[Z_m \in C_m\right] \leq \sum_{m=n+1}^{\infty} \Phi \left(\frac{a}{\sqrt{m}} - \theta \sqrt{m} \right) =$$

$$\sum_{m=n+1}^{\infty} \left(1 - \Phi \left(\theta \sqrt{m} - \frac{a}{\sqrt{m}} \right) \right) \quad .$$

Now by (2.25), $\quad \Phi(x) \geq \frac{1}{2} \left[1 + \left(1 - e^{-\frac{x^2}{2}}\right)^{\frac{1}{2}} \right]$ for $x \geq o$

and thus

$$\sum_{m=n+1}^{\infty} P\left[Z_m \in C_m\right] \leq \sum_{m=n+1}^{\infty} \left(\frac{1}{2} - \frac{1}{2} \left(1 - e^{-\frac{1}{2}\left(\theta \sqrt{m} - \frac{a}{\sqrt{m}} \right)^2} \right)^{\frac{1}{2}} \right)$$

$$\leq \frac{1}{2} \int_n^\infty (1-(1-e^{-\frac{1}{2}(\theta\sqrt{m} - \frac{a}{\sqrt{m}})^2})^{\frac{1}{2}})\, d\, m$$

$$\leq \frac{1}{2} \int_n^\infty e^{-\frac{1}{2}(\theta\sqrt{m} - \frac{a}{\sqrt{m}})^2} d\, m$$

$$\leq \frac{1}{2}\, e^{\theta a} \int_n^\infty e^{-\frac{m\theta^2}{2}} dm = \frac{1}{\theta^2}\, e^{\theta a - \frac{n\theta^2}{2}} .$$

Similarly it follows that

$$\sum_{m=n+1}^{\infty} m\, P\left[Z_m \epsilon\, C_m\right] \leq \frac{1}{2}\, e^{\theta a} \int_n^\infty m\, e^{-\frac{m\theta^2}{2}} d\, m$$

$$= \frac{2}{\theta^4}\, e^{\theta a - \frac{\theta^2 n}{2}} \left(1 + \frac{n\theta^2}{2} \right) .$$

Under H_o, $\theta = -\frac{1}{2}$ and by applying (2.24) we obtain

$$8.29 \geq \mathcal{E}(N) \geq 3.97$$
$$111.68 \geq \mathcal{E}(N^2) \geq 16.74$$

while Baker [5] observed $\mathcal{E}(N) = 7.0$ and $\mathcal{E}(N^2) = 70.5$.

The merit of the bounds given in (2.23) is their general applicability. As an illustration of the case where the bounds given in the previous sections cannot be applied, the two-sided sequential t test (see Rushton ([23], [24])) is considered. Suppose a random variable x is normally distributed with an unknown mean μ and an unknown variance σ^2. We wish to perform an s.p.r.t. discriminating between

$$H_o: \quad \mu = o \quad \text{and} \quad H_1: \ |\ \mu\ | = \sigma \ .$$

By defining $\delta = \left|\frac{\mu}{\sigma}\right|$ these hypotheses become

$$H_o: \ \delta = o \quad \text{and} \quad H_1: \delta = 1 \quad .$$

Now the likelihood ratio for this test may be written as

$$\ell_n(u_n) \ = \ e^{-\frac{n}{2}} \quad M\left(\frac{n}{2}, \frac{1}{2}, \frac{u_n}{2}\right)$$

where M is a confluent hypergeometric function (see Slater [26]) and

$$u_n = \left(\sum_{i=1}^{n} x_i\right)^2 \Big/ \sum_{i=1}^{n} x_i^2 \ .$$

For the boundary values we will use Wald's approximation for the case of $\alpha = \beta = .05$; namely $- b = a = \log 19$. Now $\ell_n(u_n)$ is an increasing function of u_n and Arnold [4] has given values a_n, b_n for $\log\ \ell_n(a_n) = \log 19$ and $\log\ \ell_n(b_n) = -\log 19$. Thus

$$P\left[\log \ell_n(u_n) < \log 19\right] = P\left[u_n < a_n\right]$$

$$= P\left[\frac{nt_n^2}{n-1+t_n^2} < a_n\right] = P\left[t_n^2 < \frac{a_n(n-1)}{n-a_n}\right] =$$

$$P\left[t_n < d_n\right] - P\left[t_n < -d_n\right]$$

where $t_n = \frac{\sqrt{n}\,\bar{X}}{s}$ and $d_n = \left[\frac{a_n(n-1)}{n-a_n}\right]^{\frac{1}{2}}$.

Similarly,

$$P\left[\log \ell_n(u_n) > -\log 19\right] = P\left[t_n > e_n\right] + P\left[t_n < -e_n\right]$$

where $e_n = \left[\frac{b_n(n-1)}{n-b_n}\right]$.

TABLE 2.3

n	e_n	d_n	n	e_n	d_n	n	e_n	d_n
4	-	-	16	1.378	3.136	28	2.162	3.442
5	-	30.085	17	1.456	3.151	29	2.216	3.473
6	.126	4.976	18	1.531	3.170	30	2.269	3.503
7	.404	3.946	19	1.603	3.191	31	2.322	3.534
8	.566	3.552	20	1.673	3.215	32	2.373	3.565
9	.700	3.353	21	1.740	3.240	33	2.423	3.596
10	.818	3.242	22	1.805	3.267	34	2.473	3.627
11	.927	3.178	23	1.869	3.294	35	2.521	3.658
12	1.028	3.143	24	1.930	3.323	36	2.568	3.688
13	1.122	3.126	25	1.990	3.352	37	2.615	3.719
14	1.212	3.121	26	2.049	3.382	38	2.661	3.750
15	1.297	3.125	27	2.106	3.412	39	2.706	3.780

Thus we have

$$P\left[Z_n \in C_n\right] = P\left[t_n < d_n\right] - P\left[t_n < -d_n\right] +$$

$$P\left[\, t_n > e_n \,\right] + P\left[\, t_n < -e_n \,\right] - 1$$

where t_n has a noncentral t distribution with $n - 1$ degrees of freedom and noncentrality parameter $\delta\sqrt{n}$.

If we set $\delta = 0$, then by using tables of the central t distribution we find that

$$\sum_{i=1}^{30} P\left[\, Z_i \in C_i \,\right] = 10.76$$

and

$$\sum_{i=1}^{39} P\left[\, Z_i \in C_i \,\right] \leq 10.92 \ .$$

Now in Siskind's [25] sampling experiment N was truncated at 61 and if we may assume that $P[\, Z_i \in C_i]$ is a decreasing function of i then

$$\sum_{i=1}^{61} P\left[\, Z_i \in C_i \,\right] \leq 11.1 \ .$$

Therefore if N is truncated at 61 we find, from (2.24),

$$12.1 \geq \mathcal{E}(N) \geq 7.67 \ ,$$

while Siskind observed $\mathcal{E}(N) \doteq 9.9$. For the case when N is not truncated, we do not know of an upper bound to $\sum_{i=1}^{\infty} P[Z_i \in C_i]$. However, it is observed that

$$.865 \leq \frac{P[Z_{i+1} \in C_{i+1}]}{P[Z_i \in C_i]} \leq .882 \quad \text{for } i = 11,\ldots,30,$$

and since $P[Z_{31} \in C_{31}] = .02587$, it is probably reasonable to set

$$\sum_{i=32}^{\infty} P[Z_i \in C_i] \doteq \frac{1}{1 - .89}(.02587) = .236 \quad .$$

We then find

$$12.00 \geq \mathcal{E}(N) \geq 7.67 \quad .$$

CHAPTER III

OPERATING CHARACTERISTIC FUNCTION

3.1 Introduction

In this chapter we will be concerned with the operating characteristic function (o. c. function), of an s.p.r.t. For the case when $z_1, z_2, \ldots$ are independent and identically distributed, Wald developed the well-known approximation along with upper and lower bounds for the o.c. function. As in the previous chapter we will begin by applying these particular results to a larger class of s.p.r. tests. Then in section 3.4 we consider conditions under which $P[\text{accept } H_o]$ is increased by replacing each z_i by a stochastically larger variable. Also some discussion is given concerning the relationships between monotone likelihood ratios, stochastic ordering and monotone o.c. functions. In section 3.5 we briefly look at the effect upon $P[\text{accept } H_o]$ after the z_i's have undergone a scale change. Finally we conclude chapter three by applying the method given in section 2.6 in order to obtain upper and lower bounds for $P[\text{accept } H_o]$. However, we shall first give a brief outline of Wald's method of approximating the o.c. function.

Let $z_1, z_2, \ldots$ be independent and identically distributed with common probability density function $f(z)$. Define t_o to be the unique nonzero solution of $\varphi(t)=1$ where $\varphi(t)$ is the moment generating function of z (i.e. $\varphi(t) = \int e^{tz} f(z)dz$). Wald has given

(a) $\mathcal{E}(z)$ exists and $\mathcal{E}(z) \neq 0$

(b) the existence of $\delta > 0$ such that $P[z > \delta] > 0$ and $P[z < -\delta] > 0$

(c) $\varphi(t)$ exists for all real t

as sufficient conditions for the existence of a unique nonzero t_o. Now if we define

$$\tilde{f}(z) = \exp\{ zt_o \} f(z)$$

and assume that $t_o > 0$ then $\tilde{f}$ is a probability density function and

$$P_f[\text{accept } H_o] = \sum_{j=1}^{\infty} P_f[b < z_1 + \dots + z_i < a, \text{ for } i = 1,\dots,j-1$$

$$\text{and } z_1 + \dots + z_j \leq b] = \sum_{j=1}^{\infty} P_f[t_o b < u_1 + \dots + u_i < t_o a , \text{ for}$$

$$i = 1,\dots,j-1 \text{ and } u_1 + \dots + u_j \leq t_o b] \equiv \sum_{j=1}^{\infty} \int_{S_j} \left[\prod_{i=1}^{j} f(z_i) \right] d\underline{z}_j$$

$$\geq \sum_{j=1}^{\infty} \int_{S_j} \left[\exp\{ -t_o b \} \prod_{i=1}^{j} \tilde{f}(z_i) \right] d\underline{z}_j = \exp\{-t_o b\} P_{\tilde{f}}[\text{accept } H_o]$$

where $u_i \equiv \log \left[\dfrac{\tilde{f}(z_i)}{f(z_i)} \right]$.

Thus

(3.1) $\quad P_f[\text{accept } H_o] \geq \exp\{-t_o b\} P_{\tilde{f}}[\text{accept } H_o]$

and similarly

(3.2) $\quad P_f[\text{accept } H_1] \leq \exp\{-t_o a\} P_{\tilde{f}}[\text{accept } H_1]$.

Since we assumed a unique positive t_o, it then follows that $P_f[z = o] \neq 1$. By the definition of $\tilde{f}$ we also have $P_{\tilde{f}}[z = o] \neq 1$ and thus by Stein's Theorem [28],

$$P_f\left[N < \infty\right] = P_{\tilde{f}}\left[N < \infty\right] = 1.$$

Therefore,

$$(3.3)\quad \begin{aligned} P_{\tilde{f}}\left[\text{accept } H_o\right] + P_{\tilde{f}}\left[\text{accept } H_1\right] &= 1 \\ P_f\left[\text{accept } H_o\right] + P_f\left[\text{accept } H_1\right] &= 1 . \end{aligned}$$

Wald's approximation then consists of changing the inequalities of (3.1) and (3.2) to equalities and of solving for $P_{\tilde{f}}\left[\text{accept } H_o\right]$ with the use of (3.3). This process yields

$$(3.4)\qquad P_f\left[\text{accept } H_o\right] \doteq \frac{1-e^{a\,t_o}}{e^{bt_o} - e^{at_o}} ,$$

and the same result is obtained for $t_o < o$. An alternative approach utilizing Wald's fundamental identity is, briefly,

$$(3.5)\qquad P\left[\text{accept } H_o\right] e^{bt_o} + \left[1-P\left[\text{accept } H_o\right]\right] e^{at_o} \doteq$$

$$P\left[\text{accept } H_o\right] \mathcal{E}\left[e^{t_oZ_N} \mid Z_N \leq b\right] + P\left[\text{reject } H_o\right] \mathcal{E}\left[e^{t_oZ_N} \mid Z_N \geq a\right]$$

$$= \mathcal{E}\left[e^{t_oZ_N}\right] = \mathcal{E}\left[e^{t_oZ_N}\left[\varphi(t_o)\right]^{-N}\right] = 1.$$

3.2 Consistency

Let $f(z_1, \ldots, z_n)$ denote the probability density function of $z_1, \ldots, z_n$ and assume that there exists a unique nonzero t_n such that

$$\int e^{t_n Z_n} f(z_1, \ldots, z_n) \; d\underline{z}_n = 1$$

where $Z_n = z_1 + \ldots + z_n$. As in the previous section

(a) $\mathcal{E}(Z_n)$ exists and $\mathcal{E}(Z_n) \neq o$

(b) there exists a $\delta_n > o$ such that

$$P\left[Z_n > \delta_n \right] > o \text{ and } P\left[Z_n < -\delta_n \right] > o$$

(c) $\int e^{tZ_n} f(z_1, \ldots, z_n) \, d\,\underline{z}_n$ exists for all real t

are sufficient conditions for this assumption. Also we shall call $\tilde{f}(z_1, \ldots, z_n)$ the <u>conjugate density</u> associated with $f(z_1, \ldots, z_n)$, where

$$(3.6) \qquad \tilde{f}(z_1, \ldots, z_n) = e^{t_n Z_n} f(z_1, \ldots, z_n).$$

Let $g(x_1, \ldots, x_i)$ denote the density of the first i observed random variables, $x_1, \ldots, x_i$. Since z_i is a function of $x_1, \ldots, x_i$, we shall assume that the density $f(z_1, \ldots, z_i)$ is determined by $g(x_1, \ldots, x_i)$. Also define $\tilde{g}(x_1, \ldots, x_i)$ to be the corresponding conjugate density associated with $g(x_1, \ldots, x_i)$ (i.e. $\tilde{g}(x_1, \ldots, x_n) =$

$e^{t_n Z_n(x_1, \ldots, x_n)} g(x_1, \ldots, x_n))$. The development outlined in section 3.1 could just as well have been given in terms of g and $\tilde{g}$ instead of f and $\tilde{f}$.

It has been stated in the literature, (Bartholomew [6], Siskind [25]), that if $t_n = t$ for all n, then Wald's o.c. formula, (3.4), "holds". At first appearances this seems to be a reasonable extension of Wald's o.c. formula. However, if we examine the development outlined in the previous section, the expression $P_{\tilde{f}}\left[\text{accept } H_0\right]$ appears in equation (3.1). Now for $P_{\tilde{f}}\left[\text{accept } H_0\right]$ to be meaningful, we must have

$$(3.7) \qquad \int \tilde{f}(z_1, \ldots, z_n) \, dz_n = \tilde{f}(z_1, \ldots, z_{n-1}) \text{ for } n=2,3,\ldots \quad .$$

If (3.7) holds, we shall say that $\{\tilde{f}\}$ is a <u>consistent</u> sequence of densities. It should be noted that $\{\tilde{f}\}$ is consistent if and only if $\{\tilde{g}\}$ is consistent. By (3.6), $\{\tilde{f}\}$ will be consistent if $z_1, z_2, \ldots$ are independent and $t_n = t$ but in the dependent case it is not obvious that this is true. As an example where $t_n = t$ for all n but $\{\tilde{f}\}$ is not consistent, consider the following:

<u>Example 3.1</u>

Let $x_1, x_2, \ldots$ be a sequence of Bernoulli random variables with

$$P\left[X_1 = x_1\right] = (\epsilon + \theta)^{x_1} (1 - \epsilon - \theta)^{1-x_1}$$

$$P\left[X_1 = x_1, X_2 = x_2\right] = \epsilon^{x_1 x_2} \mu^{(1-x_1)x_2} \theta^{x_1(1-x_2)} (1-\mu-\epsilon-\theta)^{(1-x_1)(1-x_2)}$$

$$P\left[X_1=x_1, X_2=x_2, \ldots, X_n=x_n\right] = P\left[X_1=x_1, X_2=x_2\right] \prod_{i=3}^{n} \delta^{x_i} (1-\delta)^{1-x_i}$$

Now suppose we perform an s.p.r.t. discriminating between

$$H_i: \ \theta = \theta_i, \ \mu = \mu_i, \ \epsilon = \epsilon_i, \ \delta = \delta_i \qquad i = o,1 \ .$$

If we give the parameters the following values,

$$\theta_o = \frac{1}{12} \quad \mu_o = \frac{1}{6} \quad \epsilon_o = \frac{1}{12} \quad \delta_o = \frac{1}{4}$$

$$\theta_1 = \frac{1}{12} \quad \mu_1 = \frac{1}{6} \quad \epsilon_1 = \frac{1}{6} \quad \delta_1 = \frac{1}{2}$$

$$\theta = \frac{19}{288} \quad \mu = \frac{17}{720} \quad \epsilon = \frac{19}{288} \quad \delta = \frac{5}{32}$$

it can then be shown that

$$\sum_{x_1,\ldots,x_i} \left[\frac{P\,[X_1=x_1,\ \ldots,\ X_i=x_i \mid H_1]}{P\,[X_1=x_1,\ \ldots,\ X_i=x_i \mid H_o]} \right]^2 P\left[X_1=x_1,\ \ldots,\ X_i=x_i\right] = 1$$

for all i. Thus we have $t_n = 2$ for all n, however,

$$\sum_{x_2} \left[\frac{P[X_1=x_1,\ X_2=x_2 \mid H_1]}{P[X_1=x_1,\ X_2=x_2 \mid H_o]} \right]^2 P\left[X_1 = x_1,\ X_2 = x_2\right]$$

$$\neq \left[\frac{P[X_1 = x_1 \mid H_1]}{P[X_1 = x_1 \mid H_o]} \right]^2 P\left[X_1 = x_1 \right] \quad .$$

Now, since the conjugate distributions of the observations $X_1, X_2, \ldots$ are not consistent, the conjugate distributions of the variables $z_1, z_2, \ldots$ are not consistent. Therefore, by extension of Wald's development given in the previous section, it would be valid to state that if $t_n = t$ for all n and the conjugate densities are consistent, then Wald's o.c. formula, (3.4), holds. Some situations in which one of these implies the other are indicated by the following lemmas.

Lemma 3.1 A necessary and sufficient condition for $t_n = t$ $n=1,2, \ldots$ and $\{\tilde{f}\}$ to be consistent is

$$\int e^{tz_n} f(z_n \mid z_1, \ldots, z_{n-1})\, dz_n = 1$$

for $n = 2,3, \ldots$ and

$$\int e^{tz_1} f(z_1)\, dz_1 = 1 \, .$$

Proof. If $t_n = t$ for all n and $\{\tilde{f}\}$ is consistent, then

$$e^{tZ_{n-1}} \left[\int e^{tz_n} f(z_1, \ldots, z_n)\, dz_n - f(z_1, \ldots, z_{n-1} \right] =$$

$$\int e^{tZ_n} f(z_1, \ldots, z_n)\, dz_n - e^{tZ_{n-1}} f(z_1, \ldots, z_{n-1}) = o$$

and thus the necessary part follows.

Suppose

$$\int e^{tz_n} f(z_1, \ldots, z_n) dz_n = f(z_1, \ldots, z_{n-1})$$

and

$$\int e^{tZ_{n-1}} f(z_1, \ldots, z_{n-1})\, dz_{n-1} = 1 \, .$$

Then

$$1 = \int e^{tZ_{n-1}}\left[\int e^{tz_n} f(z_1, \ldots, z_n)\, dz_n\right] dz_{n-1} =$$

$$\int e^{tZ_n} f(z_1, \ldots, z_n)\, dz_n$$

and hence, by induction, it follows that $t_n = t$ for all n. Now

$$\int e^{tZ_n} f(z_1, \ldots, z_n)\, dz_n - e^{tZ_{n-1}} f(z_1, \ldots, z_{n-1})$$

$$= e^{tZ_{n-1}}\left[\int e^{tz_n} f(z_1, \ldots, z_n)\, dz_n - f(z_1, \ldots, z_{n-1})\right] = 0$$

and thus $\{\tilde{f}\}$ is consistent.

Lemma 3.2 If $\{\tilde{f}\}$ is consistent and

$$\int e^{t_n z_n} f(z_n | z_1, \ldots, z_{n-1})\, dz_n = 1$$

for n = 2, 3, ... then $t_n = t_1$ for n = 2, 3, ...

Proof We have

$$0 = \int e^{t_n z_n} f(z_1, \ldots, z_n)\, dz_n - f(z_1, \ldots, z_{n-1})$$

$$= \int e^{t_n Z_n} f(z_1, \ldots, z_n)\, dz_n - e^{t_n Z_{n-1}} f(z_1, \ldots, z_{n-1})$$

$$= e^{t_{n-1} Z_{n-1}} f(z_1, \ldots, z_{n-1}) - e^{t_n Z_{n-1}} f(z_1, \ldots, z_{n-1})$$

and thus $t_n = t_{n-1}$.

Lemma 3.3 If $t_n = t$ for all n and $\{\tilde{f}\}$ is consistent, then

$$\int e^{tZ_n} f_1(z_1) \ldots f_n(z_n)\, d\underline{z}_n = 1$$

for all n.

Proof By lemma 3.1,

$$\int e^{tz_n} f(z_n | z_1, \ldots, z_{n-1})\, dz_n = 1$$

and all n and by taking expectations of both sides we have

$$\int e^{tZ_n} f_n(z_n)\, dz_n = 1$$

for all n.

Lemma 3.4 If $z_1, z_2, \ldots$ are independent, then $\{\tilde{f}\}$ is consistent if and only if $t_n = t_1$ for all n.

Proof Since

$$\int e^{t_n Z_n} f(z_1,\ldots,z_n) dz_n = e^{t_n Z_{n-1}} f(z_1,\ldots,z_{n-1}) \int e^{t_n z_n} f(z_n) dz_n$$

we have that $\{\tilde{f}\}$ is consistent if and only if

$$e^{t_n Z_{n-1}} f(z_1,\ldots,z_{n-1}) \int e^{t_n z_n} f(z_n) dz_n = e^{t_{n-1} Z_{n-1}} f(z_1,\ldots,z_{n-1})$$

for all n. Thus $\{\tilde{f}\}$ is consistent if and only if

$$\int e^{t_n z_n} f(z_n)\, dz_n = e^{Z_{n-1}(t_{n-1}-t_n)}$$

for $n = 2, 3, \ldots$ Now the left hand side of the above equation is constant while the right hand side is not, unless $t_{n-1} - t_n = o$. Therefore $\{\tilde{f}\}$ is consistent if and only if $t_n = t_1$ for all n.

