The use of concomitant information in multivariate sequential tests of statistical hypotheses

Thomas Joseph Keefe

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
Keefe, Thomas Joseph, "The use of concomitant information in multivariate sequential tests of statistical hypotheses " (1972).
Retrospective Theses and Dissertations. 5926.
https://lib.dr.iastate.edu/rtd/5926

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
KEEFE, Thomas Joseph, 1944-
THE USE OF CONCOMITANT INFORMATION IN
MULTIVARIATE SEQUENTIAL TESTS OF
STATISTICAL HYPOTHESES.

Iowa State University, Ph.D., 1972
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan
The use of concomitant information in multivariate sequential tests of statistical hypotheses

by

Thomas Joseph Keefe

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

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1972
PLEASE NOTE:

Some pages may have indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION AND REVIEW OF THE LITERATURE</td>
<td>1</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>B. The Sequential Probability Ratio Test</td>
<td>4</td>
</tr>
<tr>
<td>C. Extensions of the SPRT</td>
<td>9</td>
</tr>
<tr>
<td>II. RESULTS PERTINENT TO THE DEVELOPMENT OF MULTIVARIATE SEQUENTIAL</td>
<td>11</td>
</tr>
<tr>
<td>TESTS OF STATISTICAL HYPOTHESES</td>
<td></td>
</tr>
<tr>
<td>A. Introduction</td>
<td>11</td>
</tr>
<tr>
<td>B. Fixed-Sample-Size Sufficiency and Invariance</td>
<td>13</td>
</tr>
<tr>
<td>C. Invariance and Sufficiency Principles Applied to Sequential Analysis</td>
<td>32</td>
</tr>
<tr>
<td>D. Results from Multivariate Statistical Analysis</td>
<td>38</td>
</tr>
<tr>
<td>III. SINGLE-SAMPLE SEQUENTIAL TESTS ABOUT MEAN VECTORS</td>
<td>45</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>45</td>
</tr>
<tr>
<td>B. Case (i): $\Sigma$ Unknown</td>
<td>45</td>
</tr>
<tr>
<td>C. Case (ii): $\Sigma$ Unknown, but Estimated Independently</td>
<td>55</td>
</tr>
<tr>
<td>IV. SINGLE-SAMPLE SEQUENTIAL TESTS ABOUT MEAN VECTORS ADJUSTED FOR</td>
<td>63</td>
</tr>
<tr>
<td>COVARIATES</td>
<td></td>
</tr>
<tr>
<td>A. Introduction</td>
<td>63</td>
</tr>
<tr>
<td>B. Case (i): $\Sigma$ Unknown</td>
<td>64</td>
</tr>
<tr>
<td>C. Case (ii): $\Sigma$ Unknown, but Estimated Independently</td>
<td>72</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>V.</td>
<td>TWO-SAMPLE SEQUENTIAL TESTS ABOUT MEAN VECTORS</td>
</tr>
<tr>
<td></td>
<td>A. Introduction</td>
</tr>
<tr>
<td></td>
<td>B. Case (i): $\Sigma$ Unknown</td>
</tr>
<tr>
<td></td>
<td>C. Case (ii): $\Sigma$ Unknown, but Estimated Independently</td>
</tr>
<tr>
<td>VI.</td>
<td>TWO-SAMPLE SEQUENTIAL TESTS ABOUT MEAN VECTORS ADJUSTED FOR COVARIATES</td>
</tr>
<tr>
<td></td>
<td>A. Introduction</td>
</tr>
<tr>
<td></td>
<td>B. Case (i): $\Sigma$ Known</td>
</tr>
<tr>
<td></td>
<td>C. Case (ii): $\Sigma$ Unknown, but Estimated Independently</td>
</tr>
<tr>
<td>VII.</td>
<td>RELATED CONSIDERATIONS AND TOPICS FOR FURTHER RESEARCH</td>
</tr>
<tr>
<td></td>
<td>A. The ASN Function</td>
</tr>
<tr>
<td></td>
<td>B. Discussion</td>
</tr>
<tr>
<td></td>
<td>1. Tables</td>
</tr>
<tr>
<td></td>
<td>2. Determination of $H_0$ and $H_1$</td>
</tr>
<tr>
<td></td>
<td>3. The OC and ASN functions</td>
</tr>
<tr>
<td></td>
<td>4. Truncated and restricted schemes</td>
</tr>
<tr>
<td></td>
<td>C. More Topics for Further Research</td>
</tr>
<tr>
<td>VIII.</td>
<td>REFERENCES</td>
</tr>
<tr>
<td>IX.</td>
<td>ACKNOWLEDGEMENTS</td>
</tr>
<tr>
<td>X.</td>
<td>APPENDIX</td>
</tr>
<tr>
<td></td>
<td>A. The Confluent Hypergeometric Function and Pertinent Formulae</td>
</tr>
<tr>
<td></td>
<td>B. Computation of $T_{n}^{2}$</td>
</tr>
</tbody>
</table>
iv

C. Derivation of Recurrence Relationships for the Sufficient Sequences \( \{S_n\} \)

1. Recurrence relationships of Chapter III 135
2. Recurrence relationships of Chapter IV 138
3. Recurrence relationships of Chapter V 141
4. Recurrence relationships of Chapter VI 142
I. INTRODUCTION AND REVIEW OF THE LITERATURE

A. Introduction

In a wide variety of practical situations, it may be highly desirable for an investigator to follow the results of an experiment closely, as the data become available, so that decisions can be made as soon as possible. As pointed out by Hajnal (1961), Armitage (1957), and others, an early termination of a medical trial—accompanied by the immediate application of the superior treatment to all persons being treated for the affliction under study—is important, not only for ethical reasons, but also on economic grounds since it leads to a reduction in the number of subjects required. This latter consideration might be considered especially important when the experimental units, say in a destructive testing situation, are very expensive or when the testing is so extensive that any saving per sample may lead to sizeable long-term savings.

Further, if it so happens that the experimental units occur rarely, as in a clinical trial investigating the effectiveness of two or more drugs in treating a rare disease, the utility of sequential experimentation, as well as the need for some form of statistical sequential analysis of the data, is certainly obvious. These considerations, together with the military testing programs of World War II, provided the
incentive for Wald's pioneering discovery, the Sequential Probability Ratio Test (SPRT), which will be described in Section B of this Chapter.

In fixed sample size experimentation, statistical analyses utilizing concomitant information have been used, as in the analysis of covariance, most notably to increase the precision of an experiment or to adjust for sources of bias in the comparison of treatments. In this thesis, we are interested in developing sequential test procedures for the comparison of two treatments, wherein the response and covariate metarameters are both vector-valued. For example, a clinical investigator may be interested in the effectiveness of a drug in reducing hypertension. The response of interest might simply be bivariate, for instance systolic and diastolic blood pressure readings; further, the response might be supplemented by such concomitant variables as the patient's age, body weight, and height.

Statistical procedures for the comparison of two treatments in fixed sample size experiments have been thoroughly documented both for the case of univariate observations and for the case of multivariate observations. Sequential statistical methods, however, are not so well developed, especially for the case of multivariate observations.

Armitage (1960) discussed the design and sequential analysis of clinical trials with special reference to the
comparison of two treatments. Roseberry (1965) and Cox and Roseberry (1966) developed and investigated empirically some sequential tests that utilize one covariate. Sampson (1968), using a different approach, subsequently developed univariate sequential test procedures that utilize a vector of covariates.

In all of the above references, the experimental units were paired and the two treatments were assigned at random to the subjects within pairs. Consequently the metameters of interest were the differences of the within-pair responses and, when used, within-pair covariates. Thus, in a sense, these tests can be considered single-sample test procedures with at least one application being the comparison of two treatments via paired-differences. For the case of multivariate responses, Jackson and Bradley (1961a) have developed single-sample sequential tests about the mean vector of a normal population, and their suggestion for a two-sample test procedure is essentially to apply their single-sample procedure to the vector-differences of paired experimental units. This test procedure will be discussed more fully in Chapter III. The sequential test of the mean vector adjusted for covariates, which we develop in Chapter IV, is also a single-sample test procedure with obvious application to the paired-comparison experimental design.

Hajnal (1961) and Sampson (1968) developed two-sample univariate sequential t-test procedures for the case of unpaired observations. Further, Sampson (1968) developed
two-sample univariate sequential tests which utilize a vector of covariates. In Chapter V, we develop two-sample sequential procedures for testing the difference of two mean vectors from independent multivariate normal populations. Chapter VI contains the derivation of the two-sample multivariate sequential procedure for testing the difference of the mean vectors adjusted for covariates.

Further, corresponding to each of the multivariate sequential tests that we have developed, as well as to the single-sample test developed by Jackson and Bradley (1961a), we have developed sequential tests utilizing a type of information that has seemingly been ignored in previous sequential test procedures, that is, the information in a preliminary independent estimate of nuisance parameters. The advantages of incorporating this type of information into the sequential tests will become apparent in subsequent chapters.

B. The Sequential Probability Ratio Test

In this section we shall review the SPRT, which was designed by Wald (1947) to discriminate sequentially between two simple hypotheses. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with probability density function (p.d.f.) $f(x|\theta)$, where $\theta \in \Omega$, the parameter space. Suppose that we wish to test the simple null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative
hypothesis $H_1: \theta = \theta_1$, based on observations $x_1, x_2, \ldots$, with the following requirements

$$P[H_1|H_0] \leq \alpha \quad (1.1)$$

$$P[H_0|H_1] \leq \beta \quad (1.2)$$

where $P[H_i|H_j]$ is the probability of accepting $H_i$ when $H_j$ is true.

It is well known that, for a fixed sample size test, the optimum solution to this testing problem is given by the Neyman - Pearson Lemma, which states that, for a given $n$, the most powerful test (that is the test with smallest $\beta$) depends on the likelihood ratio $L_n$,

$$L_n = \prod_{i=1}^{n} \frac{f(x_i|\theta_1)}{f(x_i|\theta_0)}. \quad (1.3)$$

The test decides for or against $H_0$ according as $L_n$ is less than or greater than a constant $c$, which is chosen so that the test satisfies (1.1), while $n$ can be chosen so that the test satisfies (1.2).

If the sample size is not fixed in advance but is allowed to depend on the observations as they become available, the best testing procedure, as noted in Theorem 1.6, is Wald's SPRT, which is defined by the following rules:
(i) continue sampling as long as \( L_n \) satisfies the inequality
\[
B < L_n < A, \tag{1.4}
\]
where \( B < A \) are two given constants;

(ii) stop sampling and accept \( H_0 \) as soon as
\[
L_n < B; \quad \text{and}
\]
(iii) stop sampling and accept \( H_1 \) as soon as
\[
L_n > A.
\]

Usually, the constants \( A \) and \( B \) satisfy the inequality
\[
0 < B < 1 < A \tag{1.5}
\]
and can be chosen so that the prescribed probabilities of error \( \alpha \) and \( \beta \) are approximately obtained; in this regard, Theorems 1.2 and 1.3 given below are appropriate.

**Theorem 1.1** (Wald, 1947): The SPRT as defined above terminates with probability one.

**Theorem 1.2** (Wald, 1947): The following inequalities hold.
\[
A < \frac{1 - \beta}{\alpha}, \tag{1.6}
\]
\[
B > \frac{\beta}{1 - \alpha}. \tag{1.7}
\]

**Theorem 1.3** (Wald, 1947): If the probabilities of error \( \alpha \) and \( \beta \) are small, and if \( A \) and \( B \) are chosen such that
\[
A = \frac{1 - \beta}{\alpha}, \tag{1.8}
\]
\[
B = \frac{\beta}{1 - \alpha}, \tag{1.9}
\]
then the actual error probabilities achieved by the SPRT are approximately equal to $\alpha$ and $\beta$. In fact, if the actual values of $P[H_1 | H_0]$ and $P[H_0 | H_1]$ are denoted by $\alpha'$ and $\beta'$ respectively, then

$$\alpha' + \beta' \leq \alpha + \beta.$$  (1.10)

Since the sequential tests that we develop in this thesis are based on dependent functions of the basic observations, it is important to note that Theorem 1.2 and Theorem 1.3 also hold in such cases, as is apparent from the proofs of these theorems as given by Wald (1947). Although it has been suggested (for example, Cox (1952) and David and Kruskal (1956)) that certainty of termination is required in order to use Wald boundaries (Theorem 1.3) for a SPRT, Hall, Wijsman, and Ghosh (1965; hereafter abbreviated HWG) point out that Theorem 1.2 holds regardless of the certainty of termination.

In order to relate an optimum property of the SPRT (Theorem 1.6), it is necessary to state two more well-known results.

**Theorem 1.4** (Wald, 1947): For a SPRT as defined above, the operating characteristic (OC) curve is approximately

$$P(\theta | \alpha, \beta, \theta_0, \theta_1) = \frac{A^h(\theta) - 1}{A^h(\theta) - B^h(\theta)}$$  (1.11)
where \( P(\theta) = P(\theta|\alpha, \beta, \theta_0, \theta_1) \) is the probability of accepting \( H_0 : \theta = \theta_0 \) when \( \theta \in \Omega \) and \( h(\theta) \) is the solution of
\[
\mathbb{E}_\theta \left( \frac{\bar{f}(X|\theta_1)}{\bar{f}(X|\theta_0)} \right) h(\theta) = 1 \tag{1.12}
\]

**Theorem 1.5** (Wald, 1947): An approximation to the average sample number (ASN) of the SPRT defined by (1.4), with probabilities of error \( \alpha \) and \( \beta \), is given for any parameter point \( \theta \in \Omega \) by
\[
E(n|\theta) = \frac{P(\theta) \ln B + [1 - P(\theta)] \ln A}{E(\ln \bar{f}(X|\theta_1) - \ln \bar{f}(X|\theta_0)|\theta)} \tag{1.13}
\]

where \( P(\theta) \) and \( 1 - P(\theta) \) are the probabilities that \( L_n \) takes the values \( B \) and \( A \), respectively.

**Theorem 1.6** (Wald and Wolfowitz, 1948): For all sequential tests of \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta = \theta_1 \) having probabilities of error \( \alpha \) and \( \beta \), the SPRT has the least possible values of \( E(n|\theta_0) \) and \( E(n|\theta_1) \).

It should be noted that the SPRT does not necessarily achieve the least possible values of \( E(n|\theta) \) uniformly over \( \Omega \). In fact, at values of \( \theta \in \Omega \) between \( \theta_0 \) and \( \theta_1 \), \( E(n|\theta) \) may even be greater than the sample size required by a fixed sample size plan with the same probabilities of error. For an example in which this occurs, the reader is referred to Wetherill (1966, p. 23).
C. Extensions of the SPRT

As mentioned earlier, the SPRT was developed to discriminate sequentially between two simple hypotheses. When even one of the hypotheses is composite, some extension to the SPRT is needed. One such extension is based on Wald's method of weight functions, which seems particularly well suited for composite hypotheses concerning ranges of parameters. Some relevant sequential tests based on Waldian weight functions are Wald's (1947) sequential t-test and the sequential t-tests developed by Sampson (1968). For further discussion of the theory of weight functions in sequential analysis, the reader is referred to Wald (1947) or Wetherill (1966).

For hypotheses-testing problems in which there are unknown nuisance parameters, many extensions of the SPRT have relied heavily upon fixed sample size reduction principles such as sufficiency and invariance. Goldberg, as reported by Nandi (1948), proposed a method of frequency functions which has been used considerably in developing sequential tests of composite hypotheses; in particular, those tests of special interest to us are: Rushton's (1950) sequential t-test, Hajnal's (1961) two-sample sequential t-test, and Jackson and Bradley's (1961a) sequential $\chi^2$- and $T^2$-tests. For the method of frequency functions, the "observations" are successive values of a test-statistic, which are generally not independent.
This method has been described by D. R. Cox (1952), who gave conditions under which the joint density of n terms in the sequence of test statistics factors conveniently. However, as pointed out by HWG (1965), Cox's theorem is imprecisely stated in that a vital assumption has been omitted.

A much more elegant—and often much less laborious—approach to the problem of sequential tests of composite hypotheses is a method based directly on the application of sufficiency and invariance principles to sequential analysis. The main result in this regard is an unpublished theorem due to Charles Stein; HWG (1965) give an excellent exposition on the Stein Theorem and its application to sequential tests of composite hypotheses. Since it is by this latter method that we have developed our sequential tests a detailed review and discussion of the pertinent theory will be given in Chapter II.
II. RESULTS PERTINENT TO THE DEVELOPMENT OF MULTIVARIATE SEQUENTIAL TESTS OF STATISTICAL HYPOTHESES

A. Introduction

Many hypotheses-testing problems exhibit features that provide natural restrictions on the statistical test procedures to be employed. Suppose, for example, that the independent random variables $X_1, \ldots, X_n$ are each normally distributed with mean $\mu$ and variance $\sigma^2$. Consider the problem of testing $H_0: \lambda = \frac{\mu}{\sigma} \leq \lambda_0$ against $H_1: \lambda > \lambda_0$. Since, for any positive constant $c$, the random variables $cX_1, \ldots, cX_n$ are independent, each following a normal distribution with mean $c\mu$ and variance $(c\sigma)^2$ so that $\lambda$ is unchanged, it seems natural to choose a test function, say $\phi$, with the restriction that $\phi(cX_1, \ldots, cX_n) = \phi(X_1, \ldots, X_n)$.

The feature, such as that described in the above hypotheses-testing problem, is expressed mathematically as invariance under a suitable group of transformations. For the example given above, we shall see (Example 2.2) that the invariance of the problem under the group of positive-scale transformations permits one to restrict attention to the normalized data,

$$U(\mathbf{x}) = \left\{ \frac{X_1}{z}, \ldots, \frac{X_n}{z} \right\},$$

where $z^2 = \sum_{i=1}^{n} x_i^2$. 


Accordingly, the principle of invariance is often thought of as a reduction principle in that application of it condenses the data to a few statistics which can be used for purposes of drawing statistical inferences.

Another useful reduction principle—and one that is probably more familiar—is sufficiency, by which, loosely speaking, the whole of the relevant information contained in the data is condensed in a few sufficient statistics. Often, a statistical testing problem will permit a reduction of the data by both invariance and sufficiency. For situations in which it is possible to apply both reductions in either order, an important question is the following: when is sufficiency followed by invariance equivalent to invariance followed by sufficiency? This and related questions are treated by HWG (1965) who investigate in what sense sufficiency properties are preserved under the invariance principle.

Further, their interpretation of the sufficiency of a statistic in the presence of nuisance parameters is seen to greatly facilitate the derivation of many sequential tests of composite hypotheses. In Section B and Section C of this chapter, we shall explore these considerations more fully.
B. Fixed-Sample-Size Sufficiency and Invariance

In this chapter, a probability model for a random variable X will be represented by \( X_\theta = (\mathcal{X}, \mathcal{A}, P_\theta) \) where \( \mathcal{X} \) is a sample space of points, \( \mathcal{A} \) is a given \( \sigma \)-algebra of subsets of \( \mathcal{X} \), and \( P_\theta \) is a probability measure on \( \mathcal{A} \). The class of probability models indexed by \( \theta \) is represented by \( X_\Omega = \{X_\theta : \theta \in \Omega\} \). In order to facilitate the discussion of the principles of sufficiency and invariance and their application to sequential tests of statistical hypotheses, we will now list several definitions and results, many of which are well-known.

**Definition 2.1:** A set \( G \) of elements is called a group if

(i) there is defined a binary operation which associates, with any two elements \( g_1, g_2 \in G \), a third element \( g_3 \in G \), denoted by \( g_1g_2 = g_3 \);

(ii) \( g_1(g_2g_3) = (g_1g_2)g_3 \) for any \( g_1, g_2, g_3 \in G \);

(iii) there exists an element \( g_0 \), called the identity element, such that \( gg_0 = g_0g = g \) for every \( g \in G \); and

(iv) for each \( g \in G \), there exists an element \( g^{-1} \in G \), called the inverse of \( g \), such that \( gg^{-1} = g^{-1}g = g_0 \)

(Lehmann, 1959, p. 348).

**Definition 2.2:** A class of probability models \( X_\Omega \) is invariant under a group \( G \) of one-to-one (measurable) transformations from \( \mathcal{X} \) onto itself if each \( g \in G \) induces a unique
transformation \( g \in G \) from \( \Omega \) onto itself such that \( g(\theta) = \theta' \in \Omega \) and

\[
P_\theta(g(X) \in A) = P_{g(\theta)}(X \in A), \ A \in \mathcal{A}, \ \theta \in \Omega; \quad (2.1)
\]

invariance of \( X_\Omega \) under a group \( G \) will be represented symbolically by \( gX_\Omega = X_\Omega \) (HWG, 1965, p. 578).

Although the principle of invariance has been applied to many problems in statistical theory and methodology, this thesis is concerned only with its application to hypotheses-testing problems. In that respect, the following two definitions are pertinent.

**Definition 2.3:** The problem of testing \( H_0 : \theta \in \Omega_0 \) against \( H_1 : \theta \in \Omega - \Omega_0 \) remains invariant under a group \( G \) of transformations if, in addition to the invariance of the class of models \( X_\Omega, X_{\Omega_0} \) is also invariant under \( G \); that is, \( gX_{\Omega_0} = X_{\Omega_0} \) for every \( g \in G \) (Lehmann, 1959, p. 214).

**Definition 2.4:** The problem of testing \( H_0 : \theta \in \Omega_0 \) against \( H_1 : \theta \in \Omega_1 \), where \( \Omega_0 \cup \Omega_1 \) is a proper subset of \( \Omega \), remains invariant under a group \( G \) of transformations if, in addition to the invariance of the class of models \( X_\Omega \), both \( X_{\Omega_0} \) and \( X_{\Omega_1} \) are invariant under \( G \); that is, \( gX_{\Omega_0} = X_{\Omega_0} \) and \( gX_{\Omega_1} = X_{\Omega_1} \) for every \( g \in G \) (Ghosh, 1970, p. 59).

**Definition 2.5:** A point \( x' \in \mathcal{X} \) is equivalent to a point \( x \in \mathcal{X} \) under a group of transformations \( G \) if \( x' = g(x) \) for at least one \( g \in G \) (Ferguson, 1967, p. 149).
Thus, the group $G$ partitions $\mathcal{X}$ into equivalence classes or orbits, which are defined formally as follows.

**Definition 2.6:** The orbit of $x \in \mathcal{X}$ (more precisely, the $G$-orbit of $x$) is $G(x) = \{g(x) : g \in G\}$; similarly, the $\mathcal{G}$-orbit of $\Theta \in \mathcal{G}$ is $\mathcal{G}(\Theta) = \{\tilde{g}(\Theta) : \tilde{g} \in \mathcal{G}\}$ (Wijsman, 1967a, p. 391).

**Definition 2.7:** A function $t$ on $\mathcal{X}$ is invariant under a group $G$ of transformations if $t(g(x)) = t(x)$ for all $x \in \mathcal{X}$ and $g \in G$ (Lehmann, 1959, p. 215).

**Definition 2.8:** A function $t$ on $\mathcal{X}$ is a maximal invariant with respect to a group $G$ of transformations if

(i) (Invariance) $t$ is invariant under $G$, and

(ii) (Maximality) for all $x, x' \in \mathcal{X}$, $t(x) = t(x')$ implies $x' = g(x)$ for some $g \in G$ (Ferguson, 1967, p. 243).

Therefore, from the above definitions, it follows that a function is invariant if, and only if, it is constant on each orbit and that an invariant function is a maximal invariant if, and only if, it assumes different values on different orbits. Further, it is well-known (see, for example, Ghosh, 1970, p. 59) that maximal invariants under a group $G$ always exist but that their functional representation may not be unique.

The following examples give some maximal invariants on $\mathcal{X} = \mathbb{R}^n$ (Euclidean n-space) under specific groups of transformations, the multivariate analogues of which we shall encounter in subsequent chapters.
Example 2.1: (Location invariance) Suppose that $G$ is the group of translations, an element of which adds a constant $b$ to each $x_i$,

$$g(x_1, \ldots, x_n) = (x_1 + b, \ldots, x_n + b), \quad -\infty < b < \infty.$$ 

If $n = 1$, $X$ is a single orbit so that the only invariant functions are then the constant functions $t(x_1) = b$. If $n > 1$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, then the $n$-tuple $t(x_i) = (x_i - \bar{x}, \ldots, x_n - \bar{x})$ is a maximal invariant on $X$. Since $(\bar{x} + b) = \bar{x} + b$ and since $(x_i + b) - (\bar{x} + b) = x_i - \bar{x}$ for all $b$, $t(x)$ is invariant under $G$. If $t(x) = t(x')$ so that $x_i - \bar{x} = x'_i - \bar{x}'$, then $x'_i = (x_i + b)$ where $b = (\bar{x}' - \bar{x})$ for $i = 1, \ldots, n$, thereby proving maximality.

The function $t(x)$ is only one representation of the maximal invariant. Using similar arguments, one can verify that, for $n > 1$, the vectors $(x_1 - x_2, \ldots, x_{n-1} - x_n)$ and $(x_2 - x_3, \ldots, x_n - x_1)$ are also maximal invariants under the group $G$.

Example 2.2: (Scale invariance) Suppose that $G$ is the group of positive scale changes, an element of which multiplies each $x_i$ by a specific constant $c$,

$$g(x_1, \ldots, x_n) = (cx_1, \ldots, cx_n), \quad 0 < c < \infty.$$
Let $z^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$; then, a maximal invariant with respect to $G$ is

$$t(x) = \left(\frac{x_1}{z}, \ldots, \frac{x_n}{z}\right).$$

Since $\frac{1}{n} \sum_{i=1}^{n} (cx_i)^2 = c^2 z^2$, $t(x)$ is clearly invariant. Suppose that $t(x) = t(x')$; then, $\frac{x_i}{z} = \frac{x_i'}{z'}$, for $i = 1, \ldots, n$, so that $x_i' = cx_i$ where $c = z/z' > 0$. We have ignored the possibility here that $z$ might equal zero since, in most applications, one would have $P(Z = 0) = 0$.

Quite often, it is convenient to obtain a maximal invariant in a number of steps, each step corresponding to a subgroup of the group $G$. Lehmann (1959, p. 218) gives a stepwise method of finding the maximal invariant with respect to a group of transformations $G$ which can be generated as the smallest group containing two subgroups $H$ and $K$. Lehmann's result is given in Theorem 2.1, which will be of use in Example 2.3 and later in Chapters V and VI.

**Theorem 2.1:** Let $G$ be a group of transformations generated by two subgroups $H$ and $K$. Suppose that $y = s(x)$ is maximal invariant with respect to $H$ and that, for any $k \in K$,

$$s(x) = s(x') \implies s(k(x)) = s(k(x')). \quad (2.2)$$

If $z = t(y)$ is maximal invariant under the group $K^*$ of transformations $k^*$ defined, for $k \in K$, by

$$k^*(y) = s(k(x)) \quad \text{when} \quad y = s(x), \quad (2.3)$$

then $z = t[s(x)]$ is maximal invariant with respect to $G$. 
Example 2.3: (Location and scale invariance) Let \( \mathbf{X} = \mathbb{R}^n \), with \( n > 1 \). Suppose that \( G \) is the group of location and scale changes, an element of which is

\[
g(x_1, \ldots, x_n) = (cx_1 + b, \ldots, cx_n + b), \quad -\infty < b < \infty
\]

and \( c > 0 \).

Let \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) and \( d^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \); then, a maximal invariant with respect to \( G \) is \( u(x) = \left( \frac{x_1 - \bar{x}}{d}, \ldots, \frac{x_n - \bar{x}}{d} \right) \), as verified by the following arguments.

With respect to the above theorem, location change and scale change correspond to the subgroups \( H \) and \( K \), respectively. Now, as shown in Example 2.1, a maximal invariant under \( H \) is \( y = s(x) = (x_1 - \bar{x}, \ldots, x_n - \bar{x}) \). Since the group of transformations \( K^* \) in the above theorem becomes the group of positive scale changes of the \( y_i \)'s, we have, from Example 2.2, that the maximal invariant under \( K^* \) is \( t(y) = \left( \frac{y_1}{z}, \ldots, \frac{y_n}{z} \right) \), where \( z^2 = \sum_{i=1}^{n} y_i^2 \). Thus, by Theorem 2.1, the maximal invariant under the group \( G \) is \( t[s(x)] = u(x) \), as defined above.

If a hypotheses-testing problem is invariant under a group of transformations \( G \), it is quite natural to restrict attention to those tests in which test functions are invariant under \( G \). The rationale is that, since \( G \) partitions \( \mathbf{X} \) into
equivalence classes or orbits, the orbits rather than the individual points of $\mathcal{X}$ should be the basic elements in the construction of a test of a hypothesis $H_0$. The main simplification in the application of the principle of invariance to hypotheses-testing problems is that it is possible to describe all invariant tests as tests that are functions of the maximal invariant, since these two classes of tests are equivalent. This property is stated explicitly in the following theorem, the proof of which is given, for example, by Lehmann (1959, p. 216).

\textbf{Theorem 2.2}: Let $\mathcal{X}$ be a space, let $G$ be a group of transformations on $\mathcal{X}$, and let $t(x)$ be a maximal invariant with respect to $G$. Then, a function $\varnothing(x)$ is invariant under $G$ if, and only if, there exists a function $h$ such that $\varnothing(x) = h(t(x))$ for all $x \in \mathcal{X}$.

As mentioned previously, the parameter space $\Omega$ is partitioned into orbits by the induced group of transformations $\tilde{G}$. Hence, there always exists a maximal invariant on $\Omega$ with respect to $\tilde{G}$, which we shall denote by $\lambda(\theta)$. We then have the following very useful result given, for example, by Ferguson (1967, p. 245).

\textbf{Theorem 2.3}: If $t(x)$ is invariant under a group $G$ of transformations and if $\lambda(\theta)$ is a maximal invariant with respect to the induced group $\tilde{G}$, then the distribution of $T = t(X)$ depends on $\theta$ only through $\lambda(\theta)$. 
Thus, since the distribution of any invariant statistic depends on $\lambda = \lambda(\theta)$, $\theta \in \Omega$, we will represent the probability model corresponding to an invariant statistic $T = t(X)$ by

$$T_\lambda = (\mathcal{F}, \mathcal{A}^t, \mathcal{P}^t), \quad \lambda \in \Lambda = \lambda(\Omega)$$

(2.4)

where $\mathcal{F} = \{t(x) : x \in X\}$,

$$\mathcal{A}^t = \{A^t : t^{-1}(A^t) \in \mathcal{A}\},$$

and

$$\mathcal{P}^t$$

is such that $\mathcal{P}^t[t(X) \in A^t] = \mathcal{P}_\theta[X \in t^{-1}(A^t)]$.

The class of probability models indexed by $\lambda$ is denoted by

$$T_\Lambda = \{T_\lambda : \lambda \in \Lambda\}.$$

To aid in interpreting the principle of sufficiency under an invariance reduction, we give the following definition and discussion.

**Definition 2.9:** A set $A \in \mathcal{A}$ is an invariant set (under a group of transformations $G$) if $x \in A$ implies $g(x) \in A$ for every $g \in G$ (HWG, 1965, p. 579).

Since the image of any set $A$ under a transformation $g$ can be written as $g(A) = \{g(x) : x \in A\}$, one can see that $g(A) = A$ for any invariant set $A \in \mathcal{A}$. Further, if we write an invariant set $A$ as $A = \{x : g(x) \in A, \forall g \in G\}$ and let $A' = \{x : u(x) \in A^u\}$, where $u$ is a maximal invariant with respect to $G$, then it follows from the invariance of $u$ under $G$ that $A = A'$. The importance of these considerations will
become evident when we relate an interpretation of invariant sufficiency (Definition 2.13). We now give a precise statement of the sufficiency of a statistic.

**Definition 2.10:** A statistic $S = s(X)$ on $X$ is sufficient for $X_\Omega$ if, for every $A \in \mathcal{A}$ and $s_o \in S = s(X)$, there is a version of the conditional probability $P_\theta (A | s_o) = P_\theta (X \in A | s(X) = s_o)$ which does not depend on $\theta$ (HWG, 1965, p. 579). Further, a sufficient statistic $S$ for $X_\Omega$ is a minimal sufficient statistic for $X_\Omega$ if, for any other sufficient statistic $S'$ for $X_\Omega$, $S$ is a function of $S'$, almost everywhere with respect to $\{(\mathcal{A}, P_\theta), \theta \in \Omega\}$ (Lehmann and Scheffé, 1950, p. 311).

The probability model corresponding to a sufficient statistic $S = s(X)$ is denoted by

$$ S = (S, \mathcal{A}^S, P^S_\theta), \theta \in \Omega, \quad (2.5) $$

where $S$, $\mathcal{A}^S$, and $P^S_\theta$ are respectively the sample space, the $\sigma$-algebra, and the probability measure associated with $S$. The corresponding class of probability models is denoted by $S_\Omega = \{S_\theta : \theta \in \Omega\}$. Computation of the conditional probability $P(A | s_o)$ in order to determine whether or not a statistic $S$ is sufficient for $X_\Omega$ is, at the very least, inconvenient; a much simpler check is provided by the widely-used factorization criterion, which can be found, for example, in (Lehmann, 1959, p. 50) and is stated here in Theorem 2.4.
Theorem 2.4: If the distributions \( \{ \mathbb{P}_\theta, \theta \in \Omega \} \) have densities \( \mathbb{P}_\theta = \frac{d\mathbb{P}_\theta}{d\mu} \) with respect to a \( \sigma \)-finite measure \( \mu \), then a statistic \( S \) is sufficient for \( X_\Omega \) if, and only if, there exists non-negative \( \mathcal{A}^S \)-measurable functions \( g_\theta \) on \( S = s(\mathbb{X}) \) and a non-negative \( \mathcal{A} \)-measurable function \( h \) on \( \mathbb{X} \) such that

\[
p_\theta(x) = g_\theta[s(x)]h(x),
\]

almost everywhere with respect to \( (\mathcal{A}, \mu) \).

Although the following definition is given for an arbitrary statistic \( S = s(X) \), our attention will be restricted throughout to the situation when the statistic \( S \) is sufficient for \( X_\Omega \).

Definition 2.11: Let \( G \) be a group of transformations on \( \mathbb{X} \) and let \( S = s(X) \) be any statistic on \( \mathbb{X} \) for which \( s(g(x)) = s(g(x')) \), whenever \( s(x) = s(x') \). Then, we say that \( G \) induces a group \( G^S \) of transformations \( g^S \) on \( S \), where \( g^S \) is defined by \( g^S(s') = s(g(x')) \), for \( s' \in S \) and \( x' \) satisfying \( s(x') = s' \) (HWG, 1965, p. 579).

Any sufficient statistic that we consider has the property that the group \( G \) of transformations of interest induces a group \( G^S \) of transformations on the sample space of the sufficient statistic \( S = s(X) \). Further, in each of our hypotheses-testing problems, we have that \( G^S \) on \( S \) is completely analogous to the group \( \tilde{G} \) induced by \( G \) on the parameter space \( \Omega \). The following result, given, without
proof, by HWG (1965, p. 579), allows one to define the invariance reduction on \( S_\Omega \) rather than on \( X_\Omega \).

**Lemma 2.1:** Suppose \( G \), a group of transformations on \( X \), induces a group \( G_S \) of transformations on \( S = s(X) \). Then, if \( u_s \) is invariant on \( S \) under \( G_S \), \( u = u_s(s) = u_s(s(x)) \) is invariant on \( X \) under \( G \).

This lemma may be proved as follows:

**Proof:** For each \( g \in G \), we have

\[
u(gx) = u_s(s(gx)), \text{ where } gx \equiv g(x)
\]

\[
= u_s(g_s(s)), \text{ by the definition of } g_s
\]

\[
= u_s(s), \text{ by the invariance of } u_s \text{ under } G_S
\]

\[
= u(x).
\]

Thus, \( u = u(x) \) is invariant under \( G \). q.e.d.

Summarizing the development to this point, we have a class of probability models \( X_\Omega \), a group \( G \) of one-to-one transformations on the sample space \( X \) which leaves \( X_\Omega \) invariant, and a sufficient statistic \( S = s(X) \) for which \( G \) induces a group \( G_S \) of transformations on the sample space \( S \) of \( S \). We now define invariant sufficiency.

**Definition 2.12:** A statistic \( V = v(X) \) on \( X \) is invariantly sufficient for \( X_\Omega \) under a group of transformations \( G \) if

(i) \( v \) is invariant under \( G \), and

(ii) the conditional probability of any invariant set \( A \) given \( v \) is parameter free for \( \theta \in \Omega \) (HWG, 1965, pp. 579, 580).
By Theorem 2.2, we may write \( v = v_u(u) \) where \( u \) is a maximal invariant under \( G \) and \( v_u \) is a function on \( \mathcal{U} \), the sample space of the maximal invariant. Thus, since an invariant set \( A \) has the form \( \{ x : u(x) \in A^u \} \) and since \( P^u_{\lambda(\theta)}(U \in A^u | V_u = v_o) = P_\theta(u(X) \in A^u | V(X) = v_o) \), condition (ii) of Definition 2.12 is equivalent to saying that the conditional probability of \( A^u \in A^u \) given \( v_u \) does not depend on \( \lambda = \lambda(\theta) \); hence, an equivalent condition is

\[(ii') \quad v_u \text{ is sufficient for } U_\Lambda, \quad \Lambda = \lambda(\Omega).\]

For these reasons, HWG (1965) interpret the definition of invariant sufficiency as saying that \( V \) is sufficient for \((V, T)\) where \( T \) is any invariant statistic so that, loosely speaking, \( V \) contains all the information about \( \lambda \) that is available in any invariant statistic.

As indicated by the above discussion, an invariantly sufficient statistic \( V \) may be obtained by first making a maximal invariance reduction on \( X \) and then making a sufficiency reduction on \( \mathcal{U} \), the sample space of the maximal invariant statistic \( U \). In most applications, however, the determination of a sufficient statistic for \( U_\Lambda \) may be quite difficult. In many of these situations, \( V \) can be found by first making a sufficiency reduction on \( X \) and then making a maximal invariance reduction on \( S \), the sample space of the sufficient statistic \( S \). These two routes are symbolically expressed in the following diagram:
That the outcomes of these two routes are the same under certain conditions is the content of the following two theorems, due to C. Stein and given, for example, by HWG (1965, p. 580) and Ghosh (1970, p. 296).

**Theorem 2.5:** Let \( S = s(X) \) be a sufficient statistic for \( X_\Omega \), and let \( u_S \) be a maximal invariant function on \( S = s(X) \) under \( G_S \), the group induced on \( S \) by \( G \) which leaves \( X_\Omega \) invariant. If \( S \) is a discrete random variable, then \( V = v(X) = u_S(S) \) is invariantly sufficient for \( X_\Omega \) under \( G \).

For the case when the sufficient statistic \( S \) is not discrete, the assertion of Theorem 2.5 still holds, provided one of three assumptions is satisfied. This result, stated in Theorem 2.6 is the second part of the Stein Theorem. First, however, two additional definitions are needed.

**Definition 2.13:** A function \( \varphi(x) \) on \( X \) is said to be equivalent to an invariant function under \( G \) if there exists an invariant function \( \psi \) such that \( \varphi(x) = \psi(x) \) for all \( x \in X - N \) where \( P_\theta(N) = 0, \theta \in \Omega \) (Lehmann, 1959, p. 225).

**Definition 2.14:** A function \( \varphi(x) \) on \( X \) is said to be almost invariant under \( G \) if, for each \( g \in G \), \( \varphi(gx) = \varphi(x) \) for all \( x \in X - N_g \) where \( P_\theta(N_g) = 0, \theta \in \Omega \) (Lehmann, 1959, p. 225).
Theorem 2.6: Let \( S \) and \( u_s \) be as defined in Theorem 2.5. Then, \( V = v(X) = u_s(S) \) is invariantly sufficient for \( X_\Omega \) under \( G \), provided one of the following three assumptions is satisfied:

**Assumption 2.1:** Every almost invariant function under \( G_S \) is equivalent to an invariant function under \( G_S \) on \( S \).

**Assumption 2.2:** For each \( g \in G \), there exists an invariant conditional probability \( P(A|s) = P(gA|g_s s) \), for every invariant set \( A \subset X \) and for all \( s \in S \).

**Assumption 2.3:** \( X \) has a multivariate (nonsingular) continuous distribution, for which the region of positive probability does not vary with \( \theta \) and for which the joint density function of \( X \) can be factored as \( f_\theta(s(x))h(x) \) so that, for each \( g \in G \) and \( x \in X - A^\circ \), where \( A^\circ \) is an invariant set having probability zero and satisfying \( s(x) \neq s(x') \) if \( x \in A^\circ \) but \( x' \in X - A^\circ \), the following conditions are satisfied:

(i) each transformation \( g \) is continuously differentiable with a Jacobian depending on \( x \) only through \( s(x) \);

(ii) \( s(x) \) is continuously differentiable with matrix of partial derivatives of maximal rank;

(iii) \( \frac{h(gx)}{h(x)} \) depends on \( x \) only through \( s(x) \).

As implied by the definitions, a function \( \phi(x) \) equivalent to an invariant function is also almost invariant; however, the converse may not hold. Lehmann (1959, p. 225) shows that a sufficient condition for Assumption 2.1 is the existence of
an invariant measure on $G$, which is commonly the situation in parametric problems. In order, however, to establish the existence of such an invariant measure on the group $G$, it is necessary to consider the topological structure of the group $G$ of interest. Instead, we will show that Assumption 2.1 is satisfied in each of our problems by application of Theorem 2.7, which will be given following the definition of the completeness of a sufficient statistic $S$ for $X_\Theta$.

**Definition 2.15:** A sufficient statistic $S$ for $X_\Theta$ is (boundedly) complete if, for every (bounded) real-valued function $f$, $E_\theta[f(S)] = 0$ for all $\theta \in \Omega$ implies $P_\theta[f(S) = 0] = 1$ for all $\theta \in \Omega$ (Ferguson, 1967, p. 132).

**Theorem 2.7:** If $S$ is sufficient and complete for $X_\Theta$, then every almost invariant function under $G_\Theta$ is equivalent to an invariant function under $G_\Theta$.

Berk and Bickel (1968, p. 1573) state and prove a version of Theorem 2.7 in which attention is restricted to bounded invariant functions. More recently, however, Berk (1972b) has shown that their result also holds for arbitrary invariant functions.

As pointed out by HWG (1965), Theorem 2.6 under Assumption 2.3 may be considered to be a rigorous version of D. R. Cox's (1952) Theorem, which, as originally stated, did not contain the necessary assumption of invariance of the probability model. The assumption of invariance of the
probability model is thus included—as condition (iv, c)—in the following statement of Cox's Theorem.

**Theorem 2.8:** Let \((Y_1, Y_2, \ldots, Y_n) = Y\) be random variables (possibly vectors) whose probability density function depends on unknown parameters \(\theta_1, \ldots, \theta_p, p < n\). Suppose that

(i) \(Z_1, Z_2, \ldots, Z_p\) constitute a jointly sufficient and functionally independent set of statistics for \(\theta_1, \theta_2, \ldots, \theta_p\);

(ii) the distribution of \(Z_1\) involves \(\theta_1\) but not \(\theta_2, \ldots, \theta_p\);

(iii) \(u_1, u_2, \ldots, u_m, m < n\), are functions of \(Y\), functionally independent of each other and of \(Z_1, \ldots, Z_p\);

(iv) there exists a set \(G\) of transformations of \(y = (y_1, \ldots, y_n)\) into \(y' = (y'_1, \ldots, y'_n)\) such that

(a) the functions \(z_1, u_1, \ldots, u_m\) are invariant under \(G\),

(b) the transformation of \(Z_2, \ldots, Z_p\) into \(Z'_2, \ldots, Z'_p\)

defined by each transformation in \(G\) is one-to-one,

(c) \(gY_\Omega = Y_\Omega\) for all \(g \in G\) (see Definition 2.2),

(d) if \(z_2, \ldots, z_p\) and \(z'_2, \ldots, z'_p\) are two sets of values of \(Z_2, \ldots, Z_p\) each having non-zero probability density under at least one of the distributions of \(Y, \theta \in \Omega\), then there exists a transformation in \(G\) such that if \(Z_2 = z_2, \ldots, Z_p = z_p\),
Then the joint p.d.f. of $Z_1, U_1, \ldots, U_m$ factorizes into

$$f(z_1 | \theta_1) h(u_1, u_2, \ldots, u_m, z_1) \tag{2.8}$$

where $f(z_1 | \theta_1)$ is the p.d.f. of $Z_1$ and $h(u_1, \ldots, u_m, z_1)$ does not involve $\theta_1$.

It is interesting to note that, although Jackson and Bradley (1961a) set out to test sequentially simple hypotheses in the appropriate composite parameter, their tests (as well as the tests that we develop) actually discriminate between two composite hypotheses. In order to show this in later chapters, we need to recall a well-known corollary of the Neyman-Pearson Lemma.

**Definition 2.16:** The real-parameter family of probability densities $\{p_\theta(x), \theta \in \Omega\}$ is said to have monotone likelihood ratio (MLR) if there exists a real-valued function $r(x)$ such that for any $\theta < \theta'$ the distributions $P_\theta$ and $P_{\theta'}$ are distinct and the ratio $\frac{p_{\theta'}(x)}{p_\theta(x)}$ is a nondecreasing function of $r(x)$ (Lehmann, 1959, p. 68).

**Theorem 2.9:** Let $\theta$ be real-valued, $\theta \in \Omega$, and let the random variable $X$ have probability density $p_\theta(x)$ with MLR in $t(x)$. For testing $H_0 : \theta < \theta_0$ against $H_1 : \theta > \theta_0$, there exists a UMP test which is given by
\[
\phi(x) = \begin{cases} 
1 & \text{when } t(x) > c(a) \\
\gamma(a) & \text{when } t(x) = c(a) \\
0 & \text{when } t(x) < c(a)
\end{cases}
\] (2.9)

where the constants \(c(a)\) and \(0 \leq \gamma(a) \leq 1\) are determined by

\[E_{\Theta_0} [\phi(x)] = \alpha \text{ (Lehmann, 1959, pp. 68, 69).}\]

The application of Theorem 2.9 to the restricted class of invariant tests is fairly obvious: frequently, the maximal invariants \(V = v(X)\) and \(\lambda = \lambda(\theta)\) under \(G\) and \(\bar{G}\), respectively, are real-valued, and the family of probability density functions of \(V\), say \(\{f_\lambda(v), \lambda \in \Lambda\}\), has MLR; hence, for testing \(H_0 : \lambda \leq \lambda_0\) against \(H_1 : \lambda > \lambda_0\), there exists a UMP test among those tests depending only on \(V\) and therefore there exists a UMP invariant (UMPI) test.

Before discussing the application of the results of this section to sequential tests of statistical hypotheses, we shall consider the simple testing problem alluded to in Section A of this chapter.

**Example 2.4:** Let \(X = (X_1, \ldots, X_n)\), \(n > 1\), where the \(X_i\)'s are independent normally distributed random variables with \(\theta = (\mu, \sigma) \in \Omega\), the upper half-plane. Denote the sample mean and sample standard deviation by \(\bar{x}\) and \(d\) respectively, as defined in Examples 2.1 and 2.3. Let \(G\) be the group of positive scale changes given in Example 2.2. One can easily verify that a maximal invariant under \(G\) on \(\chi = \mathbb{R}^n\) is \(u(x) = \left(\frac{x_1}{d}, \ldots, \frac{x_n}{d}\right)\) (see Example 2.2). Consider the problem of testing \(\lambda \leq \lambda_0\)
against $\lambda > \lambda_0$ at level $\alpha$, where $\lambda = \lambda(\theta) = \frac{H}{\theta}$. The induced group $\tilde{G}$ consists of all positive scale changes on $\Omega$ so that $\tilde{g}(\theta) = (c\mu, c\sigma)$ for $c > 0$; thus, $X_\Omega$ remains invariant under $G$. A maximal invariant under $\tilde{G}$ is found to be $\lambda(\theta) = \frac{H}{\theta}$ so that, clearly, the testing problem remains invariant under $G$. Thus, by Theorem 2.2, any invariant test of $H_0 : \lambda \leq \lambda_0$ depends on $u(X) = (U_1, \ldots, U_n)$. In order to find the UMPI test without appealing to Theorem 2.6, one needs to consider the joint distribution of $(U_1, \ldots, U_n)$. Instead, by application of Theorem 2.6, we can find the UMPI test much more easily as follows.

It is well known that $x$ and $d$ together form a set of sufficient statistics for $X_\omega$. The induced group $G_{(x,d)}$ also consists of all positive scale changes so that, analogously with $\tilde{G}$ on $\Omega$, a maximal invariant with respect to $G_{(x,d)}$ is $\frac{X}{d}$, or equivalently, $T = \frac{X}{d} \sqrt{n}$, which has a noncentral t-distribution with degrees of freedom $m = (n-1)$ and noncentrality parameter $\delta = \lambda \sqrt{n}$. Since Assumption 2.3 obtains, the set $A_0$ being the line on which $X_1 = \ldots = X_n$, it follows from Theorem 2.6 that $T$ is invariantly sufficient for $X_\omega$.

Also, it can be shown (Kruskal, 1954; or, Ghosh, 1970, pp. 301-302) that the family of noncentral t-distributions possesses MLR in $\frac{t}{\sqrt{n-1+t^2}}$ so that, from Theorem 2.9, there
exists a UMP test among those tests depending only on $T$. In fact, the UMPI test of $H_0$ at level $\alpha$ is defined by the rejection region

$$T = \frac{\bar{x}\sqrt{n}}{d} > t_{n-1,1-\alpha}(\lambda\sqrt{n}),$$

(2.10)

where $t_{m,1-\alpha}(\delta)$ denotes the $100(1 - \alpha)\%$ point of the non-central $t$-distribution with $m$ degrees of freedom and non-centrality parameter $\delta$.

C. Invariance and Sufficiency Principles
Applied to Sequential Analysis

In this section we discuss the concepts of sufficiency and invariance in relation to sequential experimentation, wherein the experiment may be terminated at any stage, but performance of stage $n$ implies the previous performance of stages $1, 2, \ldots, n-1$. Following HWG (1965), we shall distinguish between three types of models:

(i) the component or marginal models

$$(X_n)_\theta = X_n = (\bar{x}_n, \bar{\alpha}_n, \bar{P}_n)$$

for stage $n$ and data $X_n$ ($n = 1, 2, \ldots$);

(ii) the joint (n-fold) models $X_{(n)} = (\bar{x}_{(n)}, \bar{\alpha}_{(n)}, \bar{P}_{(n)})$

for the accumulated data $X_{(n)} = (X_1, \ldots, X_n)$; and

(iii) the sequential model $X_\theta = (\bar{x}, \bar{\alpha}, \bar{P}_\theta)$ for the entire sequence of data $X = (X_1, \ldots, X_n, \ldots)$. 
For the case of independent random variables $\{X_i\}$, $\mathcal{X}_{(n)}$ and $\mathcal{X}$ are the product sample spaces ($n$-fold and infinite, respectively) with components $\mathcal{X}_1, \ldots, \mathcal{X}_n, \ldots$; and for each $\theta \in \Omega$, $P_\theta$ is a probability measure on $(\mathcal{X}, \mathcal{A})$ and $P_{(n)\theta}$ and $P_{n\theta}$ are the corresponding joint and marginal probability measures derived therefrom. We now make the following definitions:

**Definition 2.17**: If, for each $n$, $S_n$ is a sufficient statistic for the class of joint models $X_{(n)\Omega} = \{X_{(n)}\theta : \theta \in \Omega\}$, then $S = (S_1, S_2, \ldots)$ is called a sufficient sequence for $X_\Omega$ (HWG, 1965, p. 583).

**Definition 2.18**: For each $n$, suppose $t_n$ is a function on $\mathcal{X}_{(n)}$ so that $T_n = t_n(X_{(n)})$. If, for all $\theta$ and each $n$, the conditional distribution of $T_{n+1}$ given $X_{(n)}$ is identical with the conditional distribution of $T_n$ given $T_n$ — that is,

$$P_{\theta}(T_{n+1} \leq t|X_{(n)}) = P_{\theta}(T_{n+1} \leq t|T_n),$$  \hspace{1cm} (2.11)

then $T = (T_1, \ldots, T_n, \ldots)$ is said to be a transitive sequence for $X_\Omega$ (HWG, 1965, p. 583).

Thus, transitivity implies that all the information about $T_{n+1}$ contained in $X_{(n)}$ is carried by $T_n = t_n(X_{(n)})$.

Whereas the concepts of sufficiency and transitivity in the sequential model are defined in terms of the sequence of joint models $X_{(n)\theta}$, invariance is more suitably defined in terms of the sequential model $X_\theta$. 
Let $G$ be a group of transformations $g$ on the sequential sample space $\mathcal{X}$ for which $g(X_\omega) = X_\omega$ with maximal invariant $\lambda$ on $\Omega$. We shall also assume that each $g$ induces a transformation $g(n)$ on the joint sample space $\mathcal{X}(n)$ so that $g(n)X(n)\Omega = X(n)\Omega'$, that is, the joint models are also invariant. Let $u_n$ denote the maximal invariant on $\mathcal{X}(n)$ under $G(n)$. Hence, as noted by HWG (1965, p. 583), $G(n)$ induces $G(m)$ for $m < n$ so that $u_m$ considered as a function on $\mathcal{X}(n)$ is invariant under $G(n)$.

Therefore, by Theorem 2.2, since $u_n$ is maximal invariant on $\mathcal{X}(n)$ under $G(n)$, $u_m$ is a function of $u_n$. Thus, knowledge of the value of one term in the sequence $u = (u_1, \ldots, u_n, \ldots)$ allows one to evaluate all prior terms. Hence, as pointed out by HWG (1965), although $u$ itself is not necessarily maximal invariant under $G$, $u$ is relevant in the sequential decision problem since a maximal invariant under $G$ would depend on the entire sequence $X = (X_1, \ldots, X_n, \ldots)$ which, of course, is not available to the decision maker.

It is for these reasons that HWG (1965) interpret the principle of invariance in the sequential case as stipulating that attention be confined to $u$-rules, that is, decision procedures that depend at stage $n$ on the value of $U_n = u_n(X(n))$. In effect, the original sequence of joint probability models $\{X(n)_{\theta}\}$ is replaced by the sequence $\{U_{n\lambda}\}$, where $U_{n\lambda}$ is the probability model for $U_n$. A sufficiency reduction on each of the components $U_n$ of $U = (U_1, U_2, \ldots)$ leads to a sequence
\( V = (V_1, V_2, \ldots) \) which may be considered as an invariently sufficient sequence for the sequential model \( X_\Omega \) under \( G \), each \( V_n \) being invariently sufficient for \( X^{(n)}_\Omega \) under \( G^{(n)} \).

**Definition 2.19**: \( v \)-rules are defined as sequential decision procedures that depend on an invariently sufficient and transitive sequence \( V = (V_1, \ldots, V_n, \ldots) \) (HWG, 1965, p. 584).

Hence, when invoking the principle of invariance in sequential decision problems, Theorem 2.6 provides justification for the restriction to \( v \)-rules so long as \( V \) is transitive for the sequence of probability models \( U_{\Lambda}^{(n)} \). Theorem 2.6 also provides an alternative means of reduction from the sequence \( X = (X_1, X_2, \ldots) \) to the sequence \( V \), assuming that \( G^{(n)} \) induces a group of transformations on the sample space of \( S_n \). There remains, however, the problem of verifying the transitivity of the sequence \( V \). Fortunately, HWG prove that the sequence \( V = (V_1, V_2, \ldots) \) is transitive if the corresponding sequence of sufficient statistics \( S = (S_1, S_2, \ldots) \) is transitive (HWG, 1965, p. 603). Thus, if we append subscripts \( (n) \) to \( X \) and \( U \) and subscripts \( n \) to \( S \) and \( U \) in diagram (2.7), we have that the upper route in this diagram is completely justified when \( S \) is a transitive sequence.

In summary, then, in order to obtain an invariently sufficient and transitive sequence \( V \), we may take the following steps:
(i) make a sufficiency reduction from $X_{(n)}$ to $S_n$;
(ii) verify the transitivity of the sequence $S = \{S_n\}$;
and
(iii) make a maximal invariance reduction from $S_n$ to $V_n$.

The following theorem and corollary provide very useful techniques for verifying the transitivity of a sufficient sequence $S$; a proof of Theorem 2.10 is given by HWG (1965, p. 603) in the $\sigma$-algebra mode, as well as by Ferguson (1967, p. 335) in the random variable mode.

**Theorem 2.10:** Suppose that $X_1, X_2, \ldots$ are mutually independent random variables and that $S = \{S_n\}$ is a sufficient sequence for $X_\Omega$. Then, if there exists a function $h_n$ such that

$$S_{n+1} = h_n(S_n, X_{n+1}) \quad \text{for each } n \geq 1,$$

then $S$ is a transitive sequence for $X_\Omega$.

**Corollary 2.1:** Suppose that $X_1, X_2, \ldots$ are mutually independent random variables and that $S = \{S_n\}$ is a sufficient sequence for $X_\Omega$, where $S_n = (S_{1n}, \ldots, S_{nn})$. Then, if there exist functions $h_{1n}, \ldots, h_{nn}$ such that, for each $n \geq 1$,

$$S_{1,n+1} = h_{1n}[S_n, X_{n+1}]$$

$$S_{j,n+1} = h_{jn}[S_n, (S_{1,n+1}, \ldots, S_{j-1,n+1}), X_{n+1}]$$

then $S$ is a transitive sequence for $X_\Omega$. 
Proof: Clearly, if there exist functions \( h_{1n}, \ldots, h_{mn} \) such that (2.13) and (2.14) hold for each \( n \geq 1 \), then \( S_{n+1} \) can be written as a single function, say \( h_n \), of \( S_n \) and \( X_{n+1} \), in which case the sequence \( S \) is transitive for \( X_\Omega \) by Theorem 2.10.

Consider now the problem of discriminating sequentially between the two hypotheses

\[ H_o : \lambda = \lambda_o, \quad H_1 : \lambda = \lambda_1, \quad \lambda_o \neq \lambda_1, \quad (2.15) \]

where \( \lambda = \lambda(\theta) \) is a maximal invariant on \( \Omega \) under the induced group of transformations \( \bar{G} \). Since \( V_n \) is sufficient for the distribution of \( V_{(n)} = (V_1, \ldots, V_n) \), it follows from Theorem 2.4 that

\[ \frac{f_n(v_n; \lambda_1)}{f_n(v_n; \lambda_o)} = \frac{f(n)(v_1, \ldots, v_n; \lambda_1)}{f(n)(v_1, \ldots, v_n; \lambda_o)} \quad (2.16) \]

where \( f(n) \) is the joint density of \( V_{(n)} \) and \( f_n \) is the marginal density of \( V_n \). Hence, a SPRT based on the sequence \( V = (V_1, V_2, \ldots) \) depends only on \( V_n \), not \( V_{(n)} \), at stage \( n \) and is therefore a \( v \)-rule. Furthermore, if \( \lambda \) is real-valued, and the density of \( V_n \) possesses MLR, then by Theorem 2.9 a SPRT of the two hypotheses in (2.15) effectively tests

\[ H_o : \lambda \leq \lambda_o, \quad H_1 : \lambda \geq \lambda_1, \quad \lambda_o < \lambda_1. \quad (2.17) \]
However, since the $V_n$'s are not usually independent, SPRT's applied to them do not generally have any known optimum property, such as that described in Theorem 1.6. For a discussion of the properties that are known in general, the reader is referred to HWG (1965, pp. 586, 587). These properties, as relative to our sequential tests, will be discussed in Chapter VII.

D. Results from Multivariate Statistical Analysis

Before proceeding to the development of the various multivariate sequential test procedures, we need to catalogue some standard results from multivariate statistical analysis. Further, we derive a convenient form of the probability density of Hotelling's $T^2$-statistic, as well as verify the MLR property of the corresponding family of densities.

Definition 2.20: If the $p$-dimensional random vector $\mathbf{X}$ has probability density given by

$$f(\mathbf{x}) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left( -\frac{1}{2} (\mathbf{x} - \mathbf{\mu})' \Sigma^{-1} (\mathbf{x} - \mathbf{\mu}) \right) \quad (2.18)$$

where $\Sigma$ is a $p \times p$ positive definite matrix, then we shall say that $\mathbf{X}$ has a multivariate (nonsingular) normal distribution with mean vector $\mathbf{\mu}$ and covariance matrix $\Sigma$; the distribution of $\mathbf{X}$ will be denoted by $N_p(\mathbf{\mu}, \Sigma)$ (Anderson, 1958, p. 17).
Theorem 2.11: Let $X$ be distributed according to $N_p(\mu, \Sigma)$. Then, if $C$ is a $(q \times p)$ matrix of rank $q \leq p$ and if $b$ is an arbitrary $(q \times 1)$ vector, $Y = C(X + b)$ is distributed according to $N_q(C(\mu + b), C \Sigma C')$ (Anderson, 1958, p. 25).

Definition 2.21: Let $Z_1, \ldots, Z_m$ be mutually independent, each with distribution $N_p(0, \Sigma)$. Then, $W = \sum_{i=1}^{m} Z_i Z_i'$ has a Wishart distribution with $m$ degrees of freedom and parameter $\Sigma$, denoted by $W_p(\Sigma, m)$. If $m \geq p$, the distribution of $W$ is nonsingular and its density is given as:

$$
\frac{|W|^{(m-p+1)/2} \exp(-\frac{1}{2} \text{tr}(W \Sigma^{-1}))}{2^{mp/2} \pi^{p(p-1)/4} |\Sigma|^{m/2} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(m+1-i)\right)}
$$


Theorem 2.12: Let $W$ be distributed according to $W_p(\Sigma, m)$. Then, if $C$ is a $(q \times p)$ matrix of rank $q \leq p$, $CWC'$ is distributed according to $W_q(C\Sigma C', m)$ (Dempster, 1969, p. 296).

Theorem 2.13: If $W_1, \ldots, W_r$ are mutually independent, each $W_i$ distributed according to $W_p(\Sigma, m_i)$, then $W = \sum_{i=1}^{r} W_i$ is distributed according to $W_p(\Sigma, m)$, where $m = \sum_{i=1}^{r} m_i$ (Anderson, 1958, p. 162).

Theorem 2.14: Let $T^2 = \chi' S^{-1} \chi$, where $\chi$ is distributed according to $N_p(\mu, \Sigma)$ and $mS$ is independently distributed according to $W_p(\Sigma, m)$, $m \geq p$. Then, $\frac{(m - p + 1)}{mp} T^2$ has a non-central $F$-distribution with $p$ and $(m - p + 1)$ degrees of
freedom and noncentrality parameter $\tau^2 = \nu_\gamma \Sigma_{\nu_\gamma}^{-1}$, the probability density of which is given by

$$g(f) = \frac{p \exp\left(-\frac{1}{2} \tau^2\right)}{(m-p+1) \Gamma\left(\frac{m-p+1}{2}\right)} \sum_{i=0}^{p} \frac{(\frac{1}{2} \tau^2)^i}{i!} \frac{\left(p f_{m-p+1}\right)}{\Gamma\left(\frac{m+1}{2}\right)} \frac{\Gamma\left(\frac{m+1}{2} + i\right)}{\Gamma\left(\frac{m+1}{2}\right)};$$

(2.20)

the distribution of $T^2$ will be denoted by $T^2_p(\tau^2; m)$ (Anderson, 1958, pp. 106, 114).

Since the sequential tests that we develop are based on various $T^2$-type statistics, we now derive a form of the probability density function of the $T^2_p(\tau^2; m)$ distribution that will be convenient to use. For this, we need to introduce the confluent hypergeometric function, given as follows:

$$F(a, c; x) = \sum_{i=0}^{\infty} \frac{\Gamma(c) \Gamma(a + i)}{\Gamma(a) \Gamma(c + i)} \frac{x^i}{i!}$$

(2.21)

where $\Gamma(\cdot)$ is the Gamma function and the arguments $a, c,$ and $x$ are real-valued on $(-\infty, \infty)$. Fairly extensive tables of the confluent hypergeometric function are given, for example, by Slater (1960) and by Rushton and Lang (1954).

**Corollary 2.2:** Let $T^2$ be distributed according to $T^2_p(\tau^2; m)$, a noncentral $T^2$-distribution with $m$ degrees of freedom and noncentrality parameter $\tau^2$. Then, the probability density function of $T^2$ is given by
\[
f(t^2 | \tau^2) = \frac{\exp \left( -\frac{1}{2} \tau^2 \right) \left( t^2 \right)^{m-p+1/2}}{B \left( \frac{p}{2}, \frac{m-p+1}{2} \right) (m + t^2)^{(m+1)/2}} \, \binom{m+1}{2} \, \frac{\tau^2 t^2}{2(m + t^2)} 
\]

(2.22)

where \( B(\cdot, \cdot) \) is the beta function and \( F(\cdot, \cdot; \cdot) \) is the confluent hypergeometric function defined in (2.21).

**Proof:** Since, by Theorem 2.14, \( T^2 \) is distributed as \( \left( \frac{m-p+1}{m} \right) F \), where \( F \) has the p.d.f. given in (2.20), the p.d.f. of \( T^2 \) is derived therefrom as

\[
f(t^2) = g \left( \frac{(m-p+1)t^2}{mp} \right) \left( \frac{m-p+1}{mp} \right) 
\]

\[
= \frac{m-p+1}{mp} \frac{p \exp \left( -\frac{1}{2} \tau^2 \right)}{(m-p+1) \Gamma \left( \frac{m-p+1}{2} \right)} \sum_{i=0}^{\infty} \frac{\left( \frac{1}{2} \tau^2 \right)^i (t^2/m)^{p/2+i-i} \Gamma \left( \frac{m+1}{2} + i \right)}{i! \Gamma \left( \frac{p}{2} + i \right) (1 + t^2/m)^{\frac{m+1}{2} + i}} 
\]

\[
= \frac{\exp \left( -\frac{1}{2} \tau^2 \right) \left( t^2 \right)^{p/2-1} \Gamma \left( \frac{m+1}{2} \right)}{(m + t^2)^{m+1} \Gamma \left( \frac{m-p+1}{2} \right) \Gamma \left( \frac{p}{2} \right)} \sum_{i=0}^{\infty} \frac{\Gamma \left( \frac{p}{2} \right) \Gamma \left( \frac{m+1}{2} + i \right) \left( \frac{\tau^2 t^2}{2(m + t^2)} \right)^i}{\Gamma \left( \frac{m+1}{2} \right) \Gamma \left( \frac{p}{2} + i \right) \Gamma \left( \frac{m-p+1}{2} \right)} 
\]

\[
= \frac{\exp \left( -\frac{1}{2} \tau^2 \right) \left( t^2 \right)^{p/2-1} \Gamma \left( \frac{m+1}{2} \right)}{(m + t^2)^{m+1} \Gamma \left( \frac{m-p+1}{2} \right) \Gamma \left( \frac{p}{2} \right)} \, F \left( \frac{m+1}{2}, \frac{p}{2}; \frac{\tau^2 t^2}{2(m + t^2)} \right) . \text{ q.e.d.} 
\]
As shown in later chapters, each test that we develop will be based on the ratio of noncentral $T^2$-densities. Therefore, if we can show that the $T^2_p(\tau^2; m)$ distribution has MLR in $T^2$, an equivalent form of the sequential test can be based on the $T^2$-statistic itself—with, of course, the appropriate decision boundaries. In verifying the MLR property of the $T^2_p(\tau^2; m)$ distribution, we shall use the following lemma, given by Ghosh (1970, p. 308).

**Lemma 2.1**: Let $\{a_j\}$ and $\{b_j\}$ ($j = 0, 1, \ldots$) be two sequences of positive real numbers satisfying $b_j/a_j < b_{j+1}/a_{j+1}$ for every $j$. If $a(s) = \sum_{j=0}^{\infty} a_js^j < \infty$ and $b(s) = \sum_{j=0}^{\infty} b_js^j < \infty$ for all $s > 0$, then the function $\psi(s) = b(s)/a(s)$ is strictly increasing in $s$ on $(0, \infty)$.

**Corollary 2.3**: Let $T^2$ be defined as in Theorem 2.14. Then, the family of probability densities $\{f(t^2|\tau^2), \tau^2 > 0\}$ has monotone likelihood ratio (MLR) in $t^2$.

**Proof**: Clearly, for $\tau^2_0 < \tau^2_1$, the distributions $T^2_p(\tau^2_0; m)$ and $T^2_p(\tau^2_1; m)$ are distinct. Thus, we need to show that there exists a real-valued function—namely, $t^2$—of which the ratio $f(t^2|\tau^2_1)/f(t^2|\tau^2_0)$ is a nondecreasing function for any $\tau^2_0 < \tau^2_1$. 
Of the two cases (i) \( \tau_o^2 = 0 \) and (ii) \( \tau_o^2 > 0 \), the first is given, for example, by Anderson (1958, p. 116). For case (ii), we have from Corollary 2.2 that

\[
\frac{f(t^2|\tau_1^2)}{f(t^2|\tau_o^2)} = \frac{\exp\left(-\frac{1}{2} \tau_1^2 \right) F\left(\frac{m+1}{2}, \frac{\tau_1^2}{2(m+t^2)}\right)}{\exp(-\frac{1}{2} \tau_o^2) F\left(\frac{m+1}{2}, \frac{\tau_o^2}{2(m+t^2)}\right)} \tag{2.23}
\]

Recalling the definition of \( F(\cdots, \cdots, \cdots) \) in (2.21), we wish to establish the following correspondence with Lemma 2.1:

\[
a_j = \frac{\Gamma\left(\frac{D}{2}\right) \Gamma\left(\frac{m+1}{2} + j\right)}{\Gamma\left(\frac{m+1}{2}\right) \Gamma\left(\frac{D}{2} + j\right)} \frac{\exp\left(-\frac{1}{2} \tau_1^2 \right) (\tau_1^2/2)^j}{j!} \]

\[
b_j = \frac{\Gamma\left(\frac{D}{2}\right) \Gamma\left(\frac{m+1}{2} + j\right)}{\Gamma\left(\frac{m+1}{2}\right) \Gamma\left(\frac{D}{2} + j\right)} \frac{\exp\left(-\frac{1}{2} \tau_1^2 \right) (\tau_1^2/2)^j}{j!} \]

\[s = s(t^2) = t^2/(m + t^2)\]

\[\psi(s) \equiv f(t^2|\tau_1^2)/f(t^2|\tau_o^2)\]

Since \( 0 < \tau_1^2 < \tau_o^2 \), \( (\tau_1^2/\tau_o^2) > 0 \) so that, for all \( j \),

\[
\frac{b_j}{a_j} = \exp\left(-\frac{1}{2}(\tau_1^2 - \tau_o^2)\right) \left[\frac{\tau_1^2}{\tau_o^2}\right]^j < \exp\left(-\frac{1}{2}(\tau_1^2 - \tau_o^2)\right) \left[\frac{\tau_1^2}{\tau_o^2}\right]^{j+1} = \frac{b_{j+1}}{a_{j+1}}.
\]
Therefore, by Lemma 2.1, since the series represented by
\( F(a, c; s) \) converges for all \( a > 0, c > 0, \) and \( s > 0 \) (Slater, 1960, p. 2), \( f(t^2|\tau^2_1)/f(t^2|\tau^2_o) \) is a strictly increasing function of \( t^2/(m + t^2) \) and hence of \( t^2 \). q.e.d.
III. SINGLE-SAMPLE SEQUENTIAL
TESTS ABOUT MEAN VECTORS

A. Introduction

Let us assume that it is possible to observe p-dimensional random vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots$ from a multivariate normal population with unknown mean vector $\mathbf{\mu}$ and unknown covariance matrix $\Sigma$, where $\Sigma$ is positive-definite. Sequential tests about the mean vector $\mathbf{\mu}$ have been developed by Jackson and Bradley (1961a) through application of Cox's theorem and, more recently, by HWG (1965) and Ghosh (1970) through invariance and sufficiency principles and application of the Stein Theorem. However, neither HWG nor Ghosh verify all of the necessary conditions that lead to the sequential $T^2$-test given by Jackson and Bradley (1961a). Further, it might be noted that Jackson and Bradley's statement of Cox's theorem lacks the necessary assumption of invariance of the probability model so that their verification is also incomplete. Accordingly, for the sake of completeness, as well as for facility in the development of subsequent sequential tests, we shall include here a complete development of the sequential $T^2$-test through the theory of Chapter II.

B. Case (i): $\Sigma$ Unknown

Let $\mathbf{y}_1, \ldots, \mathbf{y}_n, \ldots$ be independent $(p \times 1)$ random vectors, each distributed according to $\mathcal{N}_p(\mathbf{\mu}, \Sigma)$. Equivalently, we shall
\[ y_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n, \ldots \]  
where the \( \varepsilon_i \)'s are i.i.d. \( \mathcal{N}_p(0, \Sigma) \).

Although one might be interested in testing the null hypothesis \( H_0 : \mu = \mu_0 \), it would be very difficult to specify a meaningful single alternative since there may be infinitely many points in \( p \)-space that are of equal importance. As pointed out by Jackson and Bradley (1961b), even a hypothesis of the type \( H_0 : (\mu - \mu_0) = \delta_0 \) would be difficult to interpret. Further, although the statements \( \mu = \mu_0 \) and \( (\mu - \mu_0)'\Sigma^{-1} (\mu - \mu_0) = 0 \) are equivalent since \( \Sigma \) is positive-definite, the quadratic form in the latter expression can be set equal to some scalar constant, say \( \lambda^2_0 \), so that the null hypothesis in the following hypotheses formulation

\[ H_0 : (\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0) = \lambda^2_0 \]  
\[ H_1 : (\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0) = \lambda^2_1, \lambda^2_0 < \lambda^2_1, \]  

represents the surface of a \( p \)-dimensional ellipsoid while the expression \( (\mu - \mu_0) = \delta_0 \) represents only a single point.

Since, in many practical situations (in particular, in the comparison of two mean vectors via sample paired-vector differences), the vector \( \mu_0 \) is taken to be the zero vector and since generality is not lost by a location transformation on the \( \gamma \)-vectors, we shall, for convenience, assume that we wish
to discriminate sequentially between the two composite hypotheses:

\[ H_0 : \lambda^2 = \lambda^2_0 \]
\[ H_1 : \lambda^2 = \lambda^2_1, \quad \lambda^2_0 < \lambda^2_1 \] \hspace{1cm} (3.3)

where \( \lambda^2 = \mu \Sigma^{-1} \mu \). The parameter space \( \Omega \) is given by

\[ \Omega = \{ \theta = (\mu, \Sigma) : -\infty < \mu_j < \infty, \ 0 < \sigma_{jj} < \infty, \]
\[ -\infty < \sigma_{jj'}, \ j' > j = 1, \ldots, p \} \] \hspace{1cm} (3.4)

For every \( n \geq 1 \), let \( G \) be a group of componentwise transformations \( g \) defined by

\[ g(x_{i j}) = Cx_{i j}, \quad i = 1, 2, \ldots \] \hspace{1cm} (3.5)

where \( C \) is an arbitrary \((p \times p)\) nonsingular matrix. By Theorem 2.11, the random vectors \( \{y_{i}^k = Cy_{i} \} \) are independent, identically distributed according to \( \mathcal{N}_p(C\mu, C\Sigma C') \) so that \( Y_{\Omega} \) remains invariant under \( G \) with the elements of the induced group \( \bar{G} \) defined by

\[ \bar{g}(\theta) = (C\mu, C\Sigma C') \] \hspace{1cm} (3.6)

Using the following lemma, we can easily show that the maximal invariant on \( \Omega \) under \( \bar{G} \) is \( \lambda^2 = \mu \Sigma^{-1} \mu \).

**Lemma 3.1:** Let \( x_{1} \) and \( x_{2} \) be \((p \times 1)\) vectors and let \( A_1 \) and \( A_2 \) be \((p \times p)\) positive-definite matrices. Then,
if, and only if, there exists a nonsingular \((p \times p)\) matrix \(B\) such that

\[
x_{\hat{2}} = Bx_{\hat{1}}
\]

and

\[
A_2 = BA_1B'.
\]

**Proof:** (i) Let \(B\) be a nonsingular \((p \times p)\) matrix such that (3.8) and (3.9) both hold. Then, by substitution,

\[
x_{\hat{2}}'A_2^{-1}x_{\hat{2}} = (Bx_{\hat{1}})'(BA_1B')^{-1}(Bx_{\hat{1}})
\]

\[
= x_{\hat{1}}'A_1^{-1}x_{\hat{1}}.
\]

(ii) Suppose equation (3.7) holds. Since \(A_1\) is positive-definite, there exists a nonsingular matrix \(E_i\) such that

\[
E_iA_1E_i' = I
\]

so that \(A_1^{-1} = E_i'E_i\) for \(i = 1, 2\) (Anderson, 1958, p. 339). Thus, if we make the transformation \(w_{\hat{i}} = E_i x_{\hat{i}}\), \(i = 1, 2\), equation (3.7) can be written as \(w_{\hat{1}}'w_{\hat{1}} = w_{\hat{2}}'w_{\hat{2}}\). Now, since there exists a \((p \times p)\) orthogonal matrix \(P\) such that

\[
w_{\hat{2}} = Pw_{\hat{1}},
\]

or equivalently such that \(E_2x_{\hat{2}} = PE_1x_{\hat{1}}\), we have that

\[
x_{\hat{2}} = E_2^{-1}P E_1 x_{\hat{1}}.
\]

Also, since \(P\) is orthogonal and since \(E_iA_iE_i' = I\) for \(i = 1, 2\), we have that \(E_2A_2E_2' = P(E_1A_1E_1')P'\) so that

\[
A_2 = E_2^{-1}P E_1 A_1 E_1'P'E_2^{-1}.
\]
Hence, (3.8) and (3.9) both hold with $B = E_2^{-1}p E_1$. q.e.d.

The invariance of $\lambda^2$ on $\Omega$ under $\bar{G}$ follows immediately from the sufficient condition of Lemma 3.1; alternately, we see directly that

$$\lambda^2(\bar{g}(\theta)) = (C_\mu)_\kappa'(C\Sigma C')^{-1}(C_\mu)_\kappa$$

$$= \mu_\kappa'\Sigma^{-1}\mu_\kappa \equiv \lambda^2(\theta).$$

The maximality of $\lambda^2$ follows from the necessary condition of Lemma 3.1: if $\lambda_1^2 = \lambda_2^2$, then there exists a nonsingular matrix $C$ such that $\mu_\kappa_2 = C\mu_\kappa_1$ and $\Sigma_2 = C\Sigma_1 C'$, thereby satisfying the second condition of a maximal invariant (see Definition 2.8).

We now wish to develop the invariant SPRT of the hypotheses in (3.3) using Theorem 2.6 (the Stein Theorem).

For $n \geq 1$, denote the sample mean vector and the sample covariance matrix at the $n$th stage of sampling by $\bar{\gamma}_n$ and $D_n$, respectively, where

$$\bar{\gamma}_n = \frac{1}{n} \sum_{i=1}^{n} \gamma_i$$

$$D_n = \frac{1}{n-1} \sum_{i=1}^{n} (\gamma_i - \bar{\gamma}_n)(\gamma_i - \bar{\gamma}_n)'$$

Clearly, $D_n$ is a positive-semidefinite matrix; and, for $n > p$, $D_n$ is positive-definite with probability one (Anderson, 1958, p. 159). Using Theorem 2.4 (the Factorization Criterion), one can easily verify (Anderson, 1958, p. 56) that $S_n = (\bar{\gamma}_n, D_n)$ is
sufficient for $Y(n)\Omega$, for each $n \geq 1$.

Analogous to the group $\bar{G}$ induced by $G$ on $\Omega$, the group $G^s$ induced by $G$ on the space of $S_n$ has elements $g_s$ defined by

$$g_s(S_n) = (C\bar{\gamma}_n, C\bar{D}_n C').$$  \hfill (3.12)

Further, if $n > p$, a maximal invariant under $G^s$ is

$$T_n^2 = V_n(S_n) = n\bar{\gamma}_n'^{-1}\bar{\gamma}_n,$$  \hfill (3.13)

the verification of which is completely analogous to that for $\chi^2$ under $\bar{G}$. The invariance of $T_n^2$ follows from the sufficient condition of Lemma 3.1 (the invariance of Hotelling’s $T^2$-statistic is well-known; see, for example, (Anderson, 1958, p. 115)). The maximality of $T_n^2$ is established by the necessary condition of Lemma 3.1 since $T_n^2(1) = T_n^2(2)$ implies that $\bar{\gamma}_n^2 = C\bar{\gamma}_n(1)$ and $D_n(2) = C\bar{D}_n(1)C'$ for some nonsingular matrix $C$; that is, there exists some $g_s \in G_s$ such that $(\bar{\gamma}_n(2), D_n(2)) = g_s(\bar{\gamma}_n(1), D_n(1))$. It might be noted that, for $n \leq p$, a maximal invariant under $G_s$ is $V_n(S_n)$ equal to an arbitrary constant; in ordinary terms, for the case $n \leq p$, there are insufficient data to obtain a nonsingular $D_n$.

Application of Theorem 2.6 can be validated by verifying Assumption 2.3, in analogy with Jackson and Bradley’s verification of the conditions of Cox’s theorem, the set $A_o$ consisting of all points in $(R^p)^n$, the $n$th product of $p$-dimensional Euclidean space, for which $D_n$ is nonsingular.
However, the verification is both tedious and unnecessary since, by the completeness of the sufficient statistic \( S_n \) and by Theorem 2.7, we have that Assumption 2.1 is satisfied. The completeness of the sufficient statistic \( S_n \) is a consequence of the well-known result (Lehmann, 1959, p. 132) that, in the \( k \)-parameter exponential family (Definition 3.1), the sufficient statistic is complete provided the parameter space contains an open set in \( \mathbb{R}^k \).

**Definition 3.1:** The exponential family of distributions is defined by probability densities of the form

\[
p_\theta(x) = K(\theta) \exp \left\{ \sum_{j=1}^{k} \phi_j(\theta) T_j(x) \right\} h(x) \tag{3.14}
\]

with respect to a \( \sigma \)-finite measure \( \nu \) over a Euclidean sample space \((\mathcal{X}, \mathcal{A})\) (Lehmann, 1959, p. 50).

Since the multivariate normal distribution belongs to the exponential family (Bildikar and Patil, 1968, p. 1316) and since the parameter space \( \Omega \) clearly contains an open set in \( \mathbb{R}^k \), \( k = p + \frac{1}{2}p(p+1) \), we have that the sufficient statistic \( S_n \) for \( Y_{(n)} \Omega \) is complete and, therefore by Theorem 2.7, Assumption 2.1 obtains.

It follows that the sequence \( \{T_n^2\} \) is an invariantly sufficient sequence for the class of sequential models \( Y_\Omega \) under \( G \). Moreover, since (see Appendix C for verification)
\[ \bar{y}_{n+1} = \frac{1}{n+1} (n \bar{y}_n + \bar{y}_{n+1}) \quad \text{and} \quad (3.15) \]

\[ D_{n+1} = \frac{n-1}{n} D_n + \frac{1}{n+1} \left( \bar{y}_{n+1} - \frac{\bar{y}_n}{n} \right) \left( \bar{y}_{n+1} - \frac{\bar{y}_n}{n} \right)' , \quad (3.16) \]

we have, by Theorem 2.10, that the sequence \( \{ S_n = (\bar{y}_n, D_n) \} \) is transitive under \( G_s \) and, therefore, that the sequence \( \{ T_n^2 \} \) is transitive. The following corollary to Theorem 2.14 establishes the distribution of \( T_n^2 \) for \( n > p \).

**Corollary 3.1:** Let \( \bar{y}_1, \ldots, \bar{y}_n \) be i.i.d. \( \mathcal{N}_p(\mu, \Sigma) \), and let \( T_n^2 = n \bar{y}_n D_n^{-1} \bar{y}_n \), for \( n > p \). Then, \( T_n^2 \) is distributed according to \( T_p^2(\tau^2; n-1) \), where \( \tau^2 = n\lambda^2 \) (Anderson, 1958, p. 107).

Now, as discussed in Section C of Chapter II, by application of Theorem 2.6, the invariant SPRT of (3.3) can be based on the ratio \( L_n \) of the p.d.f. of \( T_n^2 \) at \( n\lambda_1^2 \) and at \( n\lambda_0^2 \):

\[ L_n = \frac{f(t_n^2 | n \lambda_1^2)}{f(t_n^2 | n \lambda_0^2)} , \quad (3.17) \]

which, by Corollary 2.2, becomes

\[ L_n = \exp \left[ -\frac{1}{2n}(\lambda_1^2 - \lambda_0^2) \right] \frac{F \left( \frac{n}{2}, \frac{p}{2}, \frac{n\lambda_1^2 t_n^2}{2(n-1+t_n^2)} \right)}{F \left( \frac{n}{2}, \frac{p}{2}, \frac{n\lambda_0^2 t_n^2}{2(n-1+t_n^2)} \right)} \quad (3.18) \]
where $F(\cdot, \cdot; \cdot)$ is the confluent hypergeometric function defined in (2.21). Thus, the invariant SPRT is specified by decision rules of the form (1.4) where $L_n$ is given in (3.18) and $n > p$.

By Corollary 2.3, since $L_n = L_n(T_n^2)$ is an increasing function of $T_n^2$ on $(0, \infty)$ for every fixed set \( \{p \geq 1, 0 \leq \lambda_0^2 < \lambda_1^2; n > p\} \), an equivalent form of the invariant SPRT is:

accept or reject $H_0$ according as the lower or upper inequality in

\[
T_n^2 < T_n^2 < \overline{T}_n^2, \quad \text{for } n \geq p, \tag{3.19}
\]

is first violated, where $T_n^2$ is defined in (3.13) and the critical limits $T_n^2$ and $\overline{T}_n^2$ are the respective solutions of

\[
L_n(T_n^2) = B, \quad L_n(\overline{T}_n^2) = A. \tag{3.20}
\]

Freund and Jackson (1960) have produced extensive tables of $T_n^2$ and $\overline{T}_n^2$ for the sequential $T^2$-test in which $\lambda_0^2 = 0$. It might be noted that, in the case $\lambda_0^2 = 0$, $L_n$ given in (3.18) simplifies considerably since the confluent hypergeometric function in the denominator is then equal to unity.

For some standard formulae pertinent to the calculation of the confluent hypergeometric function, as well as some suggestions for the computation of $T_n^2$, the reader is referred to the Appendix.
Also, using the MLR property of the family of densities
for $T_n^2$, as established in Corollary 2.3, we have, by Theorem
2.9, that the SPRT given above effectively tests the hypotheses-
formulation

$$H_0 : \lambda^2 \leq \lambda^2_0$$

$$H_1 : \lambda^2 \geq \lambda^2_1, \quad \lambda^2_0 < \lambda^2_1.$$  \hspace{1cm} (3.21)

A proof that the sequential $T^2$-test terminates with
probability one is given by Jackson and Bradley (1961a), and,
more recently in slightly more general terms, by Wijsman (1967b).
Thus, as described in Section C of Chapter II, termination with
probability one allows one to use Wald boundaries with the
procedures achieving approximately the specified Type I and
Type II probabilities of error. No average sample number or
operating characteristic formulae are, however, available; nor
does it seem likely at present that theoretical analysis will
have much success in this regard.

Finally here, it might be worth noting that, for the
special case $p = 1$, $\lambda^2$ and $T_n^2$ simplify to $(\mu/\sigma)^2$ and
$(\sqrt{n} \bar{y}_n/s_n)^2$, respectively, so that the sequential $T^2$-test
coincides with the sequential $t^2$-test given by Rushton (1950).
C. Case (ii): $\Sigma$ Unknown, but Estimated Independently

Jackson and Bradley (1961a) also derive sequential $\chi^2$-tests for the case when $\Sigma$ is known. Instead, we consider a more realistic situation to be the case when there is available an independent (nonsequential) estimate of $\Sigma$ based, say, on data from a previous experiment. As in Section B of this chapter, we are led to sequential tests based on a $T^2$-statistic but with increased degrees of freedom; further, rather than delaying the test procedure until the $(p + 1)^{st}$ stage, it will be possible in this case to begin testing at the first stage of sampling.

As in Section B, we shall assume that the random vectors $y_1, \ldots, y_n, \ldots$ satisfy model (3.1). Further, we shall assume that there exists a random matrix $V$ such that $mV$ is distributed independently of the $y_i$-vectors according to $W_p(\Sigma, m)$, $m > p$; in ordinary terms, $V$ can be considered as an independent unbiased estimate of $\Sigma$ with $m$ degrees of freedom. Again, we wish to discriminate sequentially between the two hypotheses given in (3.3) where, as before, $y_{\cap 0}$ is taken to be the zero vector.

In order to apply the theory of Chapter II, we need to define modifications of the three types of probability models given in Section C of that chapter (the hat-notation is used throughout to denote these modifications):
(i) the modified component or marginal models
\[ \hat{X}_n^\theta = X_n^\theta \times V_\theta \] for stage n (that is, the product probability model of \( X_n^\theta \), as defined in Chapter II, and \( V_\theta \), the probability model corresponding to the stochastically independent random variable \( V \));

(ii) the modified joint models \( \hat{X}(n)^\theta = X(n)^\theta \times V_\theta \) for the accumulated data \( (X(n), V) \) at stage n;

(iii) the modified sequential model \( \hat{X}_\theta = X_\theta \times V_\theta \) for the entire sequence of data \( (X, V) \).

Similarly, the notation for the classes of probability models needs to be altered slightly; for example, \( \hat{X}_\Omega \) will denote the class of modified sequential models \( \hat{X}_\theta \), \( \theta \in \Omega \). It might be noted that, since \( \theta = (\nu, \Sigma) \) in this chapter, the probability model for the random matrix \( V \) could be denoted by \( V_\theta \), where \( \theta' \in \Omega' \), the subspace of \( \Omega \) corresponding to \( \Sigma \), where the parameter space \( \Omega \) is defined in (3.4).

For every \( n \geq 1 \), let \( \hat{G} \) be a group of component-wise transformations \( \hat{g} \), defined by
\[ \hat{g}(y_i^\nu, V) = (Cy_i^\nu, CVC'), \quad i = 1, 2, \ldots \] (3.22)

where \( C \) is an arbitrary \((p \times p)\) nonsingular matrix. By Theorem 2.14, \( mV^* = mCVC' \) is distributed according to \( W_p(C\Sigma C', m) \) and independently of the random vectors \( \{y_i^* = Cy_i^\nu\} \), which are i.i.d. \( N_p(C\Sigma, C\Sigma C') \). Hence, \( \hat{V}_\Omega \) remains invariant under \( \hat{G} \) with the induced group \( \bar{G} \) equal to \( \bar{G} \), the elements of
which are defined in (3.6). Thus, as in Section B, the maximal invariant on $\Omega$ under $G = \tilde{G}$ is $\lambda^2 = \mu'\Sigma^{-1}\mu$.

Again, only now utilizing the additional information on the covariance matrix $\Sigma$, we wish to develop the invariant SPRT of the hypotheses in (3.3) by means of Theorem 2.6. In ordinary terms, the following lemma establishes that $\tilde{\nu}_n$ and $\tilde{W}_n$, a pooled unbiased estimator of $\Sigma$ at stage $n$, are the jointly sufficient statistics for $\mu$ and $\Sigma$ on the basis of all the information available through stage $n$.

**Lemma 3.2:** Let $\nu_1, \ldots, \nu_n$, $n \geq 1$, be mutually independent random vectors, each distributed according to $N_p(\mu, \Sigma)$, and let $\tilde{\nu}_n$ and $D_n$ be the sample mean vector and covariance matrix defined in (3.10) and (3.11). Let $mV$ be distributed independently of $\nu_1, \ldots, \nu_n$ according to $W_p(\Sigma, m)$, $m \geq p$; and let

$$W_n = \left[ \frac{1}{m+n-1} \right] [mV + (n-1)D_n]. \tag{3.23}$$

Then, $S_n = (\tilde{\nu}_n, W_n)$ is sufficient for $\hat{\nu}_{(n)\Omega}$, for each $n \geq 1$.

**Proof:** By the mutual independence of $\nu_1, \ldots, \nu_n$ and by the independence of $mV$ and the $\nu$-vectors, we can write the joint probability density function as:

$$f(\nu_1, \ldots, \nu_n; mV|\mu, \Sigma)$$

$$= \{(2\pi)^{-pn/2} |\Sigma|^{-n/2} \exp\{-\frac{1}{2} \sum_{i=1}^{n} (\nu_i - \mu)'\Sigma^{-1}(\nu_i - \mu)\}\}$$

$$\times \{K_v|mV|^{-(m-p+1)/2} |\Sigma|^{-m/2} \exp\{-\frac{1}{2} \text{Tr}(mV\Sigma^{-1})\}\},$$
where \[ K_v^{-1} = 2^{mp/2} \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(m+i)\right). \]

Therefore, since, for matrices A, B, and C,

\[ \text{Tr}(AC) + \text{Tr}(BC) = \text{Tr}((A+B)C) \]

and since, as shown by Anderson (1958, p. 56),

\[ \sum_{i=1}^{n} (\bar{Y}_i - \mu)' \Sigma^{-1}(\bar{Y}_i - \mu) = n(\bar{Y} - \mu)' \Sigma^{-1}(\bar{Y} - \mu) + \text{Tr}\left(\{(n-1)D_n \Sigma^{-1}\right) \]

it follows that \( f(y_{n1}, \ldots, y_{nn}; mV; \mu, \Sigma) \) can be written as

\[ \left[ |\Sigma|^{-(m+n+1)/2} \exp\left[\frac{n}{2} (\bar{Y} - \mu)' \Sigma^{-1}(\bar{Y} - