SAMPSON, Charles Berlin, 1939-

CONTRIBUTIONS TO THE STUDY OF SEQUENTIAL COVARIANCE ANALYSIS.

Iowa State University, Ph.D., 1968
Statistics

University Microfilms, Inc., Ann Arbor, Michigan

CONTRIBUTIONS TO THE STUDY OF SEQUENTIAL
COVARIANCE ANALYSIS

by

Charles Berlin Sampson

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Ames, Iowa

1968

TABLE OF CONTENTS

I. INTRODUCTION

It is natural that an investigator conducting an experiment should wish to follow the results closely, as they become available, in order that decisions may be made as early as possible. The experiment may then be terminated with an economy of experimental material. Reductions in sample size to decision may be important for ethical and/or economic reasons in a wide variety of practical situations.

Armitage (3), Hajnal (26), and many other authors have pointed out that medical trials are often characterized by ethical difficulties. An early termination of a medical trial and the immediate application of a superior treatment to all persons with a particular affliction is clearly highly desirable.

In addition to such ethical considerations, it may be that the experimental units are very expensive or that the testing is so extensive and repetitive that a slight saving of observations per sample may develop into considerable long-term economic savings. These considerations, in conjunction with military programs which test to destruction, provided the impetus for Wald's development of the Sequential Probability Ratio Test.

Another motivation for sequential experimentation is that the experimental units may occur rarely. For example, a clinician may have the problem of deciding between two treatments for a rare disease. Patients with this disease may be admitted at intervals of fairly long duration. Alternatively, the preparation of an experimental unit may be time consuming. For these and other similar reasons the experimenter might not at any one time have at his disposal a group of experimental units permitting the establishment of a fixed sample size experiment. These types of situations dictate sequential experimentation and, correspondingly, where appropriate, some form of statistical sequential analysis of data.

In applications of fixed sample size theory the use of concomitant information (for example, in the analysis of covariance) has frequently resulted in an increase in the precision of the experiment. Intuitively, the use of relevant concomitant information would seem to increase the amount of information extracted from an experimental unit and result in either a reduction of the number of experimental units needed for given information or more information for a fixed number of experimental units. Correspondingly, in the context of a sequential experiment, it may be expected that the appropriate utilization of concomitant information should also result in a decision with fewer observations for given Type I and Type II errors.

In this thesis we are interested in developing sequential tests for the comparison of two treatments utilizing concomitant information. We consider some generalizations of Wald's

Sequential Probability Ratio Test (SPRT) in order to develop these tests. In Chapter II we present a definition of Wald's SPRT and describe some of its elegant properties. Chapter III contains the statement of the basic problem of the thesis, a discussion of weight-functions and prior distributions, and the development of sequential multiple covariance tests. Next, in Chapter IV, fixed sample size sufficiency is used to obtain sequential multiple covariance analyses for a reformulation of the probability model of the basic problem discussed in Chapter III. Chapter V contains a discussion of two-sample analyses and a derivation of Hajnal's two-sample t-test via weight-functions. The derivation is then extended to include a number, p say, of covariates. Finally, Chapter VI contains a discussion of some of the theoretical problems incurred in testing composite hypotheses in sequential analysis and an empirical sampling approach to the solution of these theoretical problems.

II. THE SEQUENTIAL PROBABILITY RATIO TEST

A. Introduction

Johnson and Leone (40) present a broad definition of a sequential procedure as follows

> "A sequential procedure is any procedure in which the final pattern of the data depends in some way on decisions which are based on the data themselves as they become available."

Cornfield (17) defines a sequential trial as

> "...any form of data collection in which the decision to continue or discontinue further collection depends in some sense on the information previously obtained."

Wald (69) writes

> "Sequential analysis is a method of statistical inference whose characteristic feature is that the number of observations required by the procedure is not determined in advance of the experiment."

The addition of sequential methods as defined above considerably broadens the range of experimental plans which one can use in designing an investigation. It has been shown, in some situations (for example, Wald (69, p. 57)), that by intelligent use of appropriate sequential methods, the cost in money and time of investigations can often, on the average, be reduced by introducing rules for deciding when we have enough evidence to reach a useful decision and thereby avoid the collection of superfluous data.

The definitions presented above are in order of decreasing generality. As one might expect, Wald's definition of sequential analysis is the most amenable to theoretical development. In fact, modern techniques of sequential analysis are largely inspired by the work of Wald although the first idea of sequential procedure dates back to the late 1920's when H. F. Dodge and H. G. Romig constructed a double sampling procedure. Motivated by the need to reduce the amount of effort necessary in the acceptance sampling of military supplies, Wald discovered the Sequential Probability Ratio Test (SPRT) in 1943.

In the remainder of this chapter we shall describe the SPRT and catalogue some of the SPRT's elegant properties which we find relevant to the present study. For further details the reader is referred to Wald (69), and pertinent literature as found in, for example, the bibliographies by Johnson (37) and Wetherill (74).

B. Description of the SPRT

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution which has the probability density function (p.d.f.) $f(x; \theta)$ where $\theta \epsilon [\theta; \theta = \theta_o, \theta_1]$ and θ_o and θ_1 are two known points contained in a space Ω of points $[\theta]$. Let the joint p.d.f. of $X_1, X_2, \ldots, X_n$ be denoted by

$$L(\theta,n) = f(X_1,\theta)\ f(X_2;\theta)\ \ldots\ f(X_n;\theta).$$

Let us now suppose that n is not fixed in advance and instead assume that n is a realization of a random variable N with sample space $[n; n = 1, 2, 3, \ldots]$. Let A and B be two constants such that $0<B<1<A$. Let us observe, in sequence, realizations, $x_1, x_2, x_3, \ldots, x_n, \ldots$, of mutually stochastically independent random variables, $X_1, X_2, \ldots, X_n, \ldots$, and compute the sequence

$$\frac{L(\theta_1;\ 1)}{L(\theta_o;\ 1)},\ \frac{L(\theta_1;\ 2)}{L(\theta_o;\ 2)},\ \frac{L(\theta_1;\ 3)}{L(\theta_o;\ 3)},\ \ldots .$$

The Wald SPRT procedure is then defined by the following rules.

i) The hypothesis H_o: $\theta = \theta_o$ is rejected and the hypothesis H_1: $\theta = \theta_1$ is accepted if and only if there exists a positive integer n so that the vector of realizations $(x_1, x_2, \ldots, x_n)$ is contained in C_n where

$$C_n = [(x_1, x_2, \ldots, x_n):\ B < \frac{L(\theta_1;\ j)}{L(\theta_o;\ j)} < A,\ j = 1, \ldots, n-1$$

$$\text{and } \frac{L(\theta_1;\ n)}{L(\theta_o;\ n)} \geq A].$$

ii) We shall accept the hypothesis H_o: $\theta = \theta_o$ and reject H_1: $\theta = \theta_1$ if and only if there exists a positive integer n so that $(x_1, x_2, \ldots, x_n)$ is contained in D_n where

$$D_n = [(x_1, x_2, \ldots, x_n):\ B < \frac{L(\theta_1,\ j)}{L(\theta_o,\ j)} < A,\ j = 1, 2, \ldots, n-1$$

$$\text{and } \frac{L(\theta_1;\ n)}{L(\theta_o;\ n)} \leq B]$$

iii) We continue to observe sample items as long as $(x_1, x_2, \ldots, x_n)$ is contained in the complement of $D_n \cup C_n$.

C. Some Properties of the SPRT

To facilitate discussion of sequential procedures it will be convenient to adapt in part the notation used in Johnson and Leone (40). Accordingly, we denote by $S(H_o, H_1; P(I), P(II))$ any sequential test of H_o versus H_1 with $P(I) = P[H_1 | H_o]$ and $P(II) = P[H_o | H_1]$ where $P[H_i | H_j]$ is the probability of accepting H_i if H_j is true.

Property 2.1 (Wald, 69): The SPRT as defined in Section B of this chapter terminates with probability one.

Property 2.2 (Wald, 69): The following inequalities hold.

$$A \leq \frac{1 - P(II)}{P(I)}$$

$$B \geq \frac{P(II)}{1 - P(I)}.$$

Property 2.3 (Wald, 69): If the probabilities of error P(I) and P(II) are small, and if A and B are chosen such that

$$A = \frac{1 - P(II)}{P(I)}$$

$$B = \frac{P(II)}{1 - P(I)}$$

then the actual error probabilities achieved by the SPRT are approximately equal to P(I) and P(II). In fact, if we denote the actual values of $P[H_1 | H_o]$ and $P[H_o | H_1]$ by $P'(I)$ and $P'(II)$ respectively then

$$P'(I) + P'(II) \leq P(I) + P(II).$$

A fact which may be important for some extensions of the SPRT may here be noted. This is that the Properties 2.2 and 2.3 hold even if one removes the requirements of independence of observations stated in the definition of the SPRT in Section B of this chapter. This can be verified by examining the proofs of these properties set out in Wald (69).

It is frequently suggested (for example, Cox (20) and David and Kruskal (22)) that in order to use Wald's boundaries (Property 2.2) for the SPRT one must prove termination with certainty. Hall, Wijsman, and Ghosh (28), however, point out that the requirements on the error probabilities as approximate upper bounds, rather than approximate equalities, are fulfilled regardless of the certainty of termination.

<u>Property 2.4</u> (Wald, 69): For a SPRT, say $S(H_o, H_1; P(I), P(II))$, the operating characteristic curve is approximately

$$P[\theta\colon S(H_o, H_1; P(I), P(II))] = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

where $P[\theta\colon S(H_o, H_1; P(I), P(II)]$ is the probability of deciding that the value of the parameter is θ_o when it is, in fact, $\theta \epsilon \Omega$ and $h(\theta)$ is the solution of

$$\int \left[\frac{f(x;\ \theta_1)}{f(x;\ \theta_o)}\right]^{h(\theta)} f(x;\ \theta)dx = 1.$$

<u>Property 2.5</u> (Wald, 69): An approximation to the average sample number for any parameter point $\theta \epsilon \Omega$, given a $S(H_o, H_1; P(I), P(II))$ for $H_o\colon \theta = \theta_o$, $H_1\colon \theta = \theta_1$, is

$$E[N|\theta] = \frac{P[\theta]\ln B + (1 - P[\theta])\ln A}{E[\ln f(x;\ \theta_1) - \ln f(x;\ \theta_o)|\theta]}$$

where $P[\theta]$ and $1 - P[\theta]$ are the probabilities that $\ln \frac{L(\theta_1;\ n)}{L(\theta_o;\ n)}$ takes the values $\ln B$ and $\ln A$, respectively. $P[\theta]$, for a particular $S(H_o,\ H_1;\ P(I),\ P(II))$, is computed via Property 2.4.

In other words, Property 2.5 in conjunction with Property 2.4 gives an approximate method by which the average sample number of any $S(H_o,\ H_1;\ P(I),\ P(II))$ with H_o: $\theta = \theta_o$, H_1: $\theta = \theta_1$ can be computed regardless of the actual value of θ, $\theta \epsilon \Omega$.

<u>Property</u> <u>2.6</u> (Wald and Wolfowitz, 70): For all sequential tests of H_o: $\theta = \theta_o$ versus H_1: $\theta = \theta_1$ having probabilities of error $P(I)$ and $P(II)$, the SPRT has the least possible values of $E[N|\theta_o]$ and $E[N|\theta_1]$.

The SPRT does not necessarily have least possible values of $E[N|\theta]$ for every $\theta \epsilon \Omega$, $\theta \epsilon (\theta_o,\ \theta_1)$ and, in fact, $E[N|\theta]$ is not necessarily less than the sample size required in fixed sample size plans with the same probabilities of error when $\theta \epsilon (\theta_o,\ \theta_1)$.

This last possibility may be illustrated by an example from Wetherill (74). Suppose we wish to perform a binomial SPRT for H_o: $p = .25$ versus H_1: $p = .75$ with probabilities of error $P(I) = P(II) = 0.001$. Using Property 2.5 and Property 2.4 we can construct Table 1.

Table 1. ASN for binomial $S(H_o, H_1; P(I), P(II))$ with H_o: $p = .25$ versus H_1: $p = .75$ and $P(I) = P(II) = 0.001$

θ	.25	.37	.43	.50	.57	.63	.75
ASN	12.6	22.0	32.2	39.5	32.2	22.0	12.6

Now if we design a fixed sample size experiment with 33 observations, we find by consulting tables of the cumulative probability distribution such as (29) that a rejection region of 17 or more positive responses specifies a fixed sample size test with probabilities of error $P(I) = P(II) = 0.00095 \doteq 0.001$. Thus for a range of values of p near 0.5 the ASN of the SPRT is greater than the sample size of a fixed sample size test with the same probabilities of errors.

D. Discussion

Many testing problems encountered in real life investigations will not involve simple null versus simple alternative hypothesis formulations. Thus, if the parameter of interest is the only parameter in the model, hypothesis formulations of the form

$$H_o\colon \theta \le \theta_o \text{ versus } H_1\colon \theta \ge \theta_1$$

where θ_o and θ_1 are preassigned scalars and $\theta_o < \theta_1$, or

$$H_o\colon \theta = \theta_o \text{ versus } H_1\colon \theta \ne \theta_1$$

or

$$H_o\colon |\theta| < \delta_o \text{ versus } H_1\colon |\theta| > \delta_1,$$

where δ_o and δ_1 are predetermined scalars, seem to be more relevant to real life situations than

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

The usual procedure in such cases is to apply the SPRT to a least favorable hypothesis formulation. For example, for the specification

$$H_o: \theta \leq \theta_o \text{ versus } H_1: \theta \geq \theta_1$$

the "least favorable" specification

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

may be adopted. In some instances, this can be intuitively appealing, for one might expect that a test of

$$H_o: \theta = \theta_o \text{ versus } H_1: \theta = \theta_1, \ \theta_o < \theta_1$$

would be even more efficient in terms of sample size requirements when it is actually true that $\theta < \theta_o$ or $\theta > \theta_1$. However, this advantage for $\theta < \theta_o$ or $\theta > \theta_1$ may be vitiated by a loss of efficiency if $\theta_o < \theta < \theta_1$. As exemplified earlier, $E[N|\theta]$ may be larger than the corresponding sample size needed for the fixed size sample test when

$$\theta \epsilon [\frac{\theta_o + \theta_1}{2} - \delta, \frac{\theta_o + \theta_1}{2} + \delta]$$

where δ is some number greater than zero.

The testing problems become even more complicated when nuisance parameters are present in the probability model and the "least favorable" approach has to be supplemented by other techniques in order to construct a test.

Some developments, which might be termed extensions or generalizations of the SPRT, have drawn heavily upon fixed sample size reduction principles such as sufficiency and invariance. If we have a hypothesis-testing problem in which there are unknown nuisance parameters, then we should try to construct a test statistic having a distribution not dependent on these nuisance parameters. Properties of sufficiency and invariance have been found useful in such situations.

In hypothesis testing situations that are composite because they involve ranges of the parameter(s) of interest (for example, H_0: $p \leq .3$ versus H_1: $p \geq .5$ in the binomial context), sufficiency and invariance principles do not seem to be applicable. This is so because we should not think a test statistic desirable if it did not depend upon specifications of the parameter of interest. Wald (69) introduced weight functions for the development of the sequential t-test, and weight function methods seem well suited for composite hypotheses concerning ranges of parameters.

In the following chapters we will discuss sufficiency, invariance, and weight functions more thoroughly using them to develop sequential tests for certain problems.

III. SEQUENTIAL MULTIPLE COVARIANCE ANALYSIS USING WEIGHT FUNCTIONS AND PRIOR DISTRIBUTIONS

A. Weight Functions and Prior Distributions

Given that it is possible to observe random variables X_i from a normal distribution with mean μ and variance σ^2 we consider the application of sequential tests to hypotheses about the location parameter μ regardless of the value of the unknown variance σ^2. For this problem Wald (69) suggested that the following procedure may have merit. For all μ such that $|(\mu-\mu_o)/\sigma| < k_o$, where k_o is small, it is preferred to accept the hypothesis H_o: $\mu = \mu_o$. For all μ such that $|(\mu-\mu_o)/\sigma| > k_1 > k_o$ it is preferred to accept the hypothesis H_1: $\mu \neq \mu_o$. Wald called the region in which $k_o \leq |(\mu-\mu_o)/\sigma| \leq k_1$ an indifference region. With this procedure, however, the SPRT theory outlined in Chapter II does not immediately lead to a practical test. For example, suppose we are interested in a one-sided test about the location parameter μ of a normal population with unknown variance. The regions of preference noted above depend on the quantity $\frac{\mu-\mu_o}{\sigma}$ and, using SPRT theory on the "least favorable" case (H_o: $\mu = \mu_o$ versus H_1: $\mu = \mu_o + \gamma\sigma$ where γ is a specified constant), it can be shown that we subsequently arrive at the log-likelihood ratio $n(\bar{x} - \mu_o)\frac{k}{\sigma} - \frac{nk^2}{2}$ with which to carry out the test. This log-likelihood ratio, however, still depends upon σ^2, the unknown parameter. Wald (69) introduced the theory of weight functions as one procedure to overcome this difficulty. In what follows we outline the application of Wald's weight function techniques to sequential testing and describe some examples of sequential weight function tests.

1. Weight functions

Let us assume that it is possible to observe a sequence of mutually independent random variables X_1, X_2, X_3, ... each of which has the same unknown probability density function $f(x;\ \theta)$ and where, in general, the random variable X and the parameter θ may both be vectors. Let us suppose that Ω, the parameter space of θ, can be divided into three mutually exclusive regions so that $\Omega = w_o \cup w_1 \cup w_2$ where: w_o is the region in which the null hypothesis H_o is preferred, w_1 is the region in which the alternative hypothesis H_1 is preferred, and w_2 is the region in which neither H_o or H_1 is preferred.

When statistical tests of composite hypotheses are constructed, the probabilities P(I) and P(II) of Type I and Type II errors respectively are, in general, functions of one or more of the parameters of the parameter vector $\theta = (\theta_1, \theta_2, \theta_3, \ldots, \theta_k)$. Keeping this in mind, suppose we have two weight functions for θ, $V_o(\theta)$ and $V_1(\theta)$, defined so that:

$V_o(\theta)$ is non-zero only for $\theta \epsilon w_o$, $V_o(\theta) = 0$ for $\theta \epsilon w_1$ and $\theta \epsilon w_2$, and

$$\int_{w_o} V_o(\theta)\, d\theta = 1 \qquad (3.1a)$$

and

$V_1(\theta)$ is non-zero only for $\theta \in w_1$, $V_1(\theta) = 0$ for $\theta \in w_o$ and $\theta \in w_2$, and

$$\int_{w_1} V_1(\theta)\, d\theta = 1. \tag{3.1b}$$

We also note that the integrals given by 3.1 may be multiple integrals.

Wald (69) then defined modified probability density functions which are constructed as follows:

$$\begin{aligned} g_{on}(x_1, x_2, \ldots, x_n) &= \int_{w_o} \prod_1^n f(x_i;\ \theta)\ V_o(\theta)\ d\theta \\ g_{1n}(x_1, x_2, \ldots, x_n) &= \int_{w_1} \prod_1^n f(x_i;\ \theta)\ V_1(\theta)\ d\theta. \end{aligned} \tag{3.2}$$

It is now possible to define a Wald SPRT for the hypothesis specification:

$$\begin{aligned} H_o'&:\ g(x_1, x_2, \ldots, x_n) = g_{on}(x_1, x_2, \ldots, x_n) \\ H_1'&:\ g(x_1, x_2, \ldots, x_n) = g_{1n}(x_1, x_2, \ldots, x_n) \end{aligned} \tag{3.3}$$

because H_o and H_1 are both simple hypotheses. It should here be noted that $g_{in}(x_1, x_2, \ldots, x_n)$, $i = 0, 1$, will in general not be factorizable as

$$g_{in}(x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} g_{ik}(x_k). \tag{3.4}$$

As pointed out in the discussion following Property 2.3, however, independence of the random variables is not a necessary condition for the construction of boundaries for the SPRT or for the application of Properties 2.2 and 2.3.

It is now convenient to denote the SPRT for the specification in 3.3 as $S(H_o', H_1'; P(I)', P(II)')$ where $P(I)'$ and $P(II)'$ are the probabilities of Type I and Type II errors, respectively. Then, at the n^{th} stage of sampling, $S(H_o', H_1'; P(I)', P(II)')$ induces a partitioning of the sample n-space as in 3.5.

$$\begin{aligned} C_{on} &= [(x_1, x_2, \ldots, x_n): H_o \text{ is accepted}] \\ C_{1n} &= [(x_1, x_2, \ldots, x_n): H_1 \text{ is accepted}] \\ C_{2n} &= [(x_1, x_2, \ldots, x_n): \text{take another observation}] \end{aligned} \tag{3.5}$$

Also, since $S(H_o', H_1'; P(I)', P(II)')$ is an SPRT, we know from Property 2.3 that

$$P(I)' = \sum_{n=1}^{\infty} \int_{C_{1n}} g_{on}(x_1, \ldots, x_n)\, dx \tag{3.6}$$

and

$$P(II)' = \sum_{n=1}^{\infty} \int_{C_{on}} g_{1n}(x_1, \ldots, x_n)\, dx \tag{3.7}$$

where $dx = dx_1 dx_2 \ldots dx_n$. Using definitions 3.2 we can now write

$$P(I)' = \sum_{n=1}^{\infty} \int_{C_{1n}} [\int_{w_o} \prod_{1}^{n} f(x_i;\ \theta)\ V_o(\theta)\ d\theta]\ dx. \tag{3.8}$$

If we assume that the integration and summation signs are commutative, 3.8 can be rewritten

$$P(I)' = \int_{w_o} \sum_{n=1}^{\infty} \int_{C_{1n}} \prod_{1}^{n} f(x_i;\ \theta)\ dx\ V_o(\theta)\ d\theta$$

$$= \int_{w_o} P_\theta(I)\ V_o(\theta)\ d\theta \tag{3.9a}$$

where

$$P_\theta(I) = \sum_{n=1}^{\infty} \int_{C_{1n}} \prod_{1}^{n} f(x_i;\ \theta)\ dx \tag{3.9b}$$

denotes the Type I error at any point $\theta \in w_o$. Similarly we write

$$P(II)' = \int_{w_1} P_\theta(II)\ V_1(\theta)\ d\theta \tag{3.10a}$$

where

$$P_\theta(II) = \sum_{n=1}^{\infty} \int_{C_{on}} \prod_{1}^{n} f(x_i;\ \theta)\ dx \tag{3.10b}$$

denotes the Type II error at any point $\theta \in w_1$.

The end result is then that we have a procedure for sequentially testing H_o versus H_1 where the approximate error rates are the weighted quantities $P(I)'$ and $P(II)'$.

2. Optimal weight functions

In the absence of a priori bases for selecting weight functions, the choice could be made to satisfy some sort of "optimality" criterion involving a restriction to sequential tests which have certain "best" properties. We would then search the class of sequential tests $S(H_o', H_1'; P(I)', P(II)')$, where H_o' and H_1' are given as in 3.3, for those tests (if they exist) which have these properties. If we restrict ourselves to a class of tests which have desirable properties, however, there usually is no method of choosing weight functions which generate the appropriate sequential tests.

Some cases have been reported for which specialized methods of choosing weight functions to derive "optimal" sequential tests do exist. For example Wald (69) considered the class of all sequential tests $S(H_o', H_1'; P(I)', P(II)')$ derived via weight functions and sought to choose weight functions which induce sequential tests that satisfy the following "optimum" restrictions:

$$\begin{aligned} P_\theta(I) &\leq P(I) \quad \text{for} \quad \theta \in w_o \\ P_\theta(II) &\geq P(II) \quad \text{for} \quad \theta \in w_1 \end{aligned} \tag{3.11}$$

where there exists at least one $\theta \in w_o$, say θ_o, such that $P_{\theta_o}(I) = P(I)$ and similarly there exists at least one $\theta \in w_1$, say θ_1, such that $P_{\theta_1}(II) = P(II)$. The $P_\theta(I)$ and $P_\theta(II)$ of 3.11 are defined in 3.9b and 3.10b respectively.

It is evident from

i) the relations given in 3.9 and 3.10

ii) the fact that the regions C_{on}, C_{1n}, and C_{2n} are in effect defined by

$$\frac{P(II)'}{1-P(I)'}, \ \frac{1-P(II)'}{P(I)'}, \ V_o(\theta), \text{ and } V_1(\theta)$$

iii) the fact that knowledge of $\frac{P(II)'}{1-P(I)'}$ and $\frac{1-P(II)'}{P(I)'}$ implies knowledge of $P(I)'$ and $P(II)'$

that the following relations hold

$$\begin{aligned} \max_{\theta \in w_o} P_\theta(I) &= h_1(P(I)', P(II)', V_o(\theta), V_1(\theta)) \\ \max_{\theta \in w_1} P_\theta(II) &= h_2(P(I)', P(II)', V_o(\theta), V_1(\theta)) \end{aligned} \tag{3.12}$$

where h_1 and h_2 are functions of arguments as indicated in 3.12. For given weight functions it follows that P(I)' and P(II)' can, in principle, be chosen so that $\max_{\theta \varepsilon w_o} P_\theta(I)$ and $\max_{\theta \varepsilon w_1} P_\theta(II)$ take on their desired values.

Wald showed that if he considered the subclass of weight functions which serve to simultaneously minimize the two maximums given in 3.12 for fixed P(I)' and P(II)' he would then have weight functions which generate sequential tests for which the Type I and Type II errors satisfy 3.11. The following theorem due to Wald sets out sufficient conditions which sequential weight functions tests must satisfy in order for 3.11 to hold.

Theorem 3.1 Let us suppose that the parameter space Ω can be written as $\Omega = w_o \cup w_1 \cup w_2$ where $w_i \cap w_j = \emptyset$, i, j = 0, 1, 2, $i \neq j$. As before, w_o is the region of the parameter space where H_o is preferred, w_1 is the region where H_1 is preferred and w_2 is the region where neither H_o or H_1 is preferred. Further, let us assume that the boundary of w_1 is a surface, say S_1. We suppose then that it is possible to find two weight functions $V_o(\theta)$ and $V_1(\theta)$ such that

$$\int_{w_o} V_o(\theta)\, d\theta = 1$$
$$\int_{S_1} V_1(\theta)\, d\theta = 1, \qquad (3.13)$$

and such that the SPRT based on the ratio

$$\frac{\int_{S_1} V_1(\theta) \prod_1^n f(x_i;\ \theta)\ dS_1}{\int_{w_o} V_o(\theta) \prod_1^n f(x_i;\ \theta)\ d\theta} \tag{3.14}$$

satisfies the following conditions for any values of the upper and lower boundaries of the test procedure:

i) $P_\theta(I)$ is constant in w_o

ii) $P_\theta(II)$ is constant over S_1

iii) for any point θ in the interior of w_1, the value of $P_\theta(II)$ does not exceed the constant value of $P_\theta(II)$ on S_1.

Also if the lower boundary is taken to be $\frac{P(II)}{1-P(I)}$ and if the upper boundary is taken as $\frac{1-P(II)}{P(I)}$ then we have that

$$\max_{\theta\ \epsilon\ w_o} P_\theta(I) = P(I)$$

and

$$\max_{\theta\ \epsilon\ w_1} P_\theta(II) = P(II).$$

Proof: The proof may be found in Wald (69).

3. Application of weight functions

As an application of Theorem 3.1 Wald derived a type of sequential t-test for sequentially testing hypotheses about the mean μ of a normal population with unknown variance σ^2 and, in particular, for testing that $\left|\frac{\mu-\mu_o}{\sigma}\right|$ is small relative to some value μ_o. The parameter space $\Omega = [(\mu,\sigma)$: $-\infty<\mu<\infty,\ 0<\sigma<\infty]$ is partitioned as $\Omega = w_o \cup w_1 \cup w_2$ where

$$w_o = [(\mu,\sigma): \left|\frac{\mu-\mu_o}{\sigma}\right| = 0, \ 0<\sigma<\infty]$$
$$w_1 = [(\mu,\sigma): \left|\frac{\mu-\mu_o}{\sigma}\right| > k_1, \ k_1 > 0, \ 0<\sigma<\infty] \tag{3.15}$$

and

$$w_2 = \Omega - w_o - w_1.$$

The boundary S_1 of w_1 is given by

$$S_1 = [(\mu,\sigma): |\mu-\mu_o| = k_1\sigma, \ k_1 > 0, \ 0<\sigma<\infty] .$$

Wald shows that the weight functions which satisfy the conditions of Theorem 3.1 are

$$V_o(\mu,\sigma) = 1/c; \quad 0 \leq \sigma \leq c, \ \mu = \mu_o$$
$$= 0 \quad \text{otherwise}$$
$$V_1(\mu,\sigma) = \frac{1}{2c}; \quad 0 \leq \sigma \leq c, \ \mu = \mu_o \pm k_1\sigma$$
$$= 0 \quad \text{otherwise.}$$

That is, Wald shows that the likelihood ratio

$$\lim_{c\to\infty} \frac{\int_{S_1} V_1(\mu,\sigma) \prod_1^n f(x_i;\mu,\sigma)\, d\sigma\, d\mu}{\int_{w_o} V_o(\mu,\sigma) \prod_1^n f(x_i;\mu,\sigma)\, d\sigma\, d\mu} \tag{3.17}$$

can be computed as an SPRT with Type I and Type II probabilities of error, P(I) and P(II), so that we may expect the actual hypothesis specification

$$H_o: \mu = \mu_o \quad \text{versus} \quad H_1: |(\mu-\mu_o)/\sigma| = k_1\sigma$$

to be tested with approximately these probabilities of error.

The following additional applications of weight function techniques may be noted. For the situation where $X_1, X_2, \ldots, X_n$ are independent normally distributed random variables with mean and variance σ^2 unknown, Wetherill (74) presents a weight function approach to the problem of testing $H_o\colon \sigma = \sigma_o$ versus $H_1\colon \sigma = \sigma_1 > \sigma_o$. Hoel (30) has obtained a sequential test for the canonical form of the general linear hypothesis using weight functions. Roseberry (58) derived some weight function test procedures for the comparison of two treatments using one covariate based on a bivariate normal model.

No results similar to those of Properties 2.4, 2.5, and 2.6 are available for any of the above examples of weight functions in sequential testing. In addition, Hall, Wijsman, and Ghosh (28) mention that they do not consider Wald's proof of the inequalities on the two error probabilities for Wald's sequential t-test to be adequate (see Property 2.3) and further suggest that the type of arguments necessary for demonstration of these inequalities are those given by Barnard (5).

In many practical situations the weight function approach may attract criticism from experimenters because of its arbitrariness and the possible difficulty in interpreting the functional relationships exhibited in 3.9 and 3.10 and from theoreticians because it is usually intractable to handle the properties of the tests analytically. It is hoped that the topic to be developed in the next section may alleviate the intensity of the first criticism, while on the second point the properties of weight function tests can be investigated by Monte Carlo techniques pending the development of appropriate analytic techniques.

4. Prior distributions

We feel that the initial work of Wald with weight functions and the subsequent application of weight functions for the construction of sequential tests by various authors has a close association with the growing body of statistical literature concerning Bayesian statistics. In this thesis, the term Bayesian refers to any use of prior densities on a parameter space with the associated application of Bayes' theorem in the analysis of a statistical problem.

In order to discuss weight functions in the context of prior distributions we shall adopt some definitions given by Raiffa and Schlaifer (53). They use the word likelihood to denote the value $l(z|\theta)$ taken on by the mass or density function for a given outcome z and a given parameter θ. The marginal likelihood of the experimental outcome z given a particular prior density $g(\theta)$, defined on a parameter space Ω, is defined as

$$l(z|g) = \int_{\Omega} l(z|\theta)g(\theta)d\theta. \tag{3.18}$$

Comparison of formula 3.18 with 3.1 - 3.3 shows that procedures based on weight functions can be considered as the application of prior distributions to sequential testing by the formulation of the ratio of marginal likelihoods of an experimental outcome given particular prior distributions. These prior distributions are the $V_o(\theta)$ and $V_1(\theta)$, defined over the restricted regions w_o and w_1, respectively, of the parameter space Ω, the regions being selected according to the hypotheses to be tested (see for example, the discussion of Wald's sequential t-test in the previous subsection).

In some practical applications there will exist a substantial amount of empirical evidence on which to base the prior distribution of the parameters. For example, it is possible that a production process which produces normally distributed random observations may generate a different variance σ^2 each day it is run. That is, on the first day we have $X_{11}, X_{12}, \ldots, X_{1n} \ldots NI(\mu,\sigma_1^2)$ and on the second day we have $X_{21}, X_{22}, \ldots, X_{2n} \ldots NI(\mu,\sigma_2^2)$ etc. It might then be possible to describe the distribution of values of σ^2 by a fairly common probability density function which could then be utilized in making inferences about the mean μ of the process.

With reference to possible questions concerning the fitting of prior distributions to past data, Raiffa and Schlaifer (53) report, in rather strong terms, that their experiences with real life examples show that in a great many applications the method of fitting will have no great effect on the final outcome of the results of the decision problem under consideration. In other words, the experimenter and statistician need not be very preoccupied with different methods of fitting the prior distribution to the relevant data. This is not to say that prior distributions should not be used but instead implies that the statistical analysis is usually insensitive to various recommended techniques of distribution estimation.

The problem of assessing a prior distribution is more challenging in those situations where no empirical frequency basis for assessment exists. Thus the prior information may not be straightforwardly quantifiable and in these cases may simply represent the betting odds with which the responsible person wishes his final decision to be consistent. The psychological difficulty with the assignment of such odds usually results in the "true" prior distribution being described in terms of a few summary measures such as the mean, the mean deviation, or a few fractiles. It is important, therefore, that the family F of prior distributions under consideration be such that there will be a member of F capable of expressing such fractional types of prior information.

The above difficulty increases further when, in effect, there is no prior information whatsoever and objective prior densities on the parameter space are desired. If the experimenter does not have any prior knowledge about the parameter, he cannot proceed to make any prior guess about the parameter values lying over a finite set of an infinite parameter space nor can he proceed to choose a suitable prior from some family of densities.

For such circumstances what are known as prior quasi-densities have been advocated. Wallace (71) and Stone (68) define a <u>prior quasi-density</u> as any non-negative function $g(\theta)$ defined on a parameter space Ω. As an example of a prior quasi-density consider the normal distribution with mean μ and standard deviation σ. One quasi-density for σ is $g(\sigma) = 1$, $0 < \sigma < \infty$. A prior quasi-density is called <u>admissible</u> with respect to a density $f(x|\theta)$, $x \in X$, if

$$h(x) = \int_{\Omega} f(x|\theta)\, g(\theta)\, d\theta < \infty. \tag{3.19}$$

We then have

$$g^*(\theta|x) = \frac{f(x|\theta)\, g(\theta)}{h(x)}$$

well-defined and we call $g^*(\theta|x)$ a <u>weak posterior density</u>. Wallace (71) shows that, given a prior quasi-density, there exists a sequence of proper prior densities whose corresponding proper posterior densities tend to a weak posterior density for each fixed set of data. Similarly, given two prior quasi-densities, $g_1(\theta)$ and $g_2(\theta)$, and their corresponding admissible marginal likelihoods, L_1 and L_2 say, it is possible to have a sequence of marginal likelihoods, say L_{1c} and L_{2c}, induced by proper prior distributions so that

$$\lim_{c \to \infty} \frac{L_{1c}}{L_{2c}} = \frac{L_1}{L_2}.$$

To illustrate, consider $f(x|\theta)$ where $\theta = (\mu,\sigma)$, $-\infty < \mu < \infty$, and $0 < \sigma < \infty$ and let

$$g_1(\mu,\sigma) = 1 \qquad 0 < \sigma < \infty,\ \mu = \mu_1$$
$$= 0 \qquad \text{otherwise}$$

and

$$g_2(\mu,\sigma) = 1 \qquad 0 < \sigma < \infty,\ \mu = \mu_2$$
$$= 0 \qquad \text{otherwise.}$$

Then we have

$$L_1 = \int_o^\infty f(x|\mu_1,\sigma)\, d\sigma$$
$$L_2 = \int_o^\infty f(x|\mu_2,\sigma)\, d\sigma.$$

Let us define proper priors as

$$g_{1c}(\mu,\sigma) = \frac{1}{c}, \qquad 0 < \sigma < c,\ \mu = \mu_1$$
$$= 0 \qquad \text{otherwise}$$
$$g_{2c}(\mu,\sigma) = \frac{1}{c}, \qquad 0 < \sigma < c,\ \mu = \mu_2$$
$$= 0 \qquad \text{otherwise.}$$

Then

$$\lim_{c\to\infty} \frac{L_{1c}}{L_{2c}} = \lim_{c\to\infty} \frac{\int_o^c \frac{1}{c} f(x|\mu_1,\sigma)\, d\sigma}{\int_o^c \frac{1}{c} f(x|\mu_2,\sigma)\, d\sigma} = \lim_{c\to\infty} \frac{\int_o^c f(x|\mu_1,\sigma)\, d\sigma}{\int_o^c f(x|\mu_2,\sigma)\, d\sigma}$$
$$= \frac{L_1}{L_2}.$$

5. Final remarks

In the preceding sections we have attempted to assimilate the weight function techniques as introduced by Wald into the context of Bayesian statistical procedures. It is our opinion that the use of prior distributions can serve as a valuable tool in the construction of sequential tests for composite hypotheses. It is recognized that there are difficult mathematical problems associated with these approaches, but they do reduce the difficult problem of sequentially testing composite hypotheses to the more tractable case of sequentially testing simple hypotheses.

In support of this approach it may be noted that Barnard (5) and Bartholomew (7) suggested that Bayesian statistics and classical statistics may be in agreement in the context of sequential experimentation. Portions of this chapter may add credence to this conjecture.

B. Statement of the Problem

1. Motivation

Statistical analyses for the comparison of two treatments are extensively documented for fixed sample size experiments. Statistical techniques for the comparison of two treatments by sequential experimentation are, however, not so well developed and, further, most of the available sequential techniques relate to somewhat unrealistic hypothesis formulations.

Armitage (4) discussed the design and the sequential analysis of medical trials with emphasis on the comparison of two treatments. Roseberry (58) and Cox and Roseberry (18) developed and investigated empirically some sequential tests which utilize one covariate. In these references, the experimental units were paired and the two treatments then were assigned at random to the subjects within pairs. In (58) and (18) observations were assumed to be bivariate normally distributed with one variate as the response of interest and the other variate representing the concomitant information.

Maurice (47, 48), Johnson and Maurice (41), and Colton (16) approached the problem of sequentially comparing two treatments from a decision theory standpoint using loss functions and prior distributions. As in (4), (18), and (58) these authors used a design where the observations were paired and again it was assumed that the observations were normally distributed with known variance.

Hajnal (26) derived an unpaired sequential t-test for unpaired observations and this technique will be discussed more fully in Chapter V.

The situation we shall generally envisage in this thesis is that in which observations on the response and concomitant variates are sequentially obtained. For example, a clinician may have primary interest in the effectiveness of a drug for a head cold or for the relief of arthritic pain. The response of interest may then be supplemented by such concomitant information as the patient's age and blood pressure.

In this chapter we will require that the observations be made in pairs--one for each of the two treatments being compared. As in Cox and Roseberry (18) pairs are comprised of groups of two successive units, the allocation of experimental units to the two treatments being random within each pair. We shall form the signed differences within pairs of the observations and proceed to make inferences from these differences. It should be noted here that, in general, each observation will be multivariate so that the differences we refer to will be differences of vectors.

Suppose that a vector of random variables $(W, \tilde{Z}')$, where W is the scalar variate of interest and $\tilde{Z}' = (Z_1, Z_2, \ldots, Z_p)$ are the concomitant variates, is a random vector from a population having the probability density function

$$g(w, \tilde{z}; \alpha, \tilde{\mu}, \tilde{\beta}, k^2, \tilde{\theta}) =$$

$$g_1(\tilde{z}; \tilde{\mu}, \tilde{\theta})\left(\frac{1}{2\Pi k^2}\right)^{\frac{1}{2}} \exp\left[-\frac{(w-\alpha-(\tilde{z}-\tilde{\mu})'\tilde{\beta})^2}{2k^2}\right] \quad (3.21)$$

where α and k^2 are scalars and $\tilde{\mu}$, $\tilde{\beta}$ and $\tilde{\theta}$ are vectors of parameters. Let us also assume that $\tilde{\beta}$ and k^2 are not functionally related to $\tilde{\theta}$ and that the unconditional expectation, $E[\tilde{Z}] = \tilde{\mu}$. It is easily shown that the conditional distribution of W given $\tilde{Z} = \tilde{z}$ is normal with mean $\alpha + (\tilde{z}-\tilde{\mu})'\tilde{\beta}$ and variance k^2 and, in addition, that the unconditional expectation of W, E[W], is α.

If we assume that the concomitant information, represented by $\tilde{Z}' = (Z_1, \ldots, Z_p)$, is obtained before the application of the treatment, then model 3.21, with an arbitrary marginal p.d.f. for $\tilde{Z}'$, appears to be a reasonable representation of a possible real life situation in which concomitant information is used. However, for the design we are considering, attention will be mainly concentrated on cases for which $g_1(\tilde{z}; \tilde{\mu}, \tilde{\theta})$ can be assumed multivariate normally distributed. For if $(W_1, \tilde{Z}_1)$ and $(W_2, \tilde{Z}_2)$ are random variables having p.d.f.'s $g(w_1, z_1; \alpha_1, \tilde{\mu}, \tilde{\beta}, \tilde{\theta}, k^2)$ and $g(w_2, z_2; \alpha_2, \tilde{\mu}, \tilde{\beta}, \tilde{\theta}, k^2)$ respectively then it is highly desirable that the distribution of $(W_1-W_2, \tilde{Z}_1-\tilde{Z}_2)$ should also have the form 3.21. This property holds for the density in 3.21 if

$$g_1(\tilde{z};\tilde{\mu},\Phi) = \frac{1}{((2\Pi)^p|\Phi|)^{\frac{1}{2}}} \exp[-\frac{(\tilde{z}-\tilde{\mu})'\Phi^{-1}(\tilde{z}-\tilde{\mu})}{2}] \qquad (3.22)$$

where Φ is the variance-covariance matrix of the random variable $\tilde{Z}$. Some further details on this point are noted in Section A of the Appendix.

The final form of the model to be used will now be developed. If we let $\tilde{X}$ be the p x 1 vector of differenced covariates $\tilde{Z}_1-\tilde{Z}_2$, and Y be the paired difference of the responses of interest, W_1-W_2, then we shall assume that we are sampling from the population described by the probability density function

$$g(y,\ \tilde{x};\alpha,\tilde{\beta},\Sigma,\sigma^2) = \frac{1}{(2\Pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}\ \exp\left[\frac{-\tilde{x}'\Sigma^{-1}\tilde{x}}{2}\right]$$

$$\frac{1}{\sqrt{2\Pi}\ \sigma}\ \exp\left[-\ \frac{(y-\alpha-\tilde{x}'\tilde{\beta})^2}{2\sigma^2}\right]. \qquad (3.23)$$

In 3.23 tildes have been used to indicate that $\tilde{x}$ and $\tilde{\beta}$ are vectors while Σ is a matrix. It will be convenient in what follows to drop the tilde notation unless the matrix and vector quantities are not sufficiently defined by the context in which they are found.

The probability density function given in 3.21 is a multivariate nonnormal density with the conditional p.d.f. of W given Z = z being univariate normal. The p.d.f. of 3.23 is a special case of the p.d.f. given in 3.21 and is a reparameterized form of the usual multivariate normal density as given, for example, in Anderson (1). As is shown in Section A of the Appendix, (Y,X')' has a multivariate normal distribution (MVN) with mean $(\alpha,0)'$ and variance-covariance matrix

$$\begin{pmatrix} \sigma^2+\beta'\Sigma & \beta'\Sigma \\ \Sigma\beta & \Sigma \end{pmatrix} .$$

For example, if we assume one covariate, i.e. p=1, we have $(Y,X_1)'$ distributed as MVN with mean $(\alpha, 0)'$ and variance-covariance matrix

$$\begin{pmatrix} \sigma^2+\beta_1^2\sigma_{x_1}^2 & \beta\sigma_{x_1}^2 \\ \beta_1\sigma_{x_1}^2 & \sigma_{x_1}^2 \end{pmatrix} .$$

It is shown in Section A of the Appendix that the unconditional expectation of Y is α, i.e.

$$E[Y] = \alpha \tag{3.24}$$

where α reflects the population difference between the two treatments under investigation. The problem which we shall consider in this chapter is that of testing hypothesis formulations about α using appropriate sequential procedures.

2. Sequential test with known nuisance parameters

If we assume that the nuisance parameters β, σ^2, and Σ are known, we are essentially working with the normal variate $y-x'\beta$ which has mean α. Thus if we formulate a hypothesis specification as

$$\begin{aligned} H_o&: \alpha = \alpha_o \\ H_1&: \alpha = \alpha_1 \end{aligned} \tag{3.25}$$

then we can apply, without difficulty, the standard SPRT theory as summarized below.

It is here noted that in the remainder of the thesis, unless the situation dictates, we do not distinguish between a random variable and its realization.

a. Sequential procedure Let us denote the SPRT of the specification 3.25 by $S(H_o, H_1; P(I), P(II))$ where $P(I)$ and $P(II)$ are the Type I and Type II errors respectively. Let R_n denote the likelihood ratio at stage n,

$$R_n = \frac{\prod_1^n f(y_i, x_i; \alpha_1, \beta, \Sigma, \sigma^2)}{\prod_1^n f(y_i, x_i; \alpha_o, \beta, \Sigma, \sigma^2)} \tag{3.26}$$

Then from 3.23

$$\prod_1^n f(y_i, x_i; \alpha, \beta, \Sigma, \sigma^2) = \left(\frac{1}{(2\Pi)^p |\Sigma|}\right)^{\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^n x_i' \Sigma^{-1} x_i}{2}\right]$$

$$\left(\frac{1}{2\Pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\sum_1^n \frac{(y_i-\alpha-x_i'\beta)^2}{2\sigma^2}\right] \tag{3.27}$$

and $$\sum_{i=1}^n (y_i-\alpha-x'\beta)^2 = \sum_1^n (y_i-x_i'\beta)^2 - 2\alpha\sum_1^n(y_i-x_i'\beta) + n\alpha^2$$

$$= (y-X'\beta)'(y-X'\beta) - 2\alpha e'(y-X'\beta) + n\alpha^2$$

where $x_i' = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{ip})$ so that $X = (x_1, x_2, x_3, \ldots, x_n) = (x_{ij})$ is a p x n matrix of covariate observations. Each x_{ij} refers to the j^{th} observed value of the i^{th} covariate. We also are denoting $(y_1, y_2, \ldots, y_n)$ by y' and $(1, 1, \ldots, 1)$

by e'. Therefore the likelihood ratio in 3.26 becomes

$$R_n = \exp\left[\frac{2(\alpha_1-\alpha_o)\ e'\ (y-X'\beta) - n(\alpha_1^2 - \alpha_o^2)}{2\sigma^2}\right] \tag{3.28}$$

The test procedure $S(H_o, H_1; P(I), P(II))$ then specifies that:

(a) if $R_n \geq \frac{1 - P(II)}{P(I)}$, H_o is rejected and H_1 is accepted (3.29)

(b) if $R_n \leq \frac{P(II)}{1 - P(I)}$, H_o is accepted and H_1 is rejected (3.30)

(c) if $\frac{P(II)}{1 - P(I)} < R_n < \frac{1 - P(II)}{P(I)}$, we sample the $(n+1)^{th}$ pair. (3.31)

b. <u>Average sample number</u> Property 2.5 can be used to give an approximation for the average sample number given that H_o: $\alpha = \alpha_o$ is true. For from 3.23

$$\begin{aligned} E[\ln L|H_o: \alpha=\alpha_o] &= \frac{1}{\sigma^2} E[(\alpha_1-\alpha_o)(y-x'\beta) + \tfrac{1}{2}(\alpha_o-\alpha_1)(\alpha_o+\alpha_1)] \\ &= -\frac{1}{2\sigma^2}(\alpha_o-\alpha_1)^2 \end{aligned} \tag{3.32}$$

where $L = \dfrac{f(y, x; \alpha_1, \beta, \sigma^2, \Sigma)}{f(y, x; \alpha_o, \beta, \sigma^2, \ \ )}$ and $f(y, x; \alpha, \beta, \sigma^2, \Sigma)$ is the p.d.f. given in 3.23.

Similarly

$$E[\ln L|H_1: \alpha = \alpha_1] = \frac{(\alpha_1-\alpha_o)^2}{2\sigma^2} \tag{3.33}$$

Then, Property 2.5 gives the following approximate formulae

$$E[N|\alpha_o] = \frac{P(I)\ \ln \frac{1-P(II)}{P(I)} + (1-P(I))\ \ln \frac{P(II)}{1-P(I)}}{-\frac{1}{2\sigma^2}(\alpha_1-\alpha_o)^2}$$

$$E[N|\alpha_1] = \frac{(1-P(II))\ \ln \frac{1-P(II)}{P(I)} + P(II)\ \ln \frac{P(II)}{1-P(I)}}{\frac{1}{2\sigma^2}(\alpha_1-\alpha_o)^2} \qquad (3.34)$$

where $E[N|\alpha]$ is the expected sample size of the sequential procedure $S(H_o, H_1; P(I), P(II))$ when α is the actual value of the parameter of interest.

c. Operating characteristic curve In order to compute the operating characteristic curve by means of Property 2.4 it is necessary to find $h(\alpha)$ such that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\frac{\exp[-\frac{(y-\alpha_1-x'\beta)^2}{2\sigma^2}]}{\exp[-\frac{(y-\alpha_o-x'\beta)^2}{2\sigma^2}]}\right]^{h(\alpha)} f(y,\ x;\ \alpha,\ \beta,\ \Sigma,\ \sigma^2)$$

$$dydx_1 \ \ldots \ dx_p = 1 \qquad (3.35)$$

where $f(y, x; \alpha, \beta, \Sigma, \sigma^2)$ is as given in 3.23. That is, writing $h(\alpha) = h$, we must find h such that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\frac{\exp[-\frac{(y-\alpha_1-x\beta')^2}{2\sigma^2}]}{\exp[-\frac{(y-\alpha_o-x'\beta)}{2\sigma^2}]}\right]^{h} \left(\frac{1}{(2\Pi)^{p+1}|\Sigma|\ \sigma^2}\right)^{\frac{1}{2}}$$

$$\exp[-\frac{x'\Sigma^{-1}x}{2} - \frac{(y-\alpha-x'\beta)^2}{2\sigma^2}]\ dydx_1 \ \ldots \ dx_p = 1 \qquad (3.36a)$$

Rewriting we have

$$\exp\left[\frac{1}{2\sigma^2}(\alpha_o^2 - \alpha_1^2)h\right] \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{1}{(2\Pi)^{\frac{p+1}{2}} |\Sigma|^{\frac{1}{2}} \sigma} \exp\left[Q(y,x)\right]$$

$$dydx_1 \ldots dx_p = 1 \tag{3.36b}$$

where $Q(y,x) = -\dfrac{2(\alpha_o-\alpha_1)(y-x'\beta)h}{2\sigma^2} - \dfrac{x'\Sigma^{-1}x}{2} - \dfrac{(y-\alpha-x'\beta)^2}{2\sigma^2}$.

Expanding $Q(y,x)$ in $(y-x'\beta)$ and completing the square we have

$$Q(y,x) = -\frac{1}{2\sigma^2}\left[(y-x'\beta-\alpha+\alpha_o h-\alpha_1 h)^2 - (\alpha-(\alpha_o-\alpha_1)h)^2 + \alpha^2\right]$$

$$-\frac{x'\Sigma^{-1}x}{2}.$$

Integrating 3.36b we have

$$\exp\left[-\frac{1}{2\sigma^2}(\alpha^2-(\alpha-(\alpha_o-\alpha_1)h)^2 - (\alpha_o^2-\alpha_1^2)h)\right] = 1 \tag{3.37}$$

Solving 3.37 for h we finally have

$$h = \frac{\alpha_1+\alpha_o-2\alpha}{\alpha_1-\alpha_o}. \tag{3.38}$$

Let $P[\alpha | S(H_o,\ H_1;\ P(I),\ P(II)]$ be the probability of deciding for H_o: $\alpha = \alpha_o$ when any point $\alpha\varepsilon(-\infty,\infty)$ holds as fact. Then by Property 2.4 we have

$$P[\alpha | S(H_o,\ H_1;\ P(I),\ P(II))] = \frac{\left(\frac{1-P(II)}{P(I)}\right)^h - 1}{\left(\frac{1-P(II)}{P(I)}\right)^h - \left(\frac{P(II)}{1-P(I)}\right)^h}. \tag{3.39}$$

In passing it is interesting to note that formulae 3.38 and 3.34 are unchanged even if the p.d.f. given in 3.23 is relaxed to the more general p.d.f.

$$f(y,x;\alpha,\beta,\theta) = g_1(x;\theta)\ \left(\frac{1}{2\Pi\sigma^2}\right)^{\frac{1}{2}}\ \exp\left[-\ \frac{(y-\alpha-x'\beta)^2}{2\sigma^2}\right]$$

where $g_1(x;\theta)$ is the marginal distribution of x indexed by the vector parameter θ. It is noted here that for the case of fixed x's the preceding derivations also hold.

As we have seen, the sequential testing theory for our model 3.23 follows directly, with a little algebra, from Wald's (69) SPRT procedures if the nuisance parameters are assumed known and the hypotheses are of the simple versus simple type. If, however, we cannot, with subjective or frequentist probability of one, assume known values for β, σ^2, and Σ, then we are in a composite hypothesis testing situation for which Wald's elegant theory does not apply.

In what follows we accordingly consider the construction of sequential tests of composite hypotheses using prior distributions. Our primary interest is again the problem of making inferences about the α parameter in the model 3.23.

C. A Test for H_o: $\alpha = \alpha_T$ Versus H_1: $\alpha = \alpha_A$, with Nuisance Parameters Unknown

Let us suppose we are sequentially sampling random variables (Y_i, X_i), i = 1, 2, ..., n, ... from a distribution which has the probability density function 3.23. We wish to construct sequential tests for the hypothesis formulation

$$H_o: \alpha = \alpha_T$$
$$H_1: \alpha = \alpha_A \qquad (3.40)$$

where α_T and α_A are specified scalars and β, σ^2, Σ are unknown.

The region of preference for acceptance of H_o is

$w_o = [(\alpha,\sigma,\beta_1,\beta_2,\ldots,\beta_p,\sigma_{11},\ldots,\sigma_{1p},\sigma_{22},\ldots,\sigma_{pp})$:

$\alpha = \alpha_T$ and the parameters σ; β_i, $i=1,\ldots,p$;

σ_{ij}, $i \leq j=1,\ldots,$ p are unspecified].

The region of preference for acceptance of H_1 is

$w_1 = [(\alpha,\sigma,\beta_1,\beta_2,\ldots,\beta_p,\sigma_{11},\ldots,\sigma_{1p},\sigma_{22},\ldots,\sigma_{2p},\ldots,\sigma_{pp})$:

$\alpha = \alpha_A$ and the parameters σ; β_i, $i=1,\ldots,p$;

σ_{ij}, $i \leq j=1,\ldots,$ p are unspecified].

The region where neither H_o or H_1 is preferred is the complement of the region $w_o \cup w_1$.

Let us suppose that there exists no prior information about the nuisance parameters β, σ^2, and Σ. As previously discussed in Section A, Chapter III, we will assume that this situation can be represented in terms of prior quasi-densities over the parameter spaces as follows.

Let

$$V_o(\alpha,\beta,\sigma,\Sigma) = \frac{1}{2^{\frac{p(p+1)}{2}}\, c^{\frac{(p+1)(p+2)}{2}}},\ 0<\sigma_{ii}<c;-c<\beta_i<c, i=1,2,\ldots,p;$$

$$0<\sigma<c;-c<\sigma_{ij}<c, i<j=1,\ldots,p;\ \alpha = \alpha_T$$

$$= 0,\ \text{otherwise} \tag{3.41a}$$

and let

$$V_1(\alpha,\beta,\sigma,\Sigma) = \frac{1}{2^{\frac{p(p+1)}{2}}\, 2^{\frac{(p+1)(p+2)}{2}}},\ 0<\sigma_{ii}<c,-c<\beta_i<c, i=1,2,\ldots,p;$$

$$0<\sigma<c;-c<\sigma_{ij}<c, i<j=1,\ldots,p;\ \alpha = \alpha_A$$

$$= 0,\ \text{otherwise.}$$

$V_o(\alpha,\beta,\sigma,\Sigma)$ and $V_1(\alpha,\beta,\sigma,\Sigma)$ are uniform proper prior distributions set out for the express purpose of generating the admissible prior quasi-densities, $W_o(\alpha,\beta,\sigma,\Sigma)$ and $W_1(\alpha,\beta,\sigma,\Sigma)$, and their corresponding marginal likelihoods where

$$W_o(\alpha,\beta,\sigma,\Sigma) = 1,\ 0<\sigma_{ii}<\infty,-\infty<\beta_i<\infty, i=1,2,\ldots,p;$$

$$0<\sigma<\infty;-\infty<\sigma_{ij}<\infty, i<j=1,\ldots,p;\ \alpha = \alpha_T$$

$$= 0,\ \text{otherwise} \tag{3.41b}$$

$$W_1(\alpha,\beta,\sigma,\Sigma) = 1,\ 0<\sigma_{ii}<\infty,-\infty<\beta_i<\infty, i=1,2,\ldots,p;$$

$$0<\sigma<\infty;-\infty<\sigma_{ij}<\infty, i<j=1,\ldots,p;\ \alpha = \alpha_A$$

$$= 0,\ \text{otherwise}$$

It may be noted that the c's given in 3.41a need not be the same but no loss of generality occurs, for the ensuing limiting process results in the same prior quasi-densities whether or not different c's are used. The prior quasi-densities given in 3.41b represent an equal weighting of all points in the parameter spaces w_o and w_1 which may be regarded as an expression of our ignorance about the nuisance parameters.

We now construct modified p.d.f.'s as in 3.2 and form the ratio

$$R(1;\, p;\, n;\, c) = \frac{\int_\alpha \int_{-c}^{c} \cdots \int_{-c}^{c} \int_{o}^{c} \cdots \int_{-c}^{c} V_1(\alpha, \beta, \sigma, \Sigma)}{\int_\alpha \int_{-c}^{c} \cdots \int_{-c}^{c} \int_{o}^{c} \cdots \int_{-c}^{c} V_o(\alpha, \beta, \sigma, \Sigma)} - - -$$

$$\frac{\prod_1^n f(y_i, x_i;\, \alpha, \beta, \sigma^2, \Sigma)\, d\sigma d\beta d\Sigma d\alpha}{\prod_1^n f(y_i, x_i;\, \alpha, \beta, \sigma^2, \Sigma)\, d\sigma d\beta d\Sigma d\alpha} \qquad (3.43)$$

where $d\beta = d\beta_1 d\beta_2 \ldots d\beta_p$ and $d\Sigma = \prod_{j \leq k}^{p} d\sigma_{jk}$.

After substituting from 3.41a and 3.23 and simplifying, 3.43 can be written as

$R(1;\; p;\; n;\; c) =$

$$\frac{\int_{-c}^{c} \cdots \int_{-c}^{c} \int_{o}^{c} \frac{1}{\sigma^{-n}} \exp\left[- \frac{\sum_1^n (y_i - \alpha_A - x_i'\beta)^2}{2\sigma^2}\right] d\sigma d\beta}{\int_{-c}^{c} \cdots \int_{-c}^{c} \int_{o}^{c} \frac{1}{\sigma^{-n}} \exp\left[- \frac{\sum_1^n (y_i - \alpha_T - x_i'\beta)^2}{2\sigma^2}\right] d\sigma d\beta}. \qquad (3.44)$$

Following Wald (69), Hoel (30) and Wallace (71) we now take the limit as c becomes infinite and our test statistic 3.44 becomes

$R(1;\ p;\ n) =$

$$\frac{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{\sigma^{-n}}\exp\left[-\frac{\sum_1^n(y_i-\alpha_A-x_i'\beta)^2}{2\sigma^2}\right]d\beta d\sigma}{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{\sigma^{-n}}\exp\left[-\frac{\sum_1^n(y_i-\alpha_T-x_i'\beta)^2}{2\sigma^2}\right]d\beta d\sigma}. \tag{3.45}$$

If we let $u_i = y_i-\alpha$ then

$$\sum_1^n(y_i-\alpha-x_i'\beta)^2 = \sum_1^n(u_i-x_i'\beta)^2$$

$$= (u-X'\beta)'(u-X'\beta)$$

where $u' = (u_1, u_2, \ldots, u_n)$ is a $1 \times n$ vector and where $X = (x_1, x_2, \ldots, x_n)$ is a $p \times n$ matrix of covariates. We then have

$$\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{\sigma^{-n}}\exp\left[-\frac{\sum_1^n(y_i-\alpha-x_i'\beta)^2}{2\sigma^2}\right]d\beta d\sigma \tag{3.46}$$

$$= \int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{\sigma^{-n}}\exp\left[-\frac{(u-X'\beta)'(u-X'\beta)}{2\sigma^2}\right]d\beta d\sigma$$

$$= \int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{\sigma^{-n}}\exp\left[-\frac{1}{2\sigma^2}(\beta'XX'\beta-2u'X+u'u)\right]d\beta d\sigma. \tag{3.47}$$

Using the theorem proved in Section B of the Appendix, 3.47 becomes

$$\int_0^\infty \exp\left[-\frac{u'u}{2\sigma^2}\right]\sigma^{-n}(2\Pi)^{\frac{p}{2}}\left|\sigma^2(XX')^{-1}\right|^{\frac{1}{2}} \exp\left[+\frac{u'X'(XX')^{-1}Xu}{2\sigma^2}\right] d\sigma$$

$$= (2\Pi)^{\frac{p}{2}} \left|XX'\right|^{-\frac{1}{2}} \int_0^\infty \sigma^{-(n-p)} \exp\left[-\frac{u'(I-X'(XX')^{-1}X)u}{2\sigma^2}\right] d\sigma \quad (3.48)$$

It is important that XX' be nonsingular for the integration in 3.47 to be performed. This is evident because the covariates are sampled from a continuous distribution and the probability is one that $X = (x_1, x_2, \ldots, x_n)$ is of rank p if $p \leq n$. Since rank X = rank XX' we have XX' as a p x p matrix of rank p so that its inverse exists. The integral as given in 3.48 does not have finite value unless

$$n-p > 1 \text{ and } u'(I-X'(XX')^{-1}X)u > 0. \quad (3.49)$$

The requirement $n-p > 1$ implies that the computation of the test statistic given in 3.45 cannot begin until p+2 observations are taken. It is noted that this constraint is consistent with the number of nuisance parameters remaining in 3.45, that is, σ and β_i, $i = 1, \ldots, p$. Also, letting $A = I - X'(XX')^{-1}X$ we can easily verify that

$$A' = A \text{ and } AA = A$$

so that

$$u'Au = u'A'Au = (Au)'Au. \quad (3.50a)$$

Therefore it is always true that

$$u'Au \geq 0. \quad (3.50b)$$

Now, if a random variable z is sampled from a continuous multivariate distribution and is not identically equal to zero, then

$$P[z = a] = 0$$

where a is any constant. Au is a function of the random variables

$$u_i = y_i - \alpha, \; i = 1, 2, \ldots, n$$

and

$$X = (x_1, x_2, \ldots, x_n)$$

so that we can think of Au as a random variable sampled from a continuous multivariate distribution. From 3.50a we see that

$$P[u'Au = 0] = P[Au = 0].$$

Therefore $P[u'Au = 0] = 0$ and we conclude that $u'(I-X'(XX')^{-1}X)u > 0$ with probability one.

From Lemma 1 of Section C of the Appendix the result of the integration in 3.48 is

$$(2\Pi)^{\frac{p}{2}} |XX'|^{\frac{1}{2}} (\tfrac{1}{2})[u'(I-X'(XX')^{-1}X)u]^{\frac{n-p-1}{2}} \Gamma(\tfrac{n-p-1}{2}) \qquad (3.51)$$

so that finally

$$R(1, p, n) = \left[\frac{(y-\alpha_T e)'(I-M)(y-\alpha_T e)}{(y-\alpha_A e)'(I-M)(y-\alpha_A e)}\right]^{\frac{n-p-1}{2}} \qquad (3.52)$$

where $M = X'(XX')^{-1}X$, $e' = (1, 1, \ldots, 1)$, and $y' = (y_1, \ldots, y_n)$.

The behavior of the ratio R(1, p, n) given by 3.52 as the sample size becomes large will now be examined. In particular it may be asked if the test statistic R(1, p, n) must always lead to a decision in favor of H_o or in favor of H_1 or if there is a possibility that it will remain in the interval

(B,A), where $0 < B < 1 < A$, for all n. It will be demonstrated that the probability is zero that R(1, p, n) remains in the interval (B,A) indefinitely unless $\alpha = \frac{\alpha_T+\alpha_A}{2}$.

The following three definitions are required.

<u>Definition 3.1</u>: A sequence of random variables X_n is said to converge in probability to a constant d if for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} P[X_n - d < \epsilon] = 1.$$

We denote this type of convergence by $X_n \xrightarrow{P} d$.

<u>Definition 3.2</u>: A sequence of random variables X_n is said to converge in probability to a random variable X if $(X_n - X)$ converges in probability to zero.

<u>Definition 3.3</u>: A sequence of random variables X_n is said to become arbitrarily large with probability one if for any real number $d > 0$, then

$$\lim_{n\to\infty} P[X_n > d] = 1.$$

We recall that R(1, p, n) was derived on the assumption that the data were being generated by the p.d.f. given in 3.23 so that the vector $y' = (y_1, \ldots, y_n)$ has expectation $\alpha(1, \ldots, 1)$. We now show that if $\alpha < \frac{\alpha_A+\alpha_T}{2}$ then $R(1, p, n) \xrightarrow{P} 0$ and if $\alpha > \frac{\alpha_T+\alpha_A}{2}$ then R(1, p, n) becomes arbitrarily large with probability one.

For this we require the following useful theorem due to E. Slutsky and given in Cramer (21, p. 255).

Theorem 3.2: If Y_{ni} $(i = 1, \ldots, r)$ are random variables converging in probability to the constants a_i $(i = 1, \ldots, r)$ respectively, then any rational function $T(Y_{n1}, \ldots, Y_{nr})$ converges in probability to the constant $T(a_1, a_2, \ldots, a_r)$ provided the latter is finite.

Let us consider now $S(1, p, n)$, where

$$S(1, p, n) = \frac{(y-\alpha_T e)'(I-M)(y-\alpha_T e)}{(y-\alpha_A e)'(I-M)(y-\alpha_A e)} \tag{3.53}$$

and M, e', and y are as defined in 3.52. If $E[y] = \alpha e$ then it follows from Theorem 3.2 and from the fact that maximum likelihood estimators are consistent, with certain mild assumptions (see Fisz, 34), that

$$S(1, p, n) \xrightarrow{P} \frac{\sigma^2 + (\alpha-\alpha_T)^2}{\sigma^2 + (\alpha-\alpha_A)^2}. \tag{3.54}$$

To verify 3.54 we write the numerator of 3.53 as

$$(y-\alpha e)'(I-M)(y-\alpha e) + 2(\alpha-\alpha_T)e'(I-M)(y-\alpha e) + (\alpha-\alpha_T)^2 e'(I-M)e \tag{3.55}$$

and examine the asymptotic properties of each term when that term is divided by n.

From Section A of the Appendix we see that if $E[y] = \alpha e$ then $\frac{1}{n}(y-\alpha e)'(I-M)(y-\alpha e)$ is a consistent estimator of σ^2, that is,

$$\frac{1}{n}(y-\alpha e)'(I-M)(y-\alpha e) \xrightarrow{P} \sigma^2.$$

Also, since $E[X] = 0$ and we can write

$$\frac{1}{n} e'(I-M)e = 1 + (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)(\frac{XX'}{n})^{-1}(\bar{x}_1, \ldots, \bar{x}_p)$$

then

$$\frac{1}{n} e'(I-M)e \xrightarrow{P} 1 + (0, \ldots, 0)(\Sigma)^{-1}(0, \ldots, 0) = 1.$$

Finally, we can write

$$\frac{1}{n} e'(I-M)(y-\alpha e) = y - \alpha + (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)(\frac{XX'}{n})^{-1} \frac{X(y-\alpha e)}{n}.$$

Now $\bar{y} \xrightarrow{P} \alpha$, $\frac{XX'}{n} \xrightarrow{P} \Sigma$, $\frac{X(y-\alpha e)}{n} \xrightarrow{P} \text{cov}(x,y)$ and $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p) \xrightarrow{P} 0 \cdot e$ so that

$$\frac{1}{n} e' (I-M)(y-\alpha e) \xrightarrow{P} 0.$$

Thus the numerator of 3.53, when divided by n, converges in probability to $\sigma^2 + (\alpha-\alpha_T)^2$. Similarly, we can show that the denominator of 3.53, when divided by n, converges in probability to $\sigma^2 + (\alpha-\alpha_A)^2$. Therefore use of Theorem 3.2 establishes the result given in 3.54.

Writing

$$\frac{\sigma^2 + (\alpha-\alpha_T)^2}{\sigma^2 + (\alpha-\alpha_A)^2} \equiv C(\alpha),$$

for which we have established that

$$S(1, p, n) \xrightarrow{P} C(\alpha),$$

it is then easy to verify that

$$\begin{aligned} C(\alpha) &< 1 \quad \text{if} \quad \alpha < \frac{\alpha_T + \alpha_A}{2} \\ &> 1 \quad \text{if} \quad \alpha > \frac{\alpha_T + \alpha_A}{2} \end{aligned} \tag{3.56}$$

We now show that

$$[S(1, p, n)]^{f(n)}, \tag{3.57}$$

where f(n) is an increasing unbounded function of n, does not remain in the interval (B,A), $0 < B < 1 < A$, indefinitely when $\alpha \gtrless \frac{\alpha_T+\alpha_A}{2}$. This fact follows immediately from the following theorems which are proved in Section D of the Appendix.

Theorem 3.3: If $X_n \xrightarrow{P} C$, $0 < C < 1$, and if f(n) is an increasing unbounded function of n, then $[X_n]^{f(n)} \xrightarrow{P} 0$.

Theorem 3.4: If $X_n \xrightarrow{P} C$, $C > 1$, and if f(n) is an increasing unbounded function of n, then $[X_n]^{f(n)}$ becomes arbitrarily large with probability one.

It has been shown that $S(1, p, n) \xrightarrow{P} C(\alpha)$. Hence, if $\alpha < \frac{\alpha_T+\alpha_A}{2}$ so that $C(\alpha) < 1$, then taking $f(n) = \frac{n-p-1}{2}$ in Theorem 3.3 shows $R(1, p, n) \xrightarrow{P} 0$. Again if $\alpha > \frac{\alpha_T+\alpha_A}{2}$, then $C(\alpha) > 1$ and Theorem 3.4 implies that R(1, p, n) becomes arbitrarily large with probability one. If $\alpha = \frac{\alpha_T+\alpha_A}{2}$ so that $C(\alpha) = 1$, the behavior of R(1, p, n) is still an open question. In summary, the SPRT with R(1, p, n) as the test statistic decisions with probability one if $\alpha \gtrless \frac{\alpha_T+\alpha_A}{2}$.

In conclusion, the test derived in this section should be considered as a "least favorable" approach to the hypothesis formulation

$$H_o: \ \alpha \leq \alpha_T \text{ versus } H_1: \ \alpha \geq \alpha_A$$

with nuisance parameters unknown. Because of the manner in which this test is derived one would expect its performance in practical situations to be dependent on the investigator's choice of α_T and α_A. In particular, poor specifications of α_A and/or α_T might result in excessively large average sample numbers. These topics will therefore be considered further in Chapter VI.

D. A Test for H_o: $\alpha=\alpha_T$ Versus H_1: $\alpha=\alpha_T+\gamma\sigma$ with γ and α_T Specified, and Nuisance Parameters Unknown

A number of developments in the sequential testing of hypotheses have been based on analogies with fixed sample size methodology. For example, it is well known that Neyman-Pearson testing theory, when applied to the least favorable hypothesis formulation

$$\begin{aligned} H_o&: \frac{\mu-\mu_o}{\sigma} = 0 & \sigma\varepsilon(0,\infty) \\ H_1&: \frac{\mu-\mu_o}{\sigma} = \delta, & \delta > 0,\ \sigma\varepsilon(0,\infty) \end{aligned} \tag{3.58}$$

provides a uniformly most powerful unbiased test and also a uniformly most powerful invariant test for the hypothesis formulation

$$\begin{aligned} H_o&: \frac{\mu-\mu_o}{\sigma} \leq 0 & \sigma\varepsilon(0,\infty) \\ H_1&: \frac{\mu-\mu_o}{\sigma} > 0 & \sigma\varepsilon(0,\infty) \end{aligned} \tag{3.59}$$

and also for the ultimate formulation

$$H_o: \mu \leq \mu_o \qquad \sigma\varepsilon(0,\infty)$$
$$H_1: \mu > \mu_o \qquad \sigma\varepsilon(0,\infty) \tag{3.60}$$

The test alluded to above is the well known Student's t-test.

Wald (69), Rushton (59, 60), Cox (20) and Hajnal (26) developed what are termed sequential t-tests and t^2-tests for hypotheses of the type given in 3.58. Hoel (30), Johnson (39), and Ray (55) developed and discussed sequential F-tests based on hypothesis formulations similar to those of 3.58. In sequential analysis, however, we must specify the alternative and consequently we have difficulty in developing tests for hypothesis formulations of the type given in 3.60 from those given in 3.58. For, in 3.58, we are confronted with the alternative H_1: $\mu-\mu_o = \delta\sigma$, where although δ is known, σ and therefore $\delta\sigma$ are unknown. Thus the magnitude of the difference which can be detected with a given power is unknown. However, if we have a prior estimate of σ or if we are only interested in the detection of a difference $\mu-\mu_o$ scaled in standard deviation units we may nevertheless develop a sequential test for the hypothesis formulation as given in 3.58.

As an application to the problem as stated in Section B of this chapter, it will be shown that the weight function approach provides a test statistic which would be intuitively expected as a generalization of the results of Wald (69) and Rushton (59, 60).

Let us suppose we are sequentially sampling random variables (Y_i, X_i), $i=1, 2, \ldots, n, \ldots$ from a distribution which has the probability density function 3.23. Let us concern ourselves with constructing a sequential test for the least favorable hypothesis formulation

$$H_o: \alpha = \alpha_T$$
$$H_1: \alpha = \alpha_T + \gamma\sigma \qquad (3.61)$$

where α_T and γ are specified numbers and σ, β, Σ are assumed unknown.

The region of preference for acceptance of H_o is

$w_o = [(\alpha, \sigma, \beta_1, \beta_2, \ldots, \beta_p, \sigma_{11}, \ldots, \sigma_{1p}, \sigma_{22}, \ldots, \sigma_{2p}, \ldots, \sigma_{pp})$:
$\alpha = \alpha_T$ and the parameters σ; β_i, $i=1, \ldots, p$; σ_{ij},
$i \leq j=1, \ldots, p$ are unspecified].

The region of preference for acceptance of H_1 is

$w_1 = [(\alpha, \sigma, \beta_1, \beta_2, \ldots, \beta_p, \sigma_{11}, \ldots, \sigma_{1p}, \sigma_{22}, \ldots, \sigma_{2p}, \ldots, \sigma_{pp})$:
$\alpha = \alpha_T + \gamma\sigma$ and the parameters σ; β_i, $i=1, \ldots, p$; σ_{ij},
$i \leq j=1, \ldots, p$ are unspecified].

The region where neither H_o or H_1 is preferred is the complement of the region $w_o \cup w_1$.

The prior quasi-densities we shall adopt are

$$W_o(\alpha, \beta, \sigma, \Sigma) = 1, \; 0<\sigma_{ii}<\infty, \; -\infty<\beta_i<\infty, \; i=1, \ldots, p;$$
$$0<\sigma<\infty; \; -\infty<\sigma_{ij}<\infty, \; i<j=1, \ldots, p; \; \alpha = \alpha_T$$
$$= 0 \quad \text{otherwise} \qquad (3.62)$$

$$W_1(\alpha, \beta, \sigma, \Sigma)=1, 0<\sigma_{ii}<\infty; -\infty<\beta_i<\infty, i=1, \ldots, p;$$
$$0<\sigma<\infty; -\infty<\sigma_{ij}<\infty, i\leq j=1, \ldots, p; \alpha = \alpha_T+\gamma\sigma$$
$$= 0 \quad \text{otherwise.}$$

Using these admissible prior quasi-densities (see 3.19) to form the ratio of their corresponding marginal likelihoods we subsequently arrive at the ratio

$R(2, p, n) =$

$$\frac{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{\sum_1^n (y_i-\alpha_A-\gamma\sigma-x_i'\beta)^2}{2\sigma^2}\right] d\beta d\sigma}{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{\sum_1^n (y_i-\alpha_T-x_i'\beta)^2}{2\sigma^2}\right] d\beta d\sigma} \tag{3.63}$$

where $d\beta = d\beta_1 d\beta_2 \ldots d\beta_p$.

If we let $e' = (1, \ldots, 1)$, $X = (x_1, \ldots, x_n)$ $z_i = y_i - \alpha_T$, and $z' = (z_1, z_2, \ldots, z_n)$ we can rewrite 3.63 as

$$R(2,p,n) = \frac{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{(z-\gamma\sigma e-X'\beta)'(z-\gamma\sigma e-X'\beta)}{2\sigma^2}\right] d\beta d\sigma}{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{(z-X'\beta)'(z-X'\beta)}{2\sigma^2}\right] d\beta d\sigma} \tag{3.64}$$

$$= \frac{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{\beta'XX'\beta-2(z-\gamma\sigma e)'X'\beta+(z-\gamma\sigma e)'(z-\gamma\sigma e)}{2\sigma^2}\right] d\beta d\sigma}{\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{\beta'XX'\beta-2z'X'\beta+z'z}{2\sigma^2}\right] d\beta d\sigma} \tag{3.65}$$

It is shown in Section B of the Appendix that if A is a real p x p matrix of rank p, β is a p x 1 vector of real valued variables and u' is a 1 x p vector of arbitrary real numbers then

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}\beta' A\beta + u'\beta\right] d\beta_1 d\beta_2 \ldots d\beta_p$$

$$= (2\Pi)^{\frac{p}{2}} |A|^{-\frac{1}{2}} \exp\left[\frac{u'A^{-1}u}{2}\right] \tag{3.66}$$

and hence with $A = \frac{XX'}{\sigma^2}$ and $u = \frac{(z-\gamma\sigma e)'X'}{\sigma^2}$ we have

$$R(2,p,n) = \frac{\int_0^\infty \sigma^{-n} (2\Pi)^{\frac{p}{2}} |\sigma^2(XX')^{-1}|^{\frac{1}{2}} \exp\left[-\frac{(z-\gamma\sigma e)'(I-M)(z-\gamma\sigma e)}{2\sigma^2}\right] d\sigma}{\int_0^\infty \sigma^{-n} (2\Pi)^{\frac{p}{2}} |\sigma^2(XX')^{-1}|^{\frac{1}{2}} \exp\left[-\frac{z'(I-M)z}{2\sigma^2}\right] d\sigma} \tag{3.67}$$

$$= \frac{\int_0^\infty \sigma^{-(n-p)} \exp\left[-\frac{z'(I-M)z-2\gamma\sigma e'(I-M)z+\gamma^2\sigma^2 e'(I-M)e}{2\sigma^2}\right] d\sigma}{\int_0^\infty \sigma^{-(n-p)} \exp\left[-\frac{z'(I-M)z}{2\sigma^2}\right] d\sigma} \tag{3.68}$$

where, as before, $M = X'(XX')^{-1}X$.

Further simplifications ensue by application of the following theorem which is proved in Section C of the Appendix.

<u>Theorem 3.5</u>: If $s > 1$, $\delta > 0$, and $F(p;\, q;\, x) = \sum_{i=0}^{\infty} \frac{\Gamma(p+i)\,\Gamma(q)}{\Gamma(p)\,\Gamma(q+i)} \frac{x^i}{i!}$ then

$$\frac{\int_0^\infty \sigma^{-s} \exp[-\frac{\delta}{2}\sigma^{-2} - \frac{\theta}{2}\sigma^{-1}]\, d\sigma}{\int_0^\infty \sigma^{-s} \exp[-\frac{\delta}{2}\sigma^{-2}]\, d\sigma} =$$

$$F(\tfrac{s-1}{2};\ \tfrac{1}{2};\ \tfrac{\delta^2}{8\theta}) - \frac{\delta}{(2\theta)^{\frac{1}{2}}} \frac{\Gamma(\frac{s}{2})}{\Gamma(\frac{s-1}{2})} F(\tfrac{s}{2};\ \tfrac{3}{2};\ \tfrac{\delta^2}{8\theta}). \tag{3.69}$$

Applying this theorem to the ratio in 3.68 we have

$R(2, p, n)$

$$= \exp\left[-\frac{\gamma^2 e'(I-M)e}{2}\right] \left[F\left(\frac{n-p-1}{2};\ \frac{1}{2};\ \frac{\gamma^2}{2}\ \frac{(e'(I-M)z)^2}{z'(I-M)z}\right) +\right.$$

$$\left.\frac{2\gamma e'(I-M)z}{[2z'(I-M)z]^{\frac{1}{2}}} \frac{\Gamma(\frac{n-p}{2})}{\Gamma(\frac{n-p-1}{2})} F\left(\frac{n-p}{2};\ \frac{3}{2};\ \frac{\gamma^2}{2}\ \frac{(e'(I-M)z)^2}{z'(I-M)z}\right)\right] \tag{3.70}$$

Again it is noted that the statistic $R(2, p, n)$ cannot be computed unless $n \geq p+1$. It may also be noted from the conditions required for Theorem 3.5 that the reduction from 3.68 to 3.69 requires that $z'(I-M)z > 0$. By an argument similar to that given in Section C of this chapter we can say, however, that $P[z'(I-M)z > 0] = 1$.

To see that the result given in 3.70 is what might be anticipated we examine the model

$$y = \alpha e + X'\beta + \epsilon \tag{3.71}$$

where $e' = (1, \ldots, 1)$, $X = (x_1, \ldots, x_n)$ is a fixed pxn matrix of known constants, α is an unknown scalar, and β is an unknown px1 vector. In addition we assume $\epsilon \sim N(o,\sigma^2 I)$ so that $E[y] = e\alpha + X'\beta$.

We will derive the t-statistic associated with a fixed sample size test of the hypothesis

$$\begin{aligned} H_o&: \alpha \leq 0 \\ H_1&: \alpha > 0 \end{aligned} \tag{3.72}$$

and show that if the X used in 3.70 is assumed to be a matrix of known constants then R(2, p, n) as given in 3.70 may be regarded as a generalization to the covariate case of the sequential t-test as set out by Wald (69), Rushton (59, 60) and Cox (20).

By standard methods it is easily shown from the normal equations for the model 3.71 that

$$\begin{aligned} \hat{\alpha} &= \frac{e'(I-M)y}{e'(I-M)e} \\ \hat{\beta} &= (XX')^{-1}(y-e\hat{\alpha}) \\ \hat{\sigma}^2 &= \frac{y'(I-M)y-\hat{\alpha}e'(I-M)y}{n-p-1} \end{aligned} \tag{3.73}$$

so that the t-test statistic for $\hat{\alpha}$ is

$$t = \frac{\hat{\alpha}}{\sqrt{\widehat{\text{Var}}\,(\hat{\alpha})}} = \frac{e'(I-M)y}{\sqrt{\hat{\sigma}^2 e'(I-M)e}}. \tag{3.74}$$

An examination of the sequential t-test as given by Rushton (59) shows us that the third argument in the confluent hypergeometric function F(_; _; _) as given in 3.70 should be

$$\frac{\gamma^2\sigma^2}{2\ \text{Var}\ (\hat{\alpha})} \quad \frac{t^2}{n-p-1+t^2} \tag{3.75}$$

in order for 3.70 to be consistent with what is expected as a generalization of Wald's t-test (59).

If we substitute the t as given in 3.74 into 3.75 we have

$$\frac{\gamma^2 e'(I-M)e}{2} \frac{t^2}{n-p-1+t^2}$$

$$= \frac{\gamma^2 e'(I-M)e}{2} \frac{[e'(I-M)y]^2}{(n-p-1)\hat{\sigma}^2 \; e'(I-M)e + [e'(I-M)y]^2}$$

$$= \frac{\gamma^2 e'(I-M)e}{2} \frac{[e'(I-M)y]^2}{[y'(I-M)y - \hat{\alpha}e'(I-M)y]e'(I-M)e + [e'(I-M)y]^2}$$

$$= \frac{\gamma^2}{2} \frac{[e'(I-M)y]^2}{y'(I-M)y}. \tag{3.76}$$

If in 3.71 we set $\alpha_T = 0$ so that $z = y - \alpha_T e = y$ we see that 3.76 becomes

$$\frac{\gamma^2 [e'(I-M)z]^2}{2z'(I-M)z} \tag{3.77}$$

which is the last argument of the confluent hypergeometric functions found in 3.70.

In this last part we have therefore shown that R(2, p, n), with controlled x's, may be intuitively expected as a generalization to the covariate case of the results of Wald (69) and Rushton (59, 60).

E. A Test for H_o: $|\alpha-\alpha_T| = \gamma_o\sigma$ Versus H_1: $|\alpha-\alpha_T| = \gamma_1\sigma$ with γ_o and γ_1 Specified and Nuisance Parameters Unknown

As in Sections C and D of this chapter, we suppose that we are sequentially sampling random variables (Y_i, X_i), $i = 1, 2, \ldots, n, \ldots$ from a distribution which has the probability density function as given in 3.23. For a generalization of the hypothesis formulation given in 3.61 we now consider the construction of a sequential test for

$$H_0: |\alpha-\alpha_T| = \gamma_0\sigma$$
$$H_1: |\alpha-\alpha_T| = \gamma_1\sigma \tag{3.78}$$

where α_T, γ_0, and γ_1 are known real numbers, $\gamma_0 < \gamma_1$, and where the nuisance parameters are unknown. We shall apply essentially the same techniques used in Section D of this chapter and also show that the results obtained are those to be intuitively expected as a generalization of the t^2-tests of Wald (69) and Rushton (59, 60).

We consider the prior quasi-densities which are as follows:

$$\begin{aligned} V_0(\alpha, \beta, \sigma, \Sigma) &= 1,\ 0<\sigma_{ii}<\infty,\ -\infty<\beta_i<\infty,\ i=1, \ldots, p; \\ &\qquad 0<\sigma<\infty;\ -\infty<\sigma_{ij}<\infty,\ i<j=1, \ldots, p \\ &\qquad \alpha=\alpha_T+\gamma_0\sigma,\ \alpha = \alpha_T - \gamma_0\sigma \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{3.79}$$

$$\begin{aligned} V_1(\alpha, \beta, \sigma, \Sigma) &= 1,\ 0<\sigma_{ii}<\infty,\ -\infty<\beta_i<\infty,\ i=1, \ldots, p; \\ &\qquad 0<\sigma<\infty;\ -\infty<\sigma_{ij}<\infty,\ i<j=1, \ldots, p \\ &\qquad \alpha=\alpha_T+\gamma_1\sigma,\ \alpha = \alpha_T - \gamma_1\sigma \\ &= 0 \quad \text{otherwise} \end{aligned}$$

Using these admissible prior quasi-densities (see 3.19) to form the ratio of their corresponding marginal likelihoods we subsequently have the quantity

$$R(3, p, n) = \frac{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty [Q(\beta,\sigma,\alpha_T+\gamma_1\sigma) + Q(\beta,\sigma,\alpha_T-\gamma_1\sigma)]\ d\beta d\sigma}{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty [Q(\beta,\sigma,\alpha_T+\gamma_0\sigma) + Q(\beta,\sigma,\alpha_T-\gamma_0\sigma)\ \ d\beta d\sigma} \tag{3.80}$$

where

$Q(\beta, \sigma, \alpha_T + a\sigma)$

$$= \sigma^{-n} \exp\left[-\frac{\sum_1^n (y_i-\alpha_T-a\sigma-x_i'\beta)^2}{2\sigma^2}\right] \tag{3.81}$$

and $a = \gamma_0,\ \gamma_1$.

Using matrix notation we can write

$Q(\beta,\ \sigma,\ \alpha_T + a\sigma)$

$$= \sigma^{-n} \exp\left[-\frac{\beta'XX'\beta - 2(z-a\sigma e)'X'\beta + (z-a\sigma e)'(z-a\sigma e)}{2\sigma^2}\right] \tag{3.82}$$

where $X = (x_1,\ x_2,\ \ldots,\ x_n)$

$z' = (z_1,\ z_2,\ \ldots,\ z_n)$

$z_i = y_i - \alpha_T,\ i = 1,\ 2,\ \ldots,\ n$

$e' = (1,\ 1,\ \ldots,\ 1)$.

By the result derived in Section B of the Appendix we have

$\int_{-\infty}^\infty \cdots \int_{-\infty}^\infty Q(\beta,\ \sigma,\ \alpha_T + a\sigma)\ d\beta$

$$= h(X,\ p)\ \sigma^{-(n-p)} \exp\left[-\frac{(z-a\sigma e)'(I-M)(z-a\sigma e)}{2\sigma^2}\right] \tag{3.83}$$

where

$$h(X,\ p) = (\sqrt{2\Pi})^p\ |XX'|^{-\frac{1}{2}}$$

$$M = X'(XX')^{-1}X.$$

Expanding the quadratic form given in 3.83 we write

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} Q(\beta,\ \sigma,\ \alpha_T + a\sigma)\ d\beta$$

$$= h(X,\ p)\ \sigma^{-(n-p)}\ \exp\ [-\ \frac{1}{2\sigma^2}(\sigma^2 a^2 e'(I-M)e\ -\ 2a\sigma e'(I-M)z$$

$$+\ z'(I-M)z)] \tag{3.84}$$

By Lemma 2 of Section C of the Appendix we can write

$$\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} Q(\beta,\ \sigma,\ \alpha_T + a\sigma)\ d\beta d\sigma$$

$$= h(X,\ p)\ \exp\ [-\ \frac{a^2 e'(I-M)e}{2}]\ (z'(I-M)z)^{-\frac{n-p-1}{2}}\ 2^{-\frac{n-p-4}{2}}$$

$$\Gamma(\frac{n-p-1}{2})[F(\frac{n-p-1}{2};\ \frac{1}{2};\ \frac{a^2}{2}\ \frac{(e'(I-M)z)^2}{z'(I-M)z})\ +\ \frac{2ae'(I-M)z}{2z'(I-M)z}$$

$$\frac{\Gamma(\frac{n-p}{2})}{\Gamma(\frac{n-p-1}{2})}\ F(\frac{n-p}{2};\ \frac{3}{2};\ \frac{a^2}{2}\ \ \frac{(e'(I-M)z)^2}{z'(I-M)z})]. \tag{3.85}$$

From 3.85 we can write

$$\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} [Q(\beta,\ \sigma,\ \alpha_T + a\sigma)\ +\ Q(\beta,\ \sigma,\ \alpha_T\ -\ a\sigma)]\ d\beta d\sigma$$

$$= h_1(X,\ z,\ n,\ p)\ \exp\ [-\ \frac{a^2 e'(I-M)e}{2}]\ F(\frac{n-p-1}{2};\ \frac{1}{2};\ \frac{a^2}{2}\ \ \frac{(e'(I-M)z)^2}{z'(I-M)z}) \tag{3.86}$$

where

$$h_1(X,\ z,\ n,\ p) = h(X,\ p)\ 2^{\frac{n-p-4}{2}}\ \Gamma(\frac{n-p-1}{2})(z'(I-M)z)^{-\frac{n-p-1}{2}}.$$

Therefore using the result given in 3.86 we have finally

$$R(3,\ p,\ n) = \exp\left[-\frac{(\gamma_1^2 - \gamma_0^2)e'(I-M)e}{2}\right]$$

$$\frac{F\left(\frac{n-p-1}{2};\ \frac{1}{2};\ \frac{\gamma_1^2}{2}\ \frac{(e'(I-M)z)^2}{z'(I-M)z}\right)}{F\left(\frac{n-p-1}{2};\ \frac{1}{2};\ \frac{\gamma_0^2}{2}\ \frac{(e'(I-M)z)^2}{z'(I-M)z}\right)}. \tag{3.87}$$

To see that the result given in 3.87 is as expected we examine the model 3.71 and exhibit the well-known fixed sample size test statistic, t^2,

$$t^2 = \frac{(\hat{\alpha})^2}{\widehat{\text{Var}}(\hat{\alpha})} \tag{3.88}$$

for testing the two-sided hypothesis formulation

$$\begin{aligned} H_0&:\ \alpha = 0 \\ H_1&:\ \alpha \neq 0 \end{aligned} \tag{3.89}$$

under the assumption of normality of ϵ. The estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$ are given by 3.73 so that 3.88 becomes

$$t^2 = \frac{(e'(I-M)y)^2}{\sigma^2 e'(I-M)e}. \tag{3.90}$$

An examination of the sequential t^2-test as given by Rushton (60) will show that the third argument of the confluent hypergeometric function $F(\_;\ \_;\ \_)$ given in 3.87 should be

$$\frac{\gamma^2\sigma^2}{2\ \text{Var}\ (\hat{\alpha})}\ \frac{t^2}{n-p-1 + t^2}. \tag{3.91}$$

As in Section D of this chapter we find that 3.91 can be written as

$$\frac{\gamma^2}{2}\,\frac{(e'(I-M)z)^2}{z'(I-M)z}$$

so that we have our expected correspondence.

F. Tests for H_o: $\alpha \leq \alpha_T$ Versus H_1: $\alpha \geq \alpha_A$, $\alpha_T < \alpha_A$

1. Derivation of tests

In Sections C and D we sequentially sampled random variables (Y_i, X_i), $i = 1, 2, \ldots, n, \ldots$ from a distribution with probability density function

$g(y, x; \alpha, \beta, \Sigma, \sigma^2) =$

$$\frac{1}{(2\Pi)^{\frac{p}{2}}}\,\frac{1}{|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{x'\Sigma^{-1}x}{2}\right]\frac{1}{(2\Pi)^{\frac{1}{2}}\sigma}\exp\left[-\frac{(y-\alpha-x'\beta)^2}{2\sigma^2}\right] \tag{3.92}$$

where x and β are p x 1 vectors; α, y and σ^2 are scalars; and Σ is a positive definite p x p matrix. In those sections we considered two hypothesis formulations which can be viewed as "least favorable" approaches to the more general hypothesis formulation

$$\begin{aligned} H_o&: \alpha \leq \alpha_T \\ H_1&: \alpha \geq \alpha_A \end{aligned} \tag{3.93}$$

where α_T and α_A are specified known numbers, $\alpha_T < \alpha_A$; and where β, σ, and Σ are assumed unknown. We considered regions of the parameter space which are degenerate with respect to the parameter of interest. For example, in Section C the preference regions for H_o and H_1 are planes in $(\frac{p(p+3)}{2} + 2)$ space perpendicular to the α-axis at $\alpha = \alpha_T$ and $\alpha = \alpha_A$ respectively.

In this section we consider alternative prior quasi-densities for the regions, w_o and w_1, of preference for H_o and H_1 respectively where

$$w_o = [(\alpha, \sigma, \beta_1, \ldots, \beta_p, \sigma_{11}, \sigma_{12}, \ldots, \sigma_{1p}, \sigma_{2p}, \ldots, \sigma_{pp}): \alpha \leq \alpha_T \text{ and the parameters } ; \beta_i, i = 1, \ldots, p; \sigma_{ij}, i \leq j = 1, \ldots, p \text{ are unspecified}] \quad (3.94)$$

$$w_1 = [(\alpha, \sigma, \beta_1, \ldots, \beta_p, \sigma_{11}, \sigma_{12}, \ldots, \sigma_{1p}, \sigma_{22}, \ldots, \sigma_{pp}): \alpha \geq \alpha_A \text{ and the parameters } \sigma; \beta_i, i = 1, \ldots, p; \sigma_{ij}, i \leq j = 1, \ldots, p \text{ are unspecified}].$$

The region where neither H_o or H_1 is preferred is the complement of the region w_o and w_1.

The first set of prior quasi-densities is used to give uniform weight to each point in w_o and w_1. We begin by constructing proper prior densities (i.e. which assign measures that are finite and equal to one) which are as follows. Let $c > \max(|\alpha_T|, |\alpha_A|)$ and

$$V_o(\alpha, \beta, \sigma, \Sigma) = \frac{1}{2^{\frac{p(p+1)}{2}} c^{\frac{(p+1)(p+2)}{2}} (\alpha_T + c)},$$

where $0 < \sigma_{ii} < c,\ -c < \beta_i < c,\ i = 1, 2, \ldots, p;$

$0 < \sigma < c;\ -c < \sigma_{ij} < c,\ i < j = 1, \ldots, p;\ -c \leq \alpha \leq \alpha_T$

$= 0$, otherwise (3.95)

$$V_1(\alpha, \beta, \sigma, \Sigma) = \frac{1}{2^{\frac{p(p+1)}{2}}\, c^{\frac{(p+1)(p+2)}{2}}\, (c - \alpha_A)},$$

where $0 < \sigma_{ii} < c,\ -c < {}_i < c,\ i = 1, \ldots, p;$

$0 < \sigma < c;\ -c < \sigma_{ij} < c,\ i < j = 1, \ldots, p;\ \alpha_A \leq \alpha \leq c$

$= 0$, otherwise.

We now construct the modified p.d.f.'s as in 3.2 and form their ratio to obtain $R(4, p, n, c)$, where

$$R(4, p, n, c) = \frac{\alpha_T + c}{c - \alpha_A}$$

$$\frac{\int_{\alpha_A}^{c}\int_{0}^{c}\int_{-c}^{c}\cdots\int_{-c}^{c} \sigma^{-n} \exp\left[-\frac{\sum_1^n (y_i-\alpha-x_i'\beta)^2}{2\sigma^2}\right] d\beta d\sigma d\alpha}{\int_{-c}^{\alpha_T}\int_{0}^{c}\int_{-c}^{c}\cdots\int_{-c}^{c} \sigma^{-n} \exp\left[-\frac{\sum_1^n (y_i-\alpha-x_i'\beta)^2}{2\sigma^2}\right] d\beta d\sigma d\alpha} \tag{3.96}$$

Now, as previously in Section C we allow c to become infinite and 3.96 becomes

$R(4, p, n) =$

$$\frac{\int_{\alpha_A}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{\sum_1^n (y_i-\alpha-x_i'\beta)^2}{2\sigma^2}\right] d\beta d\sigma d\alpha}{\int_{-\infty}^{\alpha_T}\int_{0}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-n} \exp\left[-\frac{\sum_1^n (y_i-\alpha-x_i'\beta)^2}{2\sigma^2}\right] d\beta d\sigma d\alpha} \tag{3.97}$$

In 3.97, $R(4, p, n)$ is the ratio of marginal likelihoods, the marginal likelihoods being of the experimental outcome given the admissible prior quasi-densities

$$
\begin{aligned}
W_o(\alpha, \beta, \sigma, \Sigma) &= 1, \text{ for each point in } w_o \\
&= 0, \text{ otherwise} \\
W_1(\alpha, \beta, \sigma, \Sigma) &= 1, \text{ for each point in } w_1 \\
&= 0, \text{ otherwise.}
\end{aligned} \tag{3.98}
$$

Letting $u_i = y_i - \alpha$, $u' = (u_1, u_2, \ldots, u_n)$, $X = (x_1, \ldots, x_n)$ we have

$R(4, p, n) =$

$$
\frac{\int_{\alpha_A}^{\infty} \int_{0}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sigma^{-n} \exp[-\frac{(u-X'\beta)'(u-X'\beta)}{2\sigma^2}]\, d\beta d\sigma d\alpha}{\int_{-\infty}^{\alpha_T} \int_{0}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sigma^{-n} \exp[-\frac{(u-x'\beta)'(u-X'\beta)}{2\sigma^2}]\, d\beta d\sigma d\alpha} . \tag{3.99}
$$

Using the same integration techniques as in Section C of this chapter we can write 3.99 as

$$
R(4, p, n) = \frac{\int_{\alpha_A}^{\infty} [(y-\alpha e)'(I-M)(y-\alpha e)]^{-\frac{n-p-1}{2}} d\alpha}{\int_{-\infty}^{\alpha_T} [(y-\alpha e)'(I-M)(y-\alpha e)]^{-\frac{n-p-1}{2}} d\alpha} \tag{3.100a}
$$

where $e' = (1, \ldots, 1)$ and $M = X'(XX')^{-1}X$. It is noted that the step from 3.99 to 3.100a requires that $(y-\alpha e)'(I-M)(y-\alpha e)>0$. By an argument similar to that given in Section C of this chapter we can say, however, that

$$
P[(y-\alpha e)'(I-M)(y-\alpha e) > 0] = 1. \tag{3.100b}
$$

It is also noted that $R(4, p, n)$ cannot be computed unless $n \geq p + 1$.

For convenience we now write

$$(y-\alpha e)'(I-M)(y-\alpha e) = a\alpha^2 + b\alpha + c \tag{3.101}$$

where

$$\begin{aligned} a &= e'(I-M)e \\ b &= -2e'(I-M)y \\ c &= y'(I-M)y. \end{aligned} \tag{3.102}$$

Beyer (11) gives the following results.

<u>Case I</u>: If $n-p$ is odd, then $n-p-1$ is even and $\frac{n-p-1}{2}$ is an integer so that

$$\int \frac{dx}{(ax^2+bx+c)^{\frac{n-p-1}{2}+1}} = \frac{(n-p-3)!}{\left(\frac{n-p-3}{2}!\right)^2}\left(\frac{a}{4ac-b^2}\right)^{\frac{n-p-3}{2}}$$

$$\frac{2ax+b}{4ac-b^2} \sum_{r=1}^{\frac{n-p-3}{2}} \left(\frac{4ac-b^2}{a(ax^2+bx+c)}\right)^r H(r) + \int \frac{dx}{ax^2+bx+c} \tag{3.103}$$

where

$$H(r) = \frac{(r-1)!\ r!}{(2r)!}, \quad r = 1, 2, \ldots, \frac{n-p-3}{2} \tag{3.104}$$

and

$$\int \frac{dx}{ax^2+bx+c} = \frac{2}{(4ac-b^2)^{\frac{1}{2}}} \tan^{-1} \frac{2ax+b}{(4ac-b^2)^{\frac{1}{2}}}. \tag{3.105}$$

It is noted here that 3.100b and 3.101 imply that $P[4ac-b^2 > 0] = 1$.

Case II: If n-p is even then n-p-2 and $\frac{n-p-2}{2}$ are integers and we have

$$\int \frac{dx}{(ax^2+bx+c)^{\frac{n-p-2}{2}+\frac{1}{2}}} =$$

$$\frac{(2ax+b)\left(\frac{n-p-2}{2}!\right)\left(\frac{n-p-4}{2}!\right) 4^{\frac{n-p-2}{2}} \left(\frac{4a}{4ac-b^2}\right)^{\frac{n-p-4}{2}}}{(4ac-b^2)[(n-p-2)!](ax^2+bx+c)^{\frac{1}{2}}} D(x) \qquad (3.106)$$

where

$$D(x) = \sum_{r=o}^{\frac{n-p-4}{2}} G(r) \left(\frac{4ac-b^2}{4a(ax^2+bx+c)}\right)^r \qquad (3.107)$$

and

$$G(r) = \frac{(2r)!}{(r!)^2}. \qquad (3.108)$$

Therefore

I. If n-p is odd 3.100a becomes

$R(4, p, n) =$

$$\frac{\Pi - \frac{2a\alpha_A + b}{(4ac-b^2)^{\frac{1}{2}}} \sum_{r=1}^{\frac{n-p-3}{2}} \left(\frac{4ac-b^2}{a(a\alpha_A^2+b\alpha_A+c)}\right)^r H(r) - 2\tan^{-1}\frac{2a\alpha_A+b}{(4ac-b^2)^{\frac{1}{2}}}}{\frac{2a\alpha_T + b}{(4ac-b^2)^{\frac{1}{2}}} \sum_{r=1}^{\frac{n-p-3}{2}} \left(\frac{4ac-b^2}{a(a\alpha_T^2+b\alpha_T+c)}\right)^r H(r) + 2\tan^{-1}\frac{2a\alpha_T+b}{(4ac-b^2)^{\frac{1}{2}}} + \Pi}$$

(3.109)

where H(r) is as given in 3.104.

In the equation

$$\tan^{-1} K = L$$

L has the form

$$\theta + 2s\Pi \quad (s = 0, 1, 2, \ldots,)$$

where θ is called the <u>principal</u> <u>value</u> and is defined to be such that

$$-\frac{\Pi}{2} < \theta < \frac{\Pi}{2}.$$

Representing the principal values of

$$\tan^{-1} \frac{2a\alpha_A + b}{(4ac - b^2)^{\frac{1}{2}}} \text{ by } \theta_A$$

and of

$$\tan^{-1} \frac{2a\alpha_T + b}{(4ac - b^2)^{\frac{1}{2}}} \text{ by } \theta_T$$

we now rewrite 3.109 as

R(4, p, n) =

$$\frac{\Pi - \dfrac{2a\alpha_A + b}{(4ac - b^2)^{\frac{1}{2}}} \displaystyle\sum_{r=1}^{\frac{n-p-3}{2}} \left(\frac{4ac - b^2}{a(a\alpha_A^2 + b\alpha_A + c)}\right)^r H(r) - 2\theta_A}{\dfrac{2a\alpha_T + b}{(4ac - b^2)^{\frac{1}{2}}} \displaystyle\sum_{r=1}^{\frac{n-p-3}{2}} \left(\frac{4ac - b^2}{a(a\alpha_A^2 + b\alpha_A + c)}\right)^r H(r) + 2\theta_T + \Pi} \qquad (3.110)$$

II. If n-p is even 3.100a becomes

$R(4, p, n) =$

$$\frac{2\sqrt{a} - \dfrac{2a\alpha_A + b}{(a\alpha_A^2 + b\alpha_A + c)^{\frac{1}{2}}} \displaystyle\sum_{r=o}^{\frac{n-p-4}{2}} G(r)\left(\frac{4ac - b^2}{16a(a\alpha_A^2 + b\alpha_A + c)}\right)^r}{2\sqrt{a} + \dfrac{2a\alpha_T + b}{(a\alpha_T^2 + b\alpha_T + c)^{\frac{1}{2}}} \displaystyle\sum_{r=o}^{\frac{n-p-4}{2}} G(r)\left(\frac{4ac - b^2}{16a(a\alpha_T^2 + b\alpha_T + c)}\right)^r} . \qquad (3.111)$$

As an alternative to the preceding, let us develop a class of procedures using prior quasi-densities which are quasi-densities and uniform for the unknown nuisance parameters β, σ, Σ and are proper half-normal densities for α. By proper we mean to say the densities integrate to one. Rather than constructing the ratio of marginal likelihoods and taking the limit as performed earlier in the section we immediately proceed to present the desired prior quasi-densities as follows. Let

$V_o(\alpha, \beta, \sigma, \Sigma)$

$$= \left(\frac{2k}{\Pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{k(\alpha-\alpha_T)^2}{2\sigma^2}\right], \quad 0 < \sigma_{ii} < \infty,\ -\infty < \beta_i < \infty,\ i = 1, \ldots, p;$$
$$0 < \sigma < \infty,\ -\infty < \sigma_{ij} < \infty,$$
$$i < j = 1, \ldots, p,\ -\infty < \alpha \leq \alpha_T$$

$= 0$, otherwise $\qquad (3.112)$

$V_1(\alpha, \beta, \sigma, \Sigma)$

$$= \left(\frac{2k}{\Pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{k(\alpha-\alpha_A)^2}{2\sigma^2}\right], \quad \begin{array}{l} 0<\sigma_{ii}<\infty,\ -\infty<\beta_i<\infty,\ i=1,\ \ldots,\ p; \\ 0<\sigma<\infty,\ -\infty<\sigma_{ij}<\infty, \\ i<j=1,\ \ldots,\ p,\ \alpha_A \leq \alpha < \infty \end{array}$$

$= 0$, otherwise.

In 3.112, k is a constant which is introduced to give V_o and V_1 defined in 3.112 more generality.

The ratio of the marginal likelihoods based on these admissible prior quasi-densities will be written as

$$R(5,\ p,\ n) = \frac{T(2)}{T(1)}. \tag{3.113}$$

The numerator of the ratio given in 3.113 is

$$T(2) = \int_{\alpha_A}^{\infty}\int_{o}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \sigma^{-(n+1)} \exp\left[-\frac{k(\alpha-\alpha_A)^2}{2\sigma^2}\right]$$

$$\exp\left[-\frac{\sum_1^n (y_i-\alpha-x_i'\beta)^2}{2\sigma^2}\right] d\beta d\sigma d\alpha \tag{3.114a}$$

$$= c\int_{\alpha_A}^{\infty}\int_{o}^{\infty} \sigma^{-(n-p+1)} \exp[-\frac{1}{2\sigma^2}\ (k(\alpha-\alpha_A)^2 + u'(I-M)u)]\ d\sigma d\alpha \tag{3.114b}$$

where

$$c = (2\Pi)^{\frac{p}{2}} |XX'|^{-\frac{1}{2}}$$

and is a constant term with respect to the integration, and u' and M are defined as in 3.99 and 3.100a. Since

$u'(I-M)u = \alpha^2 e'(I-M)e - 2\alpha e'(I-M)y + y'(I-M)y$ and, by Lemma 10.2 of Section C of the Appendix,

$$\int_0^\infty x^{-s} \exp[-(\tfrac{\theta}{2})x^{-2}]\, dx = \tfrac{1}{2}(\tfrac{\theta}{2})^{-\frac{s-1}{2}} \Gamma(\tfrac{s-1}{2}), \text{ for } \theta > 0,\ s > 1,$$

we may write

$$T(2) = c' \int_{\alpha_A}^\infty [d\alpha^2 + g(\alpha_A)\alpha + f(\alpha_A)]^{-\frac{n-p}{2}}\, d\alpha \tag{3.115}$$

where

$$c' = (2\Pi)^{\frac{p}{2}} |XX'|^{-\frac{1}{2}} (\tfrac{1}{2})^{-\frac{n-p-2}{2}} \Gamma(\tfrac{n-p}{2}) \tag{3.116}$$

and

$$\begin{aligned} d &= k + e'(I-M)e \\ g(\alpha_A) &= -2k\alpha_A - 2e'(I-M)y \\ f(\alpha_A) &= k\alpha_A^2 + y'(I-M)y. \end{aligned} \tag{3.117}$$

The denominator of the ratio given by 3.113 is

$$T(1) = \int_{-\infty}^{\alpha_T} \int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \sigma^{-(n+1)} \exp\left[-\frac{k(\alpha-\alpha_T)^2}{2\sigma^2}\right]$$

$$\exp\left[-\frac{\sum_1^n (y_i - \alpha - x_i'\beta)^2}{2\sigma^2}\right] d\beta\, d\sigma\, d\alpha \tag{3.118a}$$

and by reductions similar to those used for T(2) we obtain

$$T(1) = c' \int_{-\infty}^{\alpha_T} [d\alpha^2 + g(\alpha_T)\alpha + f(\alpha_T)]^{-\frac{n-p}{2}}\, d\alpha \tag{3.118b}$$

where d, $g(\alpha_T)$, $f(\alpha_T)$, and c' are defined in 3.116 and 3.117.

Noting that T(2) and T(1) have the common factor c',
we find that R(5, p, n) reduces to

$$R(5,\ p,\ n) = \frac{\int_{\alpha_A}^{\infty} \left[dz^2 + g(\alpha_A)z + \delta(\alpha_A)\right]^{-\frac{n-p}{2}} dz}{\int_{-\infty}^{\alpha_T} \left[dz^2 + g(\alpha_T)z + \delta(\alpha_T)\right]^{-\frac{n-p}{2}} dz}. \tag{3.119}$$

If n-p is odd we can use the integral given in 3.106 to write

R(5, p, n) =

$$A(n,\ p) \frac{2\sqrt{d} - I(\alpha_A) \sum_{r=o}^{\frac{n-p-3}{2}} G(r) \left[\frac{4df(\alpha_A) - g^2(\alpha_A)}{16d(d\alpha_A^2 + g(\alpha_A)\alpha_A + f(\alpha_A))}\right]^r}{2\sqrt{d} + I(\alpha_T) \sum_{r=o}^{\frac{n-p-3}{2}} G(r) \left[\frac{4df(\alpha_T) - g^2(\alpha_T)}{16d(d\alpha_T^2 + g(\alpha_T)\alpha_T + f(\alpha_T))}\right]^r} \tag{3.120}$$

where

$$A(n,\ p) = \left[\frac{4df(\alpha_T) - g^2(\alpha_T)}{4df(\alpha_A) - g^2(\alpha_A)}\right]^{\frac{n-p-1}{2}}$$

and

$$I(\alpha^*) = \frac{2d + g(\alpha^*)}{(d\alpha^{*2} + g(\alpha^*)\alpha^* + f(\alpha^*))^{\frac{1}{2}}},\ \alpha^* = \alpha_T,\ \alpha_A.$$

If n-p is even $\frac{n-p-2}{2} + 1$ is an integer and, from 3.103, we have

R(5, p, n) =

$$A(n,\ p)\ \frac{\Pi - 2\theta_A - J(\alpha_A)\ \sum_{r=1}^{\frac{n-p-2}{2}} H(r)\left[\frac{4df(\alpha_A) - g^2(\alpha_A)}{d(d\alpha_A^2 + g(\alpha_A)\alpha_A + f(\alpha_A))}\right]^r}{\Pi + 2\theta_T + J(\alpha_T)\ \sum_{r=1}^{\frac{n-p-2}{2}} H(r)\left[\frac{4df(\alpha_T) - g^2(\alpha_T)}{d(d\alpha_T^2 + g(\alpha_T)\alpha_T + f(\alpha_T))}\right]^r}, \quad (3.121)$$

where, A(n, p) is as defined in 3.120,

$$J(\alpha) = \frac{2d\alpha + g(\alpha)}{(4df(\alpha) - g^2(\alpha))^{\frac{1}{2}}},$$

θ_A = principal value of $\tan^{-1} \dfrac{2d\alpha_A + g(\alpha_A)}{(4df(\alpha_A) - g^2(\alpha_A))^{1/2}}$, and

θ_T = principal value of $\tan^{-1} \dfrac{2d\alpha_T + g(\alpha_T)}{(4df(\alpha_T) - g^2(\alpha_T))^{1/2}}$.

2. Location and scale invariance characteristics

By inspecting R(4, p, n) and R(5, p, n) we see that both are functions of the same quantities, α_T, α_A, y, X, n, p. Accordingly, for the present discussion, let us represent any such test statistic by $L(\alpha_T, \alpha_A, y, X, n, p)$. Recalling that we are testing the hypotheses

$$\begin{aligned} H_o&: \ \alpha \le \alpha_T \\ H_1&: \ \alpha \ge \alpha_A, \end{aligned} \qquad (3.122)$$

where $\alpha_T < \alpha_A$ and the nuisance parameters are unknown. It intuitively seems important that the testing procedure should not be affected by an arbitrary location change in our reference scale. That is, it is desirable that

$$L(\alpha_T + \delta, \alpha_A + \delta, y + \delta e, X, n, p) = L(\alpha_T, \alpha_A, y, X, n, p) \tag{3.123}$$

where δ is any real number. Since the elements of the X matrix of covariates are differences as described in Section B of this chapter, this matrix is unaffected by location changes.

In addition to location invariance, scale invariance is desirable, that is, we should have

$$L(\delta\alpha_T, \delta\alpha_A, \delta y, TX, n, p) = L(\alpha_T, \alpha_A, y, X, n, p) \tag{3.124}$$

where δ is any real number not equal to zero and where T is any nonsingular p x p matrix. Relations 3.123 and 3.124 may be summarized by the consolidated functional relationship

$$L(\theta\alpha_T + \delta, \theta\alpha_A + \delta, \theta y + \delta\theta, TX, n, p) = L(\alpha_T, \alpha_A, y, X, n, p) \tag{3.125}$$

where θ and δ are any real numbers, θ not equal to zero.

We now establish that the functional relationship given in 3.125 holds for $R(4, p, n)$ and $R(5, p, n)$. $R(4, p, n)$ is given in closed form in 3.110 and 3.111 but we find it more convenient to consider the form presented in 3.97. Let us write 3.97 as

$$R(4,\ p,\ n) = \frac{\Phi(\alpha_A,\ y,\ X)}{\Phi(\alpha_T,\ y,\ X)} \tag{3.126}$$

where $\Phi(\alpha_A, y, X)$ and $\Phi(\alpha_T, y, X)$ are the numerator and denominator respectively of the ratio given in 3.97.

We also denote $(y_1, y_2, \ldots, y_n)$ by y' and $(x_1, \ldots x_n)$ by X.

We now compute

$\Phi(\theta\alpha_A + \delta,\ \theta y + \delta e,\ TX) =$

$$\int_{\theta\alpha_A+\delta}^{\infty} \int_{0}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sigma^{-n} \exp\left[- \frac{\sum_{1}^{n} (\theta y_i + \delta - \alpha - x_i' T' \beta)^2}{2\sigma^2} \right] d\beta d\sigma d\alpha. \tag{3.127}$$

Using the transformation

$$\begin{aligned} \beta &\longrightarrow \theta(T')^{-1}\beta \\ \alpha &\longrightarrow \theta\alpha + \delta \\ \sigma &\longrightarrow \theta\sigma \end{aligned} \tag{3.128}$$

straightforward reduction shows that 1.127 becomes

$$\Phi(\theta\alpha_A + \delta,\ \theta y + \delta\theta,\ TX) = e^{-(n-p-2)} |T|^{-1} \Phi(\alpha_A,\ y,\ X). \tag{3.129}$$

By similar techniques we can show

$$\Phi(\theta\alpha_T + \delta,\ \theta y + \delta\theta,\ TX) = \theta^{-(n-p-2)} |T|^{-1} \Phi(\alpha_T,\ y,\ X) \tag{3.130}$$

so that

$$\frac{\Phi(\theta\alpha_A + \delta,\ \theta y + \delta\theta,\ TX)}{\Phi(\theta\alpha_T + \delta,\ \theta y + \delta\theta,\ TX)} = \frac{\Phi(\alpha_A,\ y,\ X)}{\Phi(\alpha_T,\ y,\ X)}$$

so that R(4, p, n) does satisfy the functional relationship given in 3.125.

We now establish that the relation 3.125 holds for the test statistic R(5, p, n) and again it is more convenient to deal with R(5, p, n) as presented in 3.113 rather than with its closed form. Let us write R(5, p, n) as given in 3.113 as

$$R(5, p, n) = \frac{\psi(\alpha_A, y, X)}{\psi(\alpha_T, y, X)} \qquad (3.131)$$

where $\psi(\alpha_A, y, X)$ is T(2) as defined by 3.114a and where $\psi(\alpha_T, y, X)$ is T(1) as defined by 3.118a. We again denote $(y_1, \ldots, y_n)$ by y' and $(x_1, \ldots, x_n)$ by X. Then

$$\psi(\theta\alpha_A + \delta, \theta y + \delta\theta, TX) =$$

$$\int_{\theta\alpha_A+\delta}^{\infty}\int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\sigma^{-(n+1)} \exp\left[-\frac{k(\alpha-\theta\alpha_A-\delta)^2}{2\sigma^2}\right]$$

$$\exp\left[-\frac{\sum_1^n(\theta y_i + \delta - \alpha - x_i'T'\beta)^2}{2\sigma^2}\right] d\beta d\sigma d\alpha. \qquad (3.132)$$

The identical transformation used in 3.128, namely,

$$\beta \longrightarrow \theta(T')^{-1}\beta$$
$$\alpha \longrightarrow \theta\alpha + \delta \qquad (3.133)$$
$$\sigma \longrightarrow \theta\sigma$$

then leads to

$$\psi(\theta\alpha_A + \delta, \theta y + \delta\theta, TX) = \theta^{-(n-p-3)}|T|^{-1}\psi(\alpha_A, y, X). \qquad (3.134)$$

Similarly

$$\psi(\theta\alpha_T + \delta, \theta y + \delta\theta, TX) = \theta^{-(n-p-3)}|T|^{-1}\psi(\alpha_T, y, X)$$

so that

$$\frac{\psi(\theta\alpha_A + \delta,\ \theta y + \delta\theta,\ TX)}{\psi(\theta\alpha_T + \delta,\ \theta y + \delta\theta,\ TX)} = \frac{\psi(\alpha_A,\ y,\ X)}{\psi(\alpha_T,\ y,\ X)}$$

and R(5, p, n) satisfies the equation given in 3.125.

In conclusion it is of interest to relate the procedures which have been developed here with the earlier work in the field by Roseberry (58). For testing the hypothesis as given in 3.93 Roseberry suggested an approach based on uniform weighting of the parameter points in w_o and w_1 which may be implemented as follows. If, in fact $\alpha_T < 0 < \alpha_A$ and if the following prior densities are constructed

$$\begin{aligned} U_o(\alpha, \beta, \sigma, \Sigma) &= \frac{1}{2^{\frac{p(p+1)}{2}}\, c^{\frac{(p+1)(p+2)}{2}}\, (c-1)\alpha_T} \\ &\quad \text{when } 0 < \sigma_{ii} < c,\ -c < \beta_i < c,\ i = 1, \ldots, p; \\ &\quad 0 < \sigma < c;\ -c < \sigma_{ij} < c,\ i < j = 1, \ldots, p; \\ &\quad c\alpha_T < \alpha < \alpha_T \\ &= 0, \text{ otherwise} \end{aligned}$$

$$\begin{aligned} U_1(\alpha, \beta, \sigma, \Sigma) &= \frac{1}{2^{\frac{p(p+1)}{2}}\, c^{\frac{(p+1)(p+2)}{2}}\, (c-1)\alpha_A} \\ &\quad \text{when } 0 < \sigma_{ii} < c,\ -c < \beta_i < c,\ i = 1, \ldots, p; \\ &\quad 0 < \sigma < c,\ -c < \sigma_{ij} < c,\ i < j = 1, \ldots, p; \\ &\quad \alpha_A < \alpha < c\alpha_A \\ &= 0, \text{ otherwise} \end{aligned}$$

then computations similar to those used earlier in this section in our derivation of R(4, p, n) will yield a test statistic T(4, p, n) related to R(4, p, n) by the equation

$$T(4,\ p,\ n) = -\ \frac{\alpha_T}{\alpha_A}\, R(4,\ n,\ p) \tag{3.135}$$

where $R(4, p, n)$ is given by 3.110 and 3.111.

Although Olin (49) has demonstrated empirically that $T(4, 0, n)$ and $T(4, 1, n)$ may have some desirable local properties it is easily seen that $T(4, p, n)$ does not satisfy 3.125. For we have demonstrated that $R(4, p, n)$ satisfies 3.125 so that $T(4, p, n)$ can satisfy 3.125 if and only if

$$\frac{\theta\alpha_T + \delta}{\theta\alpha_A + \delta} = -\ \frac{\alpha_T}{\alpha_A}$$

which is true if and only if $\delta = 0$ or $\alpha_T = \alpha_A$. It may accordingly be concluded that the procedures based on $T(4, p, n)$ will be of restricted rather than general application.

Thus far, in this chapter, we have developed sequential test procedures utilizing concomitant information via weight functions and prior distributions. These procedures, however, are not readily amenable to theoretical study of their properties. In particular there are at present no analytic expressions for the average sample number or operating characteristic curve as in the simple versus simple hypothesis testing problem discussed in Chapter II. Some empirical sampling results on these characteristics have, however, been obtained. These results will be discussed in Chapter VI.

IV. DERIVATION OF SEQUENTIAL t- AND F-TESTS, UTILIZING CONCOMITANT INFORMATION, VIA FIXED-SAMPLE SIZE SUFFICIENCY

A. Introduction

In Chapter III, we investigated the weight function and prior distribution approach for the derivation of sequential tests for composite hypotheses. Another approach will now be examined in this chapter which is to some extent notationally independent of the other chapters.

Suppose we have a sequence of observations z_1, z_2, ..., z_n which are realizations of random variables sampled sequentially from a population having a distribution function indexed by $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$, $p < n$. We can sometimes make inferences about one of the parameters, θ_1 say, in the presence of unknown nuisance parameters θ_2, θ_3, ..., θ_p by transforming z_1, z_2, ..., z_n to a new sequence u_1, u_2, ..., u_m $m < n$ of which the distribution is indexed by some function of θ, say $\gamma(\theta)$. Then, under certain conditions, it may happen that one of the terms of the sequence u_1, ..., u_m, say u_m, contains all of the relevant information (in some sense) about $\gamma(\theta)$ that is contained in the sequence u_1, u_2, ..., u_m so that the joint p.d.f. $f(u_1, \ldots, u_m|\gamma(\theta))$ can be written as $f_1(u_m|\gamma(\theta))f_2(u_1, u_2, \ldots, u_{m-1}, u_m)$ where $f_2(u_1, u_2, \ldots, u_{m-1}, u_m)$ does not depend upon $\gamma(\theta)$. Thus, if H_o and H_1 are null and alternative hypothesis statements in reference to $\gamma(\theta)$ then the likelihood ratio

$$\frac{f(u_1, u_2, \ldots, u_m|\gamma(\theta), H_1)}{f(u_1, u_2, \ldots, u_m|\gamma(\theta), H_o)} \tag{4.1}$$

reduces to

$$\frac{f(u_m|\gamma(\theta), H_1)}{f(u_m|\gamma(\theta), H_o)}. \tag{4.2}$$

Hall, Wijsman, and Ghosh (28), (sometimes hereafter abbreviated HWG) and Cox (20) discuss the conditions under which the factorization of $f(u_1, u_2, \ldots, u_m|\gamma(\theta))$ obtains and also discuss some applications to sequential methodology. In what follows we shall outline the pertinent theory and use this theory to derive sequential tests utilizing concomitant information.

B. Definitions and Theory

We consider the probability model $X_\theta = (\mathcal{X}, \mathcal{A}, P_\theta)$ where $\mathcal{X}$ is a sample space of points, $\mathcal{A}$ is a given σ-field of subsets of $\mathcal{X}$, and P_θ is a probability measure on $\mathcal{A}$ and we denote the class of probability models indexed by $\theta \in \Omega$ as X_Ω.

To understand the ensuing discussion the following definitions are required.

Definition 4.1: A set G of elements is called a group if

(i) there is defined an operation, say group multiplication, which, with any two elements $g_1, g_2 \in G$, associates an element g_3 of G.

Denote by $g_1 g_2 = g_3$.

(ii) $g_1(g_2 g_3) = (g_1 g_2) g_3$ for <u>any</u> g_1, g_2, $g_3 \in G$.

(iii) there exists an element $g_o \in G$, called the identity such that $gg_o = g_o g = g$, and

(iv) to each element $g \in G$, there exists an element $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = g_o$.

<u>Definition 4.2</u>: A class of models X_Ω is invariant under a group G of one-to-one transformations from $\mathcal{X}$ onto itself if each $g \in G$ induces a transformation $\bar{g} \in \bar{G}$ such that $\bar{g}\theta = \theta' \in \Omega$ and $P_\theta(gx \in A) = P_{\bar{g}\theta}(x \in A)$, $A \in \mathcal{A}$, $\theta \in \Omega$. We denote this property by $gX_\Omega = X_\Omega$ (28, p. 578).

<u>Definition 4.3</u>: A function t on $\mathcal{X}$ is invariant under a group G if and only if $t(gx) = t(x)$ for all $x \in \mathcal{X}$ and $g \in G$ (45, p. 215).

<u>Definition 4.4</u>: An orbit generated by a point $x \in \mathcal{X}$ consists of the totality of points gx with $g \in G$ (45, p. 215).

<u>Definition 4.5</u>: If an invariant function t on $\mathcal{X}$ assumes a different value on each orbit then t is called a maximal invariant (28, p. 579).

<u>Definition 4.6</u>: Let the probability model corresponding to any statistic t, which is invariant according to either Definition 4.3 or Definition 4.5, be denoted by

$$T = (L, \mathcal{A}^t, P^t_\gamma) \text{ where } L = [t(x);\ x \in \mathcal{X}],$$

$$\mathcal{A}^t = [A^t;\ t^{-1}A^t \in \mathcal{A}], \text{ and } P^t_\gamma \text{ is such that}$$

$$P^t_\gamma[t(x) \in A^t] = P_\theta[x \in t^{-1}A^t].$$

Here γ: $\theta \longrightarrow \gamma(\theta)$.

As an illustrative example consider a random sample, $x_1, x_2, \ldots, x_n$, from P_θ, which here is taken as the probability measure associated with $N(\mu, \sigma^2)$. We know $t(x_1, \ldots, x_p) = \bar{x}$ is distributed as $N(\mu, \frac{\sigma^2}{n})$ so that $P_\gamma^t[x \epsilon (-3, 3)] = P_\theta[(x_1, \ldots, x_n) \epsilon t^{-1}(-3, 3)]$ where $t^{-1}(-3, 3)$ is a subset of Euclidean n-space.

Definition 4.7: A statistic s on $\mathcal{A}$ is said to be sufficient for X_Ω if for every $A \epsilon \mathcal{A}$ and $s_o \epsilon S = s(\mathcal{A})$ there is a version of the conditional probability $P_\theta[A|s_o] = P_\theta[x \epsilon A|s(x) = s_o]$ which does not depend on θ (28, p. 579).

Let us consider a family of distributions indexed by θ and, also, a group G of transformations on the sample space. Decision procedures then will not be affected by transformations of G if

1) these same transformations leave the family of distributions unchanged and if
2) the decision procedures are based on invariant functions of the sample space.

On this point Lehman (45, p. 216, 220) shows that all invariant functions are functions of the maximal invariant and, in addition, if a statistic t is invariant under G then its distribution depends only on a maximal invariant function, say γ, on Ω under G.

Subject to certain conditions, to be stated later, HWG show that

1) if a statistic s contains all the relevant information about θ and

2) if $\gamma(\theta)$ is the function of θ induced by the maximal invariant function (or equivalently any invariant function) under G of s

then a maximal invariant function of s contains all the relevant information about $\gamma(\theta)$ that is available in any invariant function (see Definition 4.7).

Although the preceding discussion is appropriate for fixed sample size experimentation we now discuss the concepts of sufficiency and invariance in relation to the sequential experiment. In sequential experimentation, the experiment may be terminated at any stage, but the performance at a stage n implies some previous performance at stages 1, 2, 3, ..., n - 1. Following HWG it is here useful to distinguish three types of probability models:

(i) the <u>component</u> or <u>marginal</u> models $X_{n\theta} = (\mathcal{X}_n, \mathcal{A}_n, P_{n\theta})$ for stage n and data $x_n (n = 1, 2, \ldots)$,

(ii) the <u>joint</u> <u>(n-fold)</u> <u>models</u> $X_{(n)\theta} = (\mathcal{X}_{(n)}, \mathcal{A}_{(n)}, P_{(n)\theta})$ for the accumulated data $x_{(n)} = (x_1, \ldots, x_n)$ through stage n, and

(iii) the <u>sequential</u> <u>model</u> $X_\theta = (\mathcal{X}, \mathcal{A}, P_\theta)$ for the whole sequence of data $x = (x_1, x_2, \ldots, x_n, \ldots)$. We make the following definitions:

Definition 4.8: For each n, if a function, s_n, of the first n observations, is sufficient for the class $X_{(n)\Omega}$ of joint models ($\theta \in \Omega$), then $s = (s_1, s_2, \ldots, s_n)$ is called a sufficient sequence for X_Ω (28, p. 583).

Definition 4.9: For each n, suppose t_n is a function of $x_{(n)} = (x_1, x_2, x_3, \ldots, x_n)$. If, for all θ and each n, the conditional distribution of t_{n+1} given $x_{(n)}$ is identical with the conditional distribution of t_{n+1} given t_n, then $t = (t_1, t_2, \ldots,)$ is said to be a transitive sequence for X_Ω (28, p. 583).

With the sequential model we need consider the whole sequence of data which, of course, is not available to the decision maker. Accordingly, it seems reasonable to concern ourselves primarily with the joint n-fold model for accumulated data $x_{(n)} = (x_1, \ldots, x_n)$ and its relationship to the sequential model. With this end in mind HWG defined a sufficient sequence for $X_{(n)}$ (see Definition 4.8) and introduced a desirable property for a sequence called transitivity (Definition 4.9). The basic idea of transitivity is that all information about a statistic t_{n+1} contained in $x_{(n)} = (x_1, \ldots, x_n)$ is carried by $t_n(x_{(n)})$. We now assume that a group G of transformations g on the sample space $\mathcal{X}$, for the sequential probability model (iii) above, has the property that each g induces a transformation $g_{(n)}$ on the n-fold sample space $\mathcal{X}_{(n)}$. Denoting the induced group of transformations by $G_{(n)}$ and the maximal invariant on $\mathcal{X}_{(n)}$ under

$G_{(n)}$ by u_n, HWG judge that the sequence $u = (u_1, u_2, \ldots)$ is the relevant invariant sequence to be used to aid in constructing a sequence of test statistics containing all the available information concerning a particular parameter of interest.

To supplement the above discussion and aid in interpretation of the forthcoming discussion we give the following definitions.

Definition 4.10: Any set $A \in \mathcal{A}$ is an invariant set if $x \in A$ implies $gx \in A$ for every $g \in G$ (28, p. 579).

Definition 4.11: A function v on $\mathcal{X}$ is invariantly sufficient for X_Ω under G if

(i) v is invariant under G

(ii) the conditional probability of any invariant set A given v is parameter free for $\theta \in \Omega$ (28, p. 579, 580).

Since invariant functions are functions of the maximal invariant we may write $v = v_u(u)$ where u is a maximal invariant, U is the sample space of the probability model of the maximal invariant, and v_u is a function on U. HWG state that any invariant set is of the form $[x\colon\ u(x) \in A^u]$ where u is a maximal invariant. We show this by the following arguments. If A is an invariant set then by definition we have that $x \in A \Rightarrow g(x) \in A$ for all $g \in G$. Thus $u(x) \in A^u$ and $u(g(x)) \in A^u$ implying that A has the form $[x\colon u(x) \in A^u]$.

Conversely if $A' = [x: \ u(x) \in A^u]$ then $x \in A' \Rightarrow u(x) \in A^u$ but $u(x) = u(g(x))$ so that $g(x) \in A'$. Therefore $gA' = A'$. Thus since $P^u_{\gamma(\theta)}[u \in A^u | v_u = v_o] = P_\theta[u(x) \in A^u | v(x) = v_o]$ we have that condition (ii) of Definition 4.11 is equivalent to saying v_u is sufficient for U_Γ; $\gamma(\Omega) = \Gamma$. For these reasons we now interpret condition (ii) of Definition 4.11 as stating that v is sufficient for the distributions of (v, t) where t is any invariant function.

We shall also require the following theorem, due to C. Stein, and given, for example, in HWG.

<u>Theorem 4.1</u>: Under certain assumptions (to be considered subsequently), if s is a sufficient statistic for X_Ω and u_s is a maximal invariant function of $s(\mathfrak{X}) = S$ under G_s (the induced group of transformations on S), then $v = u_s(s)$ is invariantly sufficient for X_Ω under G.

We consider the following definition.

<u>Definition 4.12</u>: v-rules are defined to be sequential decision procedures that depend on an invariantly sufficient and transitive sequence $v = (v_1, v_2, \ldots)$. (28, p. 584)

If we replace the original sequence of joint probability models $\{X_{(n)\theta}\}$ with the sequence of probability models of the maximal invariant, $\{U_{n\gamma}\}$, then the sufficiency condensation of each of the components u_i of the sequence $u = (u_1, u_2, \ldots)$ leads to a sequence $v = (v_1, v_2, \ldots)$ which may be called an invariantly sufficient sequence for the sequential model X_Ω

under G. As an example, consider $x_{(n)} = (x_1, \ldots, x_n)$, where $x_i \sim NI(\mu, 1)$, and let G be the group of translations $g(x_{(n)}) = (x_1+c, x_2+c, x_2+c, \ldots, x_n+c)$, $-\infty < c < \infty$. The sets ${}_1u_n = (x_1-x_n, \ldots, x_{n-1}-x_n)$ or ${}_2u_n = (x_1-\bar{x}_n, \ldots, x_n-\bar{x}_n)$ are examples of maximal invariants. A sufficiency condensation of either $({}_1u_1, {}_1u_2, \ldots)$ or $({}_2u_1, {}_2u_2, \ldots)$ is $v = (v_1, v_2, \ldots) = (\bar{x}_1, \bar{x}_2, \ldots)$ where $\bar{x}_i$ is the mean of the components of $x_{(i)}$. If v is transitive then Theorem 4.1 provides justification for using v-rules in sequential testing. HWG give some methods for verifying the transitivity of a sufficient sequence and prove that the sequence $v = (v_1, v_2, \ldots)$ is transitive if the pertinent sequence of sufficient statistics is transitive.

Restricting our attention to v-rules we have v_n as sufficient for the distributions of any invariant function of which v_n is a function. Thus v_n is sufficient for the distributions of $v_{(n)} = (v_1, v_2, \ldots, v_n)$ and the joint density of $v_{(n)}$ factors according to the Fisher-Neyman factorization theorem for sufficient statistics (28, p. 585).

HWG prove that Theorem 4.1 holds under the following conditions which we label as Assumption 4.1.

<u>Assumption 4.1</u>: We are dealing with a multivariate (nonsingular) distribution for which the region of positive density does not vary with θ and for which we can factor the joint density of x as $f_\theta(s(x))h(x)$ so that the transformations

$g \in G$, the sufficient statistic $s(x)$, and the factor $h(x)$ satisfy the following conditions for all x-values except those lying in an invariant set A_o having probability zero:

(i) each g is continuously differentiable and both the Jacobian and $\frac{h(gx)}{h(x)}$ depend only on $s(x)$ and

(ii) s is continuously differentiable with a matrix of partial derivatives of maximal rank.

HWG point out that most normal theory examples satisfy these conditions and that Theorem 4.1 under Assumption 4.1 is a rigorous version of D. R. Cox's (20) fixed sample size theorem published in 1952. This theorem of Cox has been used to develop the sequential t-test of Wald (60), the sequential F-tests (39, 55), sequential multivariate χ^2 and T^2 tests for hypotheses about multivariate means (34), and simultaneous sequential methods in hierarchical classifications (25). For completeness we now state a rigorous version of Cox's 1952 theorem.

Theorem 4.2:

(i) Let $(y_1, y_2, \ldots, y_n) = y$ be random variables whose p.d.f. depends upon unknown parameters $\theta_1, \theta_2, \ldots, \theta_p$, $p < n$. The y_i themselves may be vectors.

(ii) Let $z_1, z_2, \ldots, z_p$ be a jointly sufficient and functionally independent set of estimators for $\theta_1, \theta_2, \ldots, \theta_p$.

(iii) Let the distribution of z_1 involve θ_1 say, but not $\theta_2, \ldots, \theta_p$.

(iv) Let $u_1, u_2, \ldots, u_m$, $m < n$ be functions of y, functionally independent of each other, and of $z_1, z_2, \ldots, z_p$.

(v) If there exists a set H of transformations of $y = (y_1, y_2, \ldots, y_n)$ into $y' = (y_1', \ldots, y_n')$ such that

a) $z_1, u_1, u_2, \ldots, u_m$ are all unchanged by the transformations in H,

b) the transformation of $z_2, \ldots, z_p$ into $z_2', z_3', \ldots, z_p'$ defined by each transformation in H is one-to-one,

c) we can write $hY_\Omega = Y_\Omega$ for all $h \in H$ (see Definition 4.2) and

d) if $Z_2, Z_3, \ldots, Z_p$ and $Z_2', Z_3', \ldots, Z_p'$ are two sets of values of $z_2, z_3, \ldots, z_p$ each having non-zero probability density under at least one of the distributions of y, $\theta \in \Omega$, then there exists a transformation in H such that if

$z_2 = Z_2$, $z_3 = Z_3$, $\ldots$, $z_p = Z_p$ then
$z_2' = Z_2'$, $z_3' = Z_3'$, $\ldots$, $z_p' = Z_p'$.

Then the joint probability function of $z_1, u_1, \ldots, u_m$ factorizes into

$$f(z_1 | \theta_1) h(u_1, u_2, \ldots, u_m, z_1)$$

where $f(z_1 | \theta_1)$ is the p.d.f. of z_1 and $h(u_1, u_2, \ldots, u_m, z_1)$ does not involve θ_1.

C. Reformulation of the Basic Problem

For the probability model 3.23 under consideration thus far, the methods presented in Section B of this chapter for demonstrating the factorization

$$f(u_1, u_2, \ldots, u_m|\gamma(\theta)) = f_1(u_m|\gamma(\theta))f_2(u_1, \ldots, u_{m-1}, u_m) \tag{4.3}$$

result in a p.d.f. $f_1(u_m|\gamma(\theta))$ for which we have been unable to obtain an analytical representation. Thus the expression 4.2 is mathematically intractable in this case.

We shall accordingly reformulate the basic problem as set forth in Section B of Chapter III. The basic design is as before but we shall assume the traditional analysis of covariance for two treatments:

$$z_{ij} = \mu + \alpha_i + \beta w_{ij} + e_{ij} \qquad \begin{array}{l} i = 1, 2 \\ j = 1, 2, \ldots, 2n, \ldots \end{array} \tag{4.4}$$

which, in terms of within-pair differences, gives

$$y_j = \alpha + \beta x_j + e_j \tag{4.5}$$

where $y_j = z_{1j} - z_{2j}$, $\alpha = \alpha_1 - \alpha_2$, $x_j = w_{1j} - w_{2j}$, $e_j = e_{1j} - e_{2j}$ and $E(e_j) = 0$. We assume here that the covariates (w_{ij}) and (x_j) are controlled, the random variables being (z_{ij}), (y_j), (e_{ij}), and (e_j) with the $e_j \sim NI(0, \sigma^2)$.

More generally, we may assume a model with p covariates

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i \tag{4.6}$$

where $i = 1, 2, \ldots$ and the $(x_{i1}), (x_{i2}), \ldots, (x_{ip})$ are covariates and assumed controlled.

The interest lies in developing sequential tests which are pertinent to various hypothesis formulations about α. In the model given in 4.6 α represents the differences between the two treatments when the x's are all zero; that is, when the concomitant information on the two subjects is identical.

D. Sequential Covariance Analysis for One-Sided Hypotheses

For simplicity the following discussion will be limited to the one covariate model 4.5. It is shown later that little modification is required for application to the case of p covariates.

Based on the model

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, 2, \ldots \tag{4.7}$$

where α and β are unknown parameters, x_i, $i = 1, 2, \ldots$, are fixed known constants, and $\epsilon_i \sim NI(0, \sigma^2)$ with σ^2 unknown, we consider the sequential testing of the hypothesis formulation

$$\begin{aligned} H_o&: \frac{\alpha-\alpha_o}{\sigma} = \gamma_o \\ H_1&: \frac{\alpha-\alpha_1}{\sigma} = \gamma_1 \end{aligned} \tag{4.8}$$

where α_o, γ_o, γ_1 are completely specified and where σ is not specified. In most practical cases α_o will be taken as zero and since generality is not lost by a location translation on the y-axis, we will, for convenience, consider

$$\begin{aligned} H_o\colon\ \frac{\alpha}{\sigma} &= \gamma_o \\ H_1\colon\ \frac{\alpha}{\sigma} &= \gamma_1 \end{aligned} \tag{4.9}$$

as the hypothesis of interest.

From 4.7 we see that the joint distribution of the sequence of independent random variables $y_1, y_2, \ldots, y_n$ from a distribution $N(\alpha + \beta x_i, \sigma^2)$ is completely specified by the parameters α, β, and σ^2 and, since β and σ^2 are assumed unknown, we are in a composite hypothesis testing situation. The application of the methods discussed in Section B will now be demonstrated.

We transform the observations

$$\begin{pmatrix} y_1 \\ 1 \\ x_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ 1 \\ x_2 \end{pmatrix}, \ldots, \begin{pmatrix} y_n \\ 1 \\ x_n \end{pmatrix} \tag{4.10}$$

to

$$t_3,\ t_4,\ \ldots,\ t_n, \tag{4.11}$$

where

$$t_i = \left.\frac{\hat{\alpha}}{(\widehat{\mathrm{Var}}\ (\hat{\alpha}))^{\frac{1}{2}}}\right|_i \qquad i = 3, 4, 5, \ldots \tag{4.12}$$

and

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\sigma}^2 = \frac{1}{1-2}\left[S_{yy} - \frac{S_{xy}^{\;2}}{S_{xx}}\right]$$

$$S_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) \tag{4.13}$$

$$S_{xx} = \Sigma(x-x)^2$$

$$S_{yy} = \Sigma(x-y)^2$$

$$\widehat{\mathrm{Var}(\hat{\alpha})} = \hat{\sigma}^2 q_i^2$$

$$q_i^2 = \frac{1}{i} + \frac{\bar{x}^2}{S_{xx}}$$

where Σ denotes summation from 1 to i.

Each t_i as defined in 4.12 and 4.13 has a noncentral Student t-distribution with noncentrality parameter

$$\left.\frac{\alpha}{(\mathrm{Var}(\hat{\alpha}))^{\frac{1}{2}}}\right|_i = \frac{\alpha}{\sigma q_i}. \tag{4.14}$$

Thus at stage i the distribution of t_i does not involve the nuisance parameters β and σ except the latter in the ratio $\frac{\alpha}{\sigma}$. We now establish via Theorem 4.2 that the joint p.d.f. of $t_3, t_4, \ldots, t_n$ factorizes into

$$f_1(t_n|\tfrac{\alpha}{\sigma q_i})f_2(t_3, t_4, \ldots, t_{n-1}, t_n) \tag{4.15a}$$

where $f_2(t_3, t_4, \ldots, t_{n-1}, t_n)$ does not involve $\frac{\alpha}{\sigma}$. To this end, let us examine in turn the conditions of Theorem 4.2.

Condition i: $\begin{pmatrix} y_1 \\ 1 \\ x_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ 1 \\ x_2 \end{pmatrix}, \ldots, \begin{pmatrix} y_n \\ 1 \\ x_n \end{pmatrix}$ are realizations of random variables whose probability distribution depends upon the unknown parameters (α, β, σ). For our problem we transform the parameter space to the one of interest, $(\frac{\alpha}{\sigma}, \beta, \sigma) = (\theta_1, \theta_2, \theta_3)$.

Condition ii: $(z_1, z_2, z_3) = (t_n, \hat{\beta}, (\hat{\sigma}^2)^{\frac{1}{2}})$ is the functionally independent jointly sufficient set of estimators for $(\frac{\alpha}{\sigma}, \beta, \sigma)$. The functional independence may be verified by considering $t_n = t(y_1, \ldots, y_n, x_1, \ldots, x_n)$, $\hat{\beta} = \beta(y_1, \ldots, x_n)$, and $\hat{\sigma} = \sigma(y_1, \ldots, x_n)$ and by showing that the Jacobian

$$\frac{\partial(t_n, \hat{\beta}, \hat{\sigma})}{\partial(u_1, u_2, u_3)} \neq 0 \qquad (4.15b)$$

for the ${}_{2n}C_3$ possible combinations of distinct variables (u_1, u_2, u_3), where u_1, u_2, u_3 can assume any of the 2n variables $y_1, y_2, \ldots, y_n, x_1, \ldots, x_n$. Because of the stochastic nature ofthe Jacobian 4.15b we should say

$$\frac{\partial(t_n, \hat{\beta}, \hat{\sigma})}{\partial(u_1, u_2, u_3)} \neq 0$$

with probability one. With regard to joint sufficiency we know that $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$ where $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ are defined by 4.13 is a vector of jointly sufficient statistics for $(\alpha, \beta, \sigma^2)$. Now if T is sufficient for θ then any one-to-one function of

T, say $\Phi(T)$, is sufficient for any one-to-one function of θ, say $\psi(\theta)$. Since $\sigma > 0$ it follows that

$$\left(\frac{\hat{\alpha}}{(\hat{\sigma}^2)^{\frac{1}{2}}},\ \hat{\beta},\ (\hat{\sigma}^2)^{\frac{1}{2}}\right)$$

is sufficient for

$$\left(\frac{\alpha}{\sigma},\ \beta,\ \sigma\right).$$

Condition iii: That the distribution of t_n involves (α/σ) and not β or σ has been previously noted.

Condition iv: The t_3, t_4, t_5, ..., t_{n-1} are functions of the $(y_i, 1, x_i)'$, $i = 1, 2, \ldots, n-1$ and are functionally independent of each other and of t_n, $\hat{\beta}$, $(\hat{\sigma}^2)^{1/2}$. This can be demonstrated by the same technique as indicated in Condition ii.

Condition v: Let

$$H = \left[\begin{pmatrix} y_i \\ 1 \\ x_i \end{pmatrix} = \begin{pmatrix} k_1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & k_2 \end{pmatrix} \begin{pmatrix} y_i \\ 1 \\ x_i \end{pmatrix},\quad \begin{array}{l} k_1 > 0,\ k_2 \neq 0 \text{ and} \\ i = 1, 2, \ldots, n \end{array}\right].$$

Then

$$\hat{\beta}' = \frac{S_{x'y'}}{S_{x'x'}} = \frac{k_1 k_2}{k_2^2}\frac{S_{xy}}{S_{xx}} = \frac{k_1}{k_2}\hat{\beta}$$

$$(\hat{\sigma}^2)' = \frac{1}{n-2}\quad S_{y'y'} - \frac{S_{x'y'}^{\ 2}}{S_{x'x'}} = k_1^2\hat{\sigma}^2$$

$$\hat{\alpha}' = \bar{y}' - \hat{\beta}'\bar{x}' = k_1(\bar{y} - \beta\bar{x}) = k_1\widetilde{\alpha}$$

$$(q_n^2)' = \frac{1}{n} + \frac{\bar{x}^{2'}}{S_{x'x'}} = \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} = q_n^2$$

We now examine the four subconditions of condition v.

(a) We have

$$t_i' = \left.\frac{\hat{\alpha}'}{(\widehat{\mathrm{Var}(\hat{\alpha})'})^{\frac{1}{2}}}\right|_i = \left.\frac{k_1\hat{\alpha}}{(k_1^2\ \widehat{\mathrm{Var}(\hat{\alpha})})^{\frac{1}{2}}}\right|_i = t_i,$$

where $(z_1, u_1, u_2, u_3, \ldots, u_m) = (t_n, t_3, \ldots, t_{n-1})$.

(b) $(\hat{\beta}', ((\hat{\sigma}^2)')^{\frac{1}{2}}) = (\frac{k_1}{k_2}\hat{\beta}, k_1(\hat{\sigma}^2)^{\frac{1}{2}})$

This is clearly one-to-one as required.

(c) The probability model Y_Ω remains unchanged for every $h \in H$. We show this by comparing

$P_\theta[h(y) \in (c_1, c_2)]$ and $P_{\bar{h}\theta}[y \in (c_1, c_2)]$

where

$\bar{h}\theta = (k_1\alpha, \frac{k_1}{k_2}\beta, k_1\sigma^2)$.

Since $h(y) = k_1 y$ we have that

$$P_\theta[h(y) \in (c_1, c_2)] = \int_{\frac{c_1}{k_1}}^{\frac{c_2}{k_1}} (2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\frac{(y-\alpha-\beta x)^2}{2\sigma^2}]\ dy.$$

By definition we have

$$P_{\bar{h}\theta}[y \in (c_1, c_2)] = \int_{c_1}^{c_2} (2\pi k_1^2\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(y-k_1\alpha-k_2\beta x)^2}{2\sigma^2}\right] dy$$

and writing $y = k_1 z$, it follows that

$$P_{\overline{h}\theta}[y \varepsilon (c_1, c_2)] = \int_{\frac{c_1}{k_1}}^{\frac{c_2}{k_1}} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(z-\alpha-k_1\beta x)^2}{2\sigma^2}\right] dz$$

$$= P_\theta[h(y) \varepsilon (c_1, c_2)].$$

(d) Let $(\hat{\beta}^{(0)}, \hat{\sigma}^{(0)})$ and $(\hat{\beta}^{(2)}, \hat{\sigma}^{(2)})$ be two sets of values of $(\hat{\beta}, \hat{\sigma})$. Condition v.d essentially requires that we notice that real numbers k_1 and k_2 exist such that $\hat{\sigma}^{(2)} = k_1\hat{\sigma}^{(0)}$ and $k_2\hat{\beta}^{(2)} = k_1\hat{\beta}^{(0)}$. For then if $(\hat{\beta}, \hat{\sigma}) = (\hat{\beta}^{(0)}, \hat{\sigma}^{(0)})$ we have

$$(\hat{\beta}', \hat{\sigma}') = (\frac{k_1}{k_2}\hat{\beta}, k_1\hat{\sigma})$$

$$= (\frac{k_1}{k_2}\hat{\beta}^{(0)}, k_1\hat{\sigma}^{(0)})$$

$$= (\hat{\beta}^{(2)}, \hat{\sigma}^{(2)}).$$

To have k_1 and k_2 well defined we must have $\hat{\sigma}^{(0)}$ and $\hat{\beta}^{(2)} \neq 0$ which, as pointed out before, happens with probability one.

It is now possible to write

$$f(t_3, \ldots, t_n) = f_1(t_n|\frac{\alpha}{\sigma q_i})f_2(t_3, \ldots, t_n) \qquad (4.16)$$

where $f_2(t_2, \ldots, t_n)$ does not involve $(\frac{\alpha}{\sigma})$.

The likelihood ratio 4.1 for this application can then be written as

$$R = \frac{f_1(t_n|\frac{\gamma_1}{q_i})}{f_1(t_n|\frac{\gamma_o}{q_i})} \tag{4.17}$$

where γ_o and γ_1 are defined as in 4.9.

We shall now give an alternative demonstration of the factorization shown in 4.16 by showing that t_n is a maximal invariant of the sufficient statistics under a certain group of linear transformations implying that it is sufficient for any invariant statistics, the distributions of these invariant statistics being indexed by some parameter θ. In particular, we wish to show that t_n is sufficient for the distributions of the scale invariant statistics $(t_3, t_4, \ldots, t_n)$.

For the model under consideration, 4.7, the sufficient statistics $\hat{\alpha}$, $\hat{\beta}$, and $(\hat{\sigma}^2)^{1/2}$ are defined in 4.13. We define a group of transformations

$$G = \left[\begin{pmatrix} y_i \\ 1 \\ x_i \end{pmatrix}' = \begin{pmatrix} k_1 y_i \\ 1 \\ k_2 x_i \end{pmatrix} : \begin{matrix} k_1 \in (0, \infty),\ k_2 \in (0, \infty) \\ i = 1, 2, \ldots, n \end{matrix} \right] \tag{4.18}$$

from the sample space onto itself. The class of induced transformations on the space of the probability model induced by the sufficient statistic (see Definitions 4.6 and 4.7) is

$$G_s = \left[(\hat{\alpha}', \hat{\beta}', (\hat{\sigma}^2)^{\frac{1}{2}}) = (k_1\hat{\alpha}, \frac{k_1}{k_2}\hat{\beta}, k_1(\hat{\sigma}^2)^{\frac{1}{2}}) : (k_1, k_2) \in (0, \infty) \times (0, \infty) \right] \tag{4.19}$$

and the class of induced transformations on the parameter space,

$$\Omega = [(\alpha, \beta, \sigma): \quad (\alpha, \beta, \sigma) \in (-\infty, \infty) \times (-\infty, \infty) \times (0, \infty)] \quad (4.20)$$

is

$$\overline{G} = [(\alpha', \beta', \sigma') = (k_1\alpha, \frac{k_1}{k_2}\beta, k_1\sigma): (k_1, k_2) \in (0, \infty) \times (0, \infty)]. \quad (4.21)$$

A maximal invariant function of the probability space of the sufficient statistics under 4.18 and based on all the data through stage n is

$$t_n = \frac{\hat{\alpha}}{(\widehat{\mathrm{Var}}(\hat{\alpha}))^{\frac{1}{2}}} \quad (4.22)$$

where

$$\hat{\alpha} = \overline{y} - \beta\,\overline{x}$$

$$\widehat{\mathrm{Var}}(\hat{\alpha}) = (S_{yy} - \frac{S_{xy}^{\;2}}{S_{xx}})(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}).$$

To show this we need to demonstrate that t_n is constant on an orbit (Definitions 4.3 and 4.4) and assumes different values on distinct orbits (Definition 4.5). An orbit under the group G given in 4.18 on the space of sufficient statistics S is

$$0 = [(k_1\hat{\alpha}, (\frac{k_1}{k_2})\hat{\beta}, k_1\hat{\sigma}): \text{ for a particular } (\hat{\alpha}, \hat{\beta}, \hat{\sigma}) \in S \text{ and all } (k_1, k_2) \in (0, \infty) \times (0, \infty)]$$

where for convenience we are denoting $(\hat{\sigma}^2)^{1/2}$ by $\hat{\sigma}$. It follows that t_n is constant on an orbit since

$$t_n' = \frac{\hat{\alpha}'}{(\widehat{\mathrm{Var}}(\hat{\alpha})')^{\frac{1}{2}}} = \frac{k_1\hat{\alpha}}{(k_1^2\ \mathrm{Var}(\hat{\alpha}))^{\frac{1}{2}}} = \frac{\hat{\alpha}}{(\mathrm{Var}(\hat{\alpha}))^{\frac{1}{2}}} = t_n.$$

If 0_1 and 0_2 are disjoint orbits, i.e. $0_1 \cap 0_2 = \emptyset$, then since we know that t_n is constant on an orbit we must show that $t_n^{(1)} \neq t_n^{(2)}$ where $t_n^{(1)}$ and $t_n^{(2)}$ denote respectively the constant values that t_n takes on 0_1 and 0_2. Assuming that $t_n^{(1)} = t_n^{(2)}$ then

$$\frac{\hat{\alpha}_1}{\hat{\sigma}_1 q_n} = \frac{\hat{\alpha}_2}{\hat{\sigma}_2 q_n}$$

for all $(\hat{\alpha}_1, \hat{\beta}_1, \hat{\sigma}_1) \in 0_1$ and all $(\hat{\alpha}_2, \hat{\beta}_2, \hat{\sigma}_2) \in 0_2$. That is,

$$\hat{\alpha}_1 = c_1 q_n \hat{\sigma}_1$$

$$\hat{\alpha}_2 = c_1 q_n \hat{\sigma}_2$$

where c_1 is some real number. Now

$$(\hat{\alpha}_1, \hat{\beta}_1, \hat{\sigma}_1) = (c_1 q_n \hat{\sigma}_1, \hat{\beta}_1, \hat{\sigma}_1)$$

and there exists a $(k_1, k_2) \in (0, \infty) \times (0, \infty)$ such that

$$\begin{aligned}(k_1\hat{\alpha}_2, \frac{k_1}{k_2}\hat{\beta}_2, k_1\hat{\sigma}_2) &= (k_1 c_1 q_n \hat{\sigma}_2, \hat{\beta}_1, k_1\hat{\sigma}_2)\\ &= (c_1 q_n \hat{\sigma}_1, \hat{\beta}_1, \hat{\sigma}_1)\\ &= (\hat{\alpha}_1, \hat{\beta}_1, \hat{\sigma}_1)\end{aligned}$$

which contradicts $0_1 \cap 0_2 = \emptyset$. This completes the demonstration that t_n is a maximal invariant.

We next require the following definitions.

Definition 4.13: A function f is said to be equivalent to an invariant function if there exists an invariant function h such that f(x) = h(x) for all x except possibly on a P-null set N (45, p. 225).

Definition 4.14: A function f is said to be almost invariant with respect to a group G of transformations if $f(gx) = f(x)$ for all $x \in \mathcal{X} - N_g$, $g \in G$ and N_g is the exceptional null set permitted to depend on g (45, p. 225).

HWG prove that Theorem 4.1 holds under the following assumption.

Assumption 4.2: Every almost invariant function on the sample space of the sufficient statistic S is equivalent to an invariant function.

Since we are assuming that the underlying distribution as normal, every almost invariant function of $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ is known to be equivalent to an invariant function (28, pp. 581, 604; 45, p. 225). Thus Theorem 4.1 leads to the conclusion that $(t_3, t_4, \ldots)$ is an invariantly sufficient sequence and that t_n is sufficient for the distributions of any invariant function of which it is a function and, in particular, therefore, for the distributions of $(t_3, \ldots, t_n)$.

By the same arguments as those used to show that t_n is a maximal invariant under G_s (see 4.19) in the sample space of the sufficient statistic $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ it can be shown that $\frac{\alpha}{\sigma}$ is the maximal invariant of the parameter space under $\overline{G}$, the group of transformations on Ω induced by G on $\mathcal{X}$. Invariance, orbits, and maximal invariants on the parameter space Ω are defined as in Definitions 4.3, 4.4 and 4.5 where $\mathcal{X}$ is replaced by Ω and the $x \in \mathcal{X}$ is replaced by $\theta \in \Omega$.

Theorem 4.3: If $(\mathcal{X}, A, P_\theta)$ is the probability model and if the statistic T(x) is invariant under a group of transformations G and if $\psi(\theta)$ is a maximal invariant under the induced group $\overline{G}$ then the distribution of T(x) depends only on $\psi(\theta)$ (Lehman (45, p. 220)).

We have demonstrated that $(t_3, t_4, \ldots, t_n)$ is an invariantly sufficient sequence and that t_n is sufficient for the distributions of any invariant function of which it is a function. In particular, t_n is sufficient for the invariant statistics $(t_3, \ldots, t_{n-1})$. In addition, by Theorem 4.3 we see that the distributions of $(t_3, t_4, \ldots, t_n)$ are indexed by $\frac{\alpha}{\sigma}$ and thus the factorization given in 4.16 obtains.

E. Practical Implementation

The likelihood ratio given in 4.17 is a ratio of noncentral Student t-distributions. Existing tables of the noncentral Student t-distributions are not well suited for calculating the ratio 4.17 (see Resnikoff and Lieberman (56a)). However, 4.17 can be written in terms of certain confluent hypergeometric functions for which there are fairly extensive tables (63, 67).

Let W denote a random variable which is normally distributed with mean γ and variance 1. Let V denote a random variable that is distributed as chi-square with r degrees of freedom. If W and V are stochastically independent, the new random variable

$$T = \frac{W}{(V/r)^{\frac{1}{2}}} \tag{4.23}$$

is called the noncentral t-distribution with noncentrality parameter γ and degrees of freedom r.

$$f(t|\gamma) = \left[e^{-\frac{\gamma^2}{2}} \frac{1}{(\pi)^{\frac{1}{2}} r} \frac{\Gamma(r+1)}{\Gamma(\frac{r}{2})} \frac{1}{(1+\frac{t^2}{r})^{\frac{r+1}{2}}} \right] \sum_{i=0}^{\infty} \frac{\Gamma(\frac{r+1+i}{2})}{\Gamma(\frac{r+1}{2})} \frac{\left(\frac{\sqrt{2}\ t\gamma}{r+t^2}\right)^i}{i!} \tag{4.24}$$

$$= f(t|\gamma = 0)\ e^{-\frac{\gamma^2}{2}} \sum_{i=0}^{\infty} \frac{\left(\frac{\sqrt{2}\ t\gamma}{r+t^2}\right)^i}{i!} \frac{\Gamma(\frac{r+1}{2} + \frac{i}{2})}{\Gamma(\frac{r+1}{2})}$$

when $t \in (-\infty, \infty)$.

We now express the summation given in 4.24 in terms of certain confluent hypergeometric functions.

$$\sum_{i=0}^{\infty} \frac{x^i}{i} \frac{\Gamma(\frac{r+1}{2} + \frac{i}{2})}{\Gamma(\frac{r+1}{2})}$$

$$= \sum_{i=0}^{\infty} \frac{x^{2j}}{(2j)!} \frac{\Gamma(\frac{r+1}{2} + j)}{\Gamma(\frac{r+1}{2})} + \sum_{j=0}^{\infty} \frac{x^{2j+1}}{(2j+1)!} \frac{\Gamma(\frac{r+1}{2} + \frac{2j+1}{2})}{\Gamma(\frac{r+1}{2})} \tag{4.25}$$

$$= \sum_{j=0}^{\infty} \frac{x^{2j}}{j!} \frac{j!}{(2j)!} \frac{\Gamma(\frac{r+1}{2} + j)}{\Gamma(\frac{r+1}{2})} + x \sum_{j=0}^{\infty} \frac{x^{2j}}{j!} \frac{j!}{(2j+1)!} \frac{\Gamma(\frac{r+2}{2} + j)\Gamma(\frac{r+2}{2})}{\Gamma(\frac{r+2}{2})} \tag{4.26}$$

$$= \sum_{j=0}^{\infty} \frac{x^{2j}}{j!} \frac{\Gamma(\frac{1}{2})}{2^{2j}\Gamma(j+\frac{1}{2})} \frac{\Gamma(\frac{r+1}{2} + j)}{\Gamma(\frac{r+1}{2})}$$

$$+ x \sum_{j=0}^{\infty} \frac{(x^2)^j}{j!} \frac{\Gamma(\frac{3}{2})}{2^{2j}\Gamma(\frac{3}{2} + j)} \frac{\Gamma(\frac{r+2}{2} + j)}{\Gamma(\frac{r+2}{2})} \frac{\Gamma(\frac{r+2}{2})}{\Gamma(\frac{r+1}{2})} \tag{4.27}$$

since

$$\begin{aligned} \frac{j!}{(2j)!} &= \frac{j!}{2j(2j - 1)(2j - 2)\ldots(2j - 2j + 1)} \\ &= \frac{\Gamma(\frac{1}{2})}{2^{2j}(j - \frac{1}{2})(j - \frac{3}{2})\ldots\frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\Gamma(\frac{1}{2})} \\ &= \frac{\Gamma(\frac{1}{2})}{2^{2j}\,\Gamma(j + \frac{1}{2})} \end{aligned} \tag{4.28}$$

and

$$\begin{aligned} \frac{j!}{(2j + 1)!} &= \frac{j!}{(2j + 1)(2j!)} \\ &= \frac{\Gamma(\frac{1}{2})}{2^{2j}\Gamma(j + \frac{1}{2})(2j + 1)} \\ &= \frac{\Gamma(\frac{3}{2})}{2^{2j}\Gamma(j + \frac{3}{2})}. \end{aligned} \tag{4.29}$$

From the definition of the confluent hypergeometric function $F(p;q;u)$ as given in Section C of the Appendix we write

$$\sum_{i=0}^{\infty} \frac{x^i\,\Gamma(\frac{r+1+i}{2})}{i!\,\Gamma(\frac{r+1}{2})}$$

$$= F(\tfrac{r+1}{2};\ \tfrac{1}{2};\ \tfrac{x^2}{4}) + \frac{\Gamma(\frac{r+2}{2})}{\Gamma(\frac{r+1}{2})}\, F(\tfrac{r+2}{2};\ \tfrac{3}{2};\ \tfrac{x^2}{4}) \qquad (4.30)$$

so that

$$g(t|\gamma) = g(t|\gamma = 0)\ \exp\ (-\tfrac{\gamma^2}{2}).$$

$$\left[F(\tfrac{r+1}{2};\ \tfrac{1}{2};\ \tfrac{\gamma^2}{2}\,\tfrac{t^2}{r+t^2}) + \frac{\sqrt{2}\gamma t}{\sqrt{r+t^2}}\,\frac{\Gamma(\frac{r+2}{2})}{\Gamma(\frac{r+1}{2})}\,F(\tfrac{r+2}{2};\ \tfrac{3}{2};\ \frac{\gamma^2 t^2}{2(r+t^2)})\right] \qquad (4.31)$$

Thus to test the hypothesis

$$H_o:\ \frac{\alpha}{\sigma} = \gamma_o$$

$$H_1:\ \frac{\alpha}{\sigma} = \gamma_1$$

we use the likelihood ratio

$$L(1) = \frac{g(t_n|\frac{\gamma_1}{q_n})}{g(t_n|\frac{\gamma_o}{q_n})}$$

$$= \frac{T(\gamma_1)}{T(\gamma_o)} \qquad (4.32)$$

where

$$T(\gamma) = \exp\left[-\frac{\gamma^2}{2q_n^2}\right]\left[F\left(\frac{n-1}{2};\ \frac{1}{2};\ \frac{\gamma^2 t_n^2}{2q_n^2(n-2+t_n^2)}\right)\right.$$

$$\left. + \frac{\sqrt{2}\ \gamma t_n}{q_n\sqrt{n-2+t_n^2}}\ \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\ F\left(\frac{n}{2};\ \frac{3}{2};\ \frac{\gamma^2 t_n^2}{2q_n^2(n-2+t_n^2)}\right)\right] \tag{4.33}$$

with $\gamma = \gamma_o, \gamma_1$.

F. Discussion

Although the arguments considered in this chapter have been developed in the one covariate case, the extension to p covariates is easily accomplished. The appropriate statistic with respect to hypothesis 4.9 is

$$t = t_n(p) = \frac{\overline{y} - \sum_{i=1}^{p} \hat{\beta}_j \overline{x}_j}{(\widehat{\mathrm{Var}(y} - \sum_1^p \hat{\beta}_j x_j))^{\frac{1}{2}}} \tag{4.34}$$

where $\overline{y} = \dfrac{\sum_1^n y_i}{n}$, $\overline{x}_j = \dfrac{\sum_1^n x_{ij}}{n_j}$, $\hat{\beta} = \dfrac{S_{x_j y_j}}{S_{x_j x_j}}$, $S_{x_j y} = \sum_1^n (x_{ij} - \overline{x}_j)(y_{ij})$,

and $S_{x_j x_j} = \sum_1^n (x_{ij} - \overline{x}_j)^2$.

If we let

$$q_n^2 = \frac{\mathrm{Var}(\overline{y} - \sum_{j=1}^{p} \hat{\beta}_j x_j)}{\sigma^2}$$

then we write the likelihood ratio

$$L(p) = \frac{T_p(\gamma_1)}{T_p(\gamma_o)} \tag{4.35}$$

where

$$T_p(\gamma) = \exp\left[-\frac{\gamma^2}{2q_n^2}\right]\left[F\left(\frac{n-p}{2};\ \frac{1}{2};\ \frac{\gamma^2 t^2}{n-p-1+t^2}\right)\right.$$
$$\left. + \frac{\sqrt{2}\ t\gamma}{q_n(n-p-1+t^2)^{\frac{1}{2}}}\ \frac{\Gamma(\frac{n-p-1}{2})}{\Gamma(\frac{n-p}{2})}\ F\left(\frac{n-p+1}{2};\ \frac{3}{2};\ \frac{\gamma^2 t^2}{2q_n^2(n-p-1+t^2)}\right)\right] \tag{4.36}$$

with $\gamma = \gamma_1, \gamma_o$.

Similar arguments obtain in deriving a sequential test for the two-sided hypothesis formulation

$$\begin{aligned} H_o:&\quad \left|\frac{\alpha}{\sigma}\right| = \gamma_o \\ H_1:&\quad \left|\frac{\alpha}{\sigma}\right| = \gamma_1. \end{aligned} \tag{4.37}$$

The appropriate test statistic is t^2 where t is given by 4.34. The likelihood ratio reduces by sufficiency arguments to the ratio of noncentral F-distributions with noncentrality parameters γ_1^2/q_n^2 and γ_o^2/q_n^2 where q_n^2 is as defined previously. When expressed in terms of confluent hypergeometric functions the ratio becomes

$$\frac{\exp\left[-\frac{\gamma_1^2}{2q_n^2}\right] F\left(\frac{n-p}{2};\ \frac{1}{2};\ \frac{\gamma_1^2 t^2}{2q_n^2(n-p-1+t^2)}\right)}{\exp\left[-\frac{\gamma_o^2}{2q_n^2}\right] F\left(\frac{n-p}{2};\ \frac{1}{2};\ \frac{\gamma_o^2 t^2}{2(n-p-1+t^2)q_n^2}\right)}. \tag{4.38}$$

The test statistics given by 4.33, 4.36, and 4.38 can be used with the operating procedure given in Section B of

Chapter II. However, only Properties 2.1, 2.2, and 2.3 of Section C, Chapter II are known to hold. Since the test statistics given by 4.33, 4.36, and 4.38 are sequential t- or t^2-statistic type of procedures, David and Kruskal's (22) result proves termination with probability one. The remark following Properties 2.2 and 2.3 in Section C of Chapter II allows us to use Wald's boundaries with the procedures achieving approximately the specified Type I and Type II probabilities of error. No results similar to Properties 2.4, 2.5, or 2.6 have been proved for any kind of sequential t-test however.

This completes our consideration of one-sample sequential t-tests utilizing concomitant information. In the next chapter we consider the possibilities of a two-sample analysis.

V. THE SEQUENTIAL TWO-SAMPLE t-TEST UTILIZING CONCOMITANT INFORMATION

A. Derivation of Hajnal's Two-Sample Sequential t^2- Test via Prior Distributions

In Chapters III and IV we were concerned with developing sequential tests of hypotheses for a model based on within-pair differences of observations. Thus, although we essentially began with a two-sample situation, we considered paired observations and constructed what are, in effect, one-sample sequential tests. In this chapter we investigate the possibilities of a two-sample analysis which does not require the pairing restriction.

Such a procedure has, in fact, been developed by Hajnal (26) and is termed the two-sample sequential t^2-test. For this procedure observations are taken from two normal populations with unknown means, α_1 and α_2, and common unknown variance σ^2. Based on these assumptions Hajnal presented a procedure for sequentially testing

$$\begin{aligned} H_o&: \ \alpha_1 = \alpha_2 \\ H_1&: \ (\alpha_2 - \alpha_1)^2 = \gamma^2\sigma^2 \end{aligned} \tag{5.1}$$

where γ is specified but σ is assumed unknown. He used Cox's Theorem (20) in showing that the usual fixed-sample two-sample Student t^2-statistic, say

$$t^2 = \frac{(\bar{y}_1 - \bar{y}_2)^2 u}{s^2} \tag{5.2}$$

where $\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$ is the sample mean of n_i observations taken on treatment i and $i = 1, 2$, $u = \frac{1}{n_1} + \frac{1}{n_2}$, and

$$s^2 = \frac{\sum_1^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_1^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2},$$ is sufficient for the

parameter of interest,

$$\frac{(\alpha_2 - \alpha_1)^2}{\sigma^2} = \gamma^2. \tag{5.3}$$

Thus we may factor the joint densities of the sequence of t^2-statistics $t_1^2, t_2^2, \ldots, t_{n_1+n_2-2}^2$ under each of the hypotheses H_o and H_1 as

$$g(t_n; H_i)g_1, \; i = 1, 2 \tag{5.4}$$

where g_1 does not involve γ^2 and where $n = n_1 + n_2 - 2$. Consequently, the appropriate likelihood ratio may be written as

$$\frac{g(t_n^2 : H_1)}{g(t_n^2 : H_o)} \tag{5.5}$$

which is recognized as the ratio of the p.d.f. of a noncentral F-distribution to a p.d.f. of a central F-distribution.

We shall now show an alternative derivation of this ratio using the methods presented in Chapter III. This result itself is of some theoretical interest in relating the two approaches. Further, however, we proceed to introduce concomitant information and develop sequential tests appropriate to this more general situation.

Suppose that, at each stage of sampling, we do not restrict ourselves to only one observation from each normal population. For example, at each stage we may sample one observation from the first population and three from the second population so that at the n^{th} stage we have accumulated n and 3n observations from the first and second populations respectively. The effect of such grouping is discussed in Wald (69). Wald's general conclusions are that: 1) the realized values of the Type I and Type II errors cannot exceed the intended values except by a small amount (which, he states, may be ignored for all practical puposes), 2) the number of observations required to decision will be increased from that of sampling single observations at each stage, and 3) that the realized values of the Type I and Type II errors may be substantially smaller than intended; this may be regarded as compensation for the increase in the number of observations.

Let us assume then that at the n^{th} stage of sampling we have accumulated n_1 observations from a population distributed as $N(\alpha_1, \sigma^2)$ and n_2 observations from another population distributed as $N(\alpha_2, \sigma^2)$.

Our region of preference for acceptance of H_o is

$$w_o = [(\alpha_1, \alpha_2, \sigma): -\infty < \alpha_i < \infty,\ i = 1, 2;\ 0 < \sigma < \infty,$$
$$\text{and } \alpha_1 = \alpha_2]$$

and our region of preference for acceptance of H_1 is

$$w_1 = [(\alpha_1, \alpha_2, \sigma): -\infty < \alpha_i < \infty, i = 1, 2; 0 < \sigma < \infty,$$
$$|\alpha_1 - \alpha_2| = \gamma\sigma \quad \text{and } \gamma \text{ specified}].$$

The boundary of w_1, denoted by S_1, is

$$S_1 = [(\alpha_1, \alpha_2, \sigma): -\infty < \alpha_i < \infty, i = 1, 2; 0 < \sigma < \infty,$$
$$|\alpha_1 - \alpha_2| = \gamma\sigma \quad \text{and } \gamma \text{ specified}].$$

The likelihood of all the observations taken through the n^{th} stage from both populations is

$$L(y_1, y_2; \alpha_1, \alpha_2, \sigma) = (2\pi\sigma)^{-\frac{k}{2}} \exp[-\frac{1}{2\sigma^2} Q(\alpha_1, \alpha_2)] \qquad (5.6)$$

where $y_1' = (y_{11}, y_{12}, \ldots, y_{1n_1})$, $y_2' = (y_{21}, \ldots, y_{2n_2})$,

$k = n_1 + n_2$, and

$$Q_o(\alpha_1, \alpha_2) = \sum_1^{n_1} (y_{1i}-\alpha_1)^2 + \sum_1^{n_2} (y_{1i}-\alpha_2)^2,$$

which later, it will be convenient to write as

$$Q_o(\alpha_1, \alpha_2) = n_1\alpha_1^2 + n_2\alpha_2^2 - 2n_1\bar{y}_1\alpha_1 - 2n_2\bar{y}_2\alpha_2 + R \qquad (5.7)$$

where $R = S_{y_1y_1} + n_1\bar{y}_1^2 + S_{y_2y_2} + n_2\bar{y}_2^2$, $\bar{y}_1 = \dfrac{\sum_1^{n_1} y_{1i}}{n_1}$

$\bar{y}_2 = \dfrac{\sum_1^{n_2} y_{2i}}{n_2}$, $S_{y_1y_1} = \sum_1^{n_1} (y_{1i}-\bar{y}_1)^2$, and $S_{y_2y_2} = \sum_1^{n_2} (y_{2i}-\bar{y}_2)^2$.

We consider the following prior quasi-densities,

$$V_o(\alpha_1, \alpha_2, \sigma) = 1,\ 0 < \sigma < \infty,\ -\infty < \alpha_i < \infty,\ i = 1, 2;$$

$$|\alpha_1 - \alpha_2| = \gamma\sigma,\ \gamma = 0,$$

$$= 0, \text{ otherwise} \tag{5.8}$$

$$V_1(\alpha_1, \alpha_2, \sigma) = 1,\ 0 < \sigma < \infty,\ -\infty < \alpha_i < \infty,\ i = 1, 2$$

$$|\alpha_1 - \alpha_2| = \gamma\sigma,\ \gamma \neq 0 \text{ and specified},$$

$$= 0, \text{ otherwise}$$

on the region w_o of preference for acceptance of H_o and on the boundary S_1 of w_1 where w_1 is the region of preference for acceptance of H_1. As in Chapter III we calculate the ratio of marginal likelihoods given V_1 and V_o respectively. We denote this ratio by R(0) where

$$R(0) = \frac{\int_{S_1} V_1(\alpha_1, \alpha_2; \sigma)\, L(y_1, y_2; \alpha_1, \alpha_2, \sigma)\, d\alpha_1 d\alpha_2 d\sigma}{\int_{w_o} V_o(\alpha_1, \alpha_2, \sigma)\, L(y_1, y_2; \alpha_1, \alpha_2, \sigma)\, d\alpha_1 d\alpha_2 d\sigma} \tag{5.9}$$

The numerator of 5.9 can be written as

$$c_1 \int_0^\infty \int_{-\infty}^\infty \sigma^{-k} [\exp[-\frac{1}{2\sigma^2} Q_o(\alpha_2 + \gamma\sigma, \alpha_2)]$$

$$+ \exp[-\frac{1}{2\sigma^2} Q_o(\alpha_2 - \gamma\sigma, \alpha_2)]]\, d\alpha_1 d\sigma \tag{5.10}$$

where

$$Q_o(\alpha_2 + \gamma\sigma, \alpha_2) = k\alpha_2^2 - 2\alpha_2(n_1\bar{y}_1 + n_2\bar{y}_2 - n_1\gamma\sigma)$$

$$+ n_1\gamma^2\sigma^2 - 2n_1\bar{y}_1\gamma\sigma + R$$

$$Q_o(\alpha_2 - \gamma\sigma,\ \alpha_2) = k\alpha_2^2 - 2\alpha_2(n_1\bar{y}_1 + n_2\bar{y}_2 + n_1\gamma\sigma)$$
$$+ n_1\gamma^2\sigma^2 + 2n_1\bar{y}_1\gamma\sigma + R$$
$$c_1 = (2\pi)^{-\frac{k}{2}} \tag{5.11}$$
$$k = n_1 + n_2$$

and

R is as previously defined following 5.7.

Now

$$\int_0^\infty \int_{-\infty}^\infty \sigma^{-k} \exp[-\frac{1}{2\sigma^2} Q_o(\alpha_2 + \gamma\sigma,\ \alpha_2)]\, d\alpha_2 d\sigma$$

$$= \int_0^\infty \exp[-\frac{1}{2\sigma^2} (n_1\gamma^2\sigma^2 - 2n_1\bar{y}_1\gamma\sigma + R)] \cdot T(\sigma)d\sigma \tag{5.12}$$

where

$$T(\sigma) = \int_{-\infty}^\infty \sigma^{-k} \exp - \frac{1}{2\sigma^2}(k_1\alpha_1^2 - 2\alpha_1(n_1\bar{y}_1 + n_2\bar{y}_2 - n_1\gamma\sigma))\, d\alpha_2$$

$$= c_2\sigma^{-(k-1)} \exp\left[\frac{(n_1\bar{y}_1 + n_2\bar{y}_2 - n_1\gamma\sigma)^2}{k}\right] \tag{5.13}$$

and

$$c_2 = (2\pi)^{\frac{1}{2}}.$$

We write 5.12 as

$$c_2 \int_0^\infty \sigma^{-(k-1)} \exp - \frac{1}{2\sigma^2} (\gamma^2\sigma^2(\frac{n_1 - n_1^2}{k}) - 2n_1\bar{y}_1\gamma\sigma + R$$
$$- \frac{(n_1\bar{y}_1 + n_2\bar{y}_2)^2}{k} + \frac{2n_1\gamma\sigma(n_1\bar{y}_1 + n_2\bar{y}_2)}{k}\ d\sigma. \tag{5.14}$$

Letting $u = \frac{k}{n_1 n_2}$ and expanding and gathering terms, 5.14 then becomes

$$c_2 \int_0^\infty \sigma^{-(k-1)} \exp[-\frac{1}{2\sigma^2 u} (\gamma^2\sigma^2 - 2\gamma\sigma(\bar{y}_1-\bar{y}_2) + (\bar{y}_1-\bar{y}_2)^2$$

$$+ u(S_{y_1y_1} + S_{y_2y_2}))]\, d\sigma. \qquad (5.15)$$

By a result given in Section C of the Appendix we complete the integration in 5.15 so that 5.15 becomes

$$c_2 c_3 \exp[-\frac{\gamma^2}{2u}][F(\frac{k-2}{2};\ \frac{1}{2};\ z^2)$$

$$+ \frac{\gamma\Gamma(\frac{k-1}{2})}{\Gamma(\frac{k-2}{2})} \left(\frac{2(y_1-y_2)^2}{u(S_{y_1y_1} + S_{y_2y_2}) + (y_1-y_2)^2}\right)^{\frac{1}{2}} F(\frac{k-1}{2};\ \frac{3}{2};\ z^2)]$$

where

$$z^2 = \frac{\gamma^2}{2u} \frac{(\bar{y}_1-\bar{y}_2)^2}{u(S_{y_1y_1} + S_{y_2y_2}) + (\bar{y}_1-\bar{y}_2)^2} \qquad (5.16)$$

and

$$c_3 = \left[u(S_{y_1y_1} + S_{y_2y_2})\right]^{-\frac{k-2}{2}} 2^{-\frac{k-3}{2}} \Gamma(\frac{k-2}{2})$$

and where, as before, $F(P_1;\ P_2;\ z)$ is the confluent hypergeometric function. Similarly

$$\int_0^\infty \int_{-\infty}^\infty \sigma^{-k} \exp\left[-\frac{1}{2\sigma^2} Q_o(\alpha_2 - \gamma\sigma,\ \alpha_2)\right] d\alpha_2 d\sigma$$

$$= c_2 c_3 \exp[-\frac{\gamma^2}{2u}]\Big[F(\frac{k-2}{2}); \frac{1}{2}; Z^2)$$

$$- \frac{\gamma\Gamma(\frac{k-1}{2})}{\Gamma(\frac{k-2}{2})}\left(\frac{2(\bar{y}_1-\bar{y}_2)^2}{u(S_{y_1y_1}+S_{y_2y_2})+(\bar{y}_1-\bar{y}_2)^2}\right)^{\frac{1}{2}} F(\frac{k-1}{2}; \frac{3}{2}; Z^2)\Big] \qquad (5.17)$$

where c_2 following 5.13 and c_3 and Z^2 take the same values as in 5.16. Adding 5.16 and 5.17 and multiplying the result by c_1 we obtain the following as a closed form expression for 5.10,

$$c_1c_2c_3 \exp[-\frac{\gamma^2}{2u}]F(\frac{k-2}{2}; \frac{1}{2}; \frac{\gamma^2}{2u}\frac{(\bar{y}_1-\bar{y}_2)^2}{u(S_{y_1y_1}+S_{y_2y_2})+(\bar{y}_1-\bar{y}_2)^2}). \qquad (5.18)$$

The denominator of the ratio 5.9 may be obtained from 5.18 by simply setting $\gamma = 0$. The ratio 5.9 then becomes

$$R(0) = \exp[-\frac{\gamma^2}{2u}]F(\frac{k-2}{2}; \frac{1}{2}; \frac{\gamma^2}{2u}\frac{(\bar{y}_1-\bar{y}_2)^2}{u(S_{y_1y_1}+S_{y_2y_2})+(\bar{y}_1-\bar{y}_2)^2}). \qquad (5.19)$$

Finally, if we let

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\hat{\sigma}\, u} \qquad (5.20)$$

where

$$\hat{\sigma}^2 = \frac{S_{y_1y_1} + S_{y_2y_2}}{k-2},$$

5.19 can be written

$$R(0) = \exp[-\frac{\gamma^2}{2u}]\ F(\frac{k-2}{2}; \frac{1}{2}; \frac{\gamma^2}{2u}\frac{t^2}{k-2+t^2}). \qquad (5.21)$$

If the prior quasi-densities

$$V_o(\alpha_1, \alpha_2, \sigma) = \frac{1}{\sigma},\ 0 < \sigma < \infty;\ -\infty < \alpha_i < \infty,\ i = 1, 2;$$
$$|\alpha_1 - \alpha_2| = \gamma\sigma,\ \gamma = 0$$
$$= 0, \text{ otherwise} \tag{5.22}$$
$$V_1(\alpha_1, \alpha_2, \sigma) = \frac{1}{\sigma},\ 0 < \sigma < \infty;\ -\infty < \alpha_i < \infty,\ i = 1, 2;$$
$$|\alpha_1 - \alpha_2| = \gamma\sigma,\ \gamma \neq 0 \text{ and specified}$$
$$= 0, \text{ otherwise}$$

were used rather than those given by 5.8, the first argument of the confluent hypergeometric function given in 5.21 would be $\frac{k-1}{2}$ rather than $\frac{k-2}{2}$. The k-2 of the last argument would, however, remain unchanged. The likelihood ratio would then agree exactly with Hajnal's result (26, p. 66). We chose the prior quasi-densities given by 5.8 because they uniformly weight each point $(\alpha_1, \alpha_2, \sigma)$ of w_o and S_1 whereas the prior quasi-densities given by 5.22 weigh each point inversely proportional to σ.

We know of no detailed theoretical or empirical studies of the properties of either Hajnal's result or have any been obtained for that given in 5.21. We note, however, that since the confluent hypergeometric function is monotone increasing in the first argument (Slater, 67), 5.21 is slightly more conservative than the ratio derived by Hajnal. That is, Hajnal's test procedure will reject H_o more frequently than the test procedure with 5.21 as its likelihood ratio test statistic.

B. The Sequential Two-Sample t^2-Test with One Covariate

In this section we now apply the prior distribution technique to the analysis of a model which is basically the same as in Section A except that now one covariate is introduced. At a particular stage of sampling we accordingly assume the following model:

$$
\begin{aligned}
y_{1j} &= \alpha_1 + \beta(x_{1j}-\overline{x}_1) + \epsilon_{1j}, \quad j - 1, 2, \ldots, n_1 \\
y_{2j} &= \alpha_2 + \beta(x_{2j}-\overline{x}_2) + \epsilon_{2j}, \quad j = 1, 2, \ldots, n_2
\end{aligned}
\tag{5.23}
$$

where $\epsilon_{ij} \sim NI(0, \sigma^2)$, $i = 1, 2$; $j = 1, \ldots, n_i$, and where x_{ij} is the j^{th} covariate measurement on the i^{th} treatment and $\overline{x}_i = \frac{\sum_1^{n_i} x_{ij}}{n_i}$, $i = 1, 2$. The covariates are assumed to be controlled.

If we picture the model 5.23 in Euclidean two-space we see that the difference in treatments is the distance between the two regression lines measured along any line drawn parallel to the y-axis. Since the two lines are assumed to be parallel this distance is constant for all x and is equal to the following function of the parameters and x's:

$$\alpha_1 - \alpha_2 - \beta(\overline{x}_1-\overline{x}_2).$$

Accordingly, one hypothesis formulation that we may test is a generalization of 5.1,

$$H_o\colon\ |\alpha_1 - \alpha_2 - \beta(\bar{x}_1-\bar{x}_2)| = \gamma_1\sigma$$
$$H_1\colon\ |\alpha_1 - \alpha_2 - \beta(\bar{x}_1-\bar{x}_2)| = \gamma_2\sigma, \tag{5.24}$$

where γ_1 and γ_2 are specified and $\gamma_1 \neq \gamma_2$ and where α_1, α_2, σ, and β are unspecified. If we let $\gamma_1 = 0$ then our null hypothesis is that there is no treatment difference and the alternative hypothesis is that the treatment difference is γ_2 standard deviations.

The prior quasi-densities we consider are

$$\begin{aligned} V_o(\alpha_1, \alpha_2, \beta, \sigma) &= 1,\ 0 < \sigma < \infty,\ -\infty < \alpha_i < \infty,\ i = 1, 2; \\ &\qquad -\infty < \beta < \infty,\ |\alpha_1 - \alpha_2 - \beta(\bar{x}_1-\bar{x}_2)| = \gamma_1\sigma \\ &= 0, \text{ otherwise} \qquad\qquad (5.25) \\ V_1(\alpha_1, \alpha_2, \beta, \sigma) &= 1,\ 0 < \sigma < \infty,\ -\infty < \alpha_i < \infty,\ i = 1, 2; \\ &\qquad -\infty < \beta < \infty,\ |\alpha_1 - \alpha_2 - \beta(\bar{x}_1-\bar{x}_2)| = \gamma_2\sigma \\ &= 0, \text{ otherwise.} \end{aligned}$$

where in general $\gamma_1 < \gamma_2$.

The likelihood of the n_1+n_2 observations is

$$L(\alpha_1, \alpha_2, \beta, \sigma) = (2\pi\sigma^2)^{-\frac{n_1+n_2}{2}} \exp\left[-\frac{1}{2\sigma^2} Q_1(\alpha_1, \alpha_2, \beta)\right] \tag{5.26}$$

where

$$Q_1(\alpha_1, \alpha_2, \beta) = \sum_1^{n_1} (y_{1j}-\alpha_1-\beta(x_{1j}-\bar{x}_1))^2$$
$$\sum_1^{n_2} (y_{2j}-\alpha_2-\beta(x_{2j}-\bar{x}_2))^2. \tag{5.27}$$

As in Section A of this chapter we proceed to calculate the ratio of the marginal likelihoods of $L(\alpha_1, \alpha_2, \beta, \sigma)$ using the prior quasi-densities in 5.23, as

$$R(1) = \frac{R_1(\gamma_2)}{R_1(\gamma_1)} \tag{5.28}$$

where

$$R_1(\gamma) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty L(\alpha_1, \alpha_1 - \beta(\bar{x}_1-\bar{x}_2) + \gamma\sigma, \beta, \sigma)\, d\alpha_1 d\beta d\sigma$$

$$+ \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty L(\alpha_1, \alpha_1 - \beta(\bar{x}_1-\bar{x}_2) - \gamma\sigma, \beta, \sigma)\, d\alpha_1 d\beta d\sigma \tag{5.29}$$

for $\gamma = \gamma_1, \gamma_2$.

The first term on the right hand side of the equality sign in 5.29 is

$$\int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty (2\pi\sigma^2)^{-\frac{n_1+n_2}{2}} \exp\left[-\frac{1}{2\sigma^2} Q(\alpha_1, \alpha_1 - \beta(\bar{x}_1-\bar{x}_2) + \gamma\sigma, \beta)\right] d\alpha_1 d\beta d\sigma. \tag{5.30}$$

Now $Q_1(\alpha_1, \alpha_1 - \beta(\bar{x}_1-\bar{x}_2) + \gamma\sigma, \beta)$

$$= \sum_1^{n_1} (y_{1j}-\alpha_1-\beta(x_{1j}-\bar{x}_1))^2 + \sum_1^{n_2} (y_{2j}-\alpha_1+\beta(\bar{x}_1-\bar{x}_2) - \gamma\sigma - \beta(x_{2j}-\bar{x}_2))^2$$

$$= (\alpha_1, \beta) \begin{pmatrix} n_1+n_2 & n_2(\bar{x}_2-\bar{x}_1) \\ n_2(\bar{x}_2-\bar{x}_1) & S_{x_1x_2}+S_{x_2x_2}+n_2(\bar{x}_2-\bar{x}_1) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \beta \end{pmatrix}$$

$$- 2(\Sigma y-n_2\sigma, S_{x_1y_1}+S_{x_2y_2}+n_2(\bar{x}_2-\bar{x}_1)(\bar{y}_2-\gamma\sigma)$$

$$+ \Sigma y^2 - 2\gamma\sigma\Sigma y_2 + n_2\gamma^2\sigma^2 \tag{5.31}$$

where, in addition to the notation given in 5.23, we define

$$\Sigma y = \sum_{i=1}^{2} \sum_{j=1}^{n_i} y_{ij}, \quad \Sigma y^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} y_{ij}^2$$

$$\Sigma y_1 = \sum_{1}^{n_1} y_{1j}, \quad \Sigma y_2 = \sum_{1}^{n_2} y_{2j}, \quad \bar{y}_1 = \frac{\Sigma y_1}{n_1}, \quad \bar{y}_2 = \frac{\Sigma y_2}{n_2}$$

$S_{x_1y_1} = \sum_{1}^{n_1} (x_{1j}-\bar{x}_1)y_{1j}$, and $S_{x_1x_1}$, $S_{x_2x_2}$, $S_{x_2y_2}$ are defined in a similar manner.

Integration of 5.30 over α_1 and β then gives

$$k_1\int_0^\infty \sigma^{n_1+n_2-2} \exp\left[-\frac{1}{2\sigma^2}(a+b)\right] d\sigma \tag{5.32}$$

where

$$a = \Sigma y^2 - \gamma\sigma\Sigma y_2 + n_2\gamma^2\sigma^2$$

$$b = -h^{-1}(e,\ f)\ A\binom{e}{f}$$

$$A = \begin{pmatrix} c+n_2(x_2-x_1)^2 & -n_2(\bar{x}_2-\bar{x}_1) \\ -n_2(\bar{x}_2-\bar{x}_1) & n_1+n_2 \end{pmatrix} \tag{5.33}$$

$$c = S_{x_1x_1} + S_{x_2x_2}$$

$$h = \text{determinant of } A = |A| = (n_1+n_2)c+n_1n_2(\bar{x}_2-\bar{x}_1)$$

$$e = \Sigma y_1+n_2(\bar{y}_2-\gamma\sigma)$$

$$g = S_{x_1y_1} + S_{x_2y_2}$$

$$f = g+n_2(\bar{x}_2-\bar{x}_1)(\bar{y}_2-\gamma\sigma)$$

$$k_1 = (2\pi)^{\frac{n_1+n_2-2}{2}} |A|.$$

Expanding and gathering terms in $\gamma\sigma$ we write

$$a+b = u\gamma^2\sigma^2 - 2v\gamma\sigma + w$$

where

$$u = n_2 - \frac{n_2^2}{h}(c+n_2(\bar{x}_2-\bar{x}_1)^2) + \frac{2n_2^3(\bar{x}_2-\bar{x}_1)^2}{h} - (n_1+n_2)n_2^2(\bar{x}_2-\bar{x}_1)^2$$

$$= n_2 + \frac{-n_2^2c-n_1n_2(\bar{x}_2-\bar{x}_1)^2}{h}$$

$$= \frac{1}{\frac{1}{n_1}+\frac{1}{n_2}+\frac{(\bar{x}_2-\bar{x}_1)^2}{S_{x_1x_1}+S_{x_2x_2}}}, \qquad (5.35)$$

$$v = n_2\bar{y}_2 - \frac{n_2(n_1\bar{y}_1+n_2\bar{y}_2)(c+n_2(\bar{x}_2-\bar{x}_1))}{c+n_1n_2(x_2-x_1)^2}$$

$$+ \frac{n_2^2(\bar{x}_2-\bar{x}_1)}{c+n_1n_2(\bar{x}_2-\bar{x}_1)}(g+(\bar{x}_2-\bar{x}_1)(n_1\bar{y}_1+n_2\bar{y}_2))$$

$$- \frac{n_2(n_1+n_2)(\bar{x}_2-\bar{x}_1)}{c+n_1n_2(\bar{x}_2-\bar{x}_1)^2}(g+n_2(\bar{x}_2-\bar{x}_1)y_2)$$

$$= \frac{n_1n_2\bar{y}_2 - n_1n_2\bar{y}_1 - n_1n_2(\bar{x}_2-\bar{x}_1)(g/c)}{n_1+n_2+\frac{n_1n_2(\bar{x}_2-\bar{x}_1)^2}{c}}$$

$$= -[\bar{y}_2 - \bar{y}_1 + (g/c)(\bar{x}_2-\bar{x}_1)]u, \qquad (5.36)$$

and

$$w = \Sigma y^2 - \frac{c + n_2(\bar{x}_2-\bar{x}_1)}{h}(h_1\bar{y}_1+n_2\bar{y}_2)$$

$$+ 2\,\frac{n_2(\bar{x}_2-\bar{x}_1)(n_1\bar{y}_1+n_2\bar{y}_2)}{h}(g + n_2(\bar{x}_2-\bar{x}_1)\bar{y}_1)$$

$$- \frac{(n_1+n_2)}{h}(g + n_2\bar{y}_2(\bar{x}_2-\bar{x}_1))^2. \qquad (5.37)$$

For convenience we simplify $\frac{w}{u}$ rather than w.

$$\frac{w}{u} = \frac{1}{u}[\Sigma y^2 - n_1\bar{y}_1^2 - n_2\bar{y}_2^2 - \frac{g^2}{c}] + \frac{1}{u}[w - \Sigma y^2 + n_1\bar{y}_1^2 + n_2\bar{y}_2^2 + \frac{g^2}{c}]$$

$$= \frac{1}{u}[\Sigma y^2 - n_1\bar{y}_1^2 - n_2\bar{y}_2^2 - \frac{g^2}{c}]$$

$$+ \frac{1}{n_1n_2c}[c(n_1+n_2)(n_1\bar{y}_1^2+n_2\bar{y}_2^2+\frac{g^2}{c})$$

$$+ (n_1\bar{y}_1^2+n_2\bar{y}_2^2+\frac{g^2}{c})n_1n_2(\bar{x}_2-\bar{x}_1)^2 - c(n_1\bar{y}_1+n_2\bar{y}_2)^2$$

$$- n_2(\bar{x}_2-\bar{x}_1)^2(n_1^2\bar{y}_1^2+2n_1n_2\bar{y}_1\bar{y}_2+n_2^2\bar{y}_2^2)$$

$$+ 2n_2(\bar{x}_2-\bar{x}_1)(n_1\bar{y}_1+n_2\bar{y}_2)g$$

$$+ 2n_2^2(\bar{x}_2-\bar{x}_1)^2(n_1\bar{y}_1+n_2\bar{y}_2)\bar{y}_2 - (n_1+n_2)g^2$$

$$- 2n_2(n_1+n_2)\bar{y}_2(\bar{x}_2-\bar{x}_1)g - (n_1+n_2)n_2^2\bar{y}_2^2(\bar{x}_2-\bar{x}_1)^2]$$

$$= \frac{1}{u}[\Sigma y^2 - n_1\bar{y}_1^2 - n_2\bar{y}_2^2 + \frac{g^2}{c}] + [\bar{y}_1 - \bar{y}_2 + \frac{g}{c}(\bar{x}_2-\bar{x}_1)]^2. \qquad (5.38)$$

For the model 5.21 it is easily shown that $\bar{y}_1$, $\bar{y}_2$, and $\frac{g}{c}$ are estimates of α_1, α_2 and β respectively, and an estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{\Sigma y^2 - n_1\bar{y}_1 - n_2\bar{y}_2^2 - \hat{\beta}g}{n_1 + n_2 - 3} \qquad (5.39)$$

where g, c, Σy^2 are defined in 5.33.

We now rewrite 5.36 and 5.38 so that

$$v = -[\bar{y}_2 - \bar{y}_1 + \hat{\beta}(\bar{x}_2-\bar{x}_1)]u \tag{5.40a}$$

and

$$w = (n-3)\sigma^2 + u[\bar{y}_1 - \bar{y}_2 + \hat{\beta}(\bar{x}_2-\bar{x}_1)]^2. \tag{5.40b}$$

Therefore, by Lemma 10.3 of Section C of the Appendix,

$$\int_0^\infty k_1\sigma^{-(n_1+n_2-2)} \exp[-\frac{1}{2\sigma^2}(u\gamma^2\sigma^2-2v\gamma\sigma + w)]\ d\sigma$$

$$= k_2 \exp\left[-\frac{\gamma^2u}{2}\right]\left[F(\frac{n_1+n_2-2}{2};\ \frac{1}{2};\ \frac{\gamma^2v^2}{2w^2})\right.$$

$$\left. + \frac{2\gamma v}{(2w)^{\frac{1}{2}}}\frac{(\frac{n_1+n_2-2}{2})}{(\frac{n_1+n_2-3}{2})} F(\frac{n_1+n_2-3}{2};\ \frac{3}{2};\ \frac{\gamma^2v^2}{2w})\right] \tag{5.41}$$

where $F(p_1;\ p_2;\ z)$ is the confluent hypergeometric function and

$$k_2 = k_1 w^{-\frac{n_1+n_2-3}{2}}\ 2^{-\frac{n_1+n_2-5}{2}}\ \Gamma\ (\frac{n_1+n_2-3}{2}).$$

It is noted here that k_2 does not depend on γ.

In the same manner, for the second term of the right hand side of the equality sign in 5.29 we have

$$\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty (2\pi\sigma^2)^{-\frac{n_1+n_2}{2}} \exp\left[-\frac{1}{2\sigma^2} Q_1(\alpha_1,\ \alpha_1-\beta(\bar{x}_1-\bar{x}_2)-\gamma\sigma,\beta)\right] d\alpha_1 d\beta d\sigma$$

$$= k_2 \exp\left[-\frac{\gamma^2u}{2}\right]\ \left[F(\frac{n_1+n_2-3}{2};\ \frac{1}{2};\ \frac{\gamma^2v^2}{2w})\right.$$

$$\left. - \frac{2\gamma v}{(2w)^{1/2}}\frac{\Gamma(\frac{n_1+n_2-2}{2})}{\Gamma(\frac{n_1+n_2-3}{2})} F(\frac{n_1+n_2-2}{2};\ \frac{3}{2};\ \frac{\gamma^2v^2}{2w})\right] \tag{5.42}$$

where k_2 is exactly as given in 5.41. From 5.29 we see that $R_1(\gamma)$ is the sum of the expressions given in 5.41 and 5.42, that is,

$$R_1(\gamma) = k_2 \exp[-\frac{\gamma^2 u}{2}]\ F(\frac{n_1+n_2-3}{2};\ \frac{1}{2};\ \frac{\gamma^2 v^2}{2w^2}) \tag{5.43}$$

Therefore

$$R(1) = \frac{R_1(\gamma_2)}{R_1(\gamma_1)}$$

may be obtained from 5.43 as

$$R(1) = \exp\left[-\frac{(\gamma_2^2-\gamma_1^2)u}{2}\right] \frac{F(\frac{n_1+n_2-3}{2};\ \frac{1}{2};\ \frac{\gamma_2^2 v^2}{2w})}{F(\frac{n_1+n_2-3}{2};\ \frac{1}{2};\ \frac{\gamma_1^2 v^2}{2w})}. \tag{5.44}$$

Recalling the definitions of u, v, and w given in 5.35, 5.40a and 5.40b respectively, we can write

$$\frac{\gamma^2 v^2}{2w} = \frac{\gamma^2 u}{2} \frac{t^2_{n_1+n_2-3}}{(n_1+n_2-3) + t^2_{n_1+n_2-3}} \tag{5.45}$$

where

$$t^2_{n_1+n_2-3} = \frac{(\bar{y}_1-\bar{y}_2+\hat{\beta}(\bar{x}_2-\bar{x}_1))^2}{(\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1-\bar{x}_2)^2}{S_{x_1x_1} + S_{x_2x_2}})\ \hat{\sigma}^2} \tag{5.46}$$

and where $\hat{\sigma}^2$ is as given by 5.39.

It may now be noticed that $t^2_{n_1+n_2-3}$ is the Student's t^2-test statistic appropriate in a fixed size sample sense for testing

$$H_1: \ |\alpha_1 - \alpha_2 + \beta(\bar{x}_2 - \bar{x}_1)| = 0$$
$$H_2: \ |\alpha_1 - \alpha_2 + \beta(\bar{x}_2 - \bar{x}_1)| \neq 0. \tag{5.47}$$

Since

$$u = \frac{\hat{\sigma}^2}{\mathrm{Var}(\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\beta}(\bar{x}_1 - \bar{x}_2))}$$

$$= \frac{1}{\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{S_{x_1x_1} + S_{x_2x_2}}}$$

and

$$\frac{\gamma^2 u}{2} = \frac{\gamma^2}{2(\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{S_{x_1x_2} + S_{x_2x_2}})}, \quad \gamma = \gamma_1, \gamma_2$$

then if we let $\gamma_1 = 0$ we have 5.46 as a natural extension (to the case where concomitant information is used) of Hajnal's (26) result.

C. The Sequential Two-Sample t^2-Test with p Covariates

1. The model and hypothesis formulation

In this section we extend the results of Section B to the case of p covariates. At the n^{th} stage of sampling we assume that we have n_1 observations from population 1 and n_2 observations from population 2 and that the following model obtains:

$$y_{1j} = \alpha_1 + \beta_1(x_{11j}-\bar{x}_{11}) + \beta_2(x_{12j}-\bar{x}_{12}) + \ldots$$
$$+ \beta_p(x_{1pj}-\bar{x}_{1p}) + \epsilon_{ij}$$
$$y_{2j} = \alpha_2 + \beta_1(x_{21j}-\bar{x}_{21}) + \beta_2(x_{22j}-\bar{x}_{22}) + \ldots \tag{5.48}$$
$$+ \beta_p(x_{2pj}-\bar{x}_{2p}) + \epsilon_{ij}$$

where $\epsilon_{ij} \sim NI(0, \sigma^2)$, $i = 1, 2$; $j = 1, \ldots, n_i$, and x_{ikj} is the j^{th} observation of the k^{th} covariate for treatment i.

The expression for differences in treatments based on model 5.48 is an extension to p covariates of the expression given for one covariate in Section B of this Chapter and is equal to the following function of the parameters and x's:

$$\alpha_1 - \alpha_2 + \beta_1(\bar{x}_{21}-\bar{x}_{11}) + \beta_2(\bar{x}_{22}-\bar{x}_{12}) + \ldots + \beta_p(\bar{x}_{2p}-\bar{x}_{1p})$$
$$= \alpha_1 - \alpha_2 + z'\beta \tag{5.49}$$

where

$z' = (z_1, z_2, \ldots, z_p)$, $z_i = \bar{x}_{2i} - \bar{x}_{1i}$, and

$\beta' = (\beta_1, \beta_2, \ldots, \beta_p)$.

The hypothesis we shall test is a generalization of 5.24 and is as follows:

$$H_o: |\alpha_1 - \alpha_2 + z'\beta| = \gamma_1\sigma$$
$$H_1: |\alpha_1 - \alpha_2 + z'\beta| = \gamma_2\sigma \tag{5.50}$$

where γ_1 and γ_2 are specified and α_1, α_2, and $(\beta_1, \beta_2, \ldots, \beta_p)$ are unspecified. The prior quasi-densities to be considered are

$$V_o(\alpha_1, \alpha_2, \beta, \sigma) = 1, \ 0 < \sigma < \infty, \ -\infty < \alpha_i < \infty, \ i = 1, 2;$$
$$-\infty < \beta_i < \infty, \ i = 1, \ldots, p;$$
$$|\alpha_1 - \alpha_2 + z'\beta| = \gamma_1\sigma$$
$$= 0, \text{ otherwise}$$

(5.51)

$$V_1(\alpha_1, \alpha_2, \beta, \sigma) = 1, \ 0 < \sigma < \infty, \ -\infty < \alpha_i < \infty, \ i = 1, 2;$$
$$-\infty < \beta_i < \infty, \ i = 1, 2, \ldots, p;$$
$$|\alpha_1 - \alpha_2 + z'\beta| = \gamma_2\sigma$$
$$= 0, \text{ otherwise.}$$

The likelihood of all the n_1+n_2 observations accumulated the n^{th} stage is

$$L(\alpha_1, \alpha_2, \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{(n_1+n_2)}{2}} \exp[-\frac{1}{2\sigma^2} Q_p(\alpha_1, \alpha_2, \beta)] \tag{5.52}$$

where

$$Q_p(\alpha_1, \alpha_2, \beta) = (y_1 - \alpha_1 e_1 - X_1'\beta)'(y_1 - \alpha_1 e - X_1'\beta) + (y_2 - \alpha_2 e_2 - X_2'\beta)'(y_2 - \alpha_2 e - X_2'\beta) \tag{5.53}$$

and

$$y_1' = (y_{11}, y_{12}, \ldots, y_{1n_1})$$
$$y_2' = (y_{21}, y_{22}, \ldots, y_{2n_2})$$
$$e_1' = (1, \ldots, 1)_{1xn_1}, \ e_2' = (1, \ldots, 1)_{1xn_2} \tag{5.54}$$

and X_1' is the n_1 x p matrix

$$X_1 = \begin{pmatrix} x_{111}-\bar{x}_{11} & x_{121}-\bar{x}_{12} & \cdots & x_{1p1}-\bar{x}_{1p} \\ x_{112}-\bar{x}_{11} & x_{122}-\bar{x}_{12} & \cdots & x_{1p2}-\bar{x}_{1p} \\ \vdots & \vdots & & \vdots \\ x_{11n_1}-\bar{x}_{11} & x_{12n_1}-\bar{x}_{12} & \cdots & x_{1pn_1}-\bar{x}_{n_1p} \end{pmatrix}$$

amd X_2 is a n_2 x p matrix similarly defined.

2. Derivation of the test statistic

We now evaluate the ratio, R(p), of the marginal likelihood of $L(\alpha_1, \alpha_2, \beta, \sigma^2)$ given $V_1(\alpha_1, \alpha_2, \beta, \sigma)$ to the marginal likelihood of $L(\alpha_1, \alpha_2, \beta, \sigma^2)$ given $V_o(\alpha_1, \alpha_2, \beta, \sigma)$ where

$$R(p) = \frac{R_p(\gamma_2)}{R_p(\gamma_1)} \tag{5.55}$$

and

$$R_p(\gamma) = \int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty L(\alpha_1, \alpha_1 + z'\beta - \gamma\sigma, \beta, \sigma)\, d\alpha_1 d\beta d\sigma$$
$$+ \int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty L(\alpha_1, \alpha_1 + z'\beta + \gamma\sigma, \beta, \sigma)\, d\alpha_1 d\beta d\sigma \tag{5.56}$$

and $\gamma = \gamma_2, \gamma_1$.

Before completing the integration as indicated in 5.56 we work out some algebraic details regarding $Q_p(\alpha_1, \alpha_2, \beta)$. We rewrite 5.53 as

$$\begin{aligned} Q_p(\alpha_1, \alpha_2, \beta) = {} & \beta'(X_1X_1' + X_2X_2')\beta + y_1'y_1 + y_2'y_2 \\ & - 2(X_1y_1 + X_2y_2)'\beta - 2\alpha_1 e_1'y_1 + \alpha_1^2 e_1'e_1 \\ & - 2\alpha_2 e_2'y_2 + \alpha_2^2 e_2'e_2. \end{aligned} \tag{5.57}$$

Also, letting $\alpha_2 = \alpha_1 + z'\beta - \gamma\sigma$ in $Q_p(\alpha_1, \alpha_2, \beta)$ we have

$$\begin{aligned} &Q_p(\alpha_1, \alpha_1 + z'\beta - \gamma\sigma, \beta) \\ &= \beta'(X_1X_1' + X_2X_2')\beta - 2(X_1y_1+X_2y_2)'\beta + y_1'y_1 + y_2'y_2 \\ &\quad - 2\alpha_1 e_1'y_1 + \alpha_1^2 e_1'e_1 - 2e_2'y_2(\alpha_1+z'\beta-\gamma\sigma) \\ &\quad + e_2'e_2(\alpha_1+z'\beta-\gamma\sigma)^2 \end{aligned} \tag{5.58}$$

$$= \begin{pmatrix}\alpha_1\\ \beta\end{pmatrix}' A\begin{pmatrix}\alpha_1\\ \beta\end{pmatrix} - 2(u_1+u_2\gamma\sigma)'\begin{pmatrix}\alpha_1\\ \beta\end{pmatrix} + f \tag{5.59}$$

where A is the $(p+1) \times (p+1)$ matrix of rank $p+1$

$$A = \begin{pmatrix} n_1+n_2 & n_2z' \\ zn_2 & X_1X_1' + X_2X_2' + n_2zz' \end{pmatrix} \tag{5.60}$$

and u_1 and u_2 are $(p+1) \times 1$ vectors defined by

$$u_1 = \begin{pmatrix} e_1'\ y_1+e_2'\ y_2 \\ X_1'\ y_1+X_2'\ y_2+n_2\bar{y}_2z \end{pmatrix}, \quad u_2 = \begin{pmatrix} n_2 \\ n_2z \end{pmatrix} \tag{5.61}$$

and f is a scalar defined by

$$f = y_1'\ y_1 + y_2'\ y_2 + 2\gamma\sigma e_2y_2 + n_2\gamma^2\sigma^2. \tag{5.62}$$

By Theorem 10.2 of Section B of the Appendix we may complete the integration on α_1 and β in the first term of the right hand side of 5.56 to obtain

$$\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty L(\alpha_1\ \alpha_1+z'\beta-\gamma\sigma, \beta, \sigma)\ d\alpha_1 d\beta d\sigma$$

$$= (2\pi)^{-\frac{n_1+n_2-p-1}{2}} \int_0^\infty \sigma^{-(n_1+n_2-p-1)} \exp[-\frac{1}{2\sigma^2}(a_1+a_2)]\ d\sigma \tag{5.63}$$

where

$$a_1 = y_1' y_1 + y_2' y_2 + 2\gamma\sigma e_2' y_2 + n_2\gamma^2\sigma^2 \tag{5.64}$$

$$a_2 = -(u_1+u_2\gamma\sigma)' A^{-1} (u_1+u_2\gamma\sigma) \tag{5.65}$$

We write a_1+a_2 as a quadratic expression in σ^2,

$$a_1+a_2 = \gamma^2 u\sigma^2 + 2\gamma v\sigma + w \tag{5.66}$$

where

$$\begin{aligned} u &= n_2 - u_2' A^{-1}u_2 \\ v &= e_2' y_2 - u_2' A^{-1}u_1 \\ w &= y_1' y_1 + y_2'y_2 - u_1' A^{-1}u_1 \end{aligned} \tag{5.67}$$

By Lemma 10.3 of Section C of the Appendix we may now write 5.63 as

$$k_3 \exp[-\frac{\gamma^2 u}{2}]\cdot\Big[F(\frac{n_1+n_2-p-2}{2};\ \frac{1}{2};\ \frac{\gamma^2 v^2}{2w}) + \frac{\sqrt{2}\gamma v}{\sqrt{w}} \frac{\Gamma(\frac{n_1+n_2-p-1}{2})}{\Gamma(\frac{n_1+n_2-p-2}{2})} F(\frac{n_1+n_2-p-1}{2};\ \frac{3}{2};\ \frac{\gamma^2 v^2}{2w})\Big] \tag{5.68}$$

where

$$k_3 = \Pi^{-\frac{n_1+n_2-p-1}{2}} w^{-\frac{n_1+n_2-p-2}{2}} 2^{-\frac{n_1+n_2-p-4}{2}}.$$

By replacing γ by $-\gamma$ in the derivation of relations 5.63-5.68, the second term of the right hand side of the equality sign in 5.56,

$$\int_0^\infty\int_{-\infty}^\infty \cdots \int_{-\infty}^\infty L(\alpha_1,\ \alpha_1+z'\beta+\gamma\sigma,\ \beta,\ \sigma)\ d\alpha_1 d\beta d\sigma, \tag{5.69}$$

can be shown to be equal to

$$k_3 \exp[-\frac{\gamma^2 u}{2}] \Big[F(\frac{n_1+n_2-p-2}{2};\ \frac{1}{2};\ \frac{\gamma^2 v^2}{2w}) - \frac{2\gamma v}{w} \frac{\Gamma(\frac{n_1+n_2-p-1}{2})}{\Gamma(\frac{n_1+n_2-p-2}{2})}.$$

$$F(\frac{n_1+n_2-p-1}{2};\ \frac{3}{2};\ \frac{\gamma^2 v^2}{2w})\Big]. \qquad (5.70)$$

Adding 5.68 and 5.70 we then have

$$R_p(\gamma) = k_3 \exp[-\frac{\gamma^2 u}{2}]\ F(\frac{n_1+n_2-p-2}{2};\ \frac{1}{2};\ \frac{\gamma^2 v}{2w}) \qquad (5.71)$$

where k_3 is as defined in 5.68 so that, from 5.55, we finally write

$$R(p) = \exp[-\frac{(\gamma_2^2-\gamma_1^2)}{2}u] \frac{F(\frac{n_1+n_2-p-2}{2};\ \frac{1}{2};\ \frac{\gamma_2^2 v^2}{2w})}{F(\frac{n_1+n_2-p-2}{2};\ \frac{1}{2};\ \frac{\gamma_1^2 v^2}{2w})} \qquad (5.72)$$

where u, v, and w are given by 5.67.

3. R(p) as a generalization of previous results

a. Least squares estimates We now demonstrate that 5.72 is the generalization of Hajnal's t^2-test which would be expected if p covariates were used. We begin by deriving least squares estimates of α_1, α_2, β, and σ^2, assuming the model 5.48. For this derivation we write the model 5.48 in vector form consistent with the notation of 5.53 and 5.54. Thus

$$\begin{aligned} y_1 &= \alpha_1 e_1 + X_1'\beta + \varepsilon_1 \\ y_2 &= \alpha_2 e_2 + X_2'\beta + \varepsilon_2 \end{aligned} \qquad (5.73)$$

where we assume X_1 and X_2 are controlled, $\epsilon_1 = (\epsilon_{11}, \epsilon_{12}, \ldots, \epsilon_{1n_1})$ and $\epsilon_2' = (\epsilon_{21}, \epsilon_{22}, \ldots, \epsilon_{2n_2})$. Combining both equations we then have

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} e_1 & \tilde{0} & X_1' \\ \tilde{0} & e_2 & X_2' \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{pmatrix} + \epsilon \qquad (5.74)$$

where y_1, y_2, e_1, e_2, X_1, and X_2 are given by 5.54 and $\epsilon = (\epsilon_1, \epsilon_2)$. We shall denote a 1 x m vector of zeros by $\tilde{0}'$, the size of m being determined by the context in which $\tilde{0}$ is found.

The least squares estimates required are easily shown to be

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} e_1' e_1 & 0 & \tilde{0}' \\ 0 & e_2' e_2 & \tilde{0}' \\ \tilde{0} & \tilde{0} & X_1X_1'+X_2X_2' \end{pmatrix}^{-1} \begin{pmatrix} e_1' y_1 \\ e_2' y_2 \\ X_1y_1+X_2y_2 \end{pmatrix} \qquad (5.75)$$

$$= \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ (X_1X_1'+X_2X_2')^{-1}(X_1y_1+X_2y_2) \end{pmatrix} \qquad (5.76)$$

and

$$\hat{\sigma}^2 = \frac{1}{n_1+n_2-p-2}[y_1'y_1+y_2'y_2-\hat{\alpha}_1e_1'y_1-\hat{\alpha}_2e_2'y_2-\hat{\beta}'(X_1y_1+X_2y_2)] \qquad (5.77)$$

where by definition we have $e_1'e_1 = n_1$, $e_2'e_2 = n_2$, $e_1'y_1 = n_1\bar{y}_1$, and $e_2'y_2 = n_2\bar{y}_2$. The variance-covariance matrix of $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}')'$ is

$$\begin{pmatrix} \frac{1}{n_1} & 0 & 0 \\ 0 & \frac{1}{n_2} & \tilde{0}' \\ 0 & \tilde{0} & (X_1X_1' + X_2X_2')^{-1} \end{pmatrix} \sigma^2 . \qquad (5.78)$$

If we denote the difference between treatments, $\alpha_1 - \alpha_2 + z'\beta$, by Δ we then have

$$\hat{\Delta} = \hat{\alpha}_1 - \hat{\alpha}_2 + z'\hat{\beta} \qquad (5.79)$$

and

$$\mathrm{Var}(\hat{\Delta}) = [\frac{1}{n_1} + \frac{1}{n_2} + z'(X_1X_1' + X_2X_2')^{-1} z]\ \sigma^2. \qquad (5.80)$$

b. u, v, and w as functions of the least squares estimates In order to establish R(p), as given by 5.72, as a generalization of Hajnal's t^2-test when p covariates are used we require a well known matrix result. This result is used extensively in the remainder of the chapter and a proof may be found in, for example, Anderson (1, p. 344).

Theorem 5.1: If the positive definite matrix M is partitioned as

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

so that M_{11} is square and M_{22} is nonsingular, then the determinant of M, $|M|$, can be written

$$|M| = |M_{11} - M_{12}M_{22}^{-1}M_{21}|\ |M_{22}| .$$

Recalling the definitions of u, v, and w, given by 5.67, we must show that

$$u = \frac{\sigma^2}{\text{Var}(\hat{\Delta})} \tag{5.81}$$

and

$$\frac{\gamma^2}{2}\,\frac{v^2}{w} = \frac{\gamma^2 u}{2}\;\frac{t^2_{n_1+n_2-p-2}}{n_1+n_2-p-2 + t^2_{n_1+n_2-p-2}} \tag{5.82}$$

where

$$t^2_{n_1+n_2-p-2} = \frac{(\bar{y}_1-\bar{y}_2+z'\hat{\beta})'u}{\hat{\sigma}^2}$$

is the fixed-sample size Student t^2-test statistic for testing the hypothesis formulation

$$\begin{aligned} H_0'&:\ |\Delta| = 0 \\ H_1'&:\ |\Delta| \neq 0 \end{aligned} \tag{5.83}$$

1) $\underline{u}$ To establish 5.81 we take advantage of the fact that $u = n_2 - u_2' A^{-1} u_2$ is a scalar, where u_2 is defined by 5.61, so that

$$n_2 - u_2' A^{-1} u_2 = |n_2 - u_2' A^{-1} u_2|.$$

By Theorem 5.1 we have

$$|n_2 - u_2' A^{-1} u_2| = \frac{|A|\,|n_2 - u_2' A^{-1} u_2|}{|A|}$$

$$= \frac{\begin{vmatrix} A & u_2' \\ u_2 & n_2 \end{vmatrix}}{|A|}. \tag{5.85}$$

By definition (5.60) the determinant of A, A , is

$$|A| = \begin{vmatrix} n_1+n_2 & n_2 z' \\ z n_2 & X_1X_1' + X_2X_2' + n_2 zz' \end{vmatrix} \tag{5.86}$$

so that if, in turn, the first row of $|A|$ is multiplied by the first component, z_1, of $z' = (z_1, \ldots, z_p)$ and the result subtracted from the second row of $|A|$, the first row of $|A|$ is multiplied by z_2 and the result subtracted from the third row of $|A|$, and so on until the first row of $|A|$ is multiplied by z_p and the result subtracted from the p+1 row of $|A|$, we may write

$$|A| = \begin{vmatrix} n_1+n_2 & n_2 z' \\ -n_1 z & X_1X_1' + X_2X_2' \end{vmatrix}. \tag{5.87}$$

For convenience we might have described the above row operations on $|A|$ by saying that we premultiplied the first row of $|A|$ by z and subtracted the resultant matrix (array) from the matrix (array) formed by the last p rows of $|A|$. In the future, where possible, we shall abridge the description of any set of row and/or column operations in just this manner. Using Theorem 5.1 again, 5.87 becomes

$$|A| = |X_1X_1' + X_2X_2'|\ |n_1+n_2+n_1n_2z'(X_1X_1' + X_2X_2')^{-1}z|. \tag{5.88}$$

To simplify the numerator of 5.85, we have by definition

$$\begin{vmatrix} A_2 & u_2 \\ u_2' & n_2 \end{vmatrix} = \begin{vmatrix} n_1+n_2 & n_2 z' & n_2 \\ zn_2 & X_1X_1' + X_2X_2' + n_2zz' & n_2 z \\ n_2 & n_2 z' & n_2 \end{vmatrix} \tag{5.89}$$

so that, subtracting the last row from the first and subtracting the matrix formed by premultiplying the last row by the vector z from the matrix defined by the 2^{nd} through the (p+1) rows of 5.89, we may write

$$\begin{vmatrix} A & u_2 \\ u_2' & n_2 \end{vmatrix} = \begin{vmatrix} n_1 & \tilde{0}' & 0 \\ \tilde{0} & X_1X_1' + X_2X_2' & \tilde{0} \\ n_2 & n_2 z' & n_2 \end{vmatrix}$$

$$= n_1 \begin{vmatrix} X_1X_1' + X_2X_2' & \tilde{0} \\ n_2 z' & n_2 \end{vmatrix}$$

$$= n_1 n_2 \left| X_1X_1' + X_2X_2' \right| \qquad (5.90)$$

Thus, finally, the ratio of quantities given in 5.90 and 5.88 yields

$$n_2 - u_2' A^{-1} u_2 = \frac{1}{\frac{1}{n_1} + \frac{1}{n_2} + z'(X_1X_1' + X_2X_2')^{-1} z} \qquad (5.91)$$

which is $\frac{\sigma^2}{\mathrm{Var}(\hat{\Delta})}$ as in 5.80.

2) $\underline{v}$ We now establish that v, defined in 5.67, is

$$v = (\bar{y}_1 - \bar{y}_2 + \beta' z)\, u \qquad (5.92)$$

Since v is a scalar,

$$v \equiv e_2' y_2 - u_2' A^{-1} u_1 = \left| e_2' y_2 - u_2' A^{-1} u_1 \right| \qquad (5.93)$$

so that Theorem 5.1 allows us to write

$$v = \frac{\begin{vmatrix} A & u_2' \\ u_1 & e_2' y_2 \end{vmatrix}}{|A|}. \qquad (5.94)$$

By definition,

$$\begin{vmatrix} A & u_2' \\ u_2 & e_2' y_2 \end{vmatrix} = \begin{vmatrix} n_1+n_2 & n_2 z' & n_2 \\ n_2 z & X_1X_1'+X_2X_2'+n_2 zz' & n_2 z \\ n_1\bar{y}_1+n_2\bar{y}_2 & y_1'X_1'+y_2'X_2'+n_2\bar{y}_2 z' & n_2\bar{y}_2 \end{vmatrix} \qquad (5.95)$$

so that if we, in sequence,

i) subtract the last column from the first column,

ii) postmultiply the last column by the vector z and subtract the resultant matrix of elements from the matrix formed by column 2 through column p+1,

iii) multiply the first row of the resultant determinant by $\bar{y}_2$ and subtract it from the last row,

then 5.95 can be written as

$$\begin{vmatrix} n_1 & \tilde{0}' & n_2 \\ \tilde{0} & X_1X_1' + X_2X_2' & n_2 z \\ n_1(\bar{y}_1-\bar{y}_2) & y_1X_1' + y_2X_2' & 0 \end{vmatrix}. \qquad (5.96)$$

We expand 5.96 by cofactors and have

$$n_1 \begin{vmatrix} X_1X_1' + X_2X_2' & n_2 z \\ y_1'X_1' + y_2'X_2' & 0 \end{vmatrix} - n_1(\bar{y}_1-\bar{y}_2)\begin{vmatrix} \tilde{0}' & n_2 \\ X_1X_1' + X_2X_2' & n_2 z \end{vmatrix} \qquad (5.97)$$

and by Theorem 5.1

$$= n_1 |X_1X_1' + X_2X_2'| \; [0+n_2(y_1'X_1' + y_2'X_2')(X_1X_1' + X_2X_2')^{-1}z]$$

$$+ n_1n_2(\bar{y}_1-\bar{y}_2) |X_1X_1' + X_2X_2'| \qquad (5.98)$$

$$= n_1n_2 |X_1X_1' + X_2X_2'| \; (\bar{y}_1-\bar{y}_2 + \hat{\beta}'z). \qquad (5.99)$$

Since we have already shown (see 5.88)

$$|A| = n_1n_2 |X_1X_1' + X_2X_2'| \; |n_1+n_2+n_n n_2 z'(X_1X_1' + X_2X_2')^{-1}z|$$

then v as given in 5.94 becomes

$$v = (\bar{y}_1 - \bar{y}_2 + z'\hat{\beta})u \tag{5.100}$$

as was to be demonstrated.

3) $\underline{w}$ Finally it is necessary to evaluate w, where

$$w = y_1'y_1 + y_2'y_2 - u_1'A^{-1}u_1. \tag{5.101}$$

Recalling the definitions of A and u_1 given by 5.60 and 5.61 respectively and using Theorem 5.1 we may write

$$|w| = \frac{\begin{vmatrix} n_1+n_2 & n_2z' & n_1\bar{y}_1 + n_2\bar{y}_2 \\ n_2z & X_1X_1' + X_2X_2' + n_2zz' & X_1y_1 + X_2y_2 + n_2\bar{y}_2z \\ n_1\bar{y}_1+n_2\bar{y}_2 & y_1'X_1' + y_2'X_2' + n_2\bar{y}_2z' & y_1'y_1 + y_2'y_2 \end{vmatrix}}{|A|}. \tag{5.102}$$

The numerator of 5.102 can be written as

$$\begin{vmatrix} n_1+n_2 & n_2z' & -n_2(\bar{y}_1-\bar{y}_2) \\ -n_1z & X_1X_1' + X_2X_2' & X_1y_1 + X_2y_2 \\ n_1(\bar{y}_1-\bar{y}_2) & y_1'X_1' + y_2'X_1' & y_1'y_1 + y_2'y_2 - n_1\bar{y}_1^2 - n_2\bar{y}_2^2 \end{vmatrix} \tag{5.103}$$

by performing the following operations:

i) Multiply row 1 by $\bar{y}_2$ and subtract the result from the last row and

ii) premultiply row 1 by z and subtract the resulting array from the matrix of elements defined by row 2 through row p+1.

In addition, if we premultiply the matrix of elements defined by row 2 through row p+1 of the determinant in 5.103 by

$(X_1y_1 + X_2y_2)'(X_1X_1' + X_2X_2')^{-1}$, which is $\hat{\beta}'$, and subtract the resulting row vector from the last row, then 5.103 becomes

$$\begin{vmatrix} n_1+n_2 & n_2z & -n_2(\bar{y}_1-\bar{y}_2) \\ -n_1z & X_1X_1'+X_2X_2' & X_1y_1+X_2y_2 \\ n_1(\bar{y}_1-\bar{y}_2)+n_1\beta'z & \tilde{0}' & y_1'y_1+y_2'y_2-\hat{\alpha}_1\Sigma y_1-\hat{\alpha}_2\Sigma y_2-\hat{\beta}(X_1y_1+X_2y_2) \end{vmatrix}. \tag{5.104}$$

If we expand 5.104 by cofactors we have

$$[y_1'y_1+y_2'y_2-\hat{\alpha}_1\Sigma y_1-\hat{\alpha}_2\Sigma y_2-\hat{\beta}'(X_1y_1+X_2y_2)]\ |A|$$

$$-\ n_1(\bar{y}_1-\bar{y}_2+\hat{\beta}'z)\begin{vmatrix} n_2z & -n_2(\bar{y}_1-\bar{y}_2) \\ X_1X_1'+X_2X_2' & X_1y_1+X_2y_2 \end{vmatrix} \tag{5.105}$$

which, by application of Theorem 5.1,

$$= (n_1+n_2-p-2)\hat{\sigma}^2|A| + n_1n_2(\bar{y}_1-\bar{y}_2+\hat{\beta}'z)^2|X_1X_1'+X_2X_2'| \tag{5.106}$$

where $\hat{\sigma}^2$ is defined by 5.77. Thus $|w|$, as defined by 5.102, equals

$$(n_1+n_2-p-2)\hat{\sigma}^2 + (\bar{y}_1-\bar{y}_2+\hat{\beta}'z)^2u. \tag{5.107}$$

c. <u>Evaluation of $\gamma^2v^2/2w$</u> From 5.82, 5.91, and 5.107 we may now write

$$\frac{\gamma^2}{2}\frac{v^2}{w} = \frac{\gamma^2}{2}\frac{(\bar{y}_1-\bar{y}_2+\hat{\beta}'z)^2u^2}{(n_1+n_2-p-2)\hat{\sigma}^2 + (\bar{y}_1-\bar{y}_2+\hat{\beta}'z)^2u} \tag{5.108}$$

$$= \frac{\gamma^2u}{2}\ \frac{\dfrac{(\bar{y}_1-\bar{y}_2+\hat{\beta}'z)^2u}{\hat{\sigma}^2}}{n_1+n_2-p-2+\dfrac{(\bar{y}_1-\bar{y}_2+\hat{\beta}'z)^2u}{\hat{\sigma}^2}}$$

$$= \frac{\gamma^2 u}{2} \frac{t^2_{n_1+n_2-p-2}}{n_1+n_2-p-2 + t^2_{n_1+n_2-p-2}} \tag{5.109}$$

and thus conclude that 5.82 holds.

If, for convenience, we write $t^2 = t^2_{n_1+n_2-p-2}$ then R(p) may be written as

$$R(p) = \exp[-\frac{(\gamma_2^2-\gamma_1^2)}{2} u] \frac{F(\frac{n_1+n_2-p-2}{2}; \frac{1}{2}; \frac{\gamma_2^2 u}{2} \frac{t^2}{n_1+n_2-p-2+t^2})}{F(\frac{n_1+n_2-p-2}{2}; \frac{1}{2}; \frac{\gamma_1^2 u}{2} \frac{t^2}{n_1+n_2-p-2+t^2})}. \tag{5.110}$$

Since the tests in this chapter are sequential t^2-tests we may conclude that they terminate with probability one (see David and Kruskal, 22). However, no average sample number or operating characteristic formulae are available, and because of the complexity of the mathematics involved, it appears at present unlikely that theoretical analysis will be successful on these points. We note, however, that the pros and cons of paired versus independent samples in fixed sample size experimentation also pertain when the context is sequential experimentation. In particular, the gain in degrees of freedom for estimating the standard error of the difference of two means using two independent samples is oftentimes offset by the advantages of pairing when there is a high positive correlation between observations within pairs.

VI. NUMERICAL INVESTIGATIONS

This chapter contains the results of a sampling experiment on three of the test statistics proposed in Chapter III. Before detailing the implementation of these test statistics and discussing the experimental results we shall outline the Monte Carlo procedure.

A. Monte Carlo Procedure

The Monte Carlo procedure associated with many empirical sequential trials is as follows. At each stage an observation is generated in a random manner from a specified distribution. We have a test statistic which is a function of all observations accumulated at any stage. This statistic is compared with the decision boundaries specified by the procedure. If the decision is to stop sampling and accept either H_0 or H_1, the stage number is recorded and the reason for curtailment is noted. If, on the other hand, the decision is to continue sampling, another observation is independently generated and we again compute the test statistic and compare it with the decision boundaries. We shall call the sequence of observations leading to a decision a trial.

In sequential analysis, even though a sequential procedure may decision with probability one, it is not a rare occurrence that a particular trial does not decision until a large number of observations have been taken. Accordingly, in order to utilize the available computer funds to best advantage in our empirical investigation, it was considered desirable to set an upper limit on the number of observations per trial.

As a time-saving device, we found it useful to compute simultaneously at each stage some or all of the test statistics under consideration, realizing then that for each trial the stage numbers for decision corresponding to each statistic would not be independent. Also, the computer program was written so that all of the statistics would be simultaneously computed until each had decisioned on any or all of three distinct boundary pairs, the boundary pairs specified before each set of trials was performed. In these cases, then, a trial consists of the sequential generation of observations resulting in one to six test statistics (three test statistics with and without covariance) decisioning on one to three boundary pairs or reaching the upper limit on total observations per trial as discussed earlier.

We shall call a set of trials a run. On each run we recorded the number of decisions, the number of correct decisions, the observed average sample number to decision, and the observed standard deviation of the sample number to decision for each statistic on each boundary pair.

We limited ourselves to an evaluation of the relative merits of using one covariate in addition to the variate of interest in the analysis. In accordance with the basic assumptions of Chapter III we generated bivariate normal data with prespecified means, variances, and correlation coefficient. This was accomplished by having the IBM 360/65 computer generate two independent uniformly distributed variates, say u_1, and u_2, over the interval (0, 1) and then using the following transformations, given by Box and Muller (13b), to obtain two independent univariate normal variates, z_1 and z_2, each having mean zero and variance one;

$$z_1 = (-2 \log u_1)^{\frac{1}{2}} \cos (2\pi u_2)$$

$$z_2 = (-2 \log u_1)^{\frac{1}{2}} \sin (2\pi u_2).$$

Specifying α, β_1, σ, and σ_{x_1} we then formed the variates,

$$e = \sigma z_1$$

$$x_1 = \sigma_{x_1} z_2$$

$$y_1 = \alpha + \beta_1 x_1 + e,$$

so that (x_1, y_1) is bivariately normally distributed with $\mu_{x_1} = 0$, $\mu_{y_1} = \alpha$, $Var(x_1) = \sigma^2_{x_1}$, $Var(y_1) = \beta_1^2\sigma^2_{x_1} + \sigma^2$, and $cov(x_1, y_1) = \beta\sigma^2_{x_1}$.

To summarize, after specifying a parameter configuration and a hypothesis formulation, we generated pairs (x, y) of observations and calculated test statistics for use in a sequential procedure. From repeated independent trials, we recorded estimates of Type I or Type II error, of expected sample numbers, and of sample number variances.

B. Implementation of Selected Test Statistics

These sampling experiments were primarily designed to investigate the advantages of utilizing concomitant information in sequential test procedures. Monte Carlo results were obtained for each of the test statistics R(1, p, n), R(2, p, n) and R(5, p, n) given, respectively, by 3.52, 3.70, and 3.120-3.121. In particular, we compared the sampling results when one covariate is used in the analysis with those obtained when the covariate was ignored. We now present the form of each statistic under consideration as it was coded for computer execution, and we shall point out several timesaving devices and approximations that were found useful.

R(1, p, n), in general form, is given by 3.52 and is presented here for zero covariates as

$$R(1, 0, n) = \left[\frac{\sum_1^n (y_i - \alpha_T)^2}{\sum_1^n (y_i - \alpha_A)^2}\right]^{\frac{n-1}{2}}, \qquad (6.3)$$

and for one covariate as,

$$R(1, 1, n) = \frac{\sum_1^n x_i^2 \sum_1^n (y_i - \alpha_T)^2 - \left[\sum_1^n (y_i - \alpha_T) x_i\right]^2}{\sum_1^n x_i^2 \sum_1^n (y_i - \alpha_A)^2 - \left[\sum_1^n (y_i - \alpha_A) x_i\right]^2}. \qquad (6.4)$$

We next present R(5, p, n) in general form for the cases when n-p is odd and when n-p is even. If n-p is odd then

R(5, p, n)

$$= A(n,\ p)\ \frac{2\sqrt{d} - I(\alpha_A) \displaystyle\sum_{r=0}^{\frac{n-p-3}{2}} G'(r)\left(\frac{4df(\alpha_A)-g^2(\alpha_A)}{dK(\alpha_A)}\right)^r}{2\sqrt{d} + I(\alpha_T) \displaystyle\sum_{r=0}^{\frac{n-p-3}{2}} G'(r)\left(\frac{4df(\alpha_T)-g^2(\alpha_T)}{dK(\alpha_T)}\right)^r} \qquad (6.5)$$

and if n-p is even then

R(5, p, n)

$$= A(n,\ p)\ \frac{\pi - 2\theta_A - J(\alpha_A) \displaystyle\sum_{r=1}^{\frac{n-p-2}{2}} H(r)\left(\frac{4df(\alpha_A)-g^2(\alpha_A)}{dK(\alpha_A)}\right)^r}{\pi + 2\theta_T + J(\alpha_T) \displaystyle\sum_{r=1}^{\frac{n-p-2}{2}} H(r)\left(\frac{4df(\alpha_T)-g^2(\alpha_T)}{dK(\alpha_T)}\right)^r} \qquad (6.6)$$

where in both cases

$$A(n,\ p) = \left[\frac{4df(\alpha_T)-g^2(\alpha_T)}{4df(\alpha_A)-g^2(\alpha_A)}\right]^{\frac{n-p-1}{2}}$$

$$I(\alpha) = \frac{2d + g(\alpha)}{(d\alpha^2 + g(\alpha)\alpha + f(\alpha))^{\frac{1}{2}}},\ \alpha = \alpha_T,\ \alpha_A$$

$$J(\alpha) = \frac{2d + g(\alpha)}{(4df(\alpha)-g^2(\alpha))^{\frac{1}{2}}},\ \alpha = \alpha_T,\ \alpha_A$$

$d = k + e'(I-m)e$, k any prespecified function of n

$g(\alpha) = -2k\alpha - 2e'(I-M)y,\ \alpha = \alpha_T,\ \alpha_A$

$$f(\alpha) = k\alpha^2 + y'(I-M)y, \quad \alpha = \alpha_T, \alpha_A \tag{6.7}$$

$$K(\alpha) = d\alpha^2 + g(\alpha)\alpha + f(\alpha), \quad \alpha = \alpha_T, \alpha_A$$

$$G'(r) = \frac{(2r)!}{2^{4r}(r!)^2}$$

$$H(r) = \frac{(r-1)!\ r!}{(2r)!}$$

$$\theta_A = \text{principal value of } \tan^{-1} \frac{2d\alpha_A + g(\alpha_A)}{\sqrt{4df(\alpha_A) - g^2(\alpha_A)}}$$

$$\theta_T = \text{principal value of } \tan^{-1} \frac{2d\alpha_T + g(\alpha_T)}{\sqrt{4df(\alpha_T) - g^2(\alpha_T)}}.$$

When zero covariates are used, i.e. when $p = 0$, then d, $g(\alpha)$ and $f(\alpha)$ are given by

$$d = k + n$$

$$g(\alpha) = -2k\alpha - 2\sum_1^n y_i \tag{6.8}$$

$$f(\alpha) = k\alpha^2 + \sum_1^n y_i^2$$

and when one covariate is used, $p = 1$, then

$$d = k + n - \frac{\left(\sum_1^n x_i\right)^2}{\sum_1^n x_i^2}$$

$$g(\alpha) = -2k\alpha - 2\,\frac{\sum_1^n x_i^2 \sum_1^n y_i - \sum_1^n x_i \sum_1^n x_i y_i}{\sum_1^n x_i^2} \tag{6.9}$$

$$f(\alpha) = k\alpha^2 + \frac{\sum_1^n x_i^2 \sum_1^n y_i^2 - (\sum_1^n x_i y_i)^2}{\sum_1^n x_i^2}.$$

Rather than computing the quantities G'(r) and H(r) each time they were required, two arrays,

$$G'(0),\ G'(1),\ \ldots,\ G'(40)$$
$$H(1),\ \ldots,\ H(40),$$

were constructed and stored in the computer prior to each run by using the following recursive formulae:

$$G'(r) = \frac{2r-1}{8r}\,G'(r-1)$$
$$H(r) = \frac{r-1}{2(2r-1)}\,H(r). \qquad (6.10)$$

The appropriate elements from each array were then called when required.

The statistic

$$R(2, p, n)$$
$$= \exp[-\frac{\gamma^2 e'(I-M)e}{2}][F(\frac{n-p-1}{2};\ \frac{1}{2};\ \frac{\gamma^2}{2}\,\frac{e'(I-M)z}{z'(I-M)z})$$
$$+ \frac{2\gamma e'(I-M)z}{(2z'(I-M)z)^{\frac{1}{2}}}\,\frac{\Gamma(\frac{n-p}{2})}{\Gamma(\frac{n-p-1}{2})}\,F(\frac{n-p}{2};\ \frac{3}{2};\ \frac{\gamma^2}{2}\,\frac{(e'(I-M)z)^2}{z'(I-M)z})] \qquad (6.11)$$

where, if $p = 0$ and $z' = (y_1-\alpha_T,\ y_2-\alpha_T,\ \ldots,\ y_n-\alpha_T)$,

$$e'(I-M)z = \sum_1^n (y_i-\alpha_T)$$

$$z'(I-M)z = \sum_1^n (y_i-\alpha_T)^2 \qquad (6.12)$$

$$e'(I-M)e = n$$

and, if p = 1,

$$e'(I-M)z = \frac{\sum_1^n x_i^2 \sum_1^n (y_i - \alpha_T) - \sum_1^n x_i \sum_1^n x_i (y_i - \alpha_T)}{\sum_1^n x_i^2}$$

$$z'(I-M)z = \frac{\sum_1^n x_i^2 \sum_1^n (y_i - \alpha_T)^2 - \left[\sum_1^n x_i (y_i - \alpha_T)\right]^2}{\sum_1^n x_i^2} \tag{6.13}$$

$$e'(I-M)e = \frac{n\sum_1^n x_i^2 - (\sum_1^n x_i)^2}{\sum_1^n x_i^2},$$

presented some computing difficulties because of the confluent hypergeometric F(r; s; x). Since F(r; s; x) is an infinite series, certain approximations were necessary for computer implementation. We followed Olin's (49) recommendations on this point. Kummer's identity,

$$F(r;\ s;\ x) = \exp(-x)\ F(s-r;\ s;\ -x)$$

allows F(r; s; x) to be written as the product of a more rapidly converging series exp(-x) and the possibly finite series F(s-r; s; -x). The series F(s-r; s; -x) is finite provided s-r is an integer such that $s \leq r$. In the computer subroutine for approximating F(r; s; x) the series was truncated either

1) naturally when $s-r(s \leq r)$ was an integer, or
2) artificially when the value in the last term in the series was less than 5×10^{-9}.

Olin (49) reported that the above procedure would result in approximations of $F(r; s; x)$ accurate to six decimal places. Also, $R(2, p, n)$, as given by 6.10, involves two confluent hypergeometric functions with the third argument in common. However, by the use of the identity,

$$sF(r; s; x) = sF(r-1; s; x) + xF(r; s+1; x),$$

given in Section C of the Appendix, we may rewrite 6.11 so that it is a function of three confluent hypergeometric functions, all having the last two arguments in common. That is, $R(2, p, n)$ may be written as

$$\begin{aligned} R(2,\ p,\ n) \\ = \exp\left[-\frac{\gamma^2 e'(I-M)e}{2}\right]\Bigg[F\left(\frac{n-p-1}{2};\ \frac{1}{2};\ \frac{\gamma^2}{2}\ \frac{e'(I-M)z}{z'(I-M)z}\right) \\ + \frac{2\gamma z'(I-M)z}{e'(I-M)z}\ \frac{\Gamma(\frac{n-p}{2})}{\Gamma(\frac{n-p-1}{2})}\left[F\left(\frac{n-p}{2};\ \frac{1}{2};\ \frac{\gamma^2}{2}\ \frac{(e'(I-M)z)^2}{z'(I-M)z}\right)\right. \qquad (6.14) \\ \left. - F\left(\frac{n-p-2}{2};\ \frac{1}{2};\ \frac{\gamma^2}{2}\ \frac{(e'(I-M)z)^2}{z'(I-M)z}\right)\right]\Bigg] . \end{aligned}$$

C. Empirical Results

We begin by giving some results that will provide guidance for what we might anticipate as the expected sample number saving when concomitant information is utilized. In Section B of Chapter III we present average sample number formulae, 3.34, for a Waldian SPRT when the underlying distribution has the p.d.f. 3.23 and when H_o and H_1 are both simple hypothesis specifications (that is, when all the nuisance parameters are assumed known and the parameter of interest is assumed to be one of only two values). The expected sample numbers under H_o and H_1, $E[N|\alpha_o]$ and $E[N|\alpha_1]$ are seen to depend on the parameter σ^2, the distance $|\alpha_1 - \alpha_o|$ and the specified Type I and Type II error rates, P(I) and P(II). If the concomitant information is ignored, then by Corollary 10.3 of Section A we see that the operative conditional variance of y given x, σ^2, becomes effectively the unconditional variance $\sigma^2 + \beta'\Sigma\beta$. Thus the ratio of $E[N|\alpha_o]_c$, the expected sample number with one covariate, to $E[N|\alpha_o]_{wc}$, the expected sample number without covariance, is

$$\frac{\sigma^2}{\sigma^2 + \beta_1^2\sigma_{x_1}^2}. \tag{6.15}$$

It can easily be shown that the ratio given by 6.15 is equal to $1-\rho^2$ where ρ is the correlation coefficient of y and x_1. A corresponding theoretical result for the composite hypothesis situation with nuisance parameters unknown has not, however, been obtained. It can be seen from Tables 2, 4, 5, 6, and 7 that the ratio of the observed average sample number with covariance to the observed average sample number without covariance is always larger than $1-\rho^2$. These results are in line with the previous evaluative experiments of Roseberry (58) and Olin (49). That is, use of one covariate in the analysis resulted in less than a ρ^2 x 100% economy in observations. For example, in Table 5 we have that $1-\rho^2 = 0.36$ and the ratios given there range from 0.43 to 0.56. In Table 4 three columns of ratios are given that are associated, from left to right respectively, with the values 0.53, 0.30, and 0.19 of $1-\rho^2$. We see that none of these values have been realized in the empirical results. However, we did have a substantial economy of observations when $\rho \geq 0.6$. This can be seen by again examining the ratio column given in Tables 4, 5, and 7. To summarize, although we did not achieve the ρ^2 x 100% economy that would be expected in the SPRT of the uncomplicated simple versus simple hypothesis formulation, we did have significant economy in the number of observations by the test statistics that utilize concomitant information.

In the development of R(5, p, n) in Section F of Chapter III we weighted α nonuniformly and the nuisance parameters uniformly over the parameter spaces under consideration. We allowed V_o and V_1, given by 3.112, a certain versatility or richness with the introduction of k which is referred to as a constant in the sense that it does not depend on the data but may be some function of n. By k, we can control the dispersion of the halfnormal weighting of α. We investigated the performance of R(5, p, n), p = 0, 1, for different k. Some results of this investigation are included in Table 2. Error rates most closely resembling those specified for this

Table 2. Performance of R(5, p, n) for different values of k, and parameter specifications: $\alpha_T = 0.0$, $\alpha_A = 1.0$, $\sigma^2 = 3.0$, $\sigma^2_{x_1} = 9.0$, $\rho = .6$, $P(I)^a = P(II) = .05$

p	k	Observed a.s.n.	Observed s.d.s.n.	Ratio[b]	Observed P(I)	Observed P(II)	Number[c] of Trials
				$\alpha = 0.0$			
0	n	19.63	14.66	.80	0.080	Not appropriate	150
1	n	15.65	9.57		0.067		150
0	$\sqrt{n}$	19.90	18.35	.85	0.152		100
1	$\sqrt{n}$	16.93	13.39		0.120		100
0	20	23.30	20.26	.71	0.160		50
1	20	16.50	13.17		0.060		50
0	10	19.28	14.49	.83	0.080		50
1	10	16.00	10.46		0.080		50
0	3	22.48	19.94	.65	0.140		50
1	3	14.70	10.68		0.180		50
				$\alpha = 1.0$			
0	n	18.38	12.94	.84	Not appropriate	0.060	200
1	n	15.39	9.29			0.020	200
0	$\sqrt{n}$	16.70	11.49	.93		0.040	100
1	$\sqrt{n}$	15.58	11.71			0.020	100
0	20	19.08	13.14	.78		0.020	50
1	20	14.84	6.50			0.020	50
0	10	22.16	17.77	.75		0.040	50
1	10	16.68	11.65			0.020	50
0	3	22.06	16.98	.68		0.080	50
1	3	14.96	8.30			0.160	50

[a]The specified Type I and Type II errors.

[b]Each entry in this column is the ratio of the observed a.s.n. for the test utilizing one covariate to the observed a.s.n. for the same test without covariance.

[c]Duplication of certain runs by mistake resulted in a larger number of trials for some tests than others.

Table 3. Performance of tests when $\alpha - \alpha_T$ and $\alpha - \alpha_A$ become large, parameter specifications: $\sigma^2 = 3.0$, $\sigma^2_{x_1} = 9.0$, $\beta_1 = 0.7698$, $\rho = .8$, $\alpha_T = 0.0$, $\alpha_A = 1.0$, $k = n$, $\gamma_o = .3464$, ${}^1\gamma_1 = .5774$, P(I)[a] = P(II) = .03, 50 trials

Test	a.s.n.	s.d.s.n.	Ratio[b]	Observed P(II)	Number[c] of Decisions
			$\alpha = 1.5$		
R(1, 0, n)	34.96	14.90	.48	0.00	48
R(1, 1, n)	16.48	4.67		0.00	50
R(2, 0, n)	33.12	11.84	.48	0.00	49
R(2, 1, n)	16.06	4.36		0.00	50
R(5, 0, n)[d]	25.12	16.52	.49	0.00	50
R(5, 1, n)[d]	12.28	3.79		0.00	50
			$\alpha = 11.0$		
R(1, 0, n)	40.84	1.66	.98	0.00	50
R(1, 1, n)	40.18	0.87		0.00	50
R(2, 0, n)	12.90	0.30	.72	0.00	50
R(2, 1, n)	9.32	0.55		0.00	50

[a]The specified Type I and Type II errors.

[b]The ratio of the observed a.s.n. for R(i, 1, n) to the observed a.s.n. for R(i, 0, n), i = 1, 2, 5.

[c]The trials that did not decision before n = 120 were disregarded.

[d]R(5, p, n) was not allowed to decision before n = 9.

Table 3. (Continued)

Test	a.s.n.	s.d.s.n.	Ratio[b]	Observed P(II)	Number[c] of Decisions
			$\alpha = 11.0$		
R(5, 0, n)[d]	9.00	0.00	1.00	0.00	50
R(5, 1, n)[d]	9.00	0.00		0.00	50
			$\alpha = -1000.0$	Observed P(I)	
R(1, 0, n)		No decisions			
R(1, 1, n)		No decisions			
R(2, 0, n)	10.00	0.0	1.0	0.00	50
R(2, 1, n)	8.18	0.482		0.00	50
R(5, 0, n)[d]	9.00	0.0	0.818	0.00	50
R(5, 1, n)[d]	9.02	0.141		0.00	50

Table 4. Observed a.s.n. for increasing ρ when $\alpha_T = 0.0$, $\alpha = \alpha_A = 1.0$, $\sigma^2 = 1.5$, $\sigma^2_{x_1} = 9.0$, $P(I)^a = P(II) = .05$, $k = n$

Test	Correlation Coefficient ρ						
	$\rho = 0.0$ Observed[b] a.s.n.	$\rho = .6871$ Observed[c] a.s.n.	Ratio[d]	$\rho = .8385$ Observed[c] a.s.n.	Ratio[d]	$\rho = .90$ Observed[e] a.s.n.	Ratio[d]
R(1, 0, n)	43.10		.66		.41		.32
R(1, 1, n)		28.61		17.76		13.65	
R(2, 0, n)	45.30		.66		.42		.30
R(2, 1, n)		29.85		18.83		13.48	
R(5, 0, n)	26.54		.73		.56		.46
R(5, 1, n)		19.47		14.88		12.19	

[a] P(I) = Type I error and P(II) = Type II error.

[b] Based on 235 trials.

[c] Based on 250 trials.

[d] The ratio of the observed a.s.n. of the covariance test statistic to the observed a.s.n. of the non-covariance test statistic.

[e] Based on 1000 trials.

Table 5. Observed a.s.n., s.d.s.n. and Type I and Type II error rates for the parameter specifications: $\alpha = \alpha_T = 0.0$, $\alpha_A = 1.0$, $\sigma^2 = 1.50$, $\sigma^2_{x_1} = 9.0$, $\beta_1 = 0.5445$, $\rho = 0.8$, $\gamma_o = .490$, $\gamma_1 = .8165$, $P(I)^a = P(II) = .03$, 250 trials, k = n

Test	Observed a.s.n.	Observed s.d.s.n.	Ratio[b]	Observed P(I)	Observed P(II)	Number[c] of Decisions
			$\alpha = 0.0$			
R(1, 0, n)	31.45	18.23	.50	0.020	Not appropriate	249
R(1, 1, n)	15.74	7.73		0.012		250
R(2, 0, n)	29.05	17.59	.44	0.012		248
R(2, 1, n)	12.89	6.87		0.016		250
R(5, 0, n)	22.12	16.27	.56	0.032		250
R(5, 1, n)	12.30	6.06		0.020		250
			$\alpha = 1.0$			
R(1, 0, n)	35.80	20.26	.43	Not appropriate	0.004	249
R(1, 1, n)	15.41	7.88			0.016	250
R(3, 0, n)	35.91	19.61	.45		0.020	248
R(3, 1, n)	16.01	7.31			0.016	250
R(5, 0, n)	23.57	16.87	.58		0.048	250
R(5, 1, n)	13.67	6.76			0.012	250

[a]The specified Type I and Type II errors.

[b]The ratio of the observed a.s.n. for R(i, 1, n) to the observed a.s.n. for R(i, 0, n), i = 1, 2, 5.

[c]The trials that did not decision before n = 120 were disregarded.

Table 6. Observed a.s.n., s.d.s.n., and Type I error rates for the parameter specifications: $\alpha = \alpha_T = 0.0$, $\alpha_A = 1.0$, $\sigma^2 = 1.50$, $\sigma^2_{x_1} = 9.0$, $\beta_1 = .3062$, $\rho = 0.6$, $\gamma_o = .6532$, $\gamma_1 = .8163$, $P(I)^a = P(II) = .05$, 250 trials, k = n

Test	Observed a.s.n.	Observed s.d.s.n.	Ratio[b]	Observed P(I)	Number of Decisions
R(1, 0, n)	17.19	9.20	.82	0.060	250
R(1, 1, n)	14.01	7.22		0.028	250
R(2, 0, n)	14.40	7.44	.91	0.076	250
R(2, 1, n)	13.04	5.72		0.032	250
R(5, 0, n)	15.05	9.07	.82	0.044	250
R(5, 1, n)	12.36	6.82		0.024	250

[a]The specified Type I and Type II error rates.

[b]The ratio of the observed a.s. for R(i, 1, n) to the observed a.s.n. for R(i, 0, n), i = 1, 2, 5.

Table 7. Observed a.s.n., s.d.s.n., and Type I error rates for the parameter specifications: $\alpha_T = 0$, $\alpha = \alpha_A = 1.0$, $\sigma^2 = 1.50$, $\sigma^2_{x_1} = 9.0$, $\beta_1 = .8428$, $\rho = 0.9$, $\gamma_o = .3560$, $\gamma_1{}^1 = .8163$, $P(I)^a = P(II) = .05$, $k = n$

Test	Observed a.s.n.	Observed s.d.s.n.	Ratio[b]	Observed P(II)	Number[c] of trials
R(1, 0, n)	43.10	23.86	.32	.063	222
R(1, 1, n)	13.60	5.75		.030	235
R(2, 0, n)	45.30	23.35	.31	.049	224
R(2, 1, n)	14.09	6.87		.026	235
R(5, 1, n)	26.54	21.75	.47	.126	230
R(5, 1, n)	12.39	6.25		.026	235

[a]The specified Type I and Type II error rates.

[b]Each entry in this column is the ratio of the observed a.s.n. for the test utilizing one covariate to the observed a.s.n. for the same test without covariance.

[c]The total possible number of trials is 235. If the test statistic did not decision before n = 120, the trial was disregarded.

particular configuration were achieved when $k = n$. However, any conclusions regarding these tables must be qualified in the sense that only a relatively few trials were run and the presentation of the approximate 95% confidence interval (.04, .06) for error rates specified as P(I) = P(II) = .05 would require 1900 trials. It was discovered that $R(5, p, n)$ has the tendency to decision at the wrong boundary more often than expected when $n - p$ is small. Several preliminary runs suggested that, in order to achieve the specified error rates, $R(5, p, n)$ should not be allowed to decision until 7 to 9 observations were taken. Although this censoring scheme was successful in most cases we have no theoretical substantiation for it. A censor number of seven was effective for small sample numbers (9-15); not allowing $R(5, p, n)$ to decision until stage 9 gave good results for larger sample numbers (20-35); see, for example, Tables 5 and 6. However, if either σ or $\sqrt{\sigma^2 + \beta_1^2\sigma^2}$ was too large relative to $|\alpha_T - \alpha_A|$, error rates much higher than expected were observed (see Table 7). A detailed examination of the output from which Table 7 was constructed revealed that of twenty-nine incorrect decisions, twelve occurred at stage 9 and three occurred at stage 10. Thus if we had not allowed $R(5, p, n)$ to decision until stage 11 we probably would have observed an error rate of about 0.06.

It may be noted from Tables 3, 4, 5, 6, and 7 that $R(5, p, n)$ almost always has an observed average sample number less than those of either $R(1, p, n)$ or $R(2, p, n)$. The difference is not pronounced, however, unless the sample numbers are large in general. An unqualified endorsement of $R(5, p, n)$ should not be made, however, until the possibility of real discrepencies from the specified error rates is resolved.

When $p = 0$, $R(2, p, n)$ reduces to the sequential t-test as put forward by Wald (69), and differs slightly from that introduced by Rushton (59). Cornfield (17) criticized the sequential t-test because in certain obvious situations the procedure required a surprisingly large number of observations before decisioning. With this criticism in mind we obtained the results given in Table 3 by generating observations from distributions with means $\alpha = 1.5$, 11.0, and -1000.0 and tested the hypothesis H_o: $\alpha = 0.0$ versus H_1: $\alpha = 1.0$ with the parameter configuration as shown in the table heading. Summarizing the results of these fifty trials, we found

1) that $R(1, p, n)$ did not decision at all when $\alpha = -1.000.0$,

2) that $R(2, p, n)$ delayed decisioning when $\alpha = -1000.0$,

3) that $R(5, p, n)$ decisioned almost every time at the stage it was censored,

4) that the ratio of observed average sample number (a.s.n.) with covariance to observed a.s.n. without covariance approached 1.00 as α increased for all tests except $R(2, p, n)$, and

5) that none of the tests made incorrect decisions.
We now discuss each of these points in turn.

An unusual feature of the study revealed that R(1, p, n) did not decision once in fifty trials when $\alpha = -1000.0$. This, however, is less surprising if we consider some of the results of Chapter III, specifically that given by 3.54 from which we see a tendency for S(1, p, n) to approach

$$\frac{\sigma^2 + (\alpha - \alpha_T)^2}{\sigma^2 + (\alpha - \alpha_A)^2} \tag{6.16}$$

as n becomes large. Now if σ^2 is small relative to $|\alpha - \alpha_T|$ and $|\alpha - \alpha_A|$ and both $|\alpha - \alpha_T|$ and $|\alpha - \alpha_A|$ are large, then the ratio given by 6.16 is close to one. For the parameter configuration shown in Table 3 we can show that the ratio given in 6.16 is equal to 0.998003 so that

$$R(1, p, n) = [S(1, p, n)]^{\frac{n-p-1}{2}} \tag{6.17}$$

converges slowly towards zero and the sample size required for $R(1, p, n) \leq 0.0309$ may be great. Actually, it may be necessary to sample until $n \geq 313+p$ for a decision since

$$[.998]^{312} \doteq .0309.$$

We now demonstrate the tendency for R(2, p, n) to delay decisioning when the true mean α is a great distance from either of the hypothesized values α_T and α_A. Let us assume that we have 15 observations having values dispersed about

10^{10} with a population variance of 75.14. If we test H_o: $\alpha = 0.0$ versus H_1: $\alpha = \gamma\sigma = 1.0$ using R(2, p, n) when p = 0, we have, using tables by Rushton and Lang (63),

$$R(2, 0, 15)$$

$$= e^{-.1}[F(f; \tfrac{1}{2}; .1) + 2\ .1\ \frac{\Gamma(7.5)}{\Gamma(7)} F(6.5; \tfrac{3}{2}; 1)]$$

$$= (.9048)\ [2.983 + 2(.316)(3.8985)(1.504)]$$

$$= 6.05,$$

which is less than the Waldian upper boundary 19.0 when P(I) = P(II) = .05. Part of the motivation for the test procedure using R(5, p, n) was to overcome this "delayed decisioning" property of R(2, p, n). The results of Table 3 indicate that we were successful in this respect, for R(5, p, n) decisioned at 9.0 (its censor number) almost every time.

As α increases we note that the entries in the ratio column also increase and we conjecture that, except for R(2, p, n), they will approach 1. The "delayed decisioning" characteristic mentioned above will probably inhibit R(2, p, n) from attaining this limit.

Finally, even though R(1, p, n) and R(2, p, n) have the tendency to delay decisioning we suspect that few errors would occur when $\alpha = 1.5$ and $\alpha = 11.0$, and no errors would result when α = -1000.0 if the trials were allowed to run without forced termination. From Table 3 we may infer that P(II) and P(I) are monotone decreasing functions of α if the nuisance parameters are fixed.

VII. SUMMARY AND TOPICS FOR FURTHER RESEARCH

A. Summary

In this thesis sequential procedures are developed for discrimination between two treatments when concomitant information is utilized. Wald's Sequential Probability Ratio Test is known to be optimal in certain senses when both the null and alternative hypotheses are simple. In most hypothesis testing situations, however, such hypothesis formulations are unrealistic because of unspecified nuisance parameters and the possibility that the region of interest for the parameter under investigation contains more than two points. Weight functions, originally put forward in Wald (69), have been applied to obtain appropriate test statistics for testing in the more realistic cases of composite hypotheses. In this context also the relationship between the weight function approach and one based on a Bayesian prior distribution framework is discussed.

In Chapter III we considered the design where the subjects were paired and received one of two treatments at random. In the general case each observation was taken to be a p + 1 vector consisting of the variate of interest plus p covariates and these were assumed to be multivariate normally distributed. In this case the useful property that the form of distribution is preserved under the differencing process which is essentially required obtains. The relation of the "basic" correlation between the differenced variates and the advantageous correlations due, for example, to the pairing properties are discussed in Cox and Roseberry (18). Then using the weight function approach, we put forward several hypothesis formulations and derived sequential tests for each. We pointed out location and scale invariance characteristics of these tests and we also presented some termination proofs.

The existence of the sequential t-test introduced by Wald (69) and further developed and examined by Rushton (59) and Barnard (5) motivated the development of the sequential t- and t^2-tests, utilizing concomitant information, as presented in Chapter IV. The technique of construction utilizes the concept of fixed sample size sufficiency and invariance to obtain a factorization of an otherwise formidable likelihood expression. To implement this approach, the problem considered in Chapter III was reformulated and we assumed the covariates were controlled as distinct from the previous assumption of a multivariate normal distribution. A test was then derived for testing H_o: $\alpha = \alpha_T$ versus H_1: $\alpha = \alpha_T + \sigma\gamma$, when α_T and γ were specified while the nuisance parameters were unspecified. Tests for two-sided formulations were also obtained. Next the restriction requiring pairing of the subjects was removed and, again employing the methods of weight functions and prior distributions, we constructed a sequential two-sample t-test utilizing concomitant information. We demonstrated that the above mentioned test is, in fact, a type of sequential t-test as put forward by Hajnal (26).

Results of a Monte Carlo study are presented for the following three test procedures proposed in Chapter III.

1) R(1, p, n), given by 3.52, is proposed as a test of H_o: $\alpha \leq \alpha_T$ versus H_1: $\alpha \geq \alpha_A$, $\alpha_T < \alpha_A$. R(1, p, n) was derived for the "least favorable" case, H_o: $\alpha = \alpha_T$ versus H_1: $\alpha = \alpha_A$, with uniform weighting of the nuisance parameters.

2) R(5, p, n), given by 3.120 - 3.121, is also proposed for testing H_o: $\alpha \leq \alpha_T$ versus H_1: $\alpha \geq \alpha_A$, $\alpha_T < \alpha_A$. Nonuniform weights were placed on α and uniform weights were placed on the nuisance parameters in the derivation of this test statistic.

3) R(2, p, n), given by 3.70, is proposed as a test of H_o: $\alpha = \alpha_T$ versus H_1: $\alpha = \alpha_T + \gamma\sigma$ where α_T and γ are specified. Uniform weighting of the nuisance parameters was used in this case.

We investigated these procedures with special emphasis on

a) the economies in sample size when concomitant information is utilized,

b) their realized error rates, and

c) their performance under certain extreme operating conditions in order to detect any unfavorable properties. In all cases the results indicated that a substantial saving in sample number was achieved when covariance was used, if the correlation coefficient exceeded 0.6. There were slight savings when ρ was close to 0.6 and results from a run when $\rho = .5$ indicated that there was very little advantage in including the covariate in the analysis.

Comparison of the average sample numbers achieved by the different tests showed that test R(5, p, n) consistently had smaller observed average sample number than either R(1, p, n) or R(2, p, n). This difference increased as the sample numbers increased.

Overall, R(1, p, n) had error rates less than those specified, whereas R(5, p, n) had error rates varying from slightly less to slightly more than the specified error rates. R(2, p, n) consistently produced observed error rates less than those specified.

To examine performances under extreme conditions, the hypothesis H_o: $\alpha = 0.0$ versus H_1: $\alpha = 1.0$ was tested using all three test procedures with observations having large means relative to α_o and α_1 and small standard deviations. It was found that each test decisioned correctly provided it had decisioned, but that

i) R(1, p, n) did not decision at all if $|\alpha - \alpha_T|$ and $|\alpha - \alpha_A|$ were both very large with respect to the standard error,

ii) R(5, p, n) decisioned almost every time at the censoring level specified for each run and

iii) R(2, p, n) had certain minimum stage numbers at which it would decision when $|\alpha - \alpha_T|$ was very large with respect to the standard error and these minimum stage numbers increased as $|\alpha - \alpha_T|$ was increased.

In all, although financial considerations limited the extent of the Monte Carlo investigations, the results indicated that the test statistics derived do permit advantages of practical importance to be obtained by the use of concomitant information in sequential trials.

B. Some Topics for Further Research

In this section we note some topics on which further research seems desirable. The difficult problems of finding distributions of sample numbers and operating characteristic functions for composite hypotheses are as yet unsolved. Bhate (12) has put forward a general conjecture which has been demonstrated empirically by some authors to give a good approximation to the expected sample number in some cases. In this regard, see Ray (55), Hajnal (26), and Jackson and Bradley (35). However, the burden, as of now, lies with the computer to provide guidance along these lines.

Another area of some concern is the prospect that in a particular case the sample number may become unusually large. To protect against such behavior, Armitage (3), and Schneiderman and Armitage (65, 66) have presented some exact and conjectured approximate restrictive (closed) procedures for a particular application. It would be interesting to see how these authors' ideas would work in conjunction with the procedures presented in this thesis. Again the investigation would undoubtedly have to be empirical because of theoretical difficulties.

Another topic that merits study is the possibility of extension of the results of this thesis to discrimination between more than two treatments. Wetherill (74) reports some results for selection of the largest (or smallest) of several means, but again the problem of nuisance parameters in addition to the mathematical intractability associated with intuitively pleasing designs might force the investigator to seek the aid of a computer.

The relationship of Bayesian and frequentist concepts within the framework of sequential analysis is a broad topic meriting further consideration. Bartholomew (7) has presented some ideas along these lines and Welch and Peers (74) have some mathematical formulae that one might find useful in relation to this topic.

It is conjectured that a derivation of a two-sample procedure for testing hypothesis formulations similar to those given in Section F of Chapter III might be useful. Also bearing on Chapter III, the concept of adjusting weight functions as the data becomes available, a type of empirical Bayes approach, seems worthy of consideration. Clutton-Brock (15) and Robbins (56b, 57) are preliminary references for an investigation concerning these "empirical" prior distributions. The criteria of test construction set out in Section A of Chapter III might also be improved upon by considering weighted ratios of weighted likelihoods rather than ratios of weighted likelihoods.

In Section A of Chapter V we showed how the weight function approach could be used to derive Hajnal's two-sample t^2-test. The analogous derivation of Jackson and Bradley's multivariate χ^2- and T^2- tests via the weight functions and prior distribution approach would be of some supplementary theoretical interest.

VIII. REFERENCES

1. Anderson, T. W. An introduction to multivariate statistical analysis. New York, N.Y., John Wiley and Sons, Inc. 1958.

2. Appleby, R. H. and Freund, R. J. An empirical evaluation of multivariate sequential procedure for testing means. Annals of Mathematical Statistics 33: 1413-1420. 1962.

3. Armitage, P. Restricted sequential procedures. Biometrika 44: 9-26. 1957.

4. Armitage, P. Sequential medical trials. Oxford, England, Blackwell Scientific Publications. c1960.

5. Barnard, G. A. The frequency justification of certain sequential tests. Biometrika 39: 144-150. 1952.

6. Barnard, G. A. The frequency justification of sequential tests - addendum. Biometrika 40: 468-469. 1953.

7. Bartholomew, D. J. A comparison of some Bayesian and frequentist inferences. Biometrika 52: 19-35. 1965.

8. Bartholomew, D. J. Hypothesis testing when the sample size is treated as a random variable. Royal Statistical Society Journal Series B, 29: 53-82. 1967.

9. Bateman, Harry. Higher transcendental functions. Vol. 2. New York, N.Y., McGraw-Hill Book Company, Inc. 1953.

10. Bateman, Harry. Tables of integral transforms. Vol. 1. New York, N.Y., McGraw-Hill Book Company, Inc. 1954.

11. Beyer, William H., ed. Handbook of tables for probability and statistics. Cleveland, Ohio, The Chemical Rubber Co. 1966.

12. Bhate, D. H. Approximation to the distribution of sample size for sequential tests. I. Tests of simple hypotheses. Biometrika 46: 130-138. 1959.

13a. Bhattacharya, S. K. Bayesian approach to life testing and reliability estimation. American Statistical Association Journal 62: 48-62. 1967.

13b. Box, G. E. P. and Muller, M. E. A note on the generation of random normal deviates. Annals of Mathematical Statistics 29: 610-611. 1958.

14. Bradley, R. A., Martin, D. C., and Wilcoxon, Frank. Sequential rank tests. I. Monte Carlo studies of the two sample procedure. Technometrics 7: 463-483. 1965.

15. Clutton-Brock, H. Using the observations to estimate the prior distribution. Royal Statistical Society Journal Series B, 27: 17-27. 1965.

16. Colton, T. A model for selecting one of two medical treatments. American Statistical Association Journal 58: 388-400. 1963.

17. Cornfield, J. A Bayesian test of some classical hypotheses--with application to sequential clinical trials. American Statistical Association Journal 61: 577-594. 1966.

18. Cox, C. P. and Roseberry, T. D. A large sample sequential test, using concomitant information for discrimination between two composite hypotheses. American Statistical Association Journal 61: 357-367. 1966.

19. Cox, D. R. Large sample sequential tests for composite hypotheses. Sankyha 25: 5-12. 1963.

20. Cox, D. R. Sequential tests for composite hypotheses. Cambridge Philosophical Society Proceedings 48: 290-299. 1952.

21. Cramer, H. Mathematical methods of statistics. Princeton, N.J., Princeton University Press. 1961.

22. David, Herbert T. and Kruskal, William H. The WAGR sequential t-test reaches a decision with probability one. Annals of Mathematical Statistics 27: 797-805. 1956.

23. De Jonge, H. Quantitative methods in pharmacology. New York, N.Y., Interscience Publishers, Inc. 1961.

24. Fisz, Marek. Probability theory and mathematical statistics. 3rd ed. New York, N.Y., John Wiley and Sons, Inc. 1963.

25. Ghosh, B. K. Simultaneous tests by sequential methods in hierarchical classifications. Biometrika 51: 439-450. 1964.

26. Hajnal, J. A two-sample sequential t-test. Biometrika 48: 65-75. 1961.

27. Hall, W. J. Methods of sequentially testing composite hypotheses with special reference to the two-sample problem. University of North Carolina Mimeo Series No. 441. 1965.

28. Hall, W. J., Wijsman, R. A., and Ghosh, J. K. The relationship between sufficiency and invariance with applications in sequential analysis. Annals of Mathematical Statistics 36: 575-614. 1965.

29. Harvard Computation Laboratory. Tables of the cumulative binomial probability distribution. Computation Laboratory of Harvard University Annals 35. 1955.

30. Hoel, P. H. On a sequential test for the general linear hypothesis. Annals of Mathematical Statistics 26: 136-139. 1955.

31. Ifram, Adnan F. Hypergeometric functions in sequential analysis. Annals of Mathematical Statistics 36: 1870-1872. 1965.

32. Ifram, Adnan F. On sample size and simplification of a class of sequential probability ratio tests. Annals of Mathematical Statistics 37: 425-434. 1966.

33. Ifram, Adnan F. On the asymptotic behavior of densities with applications to sequential analysis. Annals of Mathematical Statistics 36: 615-637. 1965.

34. Jackson, J. E. Bibliography on sequential analysis. American Statistical Association Journal 55: 561-580. 1960.

35. Jackson, J. E. and Bradley, R. A. Sequential χ^2 and T^2-tests. Annals of Mathematical Statistics 32: 1063-1077. 1961.

36. Johnson, N. L. A proof of Wald's theorem on cumulative sums. Annals of Mathematical Statistics 30: 1245-1247. 1959.

37. Johnson, N. L. Sequential analysis: a survey. Royal Statistical Society Journal Series A, 24: 372-411. 1961.

38. Johnson, N. L. Sequential procedures in certain component of variance problems. Annals of Mathematical Statistics 25: 357-366. 1954.

39. Johnson, N. L. Some notes on the application of sequential methods in the analysis of variance. Annals of Mathematical Statistics 24: 614-623. 1953.

40. Johnson, N. L. and Leone, F. Statistics and experimental design: in engineering and the physical sciences. Vol. 2. New York, N.Y., John Wiley and Sons, Inc. 1964.

41. Johnson, N. L. and Maurice, R. J. A minimax - regret procedure for choosing between two populations using sequential sampling. Royal Statistical Society Journal Series B, 25: 297-304. 1963.

42. Kendall, Maurice G. and Stuart, Alan. The advanced theory of statistics. 2nd ed. Vol. 1. New York, N.Y., Hafner Publishing Company. 1962.

43. Kendall, Maurice G. and Stuart, Alan. The advanced theory of statistics. Vol. 2. New York, N.Y., Hafner Publishing Company. 1961.

44. Lebedev, N. N. Special functions and their applications. Englewood Cliffs, New Jersey, Prentice Hall, Inc. 1965.

45. Lehman, E. Testing statistical hypotheses. New York, N.Y., John Wiley and Sons. 1959.

46. Lindley, D. V. and Barnett, B. N. Sequential sampling: two decision problems with linear losses for binomial and normal random variables. Biometrika 52: 507-532. 1965.

47. Maurice, R. J. A different loss function for the choice between two populations using sequential sampling. Royal Statistical Society Journal, Series B, 21: 203-213. 1959.

48. Maurice, R. J. A minimax procedure for choosing between two populations using sequential sampling. Royal Statistical Society Journal, Series B, 19: 255-261. 1957.

49. Olin, James R. Numerical investigation of sequential weight-function tests. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1967.

50. Plackett, R. L. Principles of regression analysis. Oxford, England, Clarendon Press. 1960.

51. Plackett, R. L. Current trends in statistical inference. Royal Statistical Society Journal, Series A, 129: 249-267. 1966.

52. Pratt, John W. Bayesian interpretation of standard inference statements. Royal Statistical Society Journal, Series B, 27: 169-203. 1965.

53. Raiffa, Howard and Schlaifer, Robert. Applied statistical decision theory. Boston, Mass., Division of Research, Harvard Business School. 1966.

54. Ray, W. D. A proof that the sequential probability ratio test (S.P.R.T.) of the general linear hypothesis terminates with probability unity. Annals of Mathematical Statistics 28: 521-523. 1957.

55. Ray, W. D. Sequential analysis applied to certain experimental designs in the analysis of variance. Biometrika 43: 388-403. 1966.

56a. Resnikoff, George J. and Lieberman, Gerald J. Tables of the non-central t-distribution. Stanford, California, Stanford University Press. 1957.

56b. Robbins, H. The empirical Bayes approach to statistical decision problems. Annals of Mathematical Statistics 35: 1-20. 1964.

57. Robbins, H. An empirical Bayes approach to statistics. Berkeley Symposium on Statistics and Probability 3rd, Proceedings 1: 157-194. 1955.

58. Roseberry, T. D. The utilization of concomitant information in sequential procedures for the comparison of two treatments. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1965.

59. Rushton, S. On a sequential t-test. Biometrika 37: 326-333. 1950.

60. Rushton, S. On a two-sided sequential t-test. Biometrika 39: 302-308. 1952.

61. Rushton, S. On sequential tests of the equality of variances of two normal populations with known means. Sankyha 12: 63-78. 1952.

62. Rushton, S. On the confluent hypergeometric function M(α, γ, x). Sankyha 13: 369-376. 1954.

63. Rushton, S. and Lang, E. D. Tables of the confluent hypergeometric function. Sankyha 13: 377-411. 1954.

64. Savage, L. J. and others. The foundations of statistical inference and discussion. New York, N.Y., John Wiley and Sons, Inc. 1961.

65. Schneiderman, M. A. and Armitage, P. Closed sequential t-tests. Biometrika 49: 359-366. 1962.

66. Schneiderman, M. A. and Armitage, P. A family of closed sequential procedures. Biometrika 49: 41-56. 1962.

67. Slater, L. J. Confluent hypergeometric functions. London, England, Cambridge University Press. 1960.

68. Stone, M. Right Haar measure for convergence in probability to quasi posterior distributions. Annals of Mathematical Statistics 36: 440-453. 1965.

69. Wald, A. Sequential analysis. New York, N.Y., John Wiley and Sons, Inc. c1947.

70. Wald, A. and Wolfowitz, J. Optimum character of the sequential probability ratio test. Annals of Mathematical Statistics 19: 326-339. 1948.

71. Wallace, D. L. Conditional confidence level properties. Annals of Mathematical Statistics 30: 864-676. 1959.

72. Watson, A. S. A note on maximum likelihood. Sankyha 23: 303-304. 1964.

73. Welch, B. L. and Peers, H. W. On formulae for confidence points based on integrals of weighted likelihoods. Royal Statistical Society Journal, Series B, 25: 318-329. 1963.

74. Wetherill, B. Barrie. Sequential methods in statistics. New York, N.Y., John Wiley and Sons, Inc. 1966.

IX. ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Professor C. Philip Cox for his valuable suggestions and criticisms during the course of this study.

Thanks are also given to Dr. T. A. Bancroft for his counsel and to the National Institutes of Health (Grant 430-21-40) for their financial support throughout the author's graduate program at Iowa State University. The author also wishes to thank Eric West of the Numerical Analysis-Programming Group for his excellent programming support.

To his wife, Vicki, the author is grateful for help and encouragement throughout the preparation of this dissertation and for instantaneous typing of the rough draft.

X. APPENDIX

A. Distribution Results and Maximum Likelihood Estimators

Let us consider the probability density function

$$g(z;\ \alpha,\ \mu,\ \beta,\ \Sigma,\ \sigma^2)$$

$$= ((2\Pi)^{p+1}\ \sigma^2\,|\Sigma|\,)^{-\frac{1}{2}}\ \exp\left[-\ \frac{(x-\mu)\Sigma^{-1}(x-\mu)}{2}\right]\exp\left[-\ \frac{(y-\alpha-(x-\mu)'\beta)^2}{2\sigma^2}\right] \qquad (10.1)$$

where $z' = (y,\ x')$, $x' = (x_1,\ \ldots,\ x_p)$, y is a scalar, $u' = (u_1,\ u_2,\ \ldots,\ u_p)$, $\beta' = (\beta_1,\ \ldots,\ \beta_p)$, and Σ is a positive definite matrix.

Theorem 10.1: The characteristic function $C(t)$ corresponding to the p.d.f. given in 10.1 is

$C(t) = E[\exp(it'z)]$

$$\exp\left[it'\begin{pmatrix}\alpha\\ \mu\end{pmatrix} - \frac{1}{2}\ t'\ \begin{pmatrix}\sigma^2 + \beta'\Sigma\beta & \Sigma\beta\\ \beta'\Sigma & \Sigma\end{pmatrix} t\right] \qquad (10.2)$$

where $t' = (t_1,\ \ldots,\ t_{p+1})$ is any real vector and $i = \sqrt{-1}$.

Proof:

$E[e^{it'z}]$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp(it'z)g(z';\alpha,\mu,\beta,\Sigma,\sigma^2)\,dz'$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp(it'_{(p)}x)(2\Pi\sigma^2)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}\ \exp\left[-\ \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}\right]$$

$$\left[\int_{-\infty}^{\infty}(2\Pi\sigma^2)^{-\frac{1}{2}}\ \exp(it_1y)\ \exp\left[-\ \frac{(y-\alpha-(x-\mu)'\beta)^2}{2\sigma^2}\right]\ dy\right] dx \qquad (10.3)$$

where $t'_{(p)} = (t_2, \ldots, t_{p+1})$. Completing the integration with respect to y we can write 10.3 as

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(it'_{(p)}\, x)(2\Pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[- \frac{(x-\mu)\Sigma^{-1}(x-\mu)}{2}\right]$$

$$\exp\, i(\alpha+(x-\mu)'\beta)t_1) - \frac{t_1^2}{2}\sigma^2\Big] \, dx$$

$$= \exp(i\alpha t_1 - \frac{t_1^2\sigma^2}{2}) \exp(-i\mu'\beta t_1) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}$$

$$\exp\left[i(t'_{(p)} + t_1\beta)'x\right](2\Pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[- \frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right] \, dx \tag{10.4}$$

and completing the integration on x by the same technique as shown in Section B of this Appendix we have

$$= \exp(i\alpha t_1 - \frac{t_1^2}{2}\sigma^2) \exp\left[it'_{(p)}\mu - \frac{(t_{(p)}+t_1\beta)'\Sigma(t_{(p)}+t_1\beta)}{2}\right]$$

$$= \exp\left[i(t_1,\ t'_{(p)}) \begin{pmatrix}\alpha\\ \mu\end{pmatrix} - \frac{1}{2}\begin{pmatrix}t_1\\ t_{(p)}\end{pmatrix}' \begin{pmatrix}\sigma^2 + \beta'\Sigma\beta & \Sigma\beta\\ \beta'\Sigma & \Sigma\end{pmatrix} \begin{pmatrix}t_1\\ t_{(p)}\end{pmatrix}\right].$$

<u>Corollary</u> <u>10.1</u>: $C(0) = 1$

<u>Corollary</u> <u>10.2</u>: The marginal distribution of the vector x has p.d.f.

$$f_1(x; \mu, \Sigma) = (2\Pi)^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[- \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}\right].$$

<u>Proof</u>:

Letting $t_1 = 0$, we have

$$C(0, t_2, \ldots, t_{p+1}) = \exp\left[it'_{(p)}\mu - \frac{1}{2}\, t'_{(p)}\, \Sigma t_{(p)}\right].$$

Thus the marginal p.d.f. of x is multivariate normal with mean μ and variance-covariance matrix Σ.

Corollary 10.3: The marginal distribution of y is normal with mean α and variance $\sigma^2 + \beta'\Sigma\beta$.

Proof:

Letting $t_{(p)} = 0$ we have

$$C(t_1, 0, 0, \ldots, 0) = \exp\left[it_1\alpha - \frac{t_1^2(\sigma^2 + \beta'\Sigma\beta)}{2}\right].$$

Thus the marginal p.d.f. of y is univariate normal with mean α and variance $\sigma^2 + \beta'\Sigma\beta$.

Corollary 10.4: The conditional distribution of y given the vector x is univariate normal with mean $\alpha + (x-\mu)'\beta$ and variance σ^2.

Proof:

This result follows from the relationship

$$f(y\ x) = \frac{g(y, x)}{f_1(x)},$$

from 10.1 and from Corollary 10.2.

Corollary 10.5: Let z_1 be a random variable having p.d.f. $g(z_1; \alpha_1, \mu, \beta, \Sigma, \sigma^2)$ and let z_2 be a random variable having p.d.f. $g(z_2; \alpha_2, \mu, \beta, \Sigma, \sigma^2)$. If z_1 and z_2 are stochastically independent then $z_1 - z_2$ is a random variable having p.d.f. $g(z_1 - z_2; \alpha_1 - \alpha_2, 0, \beta, 2\Sigma, 2\sigma^2)$.

Proof:

$$E\left[\exp(it'(z_1-z_2))\right] = E\left[\exp(it'z_1)\right] E\left[\exp(-it'z_2)\right]$$

$$= \exp\left[it'\begin{pmatrix}\alpha_1\\ \mu\end{pmatrix} - \frac{1}{2}t'At\right] \exp\left[-it'\begin{pmatrix}\alpha_2\\ \mu\end{pmatrix} - \frac{1}{2}t'At\right]$$

$$= \exp\left[it'\begin{pmatrix}\alpha_1-\alpha_2\\ 0\end{pmatrix} - \frac{1}{2}t'(2A)t\right]$$

where

$$A = \begin{pmatrix}\sigma^2 + \beta'\Sigma\beta & \Sigma\beta\\ \beta'\Sigma & \Sigma\end{pmatrix}.$$

We now obtain the maximum likelihood estimators for the parameters β, Σ, and σ^2 of the p.d.f.

$$f(y, x; \alpha, \beta, \Sigma, \sigma^2) = ((2\Pi)^{p+1} \sigma^2 |\Sigma|)^{-\frac{1}{2}} \exp\left[-\frac{x'\Sigma^{-1}x}{2}\right]$$
$$\exp\left[-\frac{(y-\alpha-x'\beta)^2}{2\sigma^2}\right]$$

when α is assumed known. The log-likelihood of this sample may be expressed as

$$\ln L(\alpha, \beta, \Sigma, \sigma^2) = \ln \prod_1^n f(y_i, x_i; \alpha, \beta, \Sigma, \sigma^2)$$
$$= f_1(\Sigma) + f_2(\alpha, \beta, \sigma)$$

where

$$f_1(\Sigma) = -\sum_{i=1}^{n} \frac{x_i'\Sigma^{-1}x_i}{2} - \frac{np\ln 2\Pi}{2} + \frac{n}{2}\ln|\Sigma^{-1}|$$

$$f_2(\alpha, \beta, \sigma) = -n\ln\sigma - \sum_{i=1}^{\infty} \frac{(y_i-\alpha-x_i'\beta)^2}{2\sigma^2} - \frac{n}{2}\ln 2\Pi.$$

Let us utilize the fact that if $g(x, y) = g_1(x) + g_2(y)$ then the (x_o, y_o) which maximizes $g(x, y)$ is also such that x_o maximizes $g_1(x)$ and y_o maximizes $g_2(y)$. If we assume that α

is known then we may differentiate $f_2(\alpha, \beta, \sigma)$ in the usual manner to obtain maximum likelihood estimates for β and σ^2, which are as follows:

$$\hat{\beta} = (XX')^{-1}X(y-\alpha e)$$
$$\hat{\sigma}^2 = (y-\alpha e)'(I-M)(y-\alpha e) \tag{10.6}$$

where $X' = (x_1, x_2, \ldots, x_n)$, $e' = (1, \ldots, 1)$, $y' = (y_1, \ldots, y_n)$, $M = X'(XX')^{-1}X$, and I is the identity matrix. For the maximum of $f_1(\Sigma)$ we utilize the following result (1, p. 46; 72, p. 303-304).

Lemma 10.1: If C and B are given positive definite p x p matrices then the function

$$f(C) = \ln C - \text{trace } CB^{-1} \tag{10.7}$$

takes its maximum if and only if C = B.

Proof:

Since $f(B) = \ln|B| - \text{trace } I = \ln|B| - p$ (10.8)

we need to show that

$$\ln|B| - p - \ln|C| - \text{trace } CB^{-1} \geq 0 \tag{10.9}$$

and that equality holds if and only if C = B. If $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the characteristic roots of CB^{-1} then trace $CB^{-1} = \sum_1^p \lambda_i$ and $|CB^{-1}| = \prod_1^p \lambda_i$, and 10.7 may be rewritten as

$$\sum_1^p \lambda_i - p - \ln \prod_1^p \lambda_i = \sum_1^p (\lambda_i - 1 - \ln\lambda_i) \geq 0. \tag{10.10}$$

Now $\lambda - 1 - \ln\lambda = 0$ when $\lambda = 1$ and $\lambda - 1 - \ln\lambda > 0$ if $\lambda > 0$. Thus 10.9 holds and we have 10.7 attaining its maximum when C = B.

We rewrite $f_1(\Sigma)$ so that

$$f_1(\Sigma) = -\text{ trace } \frac{XX'\Sigma^{-1}}{2} - \frac{np\ln 2\Pi}{2} + \frac{n}{2}\ln|\Sigma^{-1}|.$$

By Lemma 10.1

$$\hat{\Sigma}^{-1} = \left(\frac{XX'}{n}\right)^{-1}$$

so that (10.11)

$$\hat{\Sigma} = \frac{XX'}{n}.$$

If α is assumed unknown than by arguments similar to those used above we have the maximum likelihood estimates as follows:

$$\begin{aligned} \hat{\beta} &= (XX')^{-1}X(y-e\alpha) \\ \hat{\alpha} &= [e'(I-M)e]^{-1}e'(I-M)y \\ \hat{\sigma}^2 &= \frac{y'y - \hat{\alpha}e'y - \hat{\beta}'Xy}{n}. \end{aligned} \tag{10.12}$$

In either case, α known and α unknown, the estimates are consistent. That is, $\hat{\beta} \xrightarrow{P} \beta$, $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ and $\hat{\alpha} \xrightarrow{P} \alpha$ where $X_n \xrightarrow{P} C$ denotes convergence in probability.

B. Evaluation of a Multiple Definite Integral

Theorem 10.2: Let X' be a real $n \times p$ matrix of rank p. Let β' be a vector $(\beta_1, \beta_2, \ldots, \beta_p)$ of real valued variables and let u' be a vector $(u_1, u_2, \ldots, u_p)$ of arbitrary real numbers. Then

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\beta' XX\beta + u'\beta\right] d\beta_1 d\beta_2 \ldots d\beta_p$$

$$= (2\Pi)^{\frac{p}{2}}|XX'|^{-\frac{1}{2}} \exp\left[\frac{u'(XX')^{-1}u}{2}\right]. \tag{10.13}$$

Proof:

Since XX' is a real symmetric p x p matrix of rank p there exists an orthogonal p x p matrix C such that $C'C = I$ and $C'XX'C = D$ where $D = (d_{ij})$ is such that

$$d_{ij} = \frac{1}{d_i} \quad i = j$$
$$= 0 \quad i \neq j.$$

Now XX' is a positive definite matrix implying that $d_i > 0$, $i = 1, \ldots, p$. Since C is orthogonal, then $|C'C| = |I| = |C'||C|$ and thus $|C|^2 = 1$. We now make a transformation from $\beta_1, \beta_2, \ldots, \beta_p$ to $\gamma_1, \ldots, \gamma_p$ by writing $\beta = C\gamma$. The absolute value of the Jacobian of this transformation is $||C|| = |\pm 1| = 1$. Thus we have

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\beta'XX'\beta + u'\beta\right] d\beta_1 \ldots d\beta_p$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\gamma'D\gamma + u'C\gamma\right] d\beta_1 \ldots d\beta_p. \qquad (10.14)$$

Letting $t' = u'C = (t_1, t_2, \ldots, t_p)$ we rewrite 10.14 as

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\sum_1^p \frac{\gamma_i^2}{d_i} + \sum_1^p t_i\gamma_i\right] d\gamma_1 \ldots d\gamma_p$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\sum_1^p \frac{(\gamma_i^2 - 2d_it_i + d_i^2t_i^2 - d_i^2t_i^2)}{d_i}\right] d\gamma_1 \ldots d\gamma_p$$

$$= \prod_1^p \left[\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\frac{(\gamma_i - d_it_i)^2}{d_i}\right] d\gamma_i \exp\left[\frac{1}{2} d_it_i^2\right]\right]. \qquad (10.15)$$

Recognizing the definite integral of 10.15 as the normal probability integral we write 10.15 as

$$\prod_{1}^{p} \left[(2\Pi d_i)^{\frac{1}{2}} \exp\left[\frac{1}{2} d_i t_i\right]\right]$$

$$= (2\Pi)^{\frac{p}{2}} \left(\prod_{1}^{p} d_i\right)^{\frac{1}{2}} \exp\left[\frac{1}{2} \sum_{1}^{p} d_i t_i^2\right] \tag{10.16}$$

$$= (2\Pi)^{\frac{p}{2}} |D|^{-\frac{1}{2}} \exp\left[\frac{1}{2} \sum_{1}^{p} d_i t_i^2\right]. \tag{10.17}$$

Finally, since $t' = u'C$, $(CD^{-1}C')^{-1} = C'^{-1}DC^{-1} = XX'$, and $|D^{-1}| = |C'XX'C| = |C'XX'C|^{-1} = \left[|C'| \; |XX'| \; |C|\right]^{-1} = |XX'|^{-1}$ 10.17 becomes

$$(2\Pi)^{\frac{p}{2}} |XX'|^{-\frac{1}{2}} \exp\left[\frac{u'(XX')u}{2}\right]. \tag{10.18}$$

C. The Confluent Hypergeometric Function and Pertinent Formulae

The Pochhammer-Barnes confluent hypergeometric function is the infinite series

$$F(a;\ b;\ x) = \sum_{i=0}^{\infty} \frac{(a)_i}{(b)_i} \cdot \frac{x^i}{i!} \tag{10.19}$$

where

$$(a)_i = a(a+1)(a+2)\ldots(a+i-1)$$

$$(b)_i = b(b+1)(b+2)\ldots(b+i-1)$$

and

$$b \notin [0, -1, -2, -3, \ldots]$$

The series is convergent for all finite x, real or complex. Setting $w = F(a; b; x)$, it can be shown that w satisfies the differential equation

$$\frac{xd^2w}{dx^2} + (b-x)\frac{dw}{dx} - aw = 0. \tag{10.20}$$

Usually we find that 10.19 is written in the form

$$F(a; b; x) = \sum_{i=0}^{\infty} \frac{\Gamma(a+i)}{\Gamma(a)} \frac{\Gamma(b)}{\Gamma(b+i)} \frac{x^i}{i!} \tag{10.21}$$

where $\Gamma(c)$ is the gamma function. In this form it is particularly important to recall that b cannot equal 0 or a negative integer.

A useful relation called Kummer's identity is important in hand or computer calculations and is as follows:

$$F(a; b; x) = e^{-x}F(b-a; b; -x). \tag{10.22}$$

Kummer's identity allows $F(a; b; x)$ to be written as the product of a more rapidly convergent series and a finite series $F(b-a; b; -x)$ provided b-a is an integer such that $b \leq a$. In applications considered in this thesis, $b \leq a$ will always hold and a-b will equal an integer approximately half of the time.

The following three relations are quite helpful in hand and computer calculations when evaluating the confluent hypergeometric function.

$$(a-b+1)F(a;b;x) = aF(a+1;b;x) - (b-1)F(a;b-1;x)$$

$$b(a+x)F(a;b;x) = abF(a+1;b;x) - (a-b)xF(a;b+1;x) \quad (10.23)$$

$$bF(a;b;x) = bF(a-1;b;x) + xF(a;b+1;x).$$

For further details see Bateman (9, 10), Lebedev (44), Rushton (62), and Slater (67).

We now present two Lemmas and a Theorem which aid in the expression of weight function tests in terms of confluent hypergeometric functions.

<u>Lemma 10.2</u>: If $a > 0$ and $s > 1$ then

$$\int_0^\infty x^{-s} \exp\left[-\frac{a}{2}x^{-2}\right] dx = \frac{1}{2}\left(\frac{a}{2}\right)^{-\frac{s-1}{2}} \Gamma\left(\frac{s-1}{2}\right). \quad (10.24)$$

<u>Proof</u>:

Let $x^2 = \frac{a}{2y}$, then $x = \left(\frac{a}{2y}\right)^{\frac{1}{2}}$ and $0 < y < \infty$. The absolute value of the Jacobian of this transformation is

$$|J| = \left|\frac{\partial x}{\partial y}\right| = \frac{a}{4}\left(\frac{2y}{a}\right)^{\frac{1}{2}} \frac{1}{y^2}.$$

Therefore

$$\int_0^\infty x^{-s} \exp\left[-\frac{a}{2}x^{-2}\right] dx$$

$$= \frac{1}{2}\left(\frac{2}{a}\right)^{\frac{s-1}{2}} \int_0^\infty y^{\frac{s-3}{2}} \exp[-y]\, dy$$

$$= \frac{1}{2}\left(\frac{a}{2}\right)^{\frac{s-1}{2}} \Gamma\left(\frac{s-1}{2}\right).$$

Lemma 10.3: If $a > 0$ and $s > 0$ then

$$\int_0^\infty x^{-s} \exp\left[-\frac{a}{2}x^{-2} - \frac{b}{2}x^{-1}\right] dx$$

$$= a^{-\frac{s-1}{2}} s^{\frac{s-3}{2}} \Gamma\left(\frac{s-1}{2}\right)\left[F\left(\frac{s-1}{2};\frac{1}{2};\frac{b^2}{8a}\right) - \frac{b}{(2a)^{\frac{1}{2}}} \frac{\Gamma\left(\frac{s}{2}\right)}{\Gamma\left(\frac{s-1}{2}\right)} F\left(\frac{s}{2};\frac{3}{2};\frac{b^2}{8a}\right)\right]. \tag{10.25}$$

Proof:

Let

$$\Phi(a;b;s) = \int_0^\infty x^{-s} \exp\left[-\frac{a}{2}x^{-2} - \frac{b}{2}x^{-1}\right] dx.$$

If we expand $\exp[-\frac{b}{2}x^{-1}]$ in a power series then

$$\Phi(a;b;s) = \int_0^\infty x^{-s} \sum_{i=0}^\infty \frac{\left(\frac{b}{2}\right)^i}{i!} x^{-i} \exp[-\frac{a}{2}x^{-2}]\, dx.$$

By Lemma 10.2

$$\Phi(a;b;s) = \sum_{i=0}^\infty (-1)^i \frac{\left(\frac{b}{2}\right)^i}{i!} \frac{1}{2} \left(\frac{a}{2}\right)^{-\frac{s+i-1}{2}} \Gamma\left(\frac{s+i-1}{2}\right)$$

$$= \frac{1}{2}\left(\frac{a}{2}\right)^{-\frac{s-1}{2}} \sum_{i=0}^\infty \frac{\left(\frac{b}{2}\right)^{2i}}{(2i)!} \left(\frac{a}{2}\right)^{-\frac{2i}{2}} \Gamma\left(\frac{s+2i-1}{2}\right)$$

$$- \sum_{i=0}^\infty \frac{\left(\frac{b}{2}\right)^{2i+1}}{(2i+1)!} \left(\frac{a}{2}\right)^{\frac{2i+1}{2}} \Gamma\left(\frac{s+2i}{2}\right)$$

$$= \frac{1}{2}\left(\frac{a}{2}\right)^{-\frac{s-1}{2}} \left[\sum_{i=o}^{\infty} \frac{\left(\frac{b^2}{2a}\right)^i}{i!} \frac{i!}{(2i)!} \Gamma\left(\frac{s-1}{2} + i\right) \right.$$

$$\left. - \frac{b}{(2a)^{\frac{1}{2}}} \sum_{i=o}^{\infty} \frac{\left(\frac{b^2}{2a}\right)^i}{i!} \frac{i!}{(2i+1)!} \Gamma\left(\frac{s}{2} + i\right) \right]. \qquad (10.26)$$

Since $\frac{i!}{2i!} = \frac{\Gamma(\frac{1}{2})}{2^{2i}\,\Gamma(i+\frac{1}{2})}$ and $\frac{i!}{(2i+1)!} = \frac{\Gamma(\frac{3}{2})}{2^{2i}\,\Gamma(i+\frac{3}{2})}$ 10.26 becomes

$$\Phi(a;b;s) = \frac{1}{2}\left(\frac{a}{2}\right)^{-\frac{s-1}{2}} \left[\Gamma\left(\frac{s-1}{2}\right) F\left(\frac{s-1}{2};\frac{1}{2};\frac{b^2}{8a}\right) \right.$$

$$\left. - \frac{b}{(2a)^{\frac{1}{2}}} \Gamma\left(\frac{s}{2}\right) F\left(\frac{s}{2};\frac{3}{2};\frac{b^2}{8a}\right) \right]. \qquad (10.27)$$

From Lemmas 10.2 and 10.3 we have the following theorem.

<u>Theorem 10.3</u>: If $a > 0$ and $s > 0$ then

$$\frac{\int_o^\infty x^{-s} \exp[-\frac{a}{2}x^{-2} - \frac{b}{2}x^{-1}]dx}{\int_o^\infty x^{-s} \exp[-\frac{a}{2}x^{-2}]dx} =$$

$$F\left(\frac{s-1}{2};\frac{1}{2};\frac{b^2}{8a}\right) - \frac{b}{(2a)^{\frac{1}{2}}} \frac{\Gamma(\frac{s}{2})}{\Gamma(\frac{s-1}{2})} F\left(\frac{s}{2};\frac{3}{2};\frac{b^2}{8a}\right). \qquad (10.28)$$

D. Some Theorems on Convergence in Probability

$(W, \beta(W), \mu)$ is defined to be a measure-space if W is a space of points, $\beta(W)$ is a σ-algebra of all measurable subsets of W, and μ is a completely additive measure defined on $\beta(W)$. $(W, \beta(W), \mu)$ is called a probability space and μ a probability measure if W represents all possible outcomes and if $\mu(W) = 1$. Let $X(w)$ be a μ-measurable function from $W = w$ to R_1, the extended real line. $X(w)$ is called a random variable. We symbolically represent this mapping by

$$(W, \beta(W), \mu) \xrightarrow{X(w)} (R_1, \beta(R_1), P). \qquad (10.29)$$

If B is any Borel set in $\beta(R_1)$ then we define the probability that $X(w)\epsilon B$, say $P[X(w)\epsilon B]$, by

$$P[X(w)\epsilon B] = \mu[w: w \epsilon X^{-1}(B)]. \qquad (10.30)$$

A sequence of random variables $X_n(w)$, $n = 1, =, \ldots$ is said to converge in probability to a random variable $X(w)$ if, for a given $\epsilon > 0$,

$$\lim_{n \to \infty} P[X_n - X < \epsilon] = \lim_{n \to \infty} \mu[w: |X_n(w)-X(w)| < \epsilon] = 1. \qquad (10.31)$$

An equivalent definition of convergence in probability which will be used in this Appendix is: X_n converges in probability to X if for a given $\epsilon > 0$, $\delta > 0$ there exists an N such that $n \geq N$ implies

$$\mu[w: |X_n(w)-X(w)| \geq \epsilon] < \delta.$$

Let us denote convergence in probability by

$$P - \lim X_n = X \quad \text{or} \quad X_n \xrightarrow{P} X. \qquad (10.32)$$

We now state and prove two theorems.

<u>Theorem</u> <u>10.4</u>: If $X_n \xrightarrow{P} c$ and $0 \leq c < 1$, and if $f(n)$ is

an increasing unbounded function of n then

$$[X_n]^{f(n)} \xrightarrow{P} 0.$$

<u>Proof</u>:

We need to show that for any $\epsilon > 0$, $\delta > 0$, there exists an N such that if $n \geq N$ then

$$\delta > \mu[w\colon [X_n(w)]^{f(n)} > \epsilon].$$

We first consider the case where $\epsilon < 1 - c$. Since $\epsilon + c < 1$ we know there exists an N_2 such that $c + \epsilon < (\epsilon)^{\frac{1}{f(n)}}$ for $n \geq N_2$. Let us choose some $\epsilon^* \leq \epsilon$. For ϵ^*, $\delta > 0$ we know there exists N_1 so that $n \geq N_1$ implies

$$\delta > \mu[w\colon |X(w)-c| > \epsilon^*].$$

Now $\epsilon^* \geq |X(w)-c| \geq |X(w)-c|$ so that

$$[w\colon |X_n(w)-c| \leq \epsilon^*] \quad [w\colon |X_n(w)| \leq \epsilon^* + c]$$

and

$$[w\colon |X_n(w)-c| > \epsilon^*] \quad [w\colon |X_n(w)| > \epsilon^* + c].$$

Therefore

$$\mu[w\colon |X_n(w)-c| > \epsilon^*] \quad \mu[w\colon |X_n(w)| > \epsilon^* + c].$$

When $n \geq N_2$ we know that $\epsilon + c < (\epsilon)^{\frac{1}{f(n)}}$; thus $\epsilon^* + c < (\epsilon)^{\frac{1}{f(n)}}$ and we may write

$$\mu[w\colon |X_n(w)| > \epsilon^* + c] > \mu[w\colon |X_n(w)| > \epsilon^{\frac{1}{f(n)}}].$$

If we choose $N = \max(N_1, N_2)$ then it follows that for any $\epsilon < 1-c$, $\delta > 0$, and $n \geq N$

$$\delta > \mu\ [w: |X_n(w)-c| \geq \epsilon^*] > \mu\ [w: |X_n(w)| > \epsilon^* + c]$$

$$\mu\ [w: |X_n(w)| > \epsilon^{\frac{1}{f(n)}}] = \mu\ [w: |[X_n(w)]^{f(n)}| > \epsilon].$$

For $\epsilon_2 > 1-c$ we have $\epsilon_2 > \epsilon_1$ so that

$$\mu\ [w: |[X_n(w)]^{f(n)}| > \epsilon_1] > \mu\ [w: |[X_n(w)]^{f(n)}| > \epsilon_2].$$

Thus for any $\epsilon > 0$, $\delta > 0$ there exists an N such that if $n \geq N$ we have

$$\delta > \mu\ [w: |[X_n(w)]^{f(n)}| > \epsilon].$$

A sequence of random variables X_n is said to become large with probability one if for any real number $d > 1$ we have $\lim P[X_n > d] = 1$.

<u>Theorem</u> <u>10.5</u>: If $X_n \xrightarrow{P} c$ and $c > 0$, and if $f(n)$ is an increasing unbounded function of n, then $[X_n]^{f(n)}$ becomes arbitrarily large with probability one.

<u>Proof</u>:

Since $X_n \xrightarrow{P} c$ we know that for any $\epsilon > 0$, $\lim\limits_{n \to \infty} P[|X_n-c| < \epsilon] = 1$. Now

$$P[|X_n-c| < \epsilon] = P[c-\epsilon < X_n < c+\epsilon]$$

$$< P[c-\epsilon < X_n] = P[(c-\epsilon)^{f(n)} < X_n^{f(n)}].$$

Let d be any number greater than one. If $c - \epsilon > 1$ then either $c - \epsilon > d$ or there exists an N such that $n \geq N$ implies $(c-\epsilon)^{f(n)} > d$. This follows from the assumption that $f(n)$ is an increasing function unbounded function of n. Now

$$P[(c-\epsilon)^{f(n)} < X_n^{f(n)}] < P[d < X_n^{f(n)}] \text{ for all } n > N.$$

Thus

$$\lim_{n \to \infty} P\,[X_n - c < c] < \lim_{n \to \infty} P[d < X_n^{f(n)}] = 1.$$