The foregoing lemmas could also be stated in terms of the observations and the densities g and $\tilde{g}$. The statements and proofs are quite similar, although in lemma 3.3 the condition that $x_1, x_2, \ldots$ are independent under both H_o and H_1 needs to be added. These lemmas will be helpful in determining whether or not we are justified in using Wald's o.c. formula for a given problem. Also they may prove to be useful in constructing a sequence of densities, $\{f\}$, such that $t_n = t$ for all n and $\{\tilde{f}\}$ is consistent. It should also be remarked that if under H_o, $x_1, \ldots, x_n$ have $g(x_1, \ldots, x_n)$, for all n, as their joint probability density function, then by lemma 3.1 it follows that $t_n = 1$ for all n and $\{\tilde{g}\}$ is consistent. A similar statement also holds for H_1.

Returning to the proposition of using Wald's o.c. formula without regard to consistency, we should mention that the alternative approach, (3.5), cannot be used, since we do not have a generalized form of Wald's identity available. Now, although the use of Wald's o.c. formula without consistency is not justified, it may nevertheless be a reasonable approximation. With this in mind, we shall give the following example illustrating the relationship between t and $P\left[\text{accept } H_o\right]$.

<u>Example 3.2</u>

In example 3.1 let H denote the values assigned to the parameters which yielded $t_n = 2$ for all n. Let H' represent the following distribution of $X_1, X_2, \ldots$

$$P\left[X_1 = x_1, \ldots, X_n = x_n\right] = \prod_{i=1}^{n} P\left[X_i = x_i\right]$$

for all n.

$$P\left[X_i = x_i\right] = p_i^{x_i}(1 - p_i)^{1-x_i}$$

for all i. It can then be shown that if

$$p_1 = \frac{10^t - 9^t}{15^t - 9^t}, \quad p_2 = \frac{(1-p_1)\left(1 - \left(\frac{7}{8}\right)^t\right)}{(1-p_1)\left(1 - \left(\frac{7}{8}\right)^t\right) + p_1(2^t-1)}$$

and

$$p_i = \frac{3^t - 2^t}{6^t - 2^t} \qquad \text{for } i = 3, 4, \ldots$$

then $t_n = t$ for all n. It can also be readily observed that

$$Z_1 = \begin{cases} \log \frac{9}{10} & \text{if } x_1 = 0 \\ \log \frac{3}{2} & \text{if } x_1 = 1 \end{cases}$$

$$Z_2 = \begin{cases} 0 & \text{if } (x_1, x_2) = (0,1) \text{ or } (1,0) \\ \log \frac{7}{8} & \text{if } (x_1, x_2) = (0,0) \\ \log 2 & \text{if } (x_1, x_2) = (1,1) \end{cases}$$

$$Z_j = Z_2 + \sum_{i=3}^{j} z_i \qquad \text{where } z_i = \begin{cases} \log \frac{2}{3} & \text{if } x_i = o \\ \log 2 & \text{if } x_i = 1 \end{cases} .$$

Suppose we require that $o < a < \log \frac{3}{2}$ and $\log \frac{7}{8} < b < \log \frac{9}{10}$, it then follows that $N \leq 3$ and

$$x_1 = 1 \implies \text{reject } H_o$$

$$x_1 = o,\ x_2 = o \implies \text{accept } H_o$$

$$x_1 = 0,\ x_2 = 1,\ x_3 = 1 \implies \text{reject } H_o$$

$$x_1 = o,\ x_2 = 1,\ x_3 = o \implies \text{accept } H_o .$$

Therefore, $P\left[\text{accept } H_o\right] = P\left[X_1 = o,\ X_2 = o\right] + P\left[X_1=o, X_2=1, X_3=o\right]$ and thus

$$P_H\left[\text{accept } H_o\right] = \frac{3983}{4608} \doteqdot .864 .$$

It can similarly be calculated that if

$$t = 2 \quad \text{then} \quad P_{H'}\left[\text{accept } H_o\right] \doteqdot .822$$

and if

$$t = \frac{11}{5} \quad \text{then} \quad P_{H'}\left[\text{accept } H_o\right] \doteqdot .834 .$$

From this example we can therefore draw the following two conclusions:

(1) If H and H' yield the same constant value of t_n, it does not necessarily follow that $P_H\left[\text{accept } H_o\right] = P_{H'}\left[\text{accept } H_o\right]$.

(2) Suppose H yields $t_n = t$ for all n and H' yields $t_n = t'$ for all n. If $t > t'$ it does not necessarily follow that

$$P_H\left[\text{ accept } H_o\right] \geq P_{H'}\left[\text{ accept } H_o\right] .$$

Since Wald's o.c. formula is a monotonically increasing function of t, example 3.2 and its conclusions should indicate two things; in the first place, what the error involved in neglecting consistency may be, and secondly, the errors in the approximation itself.

We now conclude this section with a conjecture of Bartholomew [6] concerning the o.c. function. Suppose that the joint density of the observed variables $x_1, x_2, \ldots x_n$ is $p(x_1, \ldots, x_n|\theta)$ for all n and that we wish to discriminate between

$$H_o: \ \theta = \theta_o \quad \text{and} \quad H_1: \ \theta = \theta_1$$

by means of an s.p.r.t. Now, the equation for determining t in Wald's o.c. approximation may then be written as

$$\int \left[\frac{p(\underline{x}_n |\ \theta_1)}{p(\underline{x}_n |\ \theta_o)} \right]^t p(\underline{x}_n |\ \theta)\ d\underline{x}_n = 1 .$$

Bartholomew suggests

$$\left[\frac{p(\underline{x}_n |\ \theta_1)}{p(\underline{x}_n |\ \theta_o)} \right]^t$$

be expanded as a bivariate Taylor series in θ_o, θ_1 about the point θ. Then if the terms of third order and higher are neglected and the relationships

$$\mathcal{E}\left[\frac{\partial \log p}{\partial \theta}\right] = o, \quad \mathcal{E}\left[-\frac{\partial^2 \log p}{\partial \theta^2}\right] = \mathcal{E}\left[\frac{\partial \log p}{\partial \theta}\right]^2$$

are used, we find that the equation for t yields

$$t = 1 - \frac{2(\theta - \theta_o)}{\theta_1 - \theta_o} .$$

This may be recognized as the t found in the s.p.r.t. discriminating between two means of a normal variable with known variance. Applying this conjecture to the s.s.s.t. given in table 1.8 of section 1.3 we find

value of n	15	16	17	18	19	20
conjectured value of P_n[accept H_o]	.912	.810	.639	.428	.244	.123
exact value of P_n[accept H_o]	.950	.584	.354	.194	.100	.050

These values serve to illustrate the possible dangers involved with the indiscriminate use of Bartholomew's conjecture. Siskind [25] has also applied the conjecture to the two-sided sequential t test. He found the approximation satisfactory only for values of the parameter which were closer to H_1 than H_o. Siskind went one step further by considering the third order terms in the Taylor expansion. He found that if $\mathcal{E}\left(\frac{\partial^3 \log p}{\partial \theta^3}\right)$ is negligible compared with $\mathcal{E}\left(\frac{\partial \log p}{\partial \theta}\right)^3$, then $t = 1 - \frac{2(\theta - \theta_o)}{\theta_1 - \theta_o}$. However, in many problems such as the sequential t test, the range of θ for which this condition is satisfied is difficult to determine.

3.3 Bounds for the o.c. function

By assigning different values to the parameters in example 3.1, it can be shown that there exist cases where t_n is not independent of n but $\{\tilde{f}\}$ is consistent. It should be noted that by lemma 3.4, we will then necessarily be restricted to $z_1, z_2, \ldots$ being dependent. We shall assume that $\{\tilde{f}\}$ is consistent and that there exists a unique nonzero t_n for all n. Define $u = \sup_n\{t_n\}$, $\ell = \inf_n\{t_n\}$ and suppose that $\ell > o$. Then by following the development given in section 3.1 we have

$$P_f[\text{accept } H_o] = \sum_{j=1}^{\infty} P_f[\, bt_i < U_i < at_i \quad \text{for} \quad i = 1,\ldots,j-1 \quad \text{and}$$

$$U_j \leq bt_j\,] \equiv \sum_{j=1}^{\infty} \int_{S_j} f(z_1,\ldots,z_j)d\underline{z}_j \geq$$

$$\sum_{j=1}^{\infty} \int_{S_j} e^{-bt_j}\, \tilde{f}(z_1,\ldots,z_j)d\underline{z}_j \geq e^{-b\ell} \sum_{j=1}^{\infty} \int_{S_j} \tilde{f}(z_1,\ldots,z_j)d\underline{z}_j =$$

$$e^{-b\ell}\, P_{\tilde{f}}[\text{accept } H_o]$$

where $$U_i \equiv \log \frac{\tilde{f}(z_1,\ldots,z_i)}{f(z_1,\ldots,z_i)} .$$

Thus similar to equations (3.1) and (3.2) we have

$$(3.8) \qquad P_f[\text{accept } H_o] \geq e^{-b\ell}\, P_{\tilde{f}}[\text{accept } H_o]$$

and in a like manner

$$P_f[\text{accept } H_1] \leq e^{-a\ell} P_{\tilde{f}}[\text{accept } H_1] \quad . \tag{3.9}$$

We further assume that

$$P_f[\text{accept } H_o] + P_f[\text{accept } H_1] = 1 \, ,$$

and by noting $P_{\tilde{f}}[\text{accept } H_o] + P_{\tilde{f}}[\text{accept } H_1] \leq 1$, it follows from (3.9) that

$$1 - P_f[\text{accept } H_o] \leq e^{-a\ell} \left\{ 1 - P_{\tilde{f}}[\text{accept } H_o] \right\} \, . \tag{3.10}$$

If we change the inequalities of (3.8) and (3.10) to equalities and solve for $P_f[\text{accept } H_o]$, we obtain

$$P_f[\text{accept } H_o] \doteq \frac{e^{a\ell} - 1}{e^{a\ell} - e^{b\ell}} \qquad \text{for } \ell > o \, . \tag{3.11}$$

In an identical fashion, it may be shown that

$$P_f[\text{accept } H_o] \doteq \frac{1 - e^{au}}{e^{bu} - e^{au}} \qquad \text{for } u < o \, . \tag{3.12}$$

To continue, suppose we assume that $\ell > o$ and we define

$$P_f[U_N \leq c_N \mid c_n, d_n]$$

as the probability of $U_N \leq c_N$ where N is the least value of n such that $c_n < U_n < d_n$ is not satisfied. Since U_i was defined to be $\log \dfrac{\tilde{f}(z_1, \ldots, z_i)}{f(z_1, \ldots, z_i)}$ it follows that

$$P_f[\text{accept } H_o] = P_f[\, U_N \leq bt_N \mid t_n b, \, t_n a \,]$$

and also by the definition of u and ℓ, $\ell b \geq bt_n \geq ub$ and $ua \geq at_n \geq \ell a$. Therefore,

$$P_f\left[U_N \leq ub \mid ub,\ \ell a\right] \leq P_f\left[U_N \leq t_N b \mid t_n b,\ t_n a\right]$$

$$\leq P_f\left[U_N \leq \ell b \mid \ell b,\ ua\right]$$

and hence

$$(3.13)\quad P_f\left[U_N \leq ub \mid ub, \ell a\right] \leq P_f\left[\text{accept } H_o\right] \leq P_f\left[U_N \leq \ell b \mid \ell b, ua\right] \text{ if } \ell > o.$$

Similarly we have

$$(3.14)\quad P_f\left[U_N \geq \ell b \mid ua, \ell b\right] \leq P_f\left[\text{accept } H_o\right] \leq P_f\left[U_N \geq ub \mid \ell a, ub\right] \text{ if } u < o.$$

Now Wald has shown quite generally

$$\alpha \leq e^{-a}$$

and thus

$$P_f\left[U_N \geq ub \mid \ell a,\ ub\right] \leq e^{-ub} \qquad \text{if } u < o$$

$$P_f\left[U_N \leq ub \mid ub,\ \ell a\right] \geq 1 - e^{-\ell a} \qquad \text{if } \ell > o.$$

Also by using the general approximations

$$e^a \doteq \frac{1-\beta}{\alpha}$$

$$e^{b} \div \frac{\beta}{1-\alpha}$$

we obtain from (3.13) and (3.14) the following approximate bounds

$$\frac{e^{\ell a}-1}{e^{\ell a}-e^{ub}} \underset{\cdot}{\overset{\cdot}{\leq}} P_f\left[\text{accept } H_o\right] \underset{\cdot}{\overset{\cdot}{\leq}} \frac{e^{ua}-1}{e^{ua}-\ell^{\ell b}} \quad \text{for } \ell > o$$

(3.15)

$$\frac{1-e^{ua}}{e^{\ell b}-e^{ua}} \underset{\cdot}{\overset{\cdot}{\leq}} P_f\left[\text{accept } H_o\right] \underset{\cdot}{\overset{\cdot}{\leq}} \frac{1-e^{\ell a}}{e^{ub}-e^{\ell a}} \quad \text{for } u < o$$

($\underset{\cdot}{\overset{\cdot}{\leq}}$ will be used to denote an approximate inequality as in the sense here). And as expected, on comparing (3.11), (3.12) and (3.15) find

$$\frac{e^{\ell a}-1}{e^{\ell a}-e^{ub}} \leq \frac{e^{\ell a}-1}{e^{\ell a}-e^{\ell b}} \leq \frac{e^{ua}-1}{e^{ua}-e^{\ell b}} \qquad \text{for } \ell > o$$

and

$$\frac{1-e^{ua}}{e^{\ell b}-e^{ua}} \leq \frac{1-e^{ua}}{e^{ub}-e^{ua}} \leq \frac{1-e^{\ell a}}{e^{ub}-e^{\ell a}} \qquad \text{for } u < o.$$

If we have an upper bound for $P_f\,[\,N \notin G\,]$ where G is a subset of the positive integers, it may be possible to improve the approximate bounds given in (3.15) as follows:

Define

$$u_o = \sup_{n \in G} \{t_n\}$$

$$\ell_o = \inf_{n \in G} \{t_n\}$$

and assume $\ell_o > o$. Then

$$P_f\left[\text{ accept } H_o \,|\, N\epsilon\, G\right] P\left[N \epsilon G\right] \leq P_f\left[U_N \leq \ell_o b \,|\, \ell_o b,\ u_o a, N \in G\right] \times$$

$$P_f\left[N \in G\right] = P_f\left[U_N \leq \ell_o b \mid \ell_o b,\ u_o a\right]$$

$$- P_f\left[U_N \leq \ell_o b \mid \ell_o b,\ u_o a,\ N \in G\right] P_f\left[N \notin G\right] \leq$$

$$P_f\left[U_N \leq \ell_o b \mid \ell_o b,\ u_o a\right]$$

and by using a similar argument for the lower bound, we have

$$P_f\left[U_N \leq u_o b \mid u_o b,\ \ell_o a\right] - P_f\left[N \notin G\right] \leq P_f\left[\text{accept } H_o\right] \leq$$

$$P_f\left[U_N \leq \ell_o b \mid \ell_o b,\ u_o a\right] + P_f\left[N \notin G\right] \qquad \text{for } \ell_o > o.$$

It can also be shown that

$$P_f\left[U_N \geq \ell_o b \mid u_o a,\ \ell_o b\right] - P_f\left[N \notin G\right] \leq P_f\left[\text{accept } H_o\right]$$

$$\leq P_f\left[U_N \geq u_o b \mid \ell_o a,\ u_o b\right] + P_f\left[N \notin G\right] \quad \text{for } u_o < o .$$

By applying the approximations which were used to obtain (3.15), we arrive at

$$\frac{e^{\ell_o a} - 1}{e^{\ell_o a} - e^{u_o b}} - P_f\left[N \notin G\right] \mathrel{\dot{\leq}} P_f\left[\text{accept } H_o\right] \mathrel{\dot{\leq}} \frac{e^{u_o a} - 1}{e^{u_o a} - e^{\ell_o b}} + P_f\left[N \notin G\right] \qquad \text{for } \ell_o > o$$

(3.16)

$$\frac{1 - e^{u_o a}}{e^{\ell_o b} - e^{u_o a}} - P_f\left[N \notin G\right] \mathrel{\dot{\leq}} P_f\left[\text{accept } H_o\right] \mathrel{\dot{\leq}} \frac{1 - e^{\ell_o a}}{e^{u_o b} - e^{\ell_o a}} + P_f\left[N \notin G\right] \text{for } u_o < o.$$

Now consider the case where $z_1, z_2, \ldots$ are independent. Let $\Psi_i(t)$ be the moment generating function of z_i and D' be that part of the t complex plane such that $\Psi_i(t)$ exists for all i. Then following Wald's proof, as pointed out by Mallows [20], we have

$$(3.17) \qquad \mathcal{E}\left[e^{tZ_N} \left[\prod_{i=1}^{N} \Psi_i(t) \right]^{-1} \right] = 1$$

for $t \epsilon D$, where D is a subset of D' for which $|\prod_{i=1}^{n} \Psi_i(t)| \geq 1$ for all n. Blom [9] has further shown that if either

(a) there exists a $\delta > o$ and an $\epsilon > o$ such that

$$P[z_i > \delta] > \epsilon \text{ for all } i$$

or

(b) there exists a $\delta > o$ and an $\epsilon > o$ such that

$$P[z_i < -\delta] > \epsilon \text{ for all } i$$

then

$$\lim_{n \to \infty} n^k P\left[N > n \right] = o \quad \text{for any } k \geq o.$$

Blom has also proved that if (a) and (b) hold and for some real t, $\{\Psi_i(t)\}$ is bounded then (3.17) is valid. Therefore if we assume (3.17) is valid, then it follows that

$$(3.18) \qquad \begin{array}{l} \Psi_i(t) \leq 1 \text{ for all } i \quad \text{implies} \quad 1 \geq \mathcal{E}[e^{tZ_N}] \\ \Psi_i(t) \geq 1 \text{ for all } i \quad \text{implies} \quad 1 \leq \mathcal{E}[e^{tZ_N}] . \end{array}$$

Mallows [20] has shown that for $t > o$

$$\eta(t)\, e^{bt} \leq \mathcal{E}\left[e^{tZ_N} \mid Z_N \leq b \right] \leq e^{bt}$$

(3.19)

$$e^{at} \leq \mathcal{E}\left[e^{tZ_N} \mid Z_N \geq a \right] \leq \delta(t)\, e^{at}$$

where

$$\eta(t) = \inf_i \inf_{o<r<a-b} \mathcal{E}\left[e^{t(z_i+r)} \mid z_i \leq -r \right]$$

$$\delta(t) = \sup_i \sup_{o<r<a-b} \mathcal{E}\left[e^{t(z_i-r)} \mid z_i \geq r \right].$$

Similarly, for $t < o$

$$e^{bt} \leq \mathcal{E}\left[e^{tZ_N} \mid Z_N \leq b \right] \leq \delta'(t)\, e^{bt}$$

(3.20)

$$\eta'(t)\, e^{at} \leq \mathcal{E}\left[e^{tZ_N} \mid Z_N \geq a \right] \leq e^{at}$$

where

$$\eta'(t) = \inf_i \inf_{o<r<a-b} \mathcal{E}\left[e^{t(z_i-r)} \mid z_i \geq r \right]$$

$$\delta'(t) = \sup_i \sup_{o<r<a-b} \mathcal{E}\left[e^{t(z_i+r)} \mid z_i \leq -r \right].$$

Now, if we let L denote P [accept H_o] then by (3.19) and (3.20),

$$\eta(t)e^{bt}L + e^{at}(1-L) \leq \mathcal{E}\left[e^{tZ_N} \right] \leq Le^{bt} + (1-L)\delta(t)e^{at} \qquad \text{if } t > o$$

(3.21)

$$Le^{bt} + \eta'(t)e^{at}(1-L) \leq \mathcal{E}\left[e^{tZ_N} \right] \leq L\delta'(t)e^{bt} + (1-L)e^{at} \qquad \text{if } t < o.$$

If we assume there exists a unique real nonzero t_i such that $\Psi_i(t_i) = 1$, then by defining $u = \sup_i \{t_i\}$ and $\ell = \inf_i \{t_i\}$ we have

$$\begin{array}{lll} u \geq o & \text{implies} & \Psi_i(u) \geq 1 \quad \text{for all } i \\ u \leq o & \text{implies} & \Psi_i(u) \leq 1 \quad \text{for all } i \\ \ell \geq o & \text{implies} & \Psi_i(\ell) \leq 1 \quad \text{for all } i \\ \ell \leq o & \text{implies} & \Psi_i(\ell) \geq 1 \quad \text{for all } i . \end{array}$$

Then by (3.18) and (3.21) it follows that

$$\begin{array}{lll} u \geq o & \text{implies} & 1 \leq Le^{bu} + (1-L)\delta(u)e^{au} \\ u \leq o & \text{implies} & 1 \geq Le^{bu} + (1-L)\eta'(u)e^{au} \\ \ell \geq o & \text{implies} & 1 \geq L\eta(\ell)e^{b\ell} + (1-L)e^{a\ell} \\ \ell \leq o & \text{implies} & 1 \leq L\delta'(\ell)e^{b\ell} + (1-L)e^{a\ell} . \end{array}$$

Now since $\delta(t) \geq 1$, $o \leq \eta(t) \leq 1$ for $t \geq o$ and $\delta'(t) \geq 1$, $o \leq \eta'(t) \leq 1$ for $t \leq o$, it follows that

$$(3.22) \quad \begin{array}{lll} u \geq o & \text{implies} & P[\text{accept } H_o] \leq \dfrac{\delta(u)e^{au} - 1}{\delta(u)e^{au} - e^{bu}} \\ u \leq o & \text{implies} & P[\text{accept } H_o] \leq \dfrac{1 - \eta'(u)e^{au}}{e^{bu} - \eta'(u)e^{au}} \\ \ell \geq o & \text{implies} & P[\text{accept } H_o] \geq \dfrac{e^{a\ell} - 1}{e^{a\ell} - \eta(\ell)e^{b\ell}} \\ \ell \leq o & \text{implies} & P[\text{accept } H_o] \geq \dfrac{1 - e^{a\ell}}{\delta'(\ell)e^{b\ell} - e^{a\ell}} . \end{array}$$

As approximate bounds we may neglect overshoot by setting $\eta = \delta = \eta'$ $= \delta' = 1$.

3.4 Stochastic ordering

In this section we will be concerned with the proof of the following intuitive proposition. Suppose $z_1, z_2, \ldots$ are independently distributed and z_i is stochastically larger (see page 73 of Lehmann [19]) under H than under H' , then

$$P_H[\text{accept } H_o] \geq P_{H'}[\text{accept } H_o] \quad .$$

From this statement it is easy to see how bounds may be determined for $P[\text{accept } H_o]$ in many problems. Simply replace $z_1, z_2, \ldots$ by a sequence of stochastically larger (smaller) random variables for which the probability of acceptance is known or may be approximated by the usual techniques.

Some necessary preliminaries for proving the above proposition are as follows:

Lemma 3.5 Let H and G be Lebesgue measurable functions such that

(a) H is nonincreasing and nonnegative

(b) G is of bounded variation, $G(-\infty) = \lim_{x \to -\infty} G(x)$ exists, and $G(-\infty) \leq G(x)$ for all x .

Then if

$$\int HdG \quad \text{exists, it is nonnegative.}$$

Proof: Define

$$E_i^n = \left\{ x \mid \frac{i-1}{2^n} \leq H(x) < \frac{i}{2^n} \right\} \quad \text{for } i=1,\ldots,m-1$$

$$E_m^n = \left\{ x \mid n \leq H(x) \right\}$$

and

$$S_n(x) = \frac{i-1}{2^n} \quad \text{if} \quad x \in E_i^n \quad \text{for} \quad i=1,\ldots,m$$

where $m = n2^n + 1$. Now by (a), $S_n(x)$ is a step function and $E_i^n = < x_{i+1}, x_i >$ where $x_i \geq x_{i+1}$ and $< >$ denotes some type of interval. Since $\{E_i^n\}$ is a disjoint cover of the real line, we have

$$\int S_n(x)dG(x) = \sum_{i=1}^{m} \frac{i-1}{2^n} \left\{ \text{Variation of } G \text{ over } E_i^n \right\}$$

$$= \frac{1}{2^n} \sum_{i=1}^{m} (i-1)[\, G(y_i) - G(y_{i+1}) \,]$$

where y_i denotes either x_i+ or x_i-. Now

$$\sum_{i=1}^{m} (i-1)[\, G(y_i) - G(y_{i+1}) \,] = \sum_{i=1}^{m} (i-1)G(y_i) - \sum_{i=2}^{m+1} (i-2)G(y_i)$$

$$= \sum_{i=2}^{m} G(y_i) - (m-1)G(y_{m+1}) = \sum_{i=2}^{m} [\, G(y_i) - G(y_{m+1}) \,]$$

which is nonnegative since $G(y_i) \geq G(y_{m+1}) = G(-\infty)$. Therefore, by dominated convergence we have

$$\int HdG = \lim_{n \to \infty} \int S_n dG \geq 0 .$$

<u>Corollary 3.5.1</u> Let H and G be Lebesgue measurable functions such that

(a) H is nondecreasing and nonnegative

(b) G is of bounded variation, $G(\infty) = \lim_{x \to \infty} G(x)$ exists, and $G(\infty) \leq G(x)$ for all x.

Then if

$\int HdG$ exists, it is nonpositive.

Proof: Make the corresponding changes in the proof of lemma 3.5 .

<u>Corollary 3.5.2</u> Let the real valued functions on the real line H, F_1 and F_2 have the following properties:

(a) F_i is a distribution function $i=1,2$

(b) H is nonincreasing

(c) $H(-\infty) = 1$ and $H(\infty) = o$.

Then $F_1(x) \geq F_2(x)$ for all x implies that

$$\int HdF_1 \geq \int HdF_2 \quad .$$

Proof: Let $G = F_1 - F_2$ and apply lemma 3.5 .

<u>Lemma 3.6</u> Let $x_1, x_2, \ldots$ be a sequence of independent random variables with x_i having distribution function F_i' and F_i'' under H' and H'' respectively. If $F_1''(x) \leq F_1'(x)$ for $b \leq x \leq a$ and $F_i'' = F_i'$ for $i=2,3,\ldots$, then

$$\sum_{i=1}^{\infty} P_{H'}\left\{ b < x_1 + \ldots + x_j < a \text{ for } j=1,\ldots,i-1 \text{ and } x_1 + \ldots + x_i \leq b \right\}$$

$$\geq \sum_{i=1}^{\infty} P_{H''}\left\{ b < x_1 + \ldots + x_j < a \text{ for } j=1,\ldots,i-1 \text{ and } x_1 + \ldots + x_i \leq b \right\}.$$

Proof: Define

$$A_1 = \left\{ (x_1,\ldots,x_n) \mid x_1 \leq b \right\}$$

$$A_i = \left\{ (x_1,\ldots,x_n) \mid b < x_1 + \ldots + x_j < a \text{ for } j=1,\ldots i-1 \text{ and } x_1 + \ldots + x_i \leq b \right\} \quad \text{for } i=2,\ldots,n$$

and

$$I(A_i) = \begin{cases} 1 & \text{if } (x_1,\ldots,x_n) \in A_i \\ 0 & \text{if } (x_1,\ldots,x_n) \notin A_i . \end{cases}$$

Let F_i denote the distribution function of x_i for $i=1,\ldots,n$ and define $F(x_2,\ldots,x_n) = F_2(x_2)\ldots F_n(x_n)$. Then we may write

$$\sum_{i=1}^{n} P\left\{ b < x_1 + \ldots + x_j < a \text{ for } j=1,\ldots,i-1 \text{ and } x_1 + \ldots + x_i \leq b \right\}$$

$$= \int_X \int_{X^{n-1}} \left\{ \sum_{i=1}^{n} I(A_i) \right\} dF dF_1 = \int_X \int_{X^{n-1}} H_n dF dF_1$$

where

$$H_n = \sum_{i=1}^{n} I(A_i) .$$

Now H_n is a nonincreasing function of x_1 since if $(x_1,\ldots,x_n) \in A_i$ and $c > 0$ then $(x_1-c, x_2,\ldots,x_n) \in A_j$ for some $j \leq i$. Further we have

$$H_n = 0 \quad \text{if } x_1 \geq a$$

$$H_n = 1 \quad \text{if } x_1 \leq b$$

and since $\int_{X^{n-1}} H_n dF$ exists, it follows that $\int_{X^{n-1}} H_n dF$ is a nonincreasing function of x_1 and

$$\int_{X^{n-1}} H_n dF = 0 \text{ if } x_1 \geq a$$

$$\int_{X^{n-1}} H_n dF = 1 \text{ if } x_1 \leq b .$$

Therefore by collary 3.5.2 we have

$$\int_X \int_{X^{n-1}} H_n dF dF_1' \geq \int_X \int_{X^{n-1}} H_n dF dF_1''$$

and since this is true for all n, the lemma follows.

Theorem 3.7 In an s.p.r.t. suppose that $z_1, z_2, \ldots$ are independent and z_i has distribution function F_i and G_i under H' and H'' respectively. Also assume that either $P_{H'}[N < \infty] = 1$ or $P_{H''}[N < \infty] = 1$. Then $F_i(z) \geq G_i(z)$ for $b-a \leq z \leq a-b$ and $i=1,2,\ldots$ implies that

$$P_{H'}[\text{accept } H_o] \geq P_{H''}[\text{accept } H_o] .$$

Proof: Suppose $P_{H'}[N < \infty] = 1$ and let H_j for $j=2,3,\ldots$ denote the hypothesis that the distribution function of z_i is F_i for $i=1,\ldots,j-2$ and G_i for $i=j-1,j,\ldots$. Since $P_{H'}[N < \infty] = 1$, for any $\epsilon > o$ there exists an n_o such that $P_{H'}[N \geq n_o - 1] < \epsilon$, and by the definition of H_j , $P_{H_{n_o}}[N < n_o - 1]$ $= P_{H'}[N < n_o - 1]$ and therefore $P_{H_{n_o}}[N \geq n_o - 1] = P_{H'}[N \geq n_o - 1]$. Also $P_{H_{n_o}}[\text{accept } H_o | N < n_o - 1] = P_{H'}[\text{accept } H_o | N < n_o - 1]$ and thus we have

$$P_{H_{n_o}}[\text{accept } H_o] - P_{H'}[\text{accept } H_o] \leq P_{H_{n_o}}[\text{accept } H_o \mid N \geq n_o - 1] \times$$
$$P_{H_{n_o}}[N \geq n_o - 1] < \epsilon .$$

Therefore we may conclude that

$$P_{H'}[\text{accept } H_o] \geq P_{H''}[\text{accept } H_o]$$

if for all $j \geq 3$

(3.23) $$P_{H_j}[\text{accept } H_o] \geq P_{H_{j-1}}[\text{accept } H_o] .$$

For $j=3$, (3.23) follows from lemma 3.6 . Next define

$$C_j = \left\{ (z_1,\ldots,z_{j-3}) \mid b < z_1 + \ldots + z_i < a \text{ for } i=1,\ldots,j-3 \right\}$$

and let F denote the joint distribution function of $z_1,\ldots,z_{j-3}$. We may then write

$$P[\text{accept } H_o] = \sum_{i=1}^{j-3} P[\, b < z_1 + \ldots + z_k < a \text{ for } k=1,\ldots i-1 \text{ and}$$

$$z_1 + \ldots + z_i \leq b \,] + \int_{X^{j-3}} I(C_j)\left\{ \sum_{i=o}^{\infty} P[\, b-Z_{j-3} < z_{j-2} + \ldots + z_{j+k-2} < \right.$$

$$\left. a-Z_{j-3} \text{ for } k=o,\ldots,i-1 \text{ and } z_{j-2} + \ldots + z_{j+i-2} \leq b-Z_{j-3} | Z_{j-3}] \right\} dF .$$

Under H_j and H_{j-1} the z_i's have the same distribution except for z_{j-2} . Therefore in the above expression for $P[\text{accept } H_o]$, only the term within the braces, call it T, differs under H_j and H_{j-1} . Now T may be considered as $P[\text{accept } H_o]$ for an s.p.r.t. in which the first observation is z_{j-2} and the stopping boundaries are $b-Z_{j-3}$ and $a-Z_{j-3}$. We may assume $b < Z_{j-3} < a$ since T is multiplied by $I(C_j)$. Therefore, by lemma 3.6 , T , when H_j is true, is greater than or equal to T when H_{j-1} is true. Hence (3.23) is satisfied for all $j \geq 3$. A similar argument holds for the case $P_{H''}[N < \infty] = 1$.

Corollary 3.7.1 In a g.s.p.r.t. suppose that $z_1, z_2, \ldots$ are independent and z_i has distribution function F_i and G_i under H' and H" respectively. Also assume that either $P_{H'}[N < \infty] = 1$ or $P_{H''}[N<\infty]=1$. Then $F_i(z) \geq G_i(z)$ for $b_i - a_{i-1} \leq z \leq a_i - b_{i-1}$ and $i=1,2,\ldots$ where $a_o = b_o = o$ implies that

$$P_{H'}\left[\text{accept } H_o\right] \geq P_{H''}\left[\text{accept } H_o\right] .$$

Proof Simply make the appropriate changes in lemma 3.6, its proof and the proof of theorem 3.7.

Before proceeding further, we should mention the connection between theorem 3.7 and a sufficient condition for the monotonicity of the o.c. function. Suppose $x_1, x_2, \ldots$ are independently distributed with density $p_\theta(x)$. Lehmann [19] has shown that $p_\theta(x)$, possessing a monotone likelihood ratio in $T(x)$, (M.L.R.), is a sufficient condition for any s.p.r.t., testing θ_o against $\theta_1(\theta_o < \theta_1$ say), to have a monotone o.c. function. Now if $p_\theta(x)$ possesses a monotone likelihood ratio in $T(x)$, we may write:

$$z_i = \log \frac{p_{\theta_1}(x_i)}{p_{\theta_o}(x_i)} = h(T(x_i))$$

where h is a nondecreasing function of $T(x)$ depending upon θ_o and θ_1. Next by a slight extension of lemma 2 on page 74 of Lehmann[19], we have that if Ψ is a nondecreasing function of $T(x)$, then $\mathcal{E}_\theta \Psi(T(x))$ is a nondecreasing function of θ. Hence by setting

$$\Psi(T) = \begin{cases} 1 & \text{if } h(T) > c \\ o & \text{if } h(T) \leq c \end{cases}$$

we see that for $\theta' > \theta$

$$P_{\theta'}\left[z \leq c\right] = P_{\theta'}\left[\Psi(T) = o\right] = \mathcal{E}_{\theta'}\left[1 - \Psi(T)\right]$$

$$\leq \mathcal{E}_\theta\left[1 - \Psi(T)\right] = P_\theta\left[\Psi(T) = o\right] = P_\theta\left[z \leq c\right].$$

Also by theorem 3.7

$$P_{\theta'}\left[\text{accept } H_o\right] \leq P_\theta\left[\text{accept } H_o\right]$$

and monotonicity therefore follows. Thus the above statements may be summarized as

(3.24) x has M.L.R. $\Rightarrow$ z stochastically ordered $\Rightarrow$ monotone o.c.

Next we shall give two examples which illustrate the fact that the above implications cannot be reversed.

<u>Example</u> 3.3 (x has M.L.R. $\not\Leftarrow$ z stochastically ordered)

Let the random variable x take on only the three distinct values x_1, x_2 and x_3 with probability $p_1(\theta)$, $p_2(\theta)$ and $p_3(\theta)$ respectively. Define

$$p_1(\theta) = \frac{1}{6} - \frac{1}{6}\,\theta$$

$$p_2(\theta) = \frac{1}{3} - \frac{1}{2}\theta \quad - \frac{1}{3}\,\theta^2$$

$$p_3(\theta) = \frac{1}{2} + \frac{2}{3}\,\theta \quad + \frac{1}{3}\,\theta^2$$

for $-\frac{3}{4} < \theta < \frac{1}{2}$ and observe that $p_1(\theta)$ and $p_2(\theta)$ are decreasing functions of θ, while $p_3(\theta)$ is increasing. Also define

$$z(\theta_1,\ \theta_2,\ x_i) \ = \ \frac{p_{\theta_1}(x_i)}{p_{\theta_2}(x_i)} \ = \ \frac{p_i(\theta_1)}{p_i(\theta_2)} \ .$$

Now we will show for $\theta_1 > \theta_2$ and $\theta' > \theta$, that for all c

$$(3.25)\quad P_{\theta'}\left[z(\theta_1,\ \theta_2,\ x) \le c\right] \ \le \ P_\theta\left[z(\theta_1,\ \theta_2,\ x) \le c\ \right] .$$

Let z_1, z_2, z_3, where $z_1 \le z_2 \le z_3$, be the three values which $z(\theta_1, \theta_2, x)$ may take. Now since $p_1(\theta)$ and $p_2(\theta)$ are decreasing and $p_3(\theta)$ is increasing, we have $z_1 < 1$, $z_2 < 1$, $z_3 > 1$ and also $z_3 = z(\theta_1, \theta_2, x_3)$. Thus for $z_2 \le c < z_3$ we have determined that (3.25) is true. Since $p_1(\theta)$ and $p_2(\theta)$ are decreasing functions of θ, (3.25) also holds for $z_1 \le c < z_2$ and is thus satisfied for all c. Next we will show that there does not exist a $T(x)$ and an $h(\theta_1,\ \theta_2,\ T)$ such that for $\theta_1 > \theta_2$, $h(\theta_1,\ \theta_2,\ T)$ is a nondecreasing function of T and

$$z(\theta_1,\ \theta_2,\ x) \ = \ h(\theta_1, \theta_2,\ T(x)\ ).$$

Now it can be shown that $z(0, -2/3, x_1) = 3/5$, $z(0,-2/3, x_2) = 9/14$, thus $T(x_1) < T(x_2)$. However, $z(1/3, -2/3, x_1) = 2/5$, $z(1/3,-2/3, x_2) = \frac{1}{4}$ and therefore $h(1/3,-2/3,T)$ is not nondecreasing in T.

<u>Example 3.4</u> (z stochastically ordered $\not\Leftarrow$ monotone o.c.)

Consider an s.p.r.t. with $a = -b$ and

$$z_i = \begin{cases} 2a & \text{with probability } 1 - \theta - \frac{1}{8\theta} \\ \frac{2}{3a} & \text{with probability } \frac{1}{8\theta} \\ -2a & \text{with probability } \theta \end{cases}$$

where $\frac{1}{4} \leq \theta \leq \frac{1}{3}$. Since $N \leq 2$ it is clear that $P_\theta\left[\text{accept } H_o\right] = \theta + \frac{1}{8}$; hence the test has a monotonic o.c. function. However, $\theta + \frac{1}{8\theta}$ is a decreasing function of θ, so

$$P_\theta\left[z \leq c \right] \leq P_{\theta'}\left[z \leq c \right]$$

is not true for $\theta \neq \theta'$.

Returning to the expression (3.24), we should point out that "z stochastically ordered" implies that Wald's approximation to the o.c. function is monotonic. This can be shown as follows: Suppose $\theta' > \theta$, then $F_{\theta'}(z) \leq F_\theta(z)$ and by lemma 3.5 and corollary 3.5.1

$$(3.25) \qquad \int e^{tz}\, d\left[F_\theta(z) - F_{\theta'}(z)\right] \leq o \qquad \text{for } t > o$$

$$(3.26) \qquad \int e^{tz} d\left[F_{\theta}(z) - F_{\theta'}(z) \right] \geq o \qquad \text{for } t < o.$$

If we define $t(\theta) \neq o$ by $\int e^{zt(\theta)} d\, F_{\theta} = 1$ then for $t(\theta') > o$ we have by (3.25),

$$1 \geq \int e^{zt(\theta')} \, d\, F_{\theta}(z),$$

which implies that $t(\theta) \geq t(\theta')$. Similarly by (3.26) we have for $t(\theta') < o$ that $t(\theta) \geq t(\theta')$. Therefore $t(\theta)$ is a monotonically decreasing function of θ, and since Wald's o.c. approximation is a monotonic function of t, we have that it is also a monotonic function of θ.

In conclusion we will briefly consider the possibility of removing the condition of independence in theorem 3.7. To begin, suppose we look at a natural generalization of the theorem obtained by removing independence and strengthening the order relation to

$$F(z_1, \ldots, z_n) \geq G(z_1, \ldots, z_n) \qquad \text{for all } n.$$

The following example will then illustrate why the theorems conclusion does not follow under these conditions.

Example 3.5 Assume that a = -b and

$$P[z_1 = t,\ z_2 = s] = \begin{cases} p_1 \text{ if } (t,s) = (-2a, -3a) \\ p_2 \text{ if } (t,s) = (-2a, -3a) \\ p_3 \text{ if } (t,s) = (a/2, -3a) \\ p_4 \text{ if } (t,s) = (a/2, 3a\ ) \end{cases}$$

where $p_i \geq o$ and $\sum_{i=1}^{4} p_i = 1$. It then follows that P[accept H_o] = $p_i + p_2 + p_3$. Now suppose that F assigns $p_1 = p_4 = \frac{1}{2}$, $p_2 = p_3 = o$ and G assigns $p_2 = p_3 = p_4 = \frac{1}{3}$, $p_2 = o$. From this it follows that $F(z_1, z_2) \geq G(z_1, z_2)$, however $P_{H'}$[accept H_o] = $\frac{1}{2}$ and $P_{H''}$[accept H_o] = 2/3.

Although simple ordering of the joint distribution functions is not a sufficient condition for a generalization of theorem 3.7, there probably exist additional assumptions which would. However, in applications it is unlikely that P_H[accept H_o] would be known when $z_1, z_2, \ldots$ are not independent. With this in mind, the following theorem is of possible use in problems where the dependency between the z_i's is relatively "weak".

<u>Theorem 3.8</u> In an s.p.r.t. suppose that $G(z_1, \ldots, z_n)$ is the joint distribution function of $z_1, \ldots, z_n$ for all n under H' and that $z_1, z_2, \ldots$ are independent, with z_i having distribution function F_i for all i under H". Assume that $P_{H'}[N < \infty] = 1$. Then $F_i(z) \geq G_i(z | Z_{i-1})$ for $b-a \leq z \leq a-b$, $b < Z_{i-1} < a$ and all i implies that

$$P_{H'}\left[\text{ accept } H_o\right] \leq P_{H''}\left[\text{accept } H_o\right].$$

And similarly $F_i(z) \leq G_i(z | Z_{i-1})$ implies that $P_{H'}$[accept H_o] $\geq$ $P_{H''}$[accept H_o].

<u>Proof</u> Let H_j, $j = 2,3, \ldots,$ denote the hypothesis that $z_1, \ldots, z_n$

have $G(z_1, \ldots, z_{j-2})\, F_{j-1}(z_{j-1}) \ldots F_n(z_n)$ as their joint distribution function for all n. As in the proof of theorem 3.7, if for all $j \geq 3$

$$P_{H_j}\left[\text{ accept } H_o\right] \leq P_{H_{j-1}}\left[\text{ accept } H_o\right] \tag{3.27}$$

then

$$P_{H'}\left[\text{ accept } H_o\right] \leq P_{H''}\left[\text{accept } H_o\right].$$

Since $Z_o = o$, (3.27) holds for $j = 3$ by lemma 3.6. Now by referring to the proof of theorem 3.7,

$$P\left[\text{ accept } H_o\right] = \sum_{i=1}^{j-3} P\Big[\, b < z_1 + \ldots + z_k < a$$

$$\text{for } j = 1, \ldots, i-1 \text{ and } z_1 + \ldots + z_i \leq b\Big] + \int_{X^{j-3}} I(C_j)\Big\{$$

$$\sum_{i=o}^{\infty} P\Big[b - Z_{j-3} < z_{j-2} + \ldots + z_{j+k-2} < a - Z_{j-3}$$

$$\text{for } k = o, \ldots, i-1 \text{ and } z_{j-2} + \ldots + z_{j+i-2} \leq b - Z_{j-3} \,\Big|\, Z_{j-3}\Big]\Big\}\, dF$$

where F denotes the joint distribution function of $z_1, \ldots, z_{j-3}$. Under H_j and H'_{j-1}, $z_1, \ldots, z_{j-3}$ have the same joint distribution function. Thus in the above expression for P[accept H_o], only the term within the braces, let us call it D, differs under H_j and H_{j-1}.

D may be considered as P[accept H_o] if the first observation is z_{j-2} and the stopping boundaries are $b - Z_{j-3}$ and $a - Z_{j-3}$ where $b < Z_{j-3} < a$. Therefore, since $z_{j-2}, z_{j-1}, \ldots, z_n$ given Z_{j-3} has as a joint distribution function $G(z_{j-2} \mid Z_{j-3})\, F_{j-1}(z_{j-1}) \ldots F_n(z_n)$ under H_j and $F_{j-2}(z_{j-2}) \ldots F_n(z_n)$ under H_{j-1}, we have, by lemma 3.6, that D when H_j is true is less than or equal to D when H_{j-1} is true. Thus (3.27) holds for all $j \geq 2$. The same line of argument will also show that $F_i(z) \geq G_i(z \mid Z_{i-1})$ implies $P_{H'}$[accept H_o] $\leq$ $P_{H''}$[accept H_o].

3.5 Change of scale

Although the results are more general, the previous section could be thought of as being concerned with the effects upon the o.c. function of an s.p.r.t. after the variables $z_1, z_2, \ldots$ have been translated. A natural question concerning the effects of a change of scale then arises. What we shall attemp to show is, intuitively: Suppose the distribution of each z_i is more favorable toward the alternative hypothesis. Then increasing the variance of z_i while holding its mean fixed should disturb the process and increase the chance of acceptance. It should be noted that as the variance of each z_i becomes very small the probability of acceptance tends to zero, while if the variance becomes very large this probability tends to one half.

Specifically, we shall assume that $z_1, z_2, \ldots$ are independently distributed, with z_i having $f_i(z_i - \mu_i)$ as its probability density under H' and having $c_i f_i(c_i(z_i - \mu_i))$ as its density under H". Also assume that $b = -a$ and either $P_{H''}[N < \infty] = 1$ or

$P_{H'}[N < \infty\ ] = 1$. We then question under what circumstances is

$$(3.28) \qquad P_{H''}\left[\text{ accept } H_o\right] \geq P_{H'}\left[\text{ accept } H_o\ \right].$$

As in the proof of theorem 3.7, a sufficient condition for (3.28) is

$$(3.29) \qquad P_{H_j}\left[\text{ accept } H_o\right] \geq P_{H_{j-1}}\left[\text{accept } H_o\right] \text{ for } j = 3,4, \ldots$$

where H_j denotes the hypothesis that z_i has density $c_i f_i(c_i(z_i - \mu_i))$ for $i = 1, \ldots, j-2$ and density $f_i(z_i - \mu_i)$ for $i > j-2$.

Before proceeding further we shall state the following preliminary results:

Definition $f(x)$ is said to be increasing for $x \in S$ if $x \in S$, $y \in S$, $y \in S$ and $x < y$ implies that $f(x) \leq f(y)$. $f(x)$ is said to be unimodel if there exists an x_o such that $f(x)$ is increasing for $x \leq x_o$ and $f(x)$ is decreasing for $x \geq x_o$.

Lemma 3.9 Let h, g, k be real valued functions on the real line such that

(a) $h(t + \mu) = -h(-t + \mu)$ for all t

(b) $g(t + \nu) = g(-t + \nu)$ for all t

(c) $k(x) = \int h(x - u)g(u)du$ exists for all x.

Then

$$k(x + \mu + \nu) = - k(-x + \mu + \nu) \quad \text{for all } x .$$

Proof: $k(x + \mu + \nu) = \int h(x + \mu + \nu - u)g(u)du =$

$\int h(x + \mu - t)g(t + \nu)dt = - \int h(-x + \mu + t)g(-t + \nu)dt =$

$$- \int h(-x + \mu - t)g(t + \nu)dt = - \int h(-x + \mu + \nu - u)g(u)du =$$
$$- k(-x + \mu + \nu) .$$

<u>Lemma 3.10</u> Suppose the conditions of lemma 3.9 are satisfied and

(a) $g(t)$ is unimodal

(b) $h(t+\mu) \geq o$ if $t \geq o$.

Then
$$k(x+\mu+\nu) \geq o \text{ if } x \geq o .$$

Proof: $k(x+\mu+\nu) = \int h(x+\mu-t)g(t+\nu)dt = \int h(u+\mu)g(x+\nu-u)du =$

$$- \int_{-\infty}^{o} h(-u+\mu)g(-x+\nu+u)du + \int_{o}^{\infty} h(u+\mu)g(x+\nu-u)du =$$

$$\int_{o}^{\infty} [g(x+\nu-u) - g(x+\nu+u)]h(u+\mu)du$$

By (b) of lemma 3.9 and (a) of lemma 3.10 we have that $g(t)$ is decreasing for $t \geq \nu$; and thus,

(i) $x \geq u \geq o$ implies $g(x+\nu-u) \geq g(x+\nu+u)$

(ii) $u \geq x \geq o$ implies $g(x+\nu-u) = g(-x+\nu+u) \geq g(x+\nu+u)$.

Therefore $[g(x+\nu-u) - g(x+\nu+u)]h(u+\mu) \geq 0$ if $x \geq o$ and $u \geq o$ and hence $k(x+\mu+\nu) \geq o$ if $x \geq o$.

<u>Lemma 3.11</u> Let f, h, k be real valued functions on the real line such that

(a) there exists a μ such that $f(t+\mu) = f(-t+\mu)$ for all t

(b) $\mu \geq o$

(c) $f(t)$ is unimodal

(d) $h(t) \geq h(-t)$ for all $t \geq o$

(e) $h(t)$ is increasing for $t \leq o$

(f) $k(x) = \int f(x-t)h(t)dt$ exists for all x .

Then

(i) $k(x)$ is increasing for $x \leq o$

(ii) $k(x) \geq k(-x)$ for all $x \geq o$.

Proof: Part (i) - Let $x \leq o$, $y \geq o$ and define $g(s+\mu+\frac{y}{2}-x) = f(x-s) - f(x-y-s)$. Then $g(t) = f(\mu+\frac{y}{2}-t) - f(\mu-\frac{y}{2}-t) = f(\mu-\frac{y}{2}+t) - f(\mu+\frac{y}{2}+t) = -g(-t)$. Now since f is unimodal, $f(t+\mu)$ is increasing for $t \leq o$ and thus

$$t \geq \frac{y}{2} \geq o \text{ implies } f(\mu+\frac{y}{2}-t) \geq f(\mu-\frac{y}{2}-t)$$

$$\frac{y}{2} \geq t \geq o \text{ implies } f(\mu+\frac{y}{2}-t) = f(\mu-\frac{y}{2}+t) \geq f(\mu-\frac{y}{2}-t).$$

Therefore $g(t) \geq o$ for $t \geq o$. Now $k(x) - k(x-y) =$

$$\int g(s+\mu+\frac{y}{2}-x)h(s)ds = \int h(t+x-\mu-\frac{y}{2})g(t)dt = \int_o^\infty h(-t+x-\mu-\frac{y}{2})g(-t)dt$$

$$+ \int_o^\infty h(t+x-\mu-\frac{y}{2})g(t)dt = \int_o^\infty [\,h(t+x-\mu-\frac{y}{2}) - h(-t+x-\mu-\frac{y}{2})\,]g(t)dt.$$

Since $x-\mu-\frac{y}{2} \leq o$ we have by (d) and (e) that $t \geq -x+\mu+\frac{y}{2} \geq o$ implies $h(t+x-\mu-\frac{y}{2}) \geq h(-t-x+\mu+\frac{y}{2}) \geq h(-t+x-\mu-\frac{y}{2})$ and $-x+\mu+\frac{y}{2} \geq t \geq o$ implies $h(t+x-\mu-\frac{y}{2}) \geq h(-t+x-\mu-\frac{y}{2})$. Therefore $[\,h(t+x-\mu-\frac{y}{2}) - h(-t+x-\mu-\frac{y}{2})\,]g(t) \geq o$ for $t \geq o$ and thus $k(x) - k(x-y) \geq o$.

Part (ii) - Let $x \geq o$ and define $g(t+\mu) = f(x-t) - f(-x-t)$. As in part (i), we can show that $g(t) = -g(-t)$ and $g(t) \geq o$ for $t \geq o$. It also happens that $k(x) - k(-x) = \int_o^\infty [h(t-\mu) - h(-t-\mu)]g(t)dt$ and $[h(t-\mu) - h(-t-\mu)]g(t) \geq o$ for $t \geq o$. Therefore, $k(x) \geq k(-x)$ for $x \geq o$.

<u>Theorem 3.12</u> Consider an s.p.r.t. with $z_1, z_2, \ldots$ independent and with $b = -a$. Suppose the probability density function of z_i is $f_i(z_i - \mu_i)$ where f_i is unimodal, $f_i(t) = f_i(-t)$, and $\mu_i \geq o$ for all i. Then the probability density function, g_n, of Z_n, given that $N \geq n$, has the following two properties:

(i) $g_n(x) \geq g_n(-x)$ for $x \geq o$

(ii) $g_n(x)$ is increasing for $x \leq o$.

Proof: The result is clearly true for $n=1$. Now let us suppose it is also true for $n=m-1$. Let

$$h_m(x) = \begin{cases} \dfrac{g_{m-1}(x)}{\int_{-a}^{a} g_{m-1}(x)dx} & \text{for } -a < x < a \\ 0 & \text{elsewhere,} \end{cases}$$

then h_m is the density of Z_{m-1}, given that $N \geq m$, and it is clear that $h_m(x) \geq h_m(-x)$ for $x \geq o$ and that $h_m(x)$ is increasing for $x \leq o$. Now $g_m(x) = \int f_m(x-t-\mu_m)h_m(t)dt$, and by lemma 3.11 g_m satisfies (i) and (ii).

Returning to the problem of establishing (3.28) and (3.29), we begin by assuming that

(a) $\mu_i \geq o$

(b) $o < c_i \leq 1$

(c) $f_i(t) = f_i(-t)$ for all t

(d) $f_i(t)$ is unimodal

(e) $f_i(t)$ is continuous (so, in view of (d) bounded)

hold for all i. Now define g_j to be the density of Z_{j-1}, given $N \geq j$, and let

$$P_j(x) = P_{H'}[\text{accept } H_o \mid N > j \text{ and } Z_j = x]$$

$$T_j(x) = I(x \leq -a) + I(-a < x < a)P_j(x)$$

$$F_j(x) = \int_{-\infty}^{x} f_j(t)dt \quad \text{and} \quad G_j(x) = \int_{-\infty}^{x} g_j(t)dt .$$

By the definition of H_j ,

(3.30) $$P_{H_{j+2}}[\text{accept } H_o \mid N \geq j] \geq P_{H_{j+1}}[\text{accept } H_o \mid N \geq j]$$

for $j=1,2,\ldots$ is equivalent to (3.29) . It can also be seen that

$$P_{H_{j+2}}[\text{accept } H_o|N \geq j] = \int T_j(x)d[\, F_j(c_j(x-\mu_j)) * G_j(x)\,]$$

and

$$P_{H_{j+1}}[\text{accept } H_o|N \geq j] = \int T_j(x)d[\, F_j(x-\mu_j) * G_j(x)\,]$$

where $F * G(t) \equiv \int F(t-u)dG(u)$. Define

$$L(c_j) = \int T_j(x)d[F_j(c_j(x-\mu_j)) * G_j(x)] - \int T_j(x)d[F_j(x-\mu_j) * G_j(x)]$$

$$= \int [F_j(x-\mu_j) - F_j(c_j(x-\mu_j))] * G_j(x)dT_j(x)$$

and $L(c_j) \geq o$ will then imply (3.30) . Since $T_j(x)$ is constant for $x \notin (-a,a)$, f_j is bounded and by assumption (e) g_j is bounded, we have

$$L'(c_j) = \frac{dL(c_j)}{dc_j} = -\int [(x-\mu_j)f_j(c_j(x-\mu_j))] * g_j(x)dT_j(x)$$

where $f * g(t) = \int f(t-u)g(u)du$. By defining

$$S_j(x) = (\mu_j - x)f_j(c_j(x - \mu_j))$$

we have

$$L'(c_j) = \int\int S_j(x-u)g_j(u)dudT_j(x) .$$

Next define

$$r_1^*(u) = \begin{cases} g_j^*(u) & \text{for } u \le o \\ g_j^*(-u) & \text{for } u \ge o \end{cases}$$

$$r_2^*(u) = g_j^*(u) - r_1^*(u)$$

where

$$g_j^*(u) = \begin{cases} \dfrac{g_{j-1}(u) * [c_{j-1}f_{j-1}(c_{j-1}(u-\mu_{j-1}))]}{\int_{-a}^{a} g_{j-1}(u) * [c_{j-1}f_{j-1}(c_{j-1}(u-\mu_{j-1}))]du} & \text{for } -a \le u \le a \\ 0 & \text{elsewhere} \end{cases}$$

By assumption (e), $g_j^*(u)$ is continuous and bounded on $[-a,a]$ (page 491 of Apostol [2]); thus, given $\epsilon > o$, there exists a step function $\sum_{i=1}^{n} k_i^*(t)$ such that

(3.31) $$\left| r_2^*(t) - \sum_{i=1}^{n} k_i^*(t) \right| < \epsilon .$$

Now if we define

$$r_1(u) = \begin{cases} g_j(u) & \text{for } u \le o \\ g_j(-u) & \text{for } u \ge o \end{cases}$$

$$r_2(u) = g_j(u) - r_1(u)$$

it then follows from (3.31) and from $g_j(u) = \begin{cases} g_j^*(u) & -a < u < a \\ 0 & \text{elsewhere} \end{cases}$

that

$$\left| r_2(t) - \sum_{i=1}^{n} k_i(t) \right| < \epsilon$$

where

$$k_i(t) = \begin{cases} k_i^*(t) & \text{for } -a < t < a \\ 0 & \text{elsewhere} \end{cases}$$

for all i . By theorem 3.12 we have

(a) $r_1(u)$ is unimodal

(b) $r_2(u) \geq o$

(c) $r_2(u) = o$ for $u \leq o$

and hence we may assume that

$$k_i(t) = \begin{cases} \alpha_i & \text{if } t \in < a_i, a_{i+1} > \\ o & \text{otherwise} \end{cases}$$

where $\alpha_i > o$ and $< a_i, a_{i+1} >$ indicates some type of interval with $o \leq a_i$. Define

$$\ell_o(x) = \int r_1(u) S_j(x-u)du$$

$$\ell_i(x) = \int k_i(u) S_j(x-u)du \qquad \text{for } i=1,\ldots,n ,$$

we then have

$$(3.32) \qquad L'(c_j) = \int \left[\int r_1(u) S_j(x-u)du + \int r_2(u) S_j(x-u)du \right] dT_j(x)$$

$$\leq \int \left\{ \ell_o(x) + \sum_{i=1}^{n} \ell_i(x) + \epsilon \int_{-a}^{a} | S_j(x-u) | du \right\} dT_j(x) \leq$$

$$\sum_{i=o}^{n} \int \ell_i(x) dT_j(x) + \epsilon M$$

where $M \leq 2a(2a + \mu_j) f_j(o)$. Since $-S_j(t+\mu_j) = S_j(-t+\mu_j)$ and $r_1(t) = r_1(-t)$ we have by lemma 3.9 that $-\ell_o(x+\mu_j) = \ell_o(-x+\mu_j)$. Also $r_1(t)$ is unimodal and $-S_j(t+\mu_j) \geq o$ if $t \geq o$ and thus by lemma 3.10 $-\ell_o(x+\mu_j) \geq o$ if $x \geq o$ and $-\ell_o(x+\mu_j) \leq o$ if $x \leq o$. Since $S_j(t)$ is continuous, we may consider $< a_i, a_{i+1} > = (a_i, a_{i+1})$ for purposes of determining properties of $\ell_i(x)$. Let $\nu_i = \frac{1}{2}(a_i + a_{i+1})$ and observe that $\nu_i \geq o$ and $k_i(u+\nu_i) = k_i(-u+\nu_i)$. Hence by lemmas

3.9 and 3.10 we have

$$- \ell_i(x+\mu_j+\nu_i) = \ell_i(-x+\mu_j+\nu_i)$$
$$- \ell_i(x+\mu_j+\nu_i) \geq o \quad \text{if} \quad x \geq o \, .$$

At this point it is necessary to make a further assumption based on some heuristic considerations.

assumption (f) : If $M \geq o$ then $T_j(x+M) + T_j(-x+M)$ is an increasing function of x for $x \geq o$.

Let $M_i = \mu_j+\nu_i \geq o$, then
$$\int \ell_i(x)dT_j(x) \leq o \quad \text{for} \quad M_i \geq a \, ,$$

since $T_j(x)$ is decreasing and $\ell_i(x) \geq o$ for $x \leq a \leq M_i$. Assume $M_i < a$, then
$$\int \ell_i dT_j(x) = \int_{-a-M_i}^{a-M_i} \ell_i(x+M_i)dT_j(x+M_i) = -\int_{a+M_i}^{o} \ell_i(x+M_i)dT_j(-x+M_i) +$$

$$\int_o^{a-M_i} \ell_i(x+M_i)dT_j(x+M_i) = \int_o^{a-M_i} \ell_i(x+M_i)d[T_j(x+M_i) + T_j(-x+M_i)] +$$

$$\int_{a-M_i}^{a+M_i} \ell_i(x+M_i)dT_j(-x+M_i) \leq o$$

since $\ell_i(x+M_i) \leq o$, and both $T_j(-x+M_i)$ and $T_j(x+M_i) + T_j(-x+M_i)$ are increasing. Therefore $\int \ell_i(x)dT_j(x) \leq o$ for $i=o,1,\ldots,n$ and by (3.32) $L'(c_j) \leq o$. Thus $L(c_j) \geq L(1) = o$ for $o < c_j \leq 1$, which implies (3.30).

Concerning the assumptions (c), (d) and (e), possibly (c) is too restrictive, however (e) is reasonable and (d) can easily be shown to be necessary. The problem is with assumption (f), since in practice we will not know how closely (f) is satisfied. However, it is conjectured that (f) will be approximately true (at least that the integral $\int_{o}^{a-M_i} \ell_i(x + M_i)\ d[T_j(x + M_i) + T_j(-x + M_i)]$ is nonpositive) if there is not a great deal of difference between the individual densitites f_i $i=1,2,\ldots$ An intuitive argument in support of (f) is as follows. Putting $a = -b$, from (a), (c) and theorem 3.7 we have $T_j(o) \leq \frac{1}{2}$. Suppose we could apply Wald's o.c. approximation. Then, since $a = -b$ and $T_j(o) \leq \frac{1}{2}$ it is reasonable to assume that $h \leq o$ where

$$P\left[\text{ accept } H_o \mid N \geq j,\ Z_{j-1} = 0\right] \doteq \frac{e^{ha} - 1}{e^{ha} - e^{-ha}} .$$

Therefore,

$$P_j(x) \doteq \frac{e^{h(a-x)} - 1}{e^{h(a-x)} - e^{-h(a+x)}} = \frac{e^{hx} - e^{ha}}{e^{-ha} - e^{ha}} ,$$

and it follows that

$$\left[T_j(M + x_2) + T_j(M - x_2)\right] - \left[T_j(M + x_1) + T_j(M - x_1)\right]$$

$$\doteq \frac{1}{e^{-ha} - e^{ha}}\left[e^{h(M + x_2)} + e^{h(M - x_2)} - e^{h(M + x_1)} - e^{h(M - x_1)}\right]$$

$$= \frac{2\,e^{hM}}{e^{-ha} - e^{ha}}\left[\cosh h\,x_2 - \cosh h\,x_1\right] \geq o \quad \text{for } a \geq x_2 \geq x_1 \geq o.$$

Therefore when conditions (a) through (e) are satisfied and the results are applied, the conclusions are assumed to be approximately correct.

3.6 General bounds for the o.c. functions

The technique for determining bounds on the moments of the sample size used in section 2.6 may also be applied to the o.c. function. As in that particular section, let us consider any nonrandomized sequential rule such that after the ith observation x_i, if $Z_i(x_1,$

$\ldots, x_i) \epsilon\, C_i$, then another observation is taken. Otherwise, sampling is discontinued. If sampling is terminated at the ith trial, we will accept H_o if $Z_i(x_1, \ldots, x_i) \epsilon\, A_i$, where A_i and C_i are disjoint for all i; otherwise we will reject H_o. For $j = 1,2, \ldots$ define

$$E_i^j = \left\{ (x_1, \ldots, x_j) \,|\, Z_i(x_1, \ldots, x_i) \epsilon\, C_i \right\} i = 1, \ldots, j-1$$

$$E_j^j = \left\{ (x_1, \ldots, x_j) \,|\, Z_j(x_1, \ldots, x_j) \quad A_j \right\}.$$

We then have

$$P\left[\text{accept } H_o \text{ and } N = j \right] = P\left[\bigcap_{i=1}^{j} E_i^j \right] = 1 - P\left[\bigcap_{i=1}^{j} \bar{E}_i^j \right].$$

Next define

$$p_i^j = P\left[\bar{E}_i^j \right], \; p_{i,k}^j = P\left[\bar{E}_i^j \bar{E}_k^j \right], \ldots$$

$$s_1^j = \sum_{i=1}^{j} p_i^j, s_2^j = \sum_{j \geq i > k \geq 1} p_{i,k}^j, \ldots .$$

It then follows that

$$P\left[\bigcup_{i=1}^{j} \bar{E}_i^j \right] = s_1^j - s_2^j + \ldots \pm s_j^j .$$

By Bonferroni's inequality

$$s_1^j - s_2^j + \ldots - s_k^j \leq P\left[\bigcup_{i=1}^{j} \bar{E}_i^j \right] \leq s_1^j - s_2^j + \ldots + s_\ell^j$$

where ℓ is odd and k is even. Since

$$P\left[E_j^j\right] \geq P\left[\text{ accept } H_o \text{ and } N = j\right]$$

we have

$$a_{jk} \geq P\left[\text{ accept } H_o \text{ and } N = j\right] \geq b_{j\ell}$$

where

$$a_{jk} = \min\left\{ P\left[E_j^j\right], 1 - s_1^j + s_2^j - \dots + s_k^j \right\}$$

$$b_{j\ell} = \max\left\{ o, 1 - s_1^j + s_2^j - \dots - s_\ell^j \right\}$$

and

$$a_{jo} = P\left[E_j^j\right].$$

Since $P\left[\text{ accept } H_o\right] = \sum_{n=1}^{\infty} P\left[\text{accept } H_o \text{ and } N = n\right]$ it follows that

$$(3.33) \qquad \sum_{n=1}^{\infty} a_{n\,k_n} \geq P\left[\text{ accept } H_o\right] \geq \sum_{n=1}^{\infty} b_{n\,\ell_n}$$

where k_n is an even integer for all n and ℓ_n is an odd integer for all n. If we set

$$F_i^j = E_i^j \quad \text{and} \quad F_j^j = \left\{(x_1, \dots, x_j) \mid Z_j(x_1, \dots, x_j) \in R_j\right\},$$

where $R_j = \bar{C}_j \cap \bar{A}_j$; we find, in a similar manner, that

$$(3.34) \qquad \sum_{n=1}^{\infty} c_{np_n} \geq P\left[\text{ reject } H_o\right] \geq \sum_{n=1}^{\infty} d_{nq_n}$$

where

$$c_{n\,p_n} = \min\left\{P\left[F_n^n\right],\ 1 - t_1^n + t_2^n - \dots + t_{p_n}^n\right\}$$

$$d_{n\,q_n} = \max\left\{o,\ 1 - t_1^n + t_2^n - \dots - t_{q_n}^n\right\},$$

and t_i^j is defined similar to s_i^j, p_n is even and q_n is odd. Combining (3.33) and (3.34), we conclude that

$$(3.35)\quad \min\left[\sum_{n=1}^{\infty} a_{n\,k_n},\ 1 - \sum_{n=1}^{\infty} d_{n\,q_n}\right] \geq P\left[\text{accept } H_o\right] \geq$$

$$\max\left[\sum_{n=1}^{\infty} b_{n\ell_n},\ 1 - \sum_{n=1}^{\infty} c_{n\,p_n}\right].$$

By setting $k_n = p_n = o$ and $q_n = \ell_n = 1$, we obtain, as a special case of (3.35) ;

(3.36)

$$\min\left\{\sum_{n=1}^{\infty} P[Z_n \in A_n],\ 1 - \sum_{n=1}^{\infty} \max\left[o,\ P[Z_n \in R_n] - \sum_{i=1}^{n-1} P[Z_i \notin C_i]\right]\right\}$$

$$\geq P[\text{accept } H_o] \geq \max\left\{1 - \sum_{n=1}^{\infty} P[Z_n \in R_n],\ \sum_{n=1}^{\infty} \max\left[o,\right.\right.$$

$$\left.\left. P[Z_n \in A_n] - \sum_{i=1}^{n-1} P[Z_i \notin C_i]\right]\right\}.$$

As an illustration of (3.36) let us consider the two examples given in section 2.6:

<u>Example 3.6</u> For the test of means from a normal distribution we obtain from (3.36)

$$.895 \leq P_{H_o}[\text{accept } H_o] \leq .9997 ,$$

while Baker[5] has observed a value

$$P_{H_o}[\text{accept } H_o] \doteq .9646 .$$

For the sequential t-test given in section 2.6, we find, from (3.36), that

$$.8666 \leq P_{H_o}[\text{accept } H_o] \leq .9958 .$$

The lower value is valid, assuming that the test was terminated at or before $n=30$. The experimental value observed by Siskind[25] is

$$P_{H_o}[\text{accept } H_o] = .960 .$$

CHAPTER IV

APPLICATIONS

4.1 Introduction

The primary purpose of this chapter is to illustrate the methods given in chapters two and three. We begin by applying these results to the s.s.s.t. described in chapter one. For this particular test, $z_1,\ldots,z_{n_o}$ were shown to be independent; so most of the methods will be applicable. We then conclude the chapter by considering an s.p.r.t. in which the $z_1,z_2,\ldots$ are multivariate normal. This appears to be a natural case in which there is dependency between the z_i's. It should also be noted that if $x_1,x_2,\ldots$ are multivariate normal, $N(\underline{\mu}, \Sigma)$, then the Wald s.p.r.t. for discriminating between

$$H_o: \underline{\mu} = \underline{\mu}_o \text{ , } \Sigma = \Sigma_o \quad \text{and} \quad H_1: \underline{\mu} = \underline{\mu}_1 \text{ , } \Sigma = \Sigma_o$$

leads to the conclusion that $z_1,z_2,\ldots$ are also multivariate normal. An additional purpose for designating $z_1,z_2,\ldots$ as multivariate normal is because of the interest in the following technique. Suppose that $z_1,z_2,\ldots$ are independent and identically distributed and that the a.s.n. is large. It may then be reasonable to set

$$z_i^* = \sum_{j=(i-1)r+1}^{ir} z_j \text{ ,}$$

and consider the s.p.r.t. with increments $z_1^*, z_2^*, \ldots$, being approximately independent normal variables.

4.2 Bounds on $\mathcal{E}(N)$ for the s.s.s.t.

In this section the results of chapter two will be applied to the s.s.s.t. Let us begin by considering the Wald type bounds for $\mathcal{E}(N)$ which were given in section 2.3.

At the end of section 1.2 it was shown that $z_1, \ldots z_{n_o}$ are independently distributed. Now if $R(n_o+1)$ is redefined by setting

$$(4.1) \qquad z_{n_o+1} = \begin{cases} a-b & \text{if } n > n_o \\ b-a & \text{if } n = n_o , \end{cases}$$

we will then be assured that $R(n_o+1) \geq A$ if $n > n_o$ and $R(n_o+1) \leq B$ if $n=n_o$. Thus we may assume that the criterion used in the s.s.s.t. can be expressed in terms of the variables $z_1, \ldots, z_{n_o+1}$ which are independent, where z_{n_o+1} is defined by (4.1). This operation then permits the application of the bounds for $\mathcal{E}(N)$ which are given by (2.10).

It was shown in section 1.2 that for $i=1,\ldots,n_o$, the distribution function of z_i is

$$F_i(z) = \begin{cases} \exp[c_i(z-a_i)] & \text{for } z \leq a_i \\ 1 & \text{for } z > a_i \end{cases}$$

where $c_i = \dfrac{n-i+1}{n_1-n_o}$ and $a_i = \log \dfrac{n_1-i+1}{n_o-i+1}$.

It then follows that

$$\mathcal{E}[\, z_i + r \mid z_i \leq -r \,] = \frac{n_o-n_1}{n-i+1} \qquad \text{for } i=1,\ldots,n_o ,$$

hence

$$\xi' \equiv \inf_{i} \inf_{a-b>r>o} \mathcal{E}[z_i + r \mid z_i \le -r] = \begin{cases} \min\{n_o - n_1,\ b-a\} & \text{if } n = n_o \\ \dfrac{n_o - n_1}{n - n_o + 1} & \text{if } n > n_o . \end{cases}$$

Similarly, for $o < r < a_i$ we have

$$\mathcal{E}[z_i - r \mid z_i \ge r] = \frac{a_i - r}{1 - e^{c_i(r-a_i)}} - \frac{1}{c_i} ,$$

which is decreasing in r. Therefore,

$$\sup_{r>o} \mathcal{E}[z_i - r \mid z_i \ge r] = \frac{a_i}{1 - e^{-a_i c_i}} - \frac{1}{c_i} \equiv d_i .$$

Finally,

$$\xi = \begin{cases} \max\{ a-b,\ d_1, \ldots, d_{n_o} \} & \text{for } n > n_o \\ \max\{ d_1, \ldots, d_{n_o} \} & \text{for } n = n_o , \end{cases}$$

and since $\mathcal{E}(z_i) = a_i - \frac{1}{c_i}$ for $i=1,\ldots n_o$, we are in a position to calculate the bounds.

Example 4.1 Let $n_o = 5$, $n_1 = 10$, $a = 2.064$ and $b = -1.661$. These values insure that $\alpha = \beta = .05$. Some of the properties of this test are given in table 1.2. The following table gives the values of $\mathcal{E}_n(z_i)$ for $n=n_o,\ldots,n_1$ and $i=1,\ldots,n_o+1$.

TABLE 4.1

		n					
		5	6	7	8	9	10
i	1	-.307	-.140	-.021	.068	.137	.193
	2	-.439	-.189	-.022	.097	.186	.255
	3	-.685	-.268	-.018	.149	.268	.357
	4	-1.247	-.414	.003	.253	.420	.539

5	-3.208	-.708	.125	.542	.792	.959
6	-3.725	3.725	3.725	3.725	3.725	3.725

Using the correct values for P[accept H_o] given in table 1.2, the bounds on $\mathcal{E}(N)$ found in section 2.3 are:

TABLE 4.2

	n					
	5	6	7	8	9	10
upper bound	20.3	-	-	66.0	37.1	28.1
lower bound	.38	-.33	.10	.32	.44	.49
lower bound based on averages	.89	-	-	1.49	1.76	1.83

Since $\mathcal{E}(z_i)$ is monotonic in i for n=5,8,9,10 , we may, as stated at the end of section 2.3, use $\frac{1}{n_o+1}\sum_{i=1}^{n_o+1} \mathcal{E}(z_i)$ in place of $\sup_i \mathcal{E}(z_i)$, or $\inf_i \mathcal{E}(z_i)$ in calculating the inequalities for these values of n (i.e. line 3 of table 4.2).

For each value of n there is a rather large amount of relative variation among the $\mathcal{E}(z_i)$. From the development of the inequalities we would expect poor results when this is indeed the case. Hence the preceding example appears to emphasize and illustrate the necessity of having only a small amount of relative variation among the $\mathcal{E}(z_i)$ before one could expect reasonably tight bounds.

We now apply Hoeffding's bounds, (2.14) and (2.15), to the same example. Since the bounds were developed in terms of the probability densities of the observations, it is necessary to discuss the density of x_{n_o+1} instead of simply defining z_{n_o+1} as a constant. Let $f(x_1,\ldots,x_{n_o}|n)$ represent the joint density of $x_1,\ldots,x_{n_o}$ when the true sample size is n. Define

$$h(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

$$g(t) = \begin{cases} \exp(b-a) & \text{if } 0 \leq t \leq \exp(a-b) \\ 0 & \text{elsewhere} \end{cases}$$

and consider the s.p.r.t. discriminating between

$$H_o\colon p(x_1,\ldots,x_{n_o+1}) = f(x_1,\ldots,x_{n_o}|n_o)g(x_{n_o+1})$$

and

$$H_1\colon p(x_1,\ldots,x_{n_o+1}) = f(x_1,\ldots,x_{n_o}|n_1)h(x_{n_o+1})$$

where $p(x_1,\ldots,x_{n_o+1})$ represents the joint density of $x_1,\ldots,x_{n_o+1}$. Let the hypothesis H represent $p(x_1,\ldots,x_{n_o+1}) = f(x_1,\ldots,x_{n_o} \mid n) \times h(x_{n_o+1})$ where n is assumed to be greater than n_o. Then employing the notation of section 2.4,

$$z_{o,n_o+1} = a - b \quad \text{with probability one under } H$$

$$z_{1,n_o+1} = 0 \quad \text{with probability one under } H .$$

It is therefore clear that the above s.p.r.t. leads to the same decision and distribution of N as the s.s.s.t. Now

$$\mathcal{E}_H(z_{oi}) = \log \frac{n-i+1}{n_o-i+1} - \frac{n-n_o}{n-i+1} \qquad i=1,\ldots,n_o$$

$$\mathcal{E}_H(z_{1i}) = \log \frac{n-i+1}{n_1-i+1} - \frac{n-n_1}{n-i+1} \qquad i=1,\ldots,n_o$$

and both $\mathcal{E}_H(z_{oi})$ and $\mathcal{E}_H(z_{1i})$ are increasing in i for $i=1,\ldots,n_o$

Typically we will have

$$a - b \geq \log(n-n_o+1) - \frac{n-n_o}{n-n_o+1} .$$

If this is not the case, we may change $g(x_{n_o+1})$ so that $\mathcal{E}_H(z_{o,n_o+1}) = \mathcal{E}_H(z_{o,n_o})$. Therefore we may assume that $\mathcal{E}_H(z_{oi})$ is increasing in for $i=1,\ldots,n_o+1$. In light of the discussion at the end of section 2.3, it then follows that

$$e_o(n) = \frac{1}{n_o+1}\left[\max\left\{a-b,\ \log(n-n_o+1) - \frac{n-n_o}{n-n_o+1}\right\} + \sum_{i=1}^{n_o}\left\{\log\frac{n-i+1}{n_o-i+1} - \frac{n-n_o}{n-i+1}\right\}\right].$$

Similarly

$$e_1(n) = \frac{1}{n_o}\sum_{i=1}^{n_o}\left\{\log\frac{n-i+1}{n_1-i+1} - \frac{n-n_1}{n-i+1}\right\}.$$

Therefore by equation (2.13),

$$(4.2) \qquad \mathcal{E}_n(N) \geq \sup_{o<c<1} \frac{-\log\{\alpha^c(1-\beta)^{1-c} + (1-\alpha)^c\beta^{1-c}\}}{ce_o(n) + (1-c)e_1(n)} \text{ for } n>n_o.$$

If we redefine

$$g(t) = \begin{cases} \exp(a-b) & \text{if } o \leq t \leq \exp(b-a) \\ 0 & \text{elsewhere}, \end{cases}$$

it follows, in a similar manner, that

$$(4.3) \qquad \mathcal{E}_{n_o}(N) \geq \frac{\alpha\log\frac{\alpha}{1-\beta} + (1-\alpha)\log\frac{1-\alpha}{\beta}}{f_1}$$

where

$$f_1 = \frac{1}{n_o+1}\left[\max\left\{a-b,\ n_1-n_o-\log(n_1-n_o+1)\right\} + \sum_{i=1}^{n_o}\left\{\log\frac{n_o-i+1}{n_1-i+1} - \frac{n_o-n_1}{n_o-i+1}\right\}\right]$$

As a special case of (4.2), we obtain from equation (2.15),

$$(4.4) \qquad \mathcal{E}_{n_1}(N) \geq \frac{\beta \log \frac{\beta}{1-\alpha} + (1-\beta) \log \frac{1-\beta}{\alpha}}{e_o(n_1)} .$$

As an illustration, let $n_o = 5$, $n_1 = 10$, $a = 2.064$ and $b = -1.661$. For these values $\alpha = \beta = .05$ and by (4.3) and (4.4)

$$\mathcal{E}_{n_o}(N) \geq 1.65$$

$$\mathcal{E}_{n_1}(N) \geq 2.64 .$$

These bounds are considerably better than those of example 4.1 and also they are true for any sequential test procedure, not exclusively an s.p.r.t.

To conclude this section, the general bounds given in section 2.6 are applied to $\mathcal{E}(N)$ and $\mathcal{E}(N^2)$. Following the notation of sections 1.2 and 2.6 we have

$$P\left[Z_i \in C_i \right] = P\left[B < R(i) < A \right] = P\left[a_i < u_i < b_i \right]$$

where u_i is the ith order statistic of a sample of size n from a [o,1] uniform distribution. Therefore, $P[Z_i \in C_i]$ may be obtained from tables of the incomplete beta function.

<u>Example 4.2</u>

Let $n_o = 5$, $n_1 = 10$, $a = 2.064$ and $b = -1.661$ which ensure that $\alpha = \beta = .05$. Also denote $p_i = P[Z_i \in C_i]$. Then for the case of $n = n_o$ we find that

$p_1 = .9051$, $p_2 = .7715$, $p_3 = .6135$, $p_4 = .4233$, $p_5 = .2285$ and

$p_j = o$ for $j > 5$. Therefore by (2.24) we obtain

$$3.94 \geq \mathcal{E}(N) \geq 2.87$$

$$27.00 \geq \mathcal{E}(N^2) \geq 9.13,$$

while the correct values are $\mathcal{E}(N) = 3.76$ and $\mathcal{E}(N^2) = 16.59$. Similarly for $n = n_1$ we find $p_1 = .9910$, $p_2 = .9825$, $p_3 = .9367$, $p_4 = .6686$, $p_5 = .3617$ and $p_j = o$ for $j > 5$. It therefore follows that

$$4.94 \geq \mathcal{E}(N) \geq 4.45$$

$$25.44 \geq \mathcal{E}(N^2) \geq 20.43,$$

while $\mathcal{E}(N) = 4.84$ and $\mathcal{E}(N^2) = 24.45$.

From examples 4.1 and 4.2 it appears that the general method given in section 2.6 will produce superior results when there is considerable variation among the distributions of the z_i's. At least we are safe in stating that the Wald type bounds are likely to yield poor results in this particular situation.

4.3 Bounds on the o.c. function for the s.s.s.t.

We begin this section by applying the bounds given by (3.22) to the s.s.s.t. In order to do so, we must first have $z_1, \ldots, z_{n_o+1}$ independent and hence $z_{n_o+1} \geq a - b$ if $n > n_o$ and $z_{n_o+1} \leq b - a$ if $n = n_o$. If this is the case, then the equation $\Psi_{n_o+1}(t) = 1$ will not have a nonzero solution. Therefore we define the s.p.r.t., S, let us say, as the same test as the s.s.s.t., except that z_i has the same distribution as z_{n_o} for $i > n_o$, and $z_1, z_2, \ldots$ are independent. It then clearly follows that

$$P\left[N > n_o\right] + P\left[\text{accept } H_o \mid S\right] \geq P\left[\text{accept } H_o\right] \geq P\left[\text{accept } H_o \mid S\right]$$

(4.5) if $n = n_o$

$$P\left[\text{accept } H_o \mid S\right] \geq P\left[\text{accept } H_o\right] \geq P\left[\text{accept } H_o \mid S\right] - P\left[N > n_o\right]$$

if $n > n_o$.

Therefore we shall apply (3.22) to the test S and in this manner obtain bounds for $P[\text{ accept } H_o]$ in terms of $P[N > n_o]$.

Now it can be shown that

$$\Psi_i(t) = \frac{c_i e^{ta_i}}{c_i + t} \quad \text{for } t > -c_i \text{ and } i = 1, \ldots, n_o$$

$$\mathcal{E}\left[e^{t(z_i + r)} \mid z_i \leq -r\right] = \frac{c_i}{c_i + t} \quad \text{for } r > o,\ t > -c_i \text{ and}$$

$$i = 1, \ldots, n_o$$

$$\mathcal{E}\left[e^{t(z_i - r)} \mid z_i \geq r\right] = \frac{c_i}{c_i + t}\left[1 - \frac{1 - e^{t(a_i - r)}}{1 - e^{-c_i(a_i - r)}}\right]$$

for $t > -c_i$, $o < r < a_i$

and $i = 1, \ldots, n_o$.

Suppose $t > -c_i$ for $i = 1, \ldots, n_o$ (i.e. $t > \frac{n - n_o + 1}{n_1 - n_o}$) then it follows that

$$\eta(t) = \frac{n - n_o + 1}{n - n_o + 1 + (n_1 - n_o)t} \quad \text{for } t \geq o$$

$$\delta'(t) = \frac{n - n_o + t}{n - n_o + 1 + (n_1 - n_o)t} \quad \text{for} \quad t \leq o .$$

It can be shown that

$$\sup_{r > o} \mathcal{E}\left[e^{t(z_i - r)} \middle| z_i \geq r \right] = \frac{c_i}{t+c_i} \left(1 - \frac{1-e^{ta_i}}{1-e^{-c_i a_i}} \right) \quad \text{for} \quad t > o .$$

Hence

$$\delta(t) = \max_{n_o \geq i \geq 1} \left\{ \frac{c_i}{t+c_i} \left(1 - \frac{1-e^{ta_i}}{1-e^{-c_i a_i}} \right) \right\} \text{for} \quad t \geq o.$$

Similarly we have

$$\eta'(t) = \min_{n_o \geq i \geq 1} \left\{ \frac{c_i}{t+c_i} \left(1 - \frac{1-e^{ta_i}}{1-e^{-c_i a_i}} \right) \right\} \quad \text{for} \quad t \leq o.$$

We are now in a position to apply (3.22) to the test S. It can be easily shown that for $i = 1, \ldots, n_o$

$$\Psi_i(-1) = 1 \qquad \text{if } n = n_1$$

$$\Psi_i(1) = 1 \qquad \text{if } n = n_o .$$

So for simplicity, we will use these two values of n as an illustration. First of all, it should be noted that the condition $t > c_i$ for $i = 1, \ldots, n_o$ is satisfied for both $t = -1$ and $t = 1$ when $n = n_1$ and $n = n_o$, respectively. Now assume $n = n_1$ and $t = -1$ and define

$$x = \frac{n_1 - i + 1}{n_1 - n_o} .$$

We may then write

$$\frac{c_i}{t+c_i}\left(1-\frac{1-e^{ta_i}}{1-e^{-c_i a_i}}\right)=\frac{1-\left(\frac{x-1}{x}\right)^{x-1}}{1-\left(\frac{x-1}{x}\right)^{x}}=1-\frac{1}{(x-1)\left[\left(\frac{x}{x-1}\right)^{x}-1\right]}.$$

Since $x > 1$, we can show that the last expression is increasing in x. Now x is decreasing in i so that

$$\frac{c_i}{t+c_i}\left(1-\frac{1-e^{ta_i}}{1-e^{-c_i a_i}}\right)$$

is decreasing in i. Hence,

$$\eta'(-1)=\frac{1-\left(\frac{1}{n_1-n_o+1}\right)^{\frac{1}{n_1-n_o}}}{1-\left(\frac{1}{n_1-n_o+1}\right)^{\frac{1}{n_1-n_o}+1}} \qquad \text{for } n=n_1,$$

and by a similar argument

$$\delta(1)=\frac{1-\left(\frac{n_1}{n_o}\right)^{-\frac{n_o}{n_1-n_o}-1}}{1-\left(\frac{n_1}{n_o}\right)^{-\frac{n_o}{n_1-n_o}}} \qquad \text{for } n=n_o .$$

We also have

$$\delta'(-1)=n_1-n_o+1 \qquad \text{for } n=n_1$$

$$\eta(1)=\frac{1}{n_1-n_o+1} \qquad \text{for } n=n_o .$$

The following table gives the upper and lower bounds for $P[\text{accept } H_o|S]$ based upon (3.22). The values of a and b are the same as those given in section 1.3, so that for the s.s.s.t. we have $\alpha = \beta = .05$ in each case.

TABLE 4.3

		$n = n_o$		$n = n_1$	
n_o	n_1	lower bound	upper bound	lower bound	upper bound
5	10	.877	.930	.028	.183
5	15	.896	.956	.018	.211
5	20	.902	.967	.015	.254
10	15	.895	.927	.022	.145
10	20	.914	.950	.012	.146
10	25	.916	.958	.009	.163
15	20	.902	.926	.020	.131
15	25	.921	.948	.011	.124
20	25	.905	.925	.019	.125

Finally we arrive at the bounds for P[accept H_o] by (4.5) and by using the correct value of $P[N > n_o]$. It should be noted, however, that in some cases $P[N > n_o]$ is quite large. Also the values in the following table have been truncated to [o,1].

TABLE 4.4

		$n = n_o$		$n = n_1$	
n_o	n_1	lower bound	upper bound	lower bound	upper bound
5	10	.877	1.000	.000	.183
5	15	.896	.989	.000	.211
5	20	.902	.976	.000	.254
10	15	.895	1.000	.000	.145
10	20	.914	.967	.000	.146
10	25	.916	.961	.001	.163
15	20	.902	1.000	.000	.131
15	25	.921	.961	.003	.124
20	25	.905	1.000	.000	.125

Although the calculations will be much more complicated, the same technique can be used for values of n other than n_o and n_1.

Bounds for the o.c. function may also be obtained by using the methods of section 3.4. As before, we require that $z_1, z_2, \ldots$ are independent. To ensure that $R(n_o+1) \geq A$ if $n > n_o$ and $R(n_o+1) \leq B$ if $n = n_o$, we define

$$z_{n_o+1} = \begin{cases} d & \text{if } n > n_o \\ -d & \text{if } n = n_o \end{cases}$$

where $d \geq a-b$. Suppose we are able to find a distribution function G such that

$$G \geq F_i \quad \text{for } i = 1, \ldots, n_o+1$$

where F_i is the distribution function of z_i. Then by theorem 3.1

$$P\left[\text{ accept } H_o \mid S \right] \geq P\left[\text{ accept } H_o \right]$$

where S denotes an s.p.r.t. with $z_1, z_2, \ldots$ independent and z_i having distribution function G. Next we may either approximate $P[\text{accept } H_o|S]$ or find an upper bound for it by the usual techniques. The same procedure also yields a lower bound for $P[\text{accept } H_o]$.

Returning to the s.s.s.t., let us suppose that $n > n_o$ and that we wish to obtain an upper bound for $P[\text{accept } H_o]$. Now d may be made arbitrarily large without changing the outcome of the procedure, so we may restrict attention to finding a G such that

$$G \geq F_i \quad \text{for } i = 1, \ldots, n_o.$$

Recall that

$$F_i(z) = \begin{cases} \exp [\, c_i(z-a_i) \,] & \text{for } z \leq a_i \\ 1 & \text{for } z > a_i \end{cases}$$

where c_i is decreasing and a_i is increasing. Thus $F_i(z)$ and $F_j(z)$, $i \neq j$, intersect at exactly one point, other than at $F_i(z) = F_j(z)=1$, namely

$$z = d_{ij} = \frac{c_i a_i - c_j a_j}{c_i - c_j} .$$

Also note that for $j \neq i$, $d_{ij} < \min (a_i, a_j)$. If we define

$$d_1 = \max_{i > 1} d_{1i}$$

then $F_1(z) \geq F_i(z)$ for $z \geq d_1$. By continuing in this fashion we may determine a G which will be composed of, at most, one segment from each F_i. As an illustration let $n_o = 5$ $n = n_1 = 10$. It may then be shown that

i	d_{1i}	d_{2i}	d_{3i}	d_{4i}
1				
2	-.370			
3	-.463	-.555		
4	-.613	-.735	-.915	
5	-.955	-1.15	-1.45	-1.98

and

$$G(z) = \begin{cases} F_1(z) & \text{if } z \geq -.37 \\ F_2(z) & \text{if } -.37 \geq z \geq -.555 \\ F_3(z) & \text{if } -.555 \geq z \geq -.915 \\ F_4(z) & \text{if } -.915 \geq z \geq -1.98 \\ F_5(z) & \text{if } -1.98 \geq z\ . \end{cases}$$

In order to approximate P[accept $H_o|S$] by Wald's o.c. formula, it is necessary to solve

$$\mathcal{E}_G\ (e^{zt}) = 1,$$

which is calculated as,

$$\frac{.013e^{-1.98t}}{t + 1.2} + \frac{.0672e^{-.915t} - .015e^{-1.98t}}{t + 1.4} + \frac{.1367e^{-.555t} - .0768e^{-.915t}}{t + 1.6}$$

$$+ \frac{.216e^{-.37t} - .1538e^{-.555t}}{t + 1.8} + \frac{2e^{.693t} - .24e^{-.37t}}{t + 2} = 1\ .$$

It can be verified that the solution t_o falls between -.65 and -.64. By taking the values of a and b which yield $\alpha = \beta = .05$ and applying Wald's formula, we obtain

$$P\left[\text{ accept } H_o \mid S \right] = .277.$$

It should be noted that if Wald's approximations for a and b at the five percent level are used, then

$$P\left[\text{ accept } H_o \mid S \right] = .130.$$

If we are interested in a lower bound, then the variable z_{n_o+1} will cause difficulties, since it may happen that $G = F_{n_o+1}$. So the best course of action in this case is to replace the test by another, as was done in the beginning of this section.

In conclusion we will illustrate the use of bounds (3.36). Using the notation of sections 1.2 and 3.6 we have

$$P\left[Z_i \in R_i\right] = P\left[A \leq R(i)\right] = P\left[a_i \geq u_i\right]$$

and, as was shown at the end of the previous section, this value may be found from tables of the incomplete beta function. Denote $q_i = P[Z_i \in R_i]$ and by taking the same parameter values as those given in example 4.2, we find that for $n = n_o$,

$$q_1 = o, \; q_2 = o, \; q_3 = .0046, \; q_4 = .0252, \; q_5 = .0312$$

and

$$q_j = o \text{ for } j > 5.$$

Now by using the values of $P[Z_i \in C_i]$ given in example 4.2, we have

$$1 \geq P\left[\text{accept } H_o\right] \geq .939.$$

Similarly, when $n = n_1$ we have

$$q_1 = o, \; q_2 = o, \; q_3 = .0411, \; q_4 = .3094, \; q_5 = .6231$$

and $q_j = 1$ for $j > 5$. Thus

$$.0859 \geq P\left[\text{accept } H_o\right] \geq .0175 .$$

4.4 Multivariate normal

In this section we will look at the problem of determining P[accept H_o] and $\mathcal{E}(N)$ for an s.p.r.t. in which $z_1, z_2, \ldots$ have a multivariate normal distribution. More specifically, assume that the distribution of $z_1, z_2, \ldots, z_n$ is $N[\underline{\mu}_n, \Sigma_n]$ for each n and that these distributions are consistent. We begin by finding the restrictions on $\underline{\mu}$ and Σ which are necessary in order to apply the bounds on P[accept H_o] which were given in section 3.3. Define

$$\varphi_n(t) = \mathcal{E}\left\{ e^{tZ_n} \right\}$$

and t_n as the nonzero solution of $\varphi_n(t) = 1$. Then, in the terminology of section 3.2, we require that the conjugate densities be consistent. Now

$$\varphi_n(t) = (2\pi)^{-\frac{n}{2}} |\Sigma_n|^{-\frac{1}{2}} \int \exp\left\{ tZ_n - \tfrac{1}{2}(\underline{z}_n - \underline{\mu}_n)' \Sigma_n^{-1} (\underline{z}_n - \underline{\mu}_n) \right\} d\underline{z}_n$$

$$= \exp\left\{ \frac{t^2}{2} \underline{J}_n' \Sigma_n \underline{J}_n + t \underline{J}_n' \underline{\mu}_n \right\}$$

where $\underline{J}' = (1, \ldots, 1)$. Since Σ is positive definite, $\underline{J}' \Sigma \underline{J}$ is positive and therefore

$$t_n = \frac{-2 \underline{J}_n' \underline{\mu}_n}{\underline{J}_n' \Sigma_n \underline{J}_n} \quad . \tag{4.6}$$

Hence $\varphi_n(t) = 1$ has a unique nonzero solution if and only if $\sum_{i=1}^{n} \mu_i$ is nonzero. Thus we shall that $\sum_{i=1}^{n} \mu_i \neq o$ for all n. Now it is clear that the conjugate density is $N[\underline{\mu}_n + t_n \Sigma_n \underline{J}_n, \Sigma_n]$. Therefore a necessary and sufficient condition for consistent conjugate densities is

$$(4.7)\qquad (t_n \Sigma_n \underline{J}_n)_1 = t_{n-1} \Sigma_{n-1} \underline{J}_{n-1} \qquad \text{for } n=2,3,\ldots$$

where $(\underline{x}_n)_1$ denotes the vector $\underline{x}_{n-1}$. Condition (4.7) may also be written as

$$(4.8)\qquad t_n \sum_{i=1}^{n} \sigma_{ji} = t_{n-1} \sum_{i=1}^{n-1} \sigma_{ji}$$

for $j=1,\ldots,n-1$ and $n=2,3,\ldots$. Thus, if we are given $\underline{\mu}$ and Σ, we may determine t_n for all n and decide if the conjugate densities are consistent by the use of (4.6) and (4.8). Now supposing the conjugate densities are consistent, (4.8) then implies that

$$(4.9)\qquad t_n \sum_{j=1}^{n-1} \sum_{i=1}^{n} \sigma_{ji} = t_{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sigma_{ji} .$$

By (4.6) we have

$$(4.10)\qquad t_{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sigma_{ji} = -2 \sum_{j=1}^{n-1} \mu_j$$

and

$$(4.11)\qquad t_n \sum_{j=1}^{n-1} \sum_{i=1}^{n} \sigma_{ji} = -2 \sum_{j=1}^{n} \mu - t_n \sum_{j=1}^{n} \sigma_{jn} .$$

Thus if the conjugates are consistent, we obtain by (4.9), (4.10) and (4.11) that if $\mu_n \neq o$ then

$$(4.12)\qquad t_n = \frac{-2\,\mu_n}{\sum_{j=1}^{n} \sigma_{jn}} ,$$

which is less complicated than (4.6).

In order to apply Wald's o.c. formula, as discussed in section 3.2, we must have $t_n = t_1$ for all n, and the consistency of the conjugate densities. Suppose Σ is diagonal and $\frac{\mu_n}{\sigma_{nn}} = \frac{\mu_1}{\sigma_{11}}$ for all n, then by (4.6) we have $t_n = t_1$ for all n and by (4.8) we have consistent conjugates. If $t_n = t_1$ for all n and the conjugate densities are consistent, then Σ is diagonal by (4.8), $\mu_n \neq o$ for all n by (4.6), and $\frac{\mu_n}{\sigma_{nn}} = \frac{\mu_1}{\sigma_{11}}$ for all n by (4.12). Therefore a necessary and sufficient condition for $t_n = t_1$ for all n, and consistent conjugates, is that Σ is diagonal and $\frac{\mu_n}{\sigma_{nn}} = \frac{\mu_1}{\sigma_{11}}$ for all n.

If we are not permitted to use Wald's o.c. formula, we may, however, apply the bounds given by (3.15) and (3.16) if consistency holds and either $t_n > o$ for all n or $t_n < o$ for all n. By (4.6), a sufficient condition for $t_n > o$ for all n is $\mu_n < o$ for all n, and similarly for $t_n < o$.

Next we shall consider the important case of $\mu_n = \mu$ for all n. If we can find conditions on Σ, such that the conjugates are consistent, then in light of the above remarks, the bounds (3.15) will be applicable.

By (4.6) and $\mu_n = \mu$ for all n, the necessary and sufficient condition for consistency given by (4.8) becomes

(4.13)
$$\frac{n \sum_{i=1}^{n} \sigma_{ij}}{\sum_{i=1}^{n} \sum_{k=1}^{n} \sigma_{ik}} = \frac{(n-1) \sum_{i=1}^{n-1} \sigma_{ij}}{\sum_{i=1}^{n-1} \sum_{k=1}^{n-1} \sigma_{ik}} \quad \text{for } j=1,\dots,n-1 \text{ and } n=2,3,\dots \quad .$$

We shall next show that

$$(4.14)\qquad \sigma_{in} = \sigma_{1n} \quad \text{for } i=1,\dots,n-1$$

$$(4.15)\qquad (n-1)\sigma_{1n} + \sigma_{nn} = \sigma_{n+1,n+1} + (n-1)\sigma_{1,n+1}$$

for all n are a sufficient, and possibly, necessary condition for (4.13). Now it can be shown by induction and (4.15) that

$$(4.16)\qquad \sigma_{nn} = \sum_{i=1}^{n-1} \sigma_{1i} - (n-2)\sigma_{1n} \qquad \text{for all } n.$$

From (4.14) we have

$$(4.17)\qquad \sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij} = \sum_{i=1}^{n}\sigma_{ii} + 2\sum_{i<j}\sigma_{ij} = \sum_{i=1}^{n}\sigma_{ii} +$$

$$2\sum_{i=2}^{n}(i-1)\sigma_{1i} = \sigma_{11} + \sum_{j=2}^{n}\left\{\sum_{i=1}^{j-1}\sigma_{1i} - (j-2)\sigma_{1j}\right\} + 2\sum_{i=2}^{n}(i-1)\sigma_{1i} =$$

$$\sigma_{11} + \sum_{j=2}^{n}\sum_{i=1}^{j-1}\sigma_{1i} + \sum_{i=2}^{n} i\sigma_{1i} = n\sum_{i=1}^{n}\sigma_{1i} \quad .$$

Also by (4.14) and (4.16),

$$(4.18)\qquad \sum_{i=1}^{n}\sigma_{ij} = \sigma_{1j} + \dots + \sigma_{j-1,j} + \sigma_{jj} + \sigma_{j+1,j} + \quad + \sigma_{nj} =$$

$$(j-1)\sigma_{1j} + \sigma_{jj} + \sum_{i=j+1}^{n}\sigma_{1i} = (j-1)\sigma_{1j} + \sum_{i=1}^{j-1}\sigma_{1i} - (j-2)\sigma_{1j} +$$

$$\sum_{i=j+1}^{n}\sigma_{1i} = \sum_{i=1}^{n}\sigma_{1i} \quad .$$

Now by (4.17) and (4.18) we obtain for $j=1,\ldots,n$,

$$\frac{n \sum_{i=1}^{n} \sigma_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}} = 1 ,$$

and (4.13) is therefore satisfied. If $\sigma_{nn} = \sigma_{11}$ for all n, then (4.14) and (4.15) imply that the off diagonal elements are equal. Thus the form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \ldots & \rho \\ \rho & 1 & \rho & \ldots & \rho \\ \rho & \rho & 1 & \ldots & \rho \\ \vdots & & & & \vdots \\ \rho & \rho & \rho & \ldots & 1 \end{pmatrix}$$

is the only type which has equal variances and satisfies (4.14) and (4.15).

Now let us consider an alternative approach to the problem by making use of the methods given in sections 3.4 and 3.5. Assume that $b = -a$ and $P[N < \infty] = 1$. Then define

$$\gamma_n = \begin{cases} \dfrac{\sum_{i=1}^{n-1} \sigma_{ni}}{\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sigma_{ij}} & \text{for } n=2,3,\ldots \\ 0 & \text{for } n = o \end{cases}$$

and

$$c_n = \begin{cases} \underline{a} & \text{if } \gamma_n > o \\ -a & \text{if } \gamma_n < o . \end{cases}$$

Let H' denote the hypothesis which specifies that $z_1, z_2, \ldots$ are independent and the distribution of z_n is $N[\alpha_n', \beta_n]$, where

$$\beta_n = \sigma_{nn} - \gamma_n \sum_{i=1}^{n-1} \sigma_{ni}$$

$$\alpha_n' = \mu_n + \gamma_n \left\{ c_n - \sum_{i=1}^{n-1} \mu_i \right\} .$$

Also let H'' denote the hypothesis which specifies that $z_1, z_2, \ldots$ are independent and the distribution of z_n is $N[\alpha_n, \beta_n]$ where

$$\alpha_n = \mu_n + \gamma_n \left\{ -c_n - \sum_{i=1}^{n-1} \mu_i \right\} .$$

Since the distribution of z_n given Z_{n-1} is

$$N \left\{ \mu_n + \gamma_n \left[Z_{n-1} - \sum_{i=1}^{n-1} \mu_i \right] , \quad \sigma_{nn} - \gamma_n \sum_{i=1}^{n-1} \sigma_{ni} \right\} ,$$

by theorem 3.2,

(4.19) $$P_{H'}[\text{accept } H_o] \leq P[\text{accept } H_o] \leq P_{H''}[\text{accept } H_o] .$$

Next let us assume that $\alpha_n \neq o$, $\alpha_n' \neq o$, $\beta_n \neq o$ and define

$$r_n = \frac{\beta_n}{\alpha_n} \quad \text{and} \quad s_n = \frac{\beta_n}{\alpha_n'} .$$

Then we have

$$H' : z_n \sim N[\alpha_n', s_n\alpha_n']$$

$$H'' : z_n \sim N[\alpha_n, r_n\alpha_n] .$$

Also define

$$\bar{r} = \sup_i \{ r_i \} \qquad \underline{r} = \inf_i \{ r_i \}$$

$$\bar{s} = \sup_i \{ s_i \} \qquad \underline{s} = \inf_i \{ s_i \} .$$

Now, if $\alpha_n > o$ for all n, we are able to apply the results of section 3.5 to obtain

$$(4.20) \qquad P_{H_\ell''}[\text{accept } H_0] \lessdot P_{H''}[\text{accept } H_o] \lessdot P_{H_u''}[\text{accept } H_o]$$

where

H_ℓ'' : $z_n \sim N[\alpha_n, \underline{r}\,\alpha_n]$ for all n and $z_1, z_2, \ldots$ independent

H_u'' : $z_n \sim N[\alpha_n, \bar{r}\,\alpha_n]$ for all n and $z_1, z_2, \ldots$ independent.

If $\alpha_n < o$ for all n, we may, by symmetry considerations, also apply the results of section 3.5 to obtain

$$P_{H_u''}[\text{accept } H_1] \lessdot P_{H''}[\text{accept } H_1] \lessdot P_{H_\ell''}[\text{accept } H_1] ,$$

and by assuming termination with probability one, we have (4.20). In a similar manner, if $\alpha_n' > o$ for all n, or $\alpha_n' < o$ for all n, it then follows that

$$(4.21) \qquad P_{H_\ell'}[\text{accept } H_o] \lessdot P_{H'}[\text{accept } H_o] \lessdot P_{H_u'}[\text{accept } H_o]$$

where

H_ℓ' : $z_n \sim N[\alpha_n', \underline{s}\,\alpha_n']$ for all n and $z_1, z_2, \ldots$ independent

H_u' : $z_n \sim N[\alpha_n', \bar{s}\,\alpha_n']$ for all n and $z_1, z_2, \ldots$ independent.

From the remarks made earlier, we may then use Wald's o.c. formula to obtain approximate values of $P[\text{accept } H_o]$ under H_ℓ' , H_u' , H_ℓ'' and H_u'' . Thus we find approximate upper and lower bounds for $P[\text{accept } H_o]$.

Suppose $\mu_i = \mu > o$ for all i , then $\alpha_1 > o$ and $\alpha_1' > o$. So, in order to apply the above technique, we require that $\alpha_n > o$ and $\alpha_n' > o$ for all n. If $\gamma_n > o$, then $\alpha_n = \mu + \gamma_n(-a-(n-1)\mu)$ and thus $\alpha_n > o$ implies that $o < \gamma_n < (n-1+ \frac{a}{\mu})^{-1}$. If $\gamma_n < o$, then $\alpha_n = \mu + \gamma_n(a-(n-1)\mu)$, and thus $\alpha_n > o$ requires that $\frac{a}{\mu} > n-1$ must imply that $o > \gamma_n > (n-1- \frac{a}{\mu})^{-1}$. This may be summarized as follows:

(4.22) $\alpha_n > o$ if and only if

and (a) $\gamma_n > o$ implies $\gamma_n < (n-1+ \frac{a}{\mu})^{-1}$

(b) $\gamma_n < o$ and $\frac{a}{\mu} > n-1$ implies $\gamma_n > (n-1+ \frac{a}{\mu})^{-1}$.

If $\gamma_n > o$, then $\alpha_n' = \mu + \gamma_n(a-(n-1)\mu)$ and thus $\alpha_n' > o$ and $\frac{a}{\mu} <$ n-1 requires that $\gamma_n < (n-1- \frac{a}{\mu})^{-1}$. If $\gamma_n < o$, then $\alpha_n' = \mu + \gamma_n(-a-(n-1)\mu)$, and therefore $\alpha_n' > o$. In summarizing this, we find that

(4.23) $\alpha_n' > o$ if and only if

(a) $\gamma_n > o$ and $\frac{a}{\mu} < n-1$ implies $\gamma_n < (n-1- \frac{a}{\mu})^{-1}$.

Similar results are obtainable for the case $\mu < o$.

To illustrate the above procedure, we shall consider three examples of common covariance matrices; in each case let us assume that $\mu_i = \mu$ for all i and $a > \mu > o$.

<u>Example 4.3</u> Let

$$\Sigma = \Sigma_1 \equiv \sigma^2 \begin{pmatrix} 1 & \rho & o & o & \dots & o \\ \rho & 1 & \rho & o & \dots & o \\ o & \rho & 1 & \rho & \dots & o \\ \vdots & & & & & \vdots \\ o & o & o & o & \dots & 1 \end{pmatrix} .$$

It then can be shown that

$$|\rho| \leq \frac{1}{2} \quad \text{implies} \quad |\Sigma_1| > o$$

and

$$\gamma_n = \left[2(n-2) + \frac{n-1}{\rho} \right]^{-1} \quad \text{for} \quad n > 1 .$$

If we assume that $\frac{1}{2} \geq \rho > o$, then $\gamma_n > o$, and by (4.22), $\alpha_n > o$ if $2(n-2) + \frac{n-1}{\rho} > n-1 + \frac{a}{\mu}$. Therefore, $\alpha_n > o$ for all n if

$$o < \rho < \frac{1}{1 + \frac{a}{\mu}} .$$

By (4.23), $\alpha_n' > o$ if $n-1 > \frac{a}{\mu}$ implies $(n-1)(1+\frac{1}{\rho}) > 2-\frac{a}{\mu}$. However, $(n-1)(1+\frac{1}{\rho}) \geq 1+\frac{1}{\rho} > 2 > 2-\frac{a}{\mu}$ for $n > 1$ and therefore $\alpha_n' > o$ for all n if $o < \rho \leq \frac{1}{2}$. Next let us assume that $-\frac{1}{2} \leq \rho < o$. Then $\gamma_2 < o$ and $\gamma_n > o$ for $n > 2$. By (4.22), $\alpha_n > o$ for $n > 2$ if $(n-1)(1+\frac{1}{\rho}) > 2 + \frac{a}{\mu}$ but $(n-1)(1+\frac{1}{\rho}) < o$. Therefore there does not exist a $\rho < o$ such that $\alpha_n > o$ for all n. By (4.23), if $\alpha_n' > o$ for all n then the following condition must be satisfied for all n:

$$\frac{a}{\mu} < n-1 \quad \text{implies} \quad (n-1)(1+\frac{1}{\rho}) > 2-\frac{a}{\mu} .$$

Since $-(n-1)(1+\frac{1}{\rho})$ can be made arbitrarily large for a sufficiently large n, $(n-1)(1+\frac{1}{\rho}) < 2-\frac{a}{\mu}$ is true for large n. Therefore there does not exist a $\rho < o$ such that $\alpha_n' > o$ for all n. The above may be summarized as follows:

$$\text{(4.24)} \quad \begin{array}{l} \alpha_n > o \quad \text{for all } n \quad \text{if and only if} \quad o \leq \rho < (1+\frac{a}{\mu})^{-1} \\ \alpha_n' > o \quad \text{for all } n \quad \text{if and only if} \quad o \leq \rho \leq \frac{1}{2} . \end{array}$$

Now for $\rho \geq o$ we have

$$\beta_n = \sigma^2(1 - \rho\,\gamma_n)$$
$$\alpha_n = \mu + \gamma_n(-a - (n-1)\mu)$$

$$\alpha_n' = \mu + \gamma_n(a - (n-1)\mu) .$$

Thus,

$$r_n = \sigma^2 \frac{2(n-2) + (n-1)\rho^{-1} - \rho}{\mu[2(n-2) + (n-1)\rho^{-1}] - [a + (n-1)\mu]}$$

for $\rho > o$ and $n > 1$, and $r_n = \frac{\sigma^2}{\mu}$ for $\rho = o$ or $n = 1$.

Since the numerator and denominator of r_n are linear in n, the sign of $\frac{dr_n}{dn}$ is independent of n for $n > 1$. Hence,

$$\bar{r} = \max\{ r_1, r_2, \lim_{n\to\infty} r_n \}$$

$$\underline{r} = \min\{ r_1, r_2, \lim_{n\to\infty} r_n \} .$$

Now $\lim_{n\to\infty} r_n = \frac{\sigma^2}{\mu}(2\rho + 1)(\rho + 1)^{-1}$, $r_2 = \frac{\sigma^2}{\mu}(1 - \rho^2)(1 - \rho - \rho \frac{a}{\mu})^{-1}$, and $r_1 = \frac{\sigma^2}{\mu}$. Since $(1-\rho^2)(1-\rho-\rho \frac{a}{\mu})^{-1} \geq (1-\rho^2)(1-\rho)^{-1} = (1+\rho)^2(1+\rho)^{-1} \geq (2\rho+1)(\rho+1)^{-1} \geq 1$, we have $r_2 \geq \lim_{n\to\infty} r_n \geq r_1$.

Therefore,

$$\bar{r} = \frac{\sigma^2}{\mu}(1 - \rho^2)(1 - \rho - \rho \frac{a}{\mu})^{-1}$$

$$\underline{r} = \frac{\sigma^2}{\mu} .$$

Using a similar argument,

$$\bar{s} = \max\{ s_1, s_2, \lim_{n\to\infty} s_n \}$$

$$\underline{s} = \min\{ s_1, s_2, \lim_{n\to\infty} s_n \}$$

where $s_1 = \frac{\sigma^2}{\mu}$, $s_2 = \frac{\sigma^2}{\mu}(1-\rho^2)(1-\rho+\rho \frac{a}{\mu})^{-1}$ and $\lim_{n\to\infty} s_n = \frac{\sigma^2}{\mu}(2\rho+1)(\rho+1)^{-1}$. Now, since $a > \mu > o$, $(1-\rho^2)(1-\rho+\rho \frac{a}{\mu})^{-1} \leq (1-\rho^2)(1-\rho+\rho)^{-1} \leq 1 \leq (1+2\rho)(1+\rho)^{-1}$. Therefore,

$$\bar{s} = \frac{\sigma^2}{\mu}(1 + 2\rho)(1 + \rho)^{-1}$$

$$\underline{s} = \frac{\sigma^2}{\mu}(1 - \rho^2)(1 - \rho + \rho \frac{a}{\mu})^{-1} .$$

If we define

$$H(t) = \frac{e^{at} - 1}{e^{at} - e^{-at}} ,$$

then by Wald's o.c. approximation,

$$P_{H'_\ell}[\text{accept } H_o] \doteq H\left\{ -\frac{2}{\underline{s}} \right\}$$

$$P_{H''_u}[\text{accept } H_o] \doteq H\left\{ -\frac{2}{\bar{r}} \right\}$$

where H'_ℓ , H''_u are as defined in (4.20) and (4.21). Thus we may conclude that:

(a) If $o \leq \rho \leq (1+ \frac{a}{\mu})^{-1}$ then

$$P[\text{accept } H_o] \lesssim H\left\{ -\frac{2\mu}{\sigma^2}(1-\rho-\rho \frac{a}{\mu})(1-\rho^2)^{-1} \right\}$$

(b) If $o \leq \rho \leq \frac{1}{2}$ then

$$P[\text{accept } H_o] \gtrsim H\left\{ -\frac{2\mu}{\sigma^2}(1-\rho+\rho \frac{a}{\mu})(1-\rho^2)^{-1} \right\} .$$

<u>Example 4.4</u> Let

$$\Sigma = \Sigma_2 \equiv \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & & & & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

and $1 > \rho \geq o$ will ensure that $|\Sigma_2| > o$. Therefore we will assume that $1 > \rho \geq 0$. Now $\gamma_n = (n-2+\frac{1}{\rho})^{-1}$ for $n > 1$ and hence, $\gamma_n > 0$ for $n > 1$. As in example 4.3, it can be shown that

$\alpha_n > 0$ for all n if and only if $0 \leq \rho < (1+\frac{a}{\mu})^{-1}$,

$\alpha_n' > 0$ for all n if and only if $0 \leq \rho < 1$.

Also, $r_n = \sigma^2[n-2+\frac{1}{\rho}-\rho][\mu(n-2+\frac{1}{\rho}) - (a+(n-1)\mu)]^{-1}$ for $n > 1$,

and as before, $\bar{r} = \max\{r_1, r_2, \lim_{n\to\infty} r_n\}$ and similarly for $\underline{r}$. Now $\lim_{n\to\infty} r_n = \infty$, $r_2 = \frac{\sigma^2}{\mu}(1-\rho^2)(1-\rho-\rho\frac{a}{\mu})^{-1}$, $\lim_{n\to\infty} s_n = \infty$ and $s_2 = \frac{\sigma^2}{\mu}(1-\rho^2)(1-\rho+\rho\frac{a}{\mu})^{-1}$. Therefore we may conclude that if $0 \leq \rho < 1$, then

$$P[\text{accept } H_o] \gtrsim H\left\{-\frac{2\mu}{\sigma^2}(1-\rho+\rho\frac{a}{\mu})(1-\rho^2)^{-1}\right\}.$$

<u>Example 4.5</u> Let

$$\Sigma = \Sigma_3 \equiv \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

It is known that $|\Sigma_3| > 0$ for $|\rho| < 1$. However, we shall only consider $0 \leq \rho < 1$. Now it can be shown that

$$\gamma_n = \frac{\rho(1-\rho)(1-\rho^{n-1})}{(n-1)(1-\rho^2) - 2\rho(1-\rho^{n-1})}.$$

Since the denominator of γ_n is greater than or equal to $(n-1)[(1-\rho^2) + 2\rho\log\rho] > 0$, we have that $\gamma_n \geq 0$ for all $n > 1$. In order to have $\alpha_n > 0$ and $\alpha_n' > 0$ for all n, it is required that for $n > 1$

$$(n-1+c)^{-1} > \rho(1-\rho)[(n-1)(1-\rho^2)(1-\rho^{n-1})^{-1} - 2\rho]^{-1}$$

whenever $n-1+c > o$ and where c is defined to be $\frac{a}{\mu}$ for the α_n case and $-\frac{a}{\mu}$ for the α_n' case. This is equivalent to

$$(4.25) \qquad (n-1)\left[\frac{1+\rho}{\rho}(1-\rho^{n-1})^{-1} - 1\right] > c + \frac{2}{1-\rho}$$

for all $n > 1$ whenever $n-1+c > o$. Next we define $x=n-1$, $d=\rho$, $e=(1+\rho)\rho^{-1}$ and rewrite (4.25) as

$$h(x) \equiv x[\, e(1-d^x)^{-1} - 1] > c + \frac{2}{1-d}$$

where $x \geq 1$, $o < d < 1$ and $e \geq 2$. Now

$$h'(x) = [1-d^x]^{-2}[e(1-d^x)-(1-d^x)^2+ed^x\log(d^x)]$$

and

$$g'(d^x) = e\log(d^x) + 2(1-d^x)$$

where $g(d^x) \equiv e(1-d^x) - (1-d^x)^2 + ed^x\log(d^x)$. Hence, $g'(d^x) \leq o$ since $e \geq 2$. Thus $g(d^x) \geq g(d)$ and since $\frac{\partial g(d)}{\partial e} = 1-d+d\log(d) \geq o$, we have $g(d^x) \geq 1-d^2+d\log(d^2) \geq o$. Therefore $h'(x) \geq o$, so that $h(x)$ is a nondecreasing function of x. Hence (4.25) is satisfied if

$$(4.26) \qquad (1+\rho^2)[\rho(1-\rho)]^{-1} > c + 2(1-\rho)^{-1}$$

which is equivalent to $\rho < (1+c)^{-1}$ for $c > -1$ and $\rho > (1+c)^{-1}$ for $c < -1$. Therefore we may conclude that

$$\alpha_n > o \text{ for all } n \text{ if and only if } o \leq \rho < (1+\tfrac{a}{\mu})^{-1}$$

$$\alpha_n' > o \text{ for all } n \text{ if and only if } o \leq \rho < 1.$$

If we define

$$r_k(n) = \frac{1 - \rho(1-\rho^{n-1})(1-\rho)^{-1}\gamma_n}{1 - (n-2-k)\gamma_n},$$

then

$$r_k(n) = \begin{cases} \frac{\mu}{\sigma^2} s_n & \text{if } k = \frac{a}{\mu} - 1 \\ \frac{\mu}{\sigma^2} r_n & \text{if } k = -\frac{a}{\mu} - 1 . \end{cases}$$

Next let us assume that $n \geq 2$. Then, since $\alpha_n' > o$ and $\alpha_n > o$ for all n, we have $1-(n-2-k)\gamma_n > o$ for the appropriate range of ρ corresponding to whether $k = \frac{a}{\mu} - 1$ or $k = -\frac{a}{\mu} - 1$. Therefore, since the denominator of γ_n is positive, we may write

$$r_k(n) = \frac{(n-1)(1-\rho^2) - 2\rho(1-\rho^{n-1}) - \rho^2(1-\rho^{n-1})^2}{(n-1)(1-\rho^2) - 2\rho(1-\rho^{n-1}) - \rho(1-\rho)(1-\rho^{n-1})(n-2-k)}$$

where the denominator is positive. We wish to show that $s_{n+1} - s_2 \geq o$ and this will follow if $r_k(n+1) - r_k(2) \geq o$ for $k > o$. Now $r_k(n+1) - r_k(2) \geq o$ if and only if $[n(1-\rho^2) - 2\rho(1-\rho^n) - \rho^2(1-\rho^n)^2] \times[1+\rho k] + [n(1-\rho^2) - 2\rho(1-\rho^n) - (n-1-k)\rho(1-\rho)(1-\rho^n)][\rho^2-1] \geq o$, which may be written as

(4.27) $$\rho[(1+k)f_n + g_n] \geq o$$

where

$$f_n = n(1-\rho^2) - (1-\rho^n)(1+\rho+\rho^3-\rho^{n+2})$$
$$g_n = (1-\rho)[\rho(1-\rho^n)(1+\rho^n) - n\rho^n(1-\rho^2)] .$$

Now $g_n = (1-\rho)(1-\rho^2)\rho^n[n(1-\rho) - \rho(1-\rho^n)]$ and suppose that we set $t_n = n(1-\rho) - \rho(1-\rho^n)$. Then $t_{n+1} - t_n = (1-\rho)(1-\rho^{n+1}) \geq o$ and therefore $g_n \geq o$ since $t_n \geq t_1 = (1-\rho)^2$. Next we wish to show that $f_n \geq g_n$. It follows from the definition of f_n and g_n that

$$f_n - g_n = n(1-\rho^2)(1+\rho^n(1-\rho)) - (1-\rho^n)[(1-\rho)^2 + \rho^3 + \rho^{n+1}(1-2\rho)].$$

Suppose $\rho \leq \frac{1}{2}$, then we have $f_n - g_n \geq n(1-\rho^2) - (1-\rho^n)[(1-\rho)^2 + \rho^3 + \rho^2(1-2\rho)] = (1-\rho)[n(1+\rho) - (1-\rho^n)(1-\rho+\rho^2)]$. If we set $t_n = n(1+\rho) - (1-\rho^n)(1-\rho+\rho^2)$ then it follows that $t_{n+1} - t_n \geq 1 + 2\rho^2 - 3\rho^3 + \rho^4 \geq o$. Therefore $f_n - g_n \geq (1-\rho)t_n \geq (1-\rho)t_1 = 2\rho + \rho(1-\rho)^2 \geq o$. The case $\rho \geq \frac{1}{2}$ may be shown in a similar manner. Since we have shown that $g_n \geq o$ and $f_n \geq g_n$, (4.27) follows. Also we have that $\rho[(1+k)f_n + g_n] \leq o$ for $k \leq -2$. Therefore $r_n - r_2 \leq o$ for $n \geq 2$ and since $r_k(2) = (1-\rho^2)(1+k\rho)^{-1}$ we have $s_2 \leq s_1$ and $r_2 \geq r_1$. Hence

$$\bar{r} = r_2 = \frac{\sigma^2}{\mu}(1-\rho^2)(1-\rho-\rho\,\frac{a}{\mu})^{-1}$$

and

$$\underline{s} = s_2 = \frac{\sigma^2}{\mu}(1-\rho^2)(1-\rho+\rho\,\frac{a}{\mu})^{-1} .$$

We finally conclude that

(a) if $o \leq \rho < (1+\frac{a}{\mu})^{-1}$, then

$$P[\text{accept } H_o] \ \lessdot\ H\left\{-\frac{2\mu}{\sigma^2}(1-\rho-\rho\,\frac{a}{\mu})(1-\rho^2)^{-1}\right\}$$

(b) if $o \leq \rho < 1$, then

$$P[\text{accept } H_o] \ \gtrdot\ H\left\{-\frac{2\mu}{\sigma^2}(1-\rho+\rho\,\frac{a}{\mu})(1-\rho^2)^{-1}\right\} .$$

Basing our arguments on intuitive ideas, we may obtain approximations to P[accept H_o] in the above examples as follows. If the a.s.n. is quite large, it may then be reasonable to obtain $\bar{r}$ and $\underline{r}$ from $r_{n_o}, r_{n_o+1}, \ldots$ instead of $r_1, r_2, \ldots$. Thus it is felt that for large a.s.n. an approximation for $P_{H''}$[accept H_o] and $P_{H'}$[accept H_o] is $H\left\{-\frac{2}{\lim_{n\to\infty} r_n}\right\}$ and $H\left\{-\frac{2}{\lim_{n\to\infty} s_n}\right\}$, respectively. If $\mu_i = \mu$ for all i, then $\lim_{n\to\infty} r_n = \lim_{n\to\infty} s_n$, so we then have an approximation for P[accept H_o]. Therefore, assuming large a.s.n. we have

(a) example 4.3 : $\quad P[\text{accept } H_o] \doteq H\left\{-\frac{2\mu}{\sigma^2}(1+\rho)(1+2\rho)^{-1}\right\}$

(b) example 4.4 : $\quad P[\text{accept } H_o] \doteq \frac{1}{2}$

(c) example 4.5 : $\quad P[\text{accept } H_o] \doteq H\left\{-\frac{2\mu}{\sigma^2}(1+\rho)^{-1}\right\}$.

Then if we set $\Sigma_o = \sigma^2 I$, it follows that

$\frac{1}{2} \doteq P_{\Sigma_2}[\text{accept } H_o] \gtrsim P_{\Sigma_3}[\text{accept } H_o] \gtrsim P_{\Sigma_1}[\text{accept } H_o] \gtrsim P_{\Sigma_o}[\text{accept } H_o]$.

This may have some intuitive appeal, remembering that $\rho > o$ and $\mu > o$.

Since $z_1, z_2, \ldots$ are not independent, our only available means for calculating the a.s.n. is the general method given in section 2.6. Now Z_n has a normal distribution with mean $\sum_{i=1}^{n}\mu_i$ and variance $\sum_{j=1}^{n}\sum_{i=1}^{n}\sigma_{ij}$. Thus

$$P[Z_n \in C_n] = \Phi\left\{\left[a - \sum_{i=1}^{n}\mu_i\right]\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\right]^{-\frac{1}{2}}\right\} -$$

$$\Phi\left\{\left[b - \sum_{i=1}^{n}\mu_i\right]\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\right]^{-\frac{1}{2}}\right\} .$$

Although an upper bound for $\sum_{n=n_o}^{\infty} P[Z_n \in C_n]$ must be found, there should not be any difficulty in applying the bounds given by (2.24). It should also be remarked that the general bounds of section 3.6 could equally be applied to $P[\text{accept } H_o]$.

CHAPTER V

TOMLINSON PROCEDURES

5.1 Introduction

In 1957, Tomlinson [30] proposed a simple sequential procedure for discriminating between

$$H_o: \ m \leq \mu \qquad \text{and} \qquad H_1: m > \mu$$

where m is the median of the univariate distribution F from which random observations are sequentially observed. Tomlinson's plan consists of choosing two points on the real line λ_ℓ and λ_u such that $\lambda_\ell < \mu < \lambda_u$ and defining four regions, $B_1, \ldots, B_4$, where $B_1 = (-\infty, \lambda_\ell]$, $B_2 = (\lambda_\ell, \mu]$, $B_3 = (\mu, \lambda_u]$ and $B_4 = (\lambda_u, \infty)$. Next, random observations are sequentially drawn from F until one of the following four events occurs:

(a) one observation falls into B_1

(b) two consecutive observations fall into B_2

(c) two consecutive observations fall into B_3

(d) one observation falls into B_4 .

Sampling is then terminated and H_o is accepted if either event (a) or (b) has occurred, otherwise H_1 is accepted. Suppose we define

$$\begin{aligned} p_1 &= F(\lambda_\ell) \\ p_2 &= F(\mu) - F(\lambda_\ell) \\ p_3 &= F(\lambda_u) - F(\mu) \\ p_4 &= 1 - F(\lambda_u) . \end{aligned}$$

Then Tomlinson has shown that

$$
(5.1) \quad
\begin{aligned}
P\left[\text{ accept } H_o\right] &= 1 - \frac{(1+p_2)\,(p_3^2 + p_4(1+p_3)\,)}{1 - p_2\, p_3} \\
\text{a.s.n.} &= \frac{(1+p_2)\,(1+p_3)}{1 - p_2 p_3}
\end{aligned}
$$

and has observed that the a.s.n. has an upper bound of three. He further suggests that $\lambda_\ell = \mu - 2\sigma$ and $\lambda_u = \mu + 2\sigma$ where σ is the standard deviation of an observation. If we use these values of λ_u, λ_ℓ and if F is a normal distribution with mean θ and variance σ^2, then it can be shown that P[accept H_o] is approximately equal to .99, .95, .05, .01 for $\theta = \mu - 2\sigma$, $\mu - \sigma$, $\mu + \sigma$, $\mu + 2\sigma$ respectively.

Tomlinson's procedure, along with others of an equally simple nature, was considered by Craig [11], who pointed out a major weakness in Tomlinson's scheme. The difficulty is, that given two values of θ and their corresponding required probabilities of acceptance, there does not appear to be a straight forward method of determining λ_ℓ and λ_u (assuming they exist) which will guarantee these probabilities. In practical applications, however, this problem may be attacked by trial-and-error methods using high speed computers.

The remaining part of this chapter is concerned with procedures of Tomlinson's type, gereralized and applied to a multiple decision problem.

5.2 A simple multiple decision procedure.

Let $\mathcal{B} = \{B_1, \ldots, B_\ell\}$ be a finite collection of disjoint Borel sets on the real line, and G_θ a distribution function depending on the parameter θ, which may be a vector. Suppose we have m distinct simple hypotheses $H_1, H_2, \ldots, H_m$, and let H_i specify that $\theta = \theta_i$ for $i = 1, \ldots, m$. Next we associate with each H_i a set $\mathcal{B}_i$, $i = 1, \ldots, m$, where the $\mathcal{B}_i$, $i = 1, \ldots, m$, are disjoint sub-collections of $\mathcal{B}$, such that $\cup\mathcal{B}_i = \mathcal{B}$. We now propose the following sequential procedure.

At the nth stage a real-valued random observation x_n is taken from G_θ and if x_n is the k_ith successive observation to fall in B_i for some $i = 1, \ldots, \ell$, then sampling is terminated, otherwise sampling continues to stage $n + 1$ with the x_{n+1}th observation from G_θ. If sampling is terminated at the nth stage with x_n falling in B_i, then the hypothesis H_j is accepted where $\mathcal{B}_j$ is the unique set to which B_i belongs.

By applying some results from recurrent event theory (see chapter 13 of Feller [12]), the mean and variance of the sample size, N, and $P[\text{ accept } H_i\,]$ may be shown to have quite simple expressions. To begin, we define $p_i = P[\,x \in B_i]$ and $q_i = 1-p_i$ where the distribution of x is G_θ. Next, let the recurrent event E_i represent the occurrence of k_i successive observations falling in the set B_i, and let $E = \cup E_i$. Then if $U_i(s)$ denotes the probability generating function of the occurrence of E_i at trial n and $U(s)$ the probability generating function of E at trial n, we have

$$U_i(s) = \frac{1-s+q_i p_i^{k_i} s^{k_i+1}}{(1-s)(1-p_i^{k_i} s^{k_i})}$$

(5.2)

$$U(s) = \sum_{i=1}^{\ell} U_i(s) - (\ell-1)\,.$$

If f_n is the probability of the first occurrence of the event E at trial n, then it follows that

$$F(s) = \sum_{i=1}^{\infty} f_i \, s^i = 1 - \frac{1}{\sum_{i=1}^{\ell} U_i(s) - (\ell-1)} .$$

Suppose we define

$$Q(s) = \frac{1-F(s)}{1-s} .$$

Then, since

$$P\Big[N = n \Big] = f_n$$

we have

$$\mathcal{E}(N) = Q(1)$$

$$\mathrm{Var}(N) = 2Q'(1) + Q(1) - Q^2(1).$$

Now,

$$Q(s) = \frac{1}{\sum_{i=1}^{\ell} U_i(s)\,(1-s) - (\ell-1)\,(1-s)}$$

and from (5.2) it follows that

$$Q(1) = \left\{ \sum_{i=1}^{\ell} \frac{1 - p_i}{(\frac{1}{p_i})^{k_i} - 1} \right\}^{-1}$$

$$Q'(1) = Q^2(1) \left\{ 1-\ell + \sum_{i=1}^{\ell} \frac{q_i p_i^{2k_i} - (k_i+1) q_i p_i^{k_i} + 1 - p_i^{k_i}}{(1 - p_i^{k_i})^2} \right\} .$$

Therefore,

$$\mathcal{E}(N) = \left\{ \sum_{i=1}^{\ell} \frac{1 - p_i}{\left(\frac{1}{p_i}\right)^{k_i} - 1} \right\}^{-1}$$

(5.3

$$\mathrm{Var}(N) = \mathcal{E}(N) + \mathcal{E}^2(N)\left\{1-2\ell + 2\sum_{i=1}^{\ell} \frac{q_i p_i^{2k_i} - (k_i+1) q_i p_i^{k_i} + 1 - p_i^{k_i}}{(1 - p_i^{k_i})^2}\right\}.$$

In order to obtain P[accept H_i], we define the recurrent event E' to be $\bigcup_{j \neq i} E_j$ and

$$t_n = P\left\{ E_i \text{ occurs for the first time at trial } n \text{ while } E' \text{ has not occurred} \right\}$$

$$r_n = P\left\{ E' \text{ occurs for the first time at trial } n \text{ while } E_i \text{ has not yet occurred} \right\}$$

$$v_n = P\left\{ E_i \text{ occurs for the first time at trial } n \right\}$$

$$w_n = P\left\{ E' \text{ occurs for the first time at trial } n \right\}.$$

Now, by letting $T(s)$ denote the probability generating function of t_n, and by using a similar notation for the others, we have

$$T(s) = V(s) - R(s)\, V(s)$$

$$R(s) = W(s) - T(s)\, W(s)$$

and therefore

$$T(s) = \frac{V(s)\,[1-W(s)\,]}{1 - V(s)\,W(s)} .$$

Since

$$V(s) = 1 - (U_i(s)\,)^{-1}$$

$$W(s) = 1 - (\sum_{j\neq i} U_j(s) - (\ell-2)\,)^{-1}$$

we arrive at

$$T(s) = \left\{ \frac{\sum_{j\neq i} (1-s)\,U_j(s) - (1-s)\,(\ell-2)}{(1-s)\,U_i(s) - (1-s)} \right\}^{-1} .$$

Thus it follows that

$$(5.4) \qquad P\left\{ E_i \text{ occurs before } E' \text{ occurs} \right\} = T(1) =$$

$$\left\{ \sum_{j\neq i} q_j\, p_j^{\,k_j} \,(1-p_j^{\,k_j}\,)^{-1} \right\}^{-1} .$$

If we define

$$c_i = \frac{1 - p_i}{(\frac{1}{p_i})^{k_i} - 1} \qquad \text{for } i = 1, \ldots, \ell$$

$$J_i = \left\{ j \,|\, B_j \in \mathcal{B}_i \right\}$$

then by (5.3) and (5.4) we have

$$\mathcal{E}(N) = \left[\sum_{i=1}^{\ell} c_i \right]^{-1}$$

(5.5)
$$P\left[\text{accept } H_i \right] = \begin{cases} \mathcal{E}(N)\left[\sum_{j\in J_i} c_j\right] & \text{if } \sum_{i=1}^{\ell} p_i \neq 0 \\ 0 & \text{if } \sum_{i=1}^{\ell} p_i = 0. \end{cases}$$

It should be noted that if $k_1 = k_4 = 1$, $k_2 = k_3 = 2$, $J_1 = \{1,2\}$ and $J_2 = \{3,4\}$, we obtain Tomlinson's formulae given in (5.1). Also, as with Tomlinson's procedure, if $\mathcal{B}$ is a cover for the real line, then $\mathcal{E}(N)$ has an upper bound depending upon ℓ and $k = \max_i k_i$. This can be shown as follows. Define

$$v(p) = \frac{1 - p}{(\frac{1}{p})^k - 1} \quad ;$$

then $c_i \geq v(p_i)$, since c_i is a decreasing function of k_i. Because we have assumed that $\mathcal{B}$ is a cover, it follows that $\max_i p_i \geq \frac{1}{\ell}$. Now, since $v(p)$ is an increasing function of p,

$$\sum_{i=1}^{\ell} c_i \geq v(\tfrac{1}{\ell})$$

and therefore

$$\mathcal{E}(N) = (\sum_{i=1}^{\ell} c_i)^{-1} \leq (v(\tfrac{1}{\ell}))^{-1} = \frac{\ell}{\ell-1}(\ell^k-1).$$

It should be remarked that this bound is not particularly good and equality can never be achieved because $\sum_{i=1}^{\ell} c_i > v(\frac{1}{\ell})$.

An immediate criticism of this procedure is that decisions will often be based upon only a small portion of the available data. It is felt that a similar procedure in which "a total of k_i observations fall in B_i" replaces "k_i successive observations fall in B_i" is preferable in many cases. Unfortunately, we would then not be able to express simply the exact properties of the test as was done in (5.5)[1/]. Thus in applying our procedure, we should consider attempting to keep k small, (3 or 4), and varying the B_i in order to obtain the desired error levels. As an example, Tomlinson has shown that his procedure, while not significantly altering the o.c. function, has a uniformly smaller a.s.n. than a modified procedure (for use where σ is unknown) in which $\lambda_\ell = -\infty$ and $\lambda_u = \infty$.

Suppose G_θ possesses a probability density function g_θ . Then the type of regions described by Armitage [3] should be helpful in determining a reasonable choice of $\mathcal{B}$. These regions are as follows. If we define

$$R_i = \left\{ x \mid \frac{g_{\theta_i}(x)}{g_{\theta_j}(x)} \geq 1 \quad \text{for } j \neq i \right\} \quad \text{for } i = 1, \ldots, m$$

then it seems plausible to associate an observation belonging to

---

1/ Tsao [31] has obtained the properties of this test for the case of $m = \ell = 2$. He also discusses the optimum choice of B_1 and B_2.

R_i with the hypothesis H_i. Thus, as a starting point for determining $\mathcal{B}$, we may require

$$\bigcup_{j \in J_i} B_j \subset R_i \qquad \text{for } i = 1, \ldots, m.$$

Finally, we shall conclude this chapter with the following illustration.

Example 5.1

Suppose G_θ is a normal distribution with mean θ and variance one and that $m = 3$ with

$$H_1: \theta=-1 \qquad H_2: \theta=0 \qquad H_3: \theta=1.$$

Although there are many reasonable choices of $\mathcal{B}$, we shall, for this example, restrict ourselves to the following class. Define

$$B_1 = (-\infty, -t)$$
$$B_2 = (-t, t)$$
$$B_3 = (t, \infty)$$

and set $\mathcal{B}_i = \{B_i\}$ for $i = 1,2,3$. If we further restrict ourselves to only those cases in which $k_1 = k_3$, then we may obtain equal error probabilities by finding that value of t such that

$$(5.6) \qquad P\left[\text{accept } H_i \mid H_i\right] = d \quad \text{for } i = 1,2,3$$

and some d. Now, by the symmetry of the problem we will have $\mathcal{E}(N|H_1) = \mathcal{E}(N|H_3)$. The following table illustrates these values for several choices of k_1 and k_2.

TABLE 5.1

k_1	k_2	t	d	$\mathcal{E}(N\|H_1)$	$\mathcal{E}(N\|H_2)$
1	1	.803	.578	1.00	1.00
1	2	1.264	.630	1.59	1.79
1	3	1.514	.658	2.17	2.63
1	4	1.680	.677	2.73	3.48
2	1	.381	.619	2.00	2.08
2	2	.750	.697	3.11	3.61
2	3	.966	.746	4.28	5.32
2	4	1.115	.779	5.49	7.13
2	5	1.227	.804	6.74	9.00
2	6	1.317	.824	8.03	10.91
3	1	.216	.655	3.26	3.84
3	2	.522	.744	5.01	6.55
3	3	.720	.801	6.98	9.81
3	4	.860	.839	9.12	13.38
3	5	.968	.867	11.41	17.15

In figure 5.1 we have plotted the values of d against $\max\{ \mathcal{E}(N|H_1), \mathcal{E}(N|H_2) \}$ for the various pairs (k_1, k_2) given in table 5.1 (a few additional pairs have also been given). In this figure we have drawn a line connecting (k_1, k_2) and (k_1, k_2+1). Now, if our criterion for selecting a procedure is to pick the one which has the smallest $\max\{ \mathcal{E}(N|H_1), \mathcal{E}(N|H_2) \}$, then figure 5.1 will be quite helpful. One simply decides which value of d he desires and then finds the lowest point in the figure which is on or to the right of the vertical line running through his choice of d.

Figure 5.2 is the same as figure 5.1 except that "a total of k_i observations in B_i" has replaced "k_i successive observations in B_i". A different value of t was used in order to ensure that (5.6) would hold. In comparing these two figures, we observe that for the larger values of d and hence larger k_1, k_2, the procedure represented by figure 5.1 is not particularly good. This seems to substantiate to some extent what was said earlier.

One further point of interest is the following. For the test in table 5.1 given by $k_1 = 2$ and $k_2 = 2$, we find that $\mathcal{E}_\theta(N)$ does not achieve its maximum at $\theta = -\frac{1}{2}$ and $\theta = \frac{1}{2}$ as in the test given by Sobel and Wald [27]. In fact, from table 5.2, it appears as though the maximum occurs at $\theta = 0$.

TABLE 5.2

θ	0	.1	.2	.3	.4	.5
$\mathcal{E}_\theta(N)$	3.611	3.605	3.589	3.561	3.523	3.475

θ	.6	.7	.8	.9	1.0
$\mathcal{E}_\theta(N)$	3.417	3.350	3.276	3.195	3.109

FIGURE 5.1

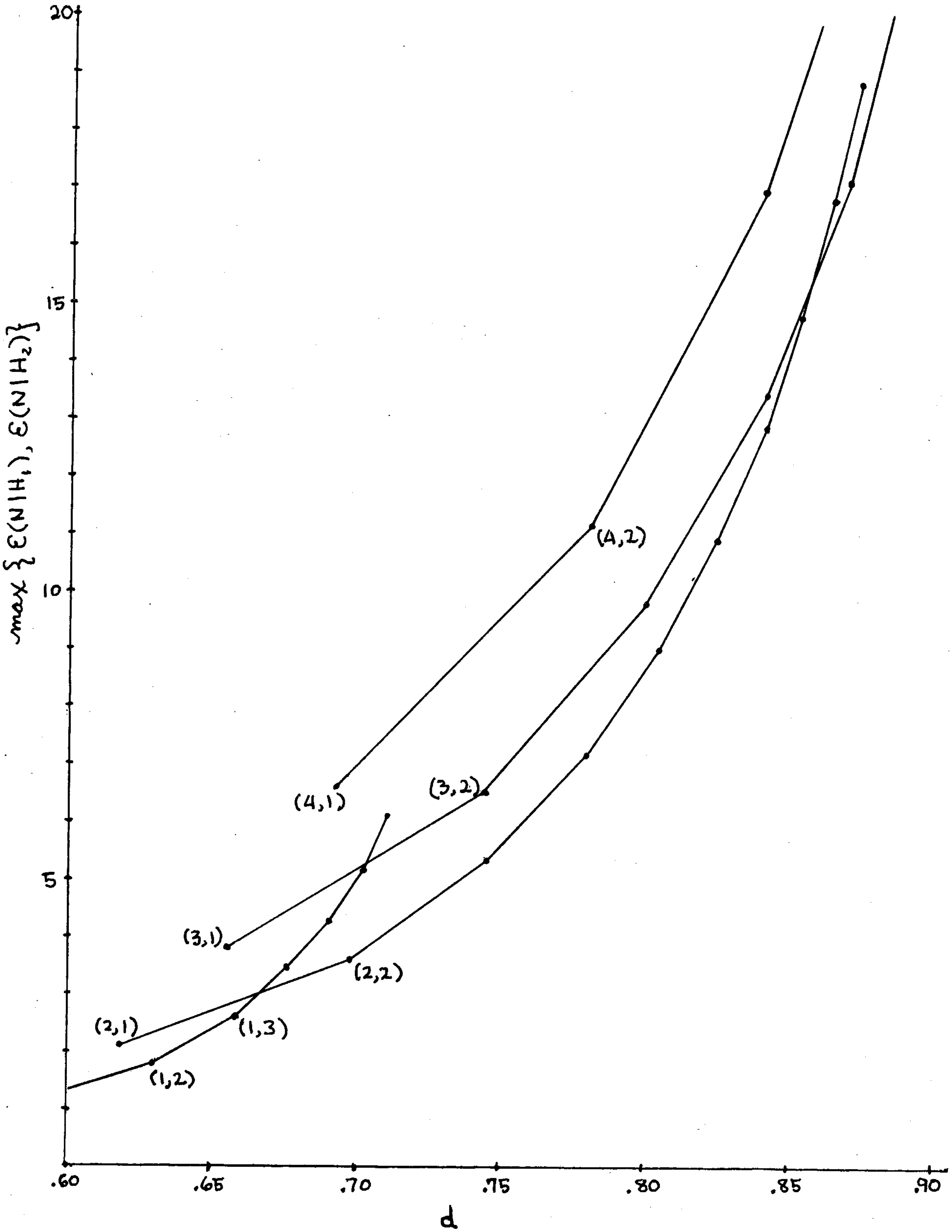
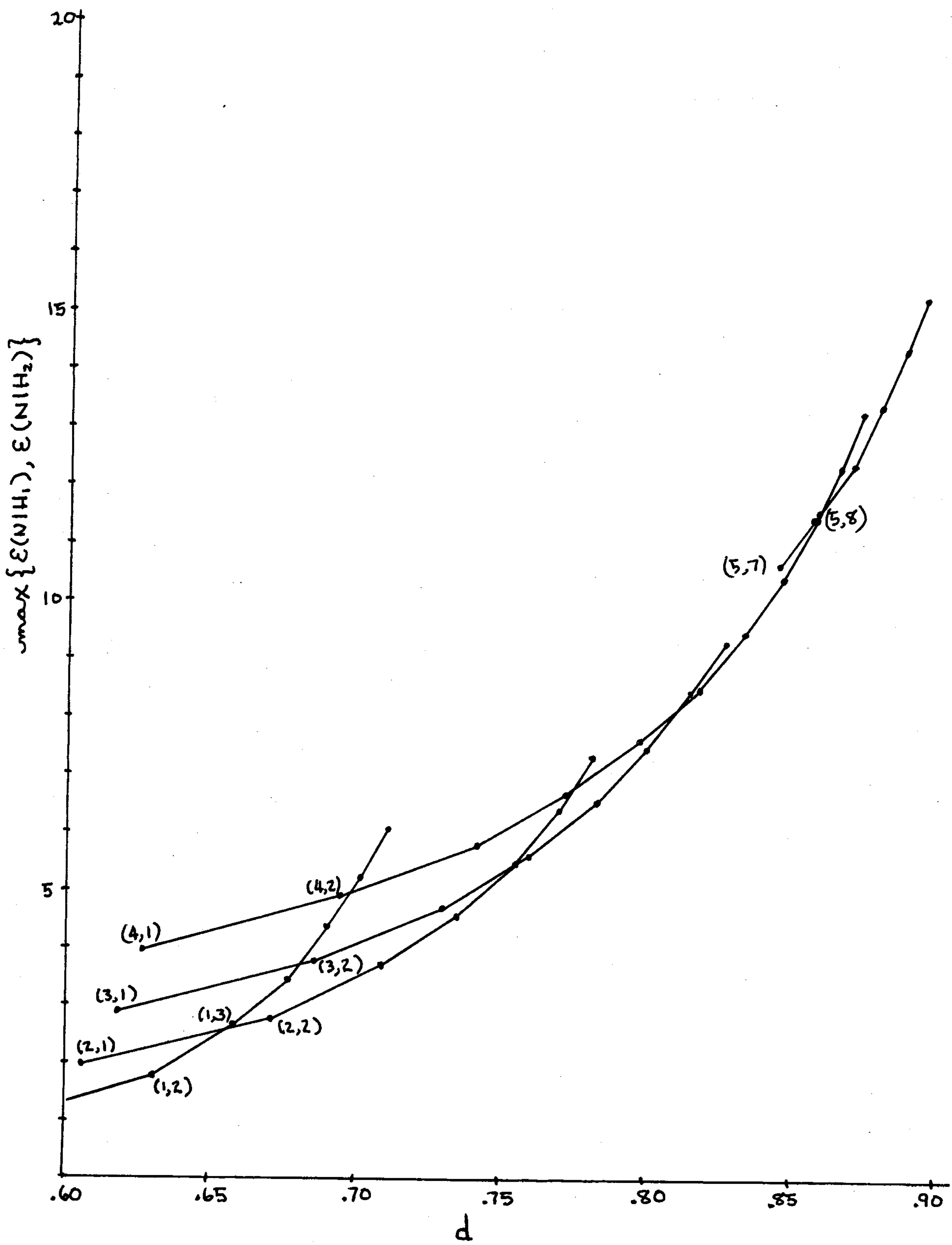


FIGURE 5.2



20
15
10
5
max {ε(N1H1), ε(N1H2)}
(5,7)
(5,8)
(4,1)
(4,2)
(3,1)
(3,2)
(1,3)
(2,2)
(2,1)
(1,2)
.60
.65
.70
.75
.80
.85
.90
d

BIBLIOGRAPHY

[1] Anderson, T. W., "A modification of the sequential probability ratio test to reduce the sample size", Annals of Mathematical Statistics, Volume 31 (1960), p.p 165-197.

[2] Apostol, T. M., Mathematical Analysis, Addison-Wesley Publishing Company, Inc., London, 1960.

[3] Armitage, P., "Sequential analysis with more than two alternative hypotheses and its relation to discriminant function analysis", Journal of the Royal Statistical Society, Series B, Volume 12 (1950), pp. 137-144.

[4] Arnold, K. J.(Ed.), Tables to facilitate Sequential t-tests, (National Bureau of Standards Applied Mathematics Series,7) Washington: U.S. Dept. of Commerce, 1951.

[5] Baker, A. G., "Properties of some tests in sequential analysis", Biometrika, Volume 37 (1950), pp. 334-346.

[6] Bartholomew, D.J., "A sequential test for randomness of intervals", Journal of the Royal Statistical Society, Series B, Volume 18 (1956), pp. 95-103.

[7] Bhate, D. H., "Sequential analysis with special reference to distributions of sample size", London University Ph.D. Thesis,1955.

[8] Bhate, D. H., "Approximations to the distribution of sample size for sequential tests, I. Tests of simple hypotheses", Biometrika, Volume 46 (1959), pp. 130-138.

[9] Blom, G., "A generalizarion of Wald's fundamental identity", Annals of Mathematical Statistics, Volume 20 (1949), pp.439-444.

[10] Chow, Y. S., Robbins, H. and Teicher, H., "Moments of randomly stopped sums", Annals of Mathematical Statistics, Volume 36(1965), pp.789-799.

[11] Craig, C. C., "On a class of simple sequential tests on means", Technometrics, Volume 4 (1962), pp. 345-359.

[12] Feller, W., An Introduction to Probability Theory and Its Applications, Volume I, Second Edition, John Wiley and Sons,Inc. New York, 1958.

[13] Hall, W. J., Wijsman, R.A. and Ghosh, J. K., "The relations between sufficiency and invariance with applications in sequential analysis", Annals of Mathematical Statistics, Volume 36 (1965), pp. 575-614.

[14] Hoeffding, W., "A lower bound for the average sample number of a sequential test", Annals of Mathematical Statistics, Volume 24 (1953), pp. 127-130.

[15] Hoeffding, W., "Lower bounds for the expected sample size and the average risk of a sequential procedure", Annals of Mathematical Statistics, Volume 31 (1960), pp. 352-368.

[16] Johnson, N. L., "A proof of Wald's theorem on cumulative sums", Annals of Mathematical Statistics, Volume 30 (1959), pp.1245-1247.

[17] Johnson, N. L., "Estimation of Sample Size", Technometrics, Volume 4 (1962), pp. 59-67.

[18] Kemp, K. W., "Formulae for calculating the operating characteristic and the average sample number of some sequential tests",

Journal of the Royal Statistical Society, Series B, Volume 20 (1958), pp. 379-386.

[19] Lehmann, E. L., Testing Statistical Hypotheses, John Wiley and Sons, Inc., New York, 1959.

[20] Mallows, C. L., "Sequential discrimination", Sankhyā, Volume 12 (1953), pp. 321-338.

[21] Page, E. S., "An improvement to Wald's approximation for some properties of sequential tests", Journal of the Royal Statistical Society, Series B, Volume 16 (1954), pp. 136-139.

[22] Ray, W. D., "Sequential analysis applied to certain experimental designs in the analysis of variance", Biometrika, Volume 43 (1956), pp. 388-403.

[23] Rushton, S., "On a sequential t-test", Biometrika, Volume 37 (1950), pp. 326-333.

[24] Rushton, S., "On a two-sided sequential t-test", Biometrika, Volume 39 (1952), pp. 302-308.

[25] Siskind, V., "On certain suggested formulae applied to the sequential t-test", Biometrika, Volume 51 (1964), pp. 97-106.

[26] Slater, L. J., Confluent Hypergeometric Functions, Cambridge University Press, Cambridge, 1960.

[27] Sobel, M. and Wald, A., "A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution", Annals of Mathematical Statistics, Volume 20 (1949), pp. 502-522.

[28] Stein, C., "A note on cumulative sums", Annals of Mathematical Statistics, Volume 17 (1946), pp. 498-499.

[29] Tallis, G. M. and Vagholkar, M K., "Formulae to improve Wald's approximation for some properties of sequential tests", Journal of the Royal Statistical Society, B, Volume 27(1965), pp 74-81.

[30] Tomlinson, R. C., "A simple sequential procedure to test whether average conditions achieve a certain standard", Applied Statistics, Volume 6 (1957), pp. 198-207.

[31] Tsao, C. K., "A simple sequential procedure for testing statistical hypotheses", Annals of Mathematical Statistics, Volume 25 (1954), pp. 687-702.

[32] Wald, A., "On cumulative sums of random variables", Annals of Mathematical Statistics, Volume 15 (1944), pp. 283-296.

[33] Wald, A., "Sequential tests of statistical hypotheses", Annals of Mathematical Statistics, Volume 16 (1945), pp. 117-186.

[34] Wald, A., "Sequential method of sampling for deciding between two courses of action", Journal of the American Statistical Association, Volume 40 (1945), pp. 277-306.

[35] Wald, A., "Some generalizations of the theory of cumulative sums of random variables", Annals of Mathematical Statistics, Volume 16 (1945), pp. 287-293.

[36] Wald, A., "Some improvements in setting limits for the expected number of observations required by a sequential probability test", Annals of Mathematical Statistics, Volume 17 (1946), pp. 466-474.

[37] Wald, A., "Differentiation under the expectation sign in the fundamental identity of sequential analysis", Annals of Mathematical Statistics, Volume 17 (1946), pp. 493-497.

[38] Wald, A., Sequential Analysis, John Wiley and Sons, Inc., New York, 1947.

[39] Weiss, L., "Testing one simple hypothesis against another", Annals of Mathematical Statistics, Volume 24 (1953), pp. 273-281.

[40] Williams, J. D., "An approximation to the probability integral", Annals of Mathematical Statistics, Volume 17 (1946), pp. 363-365.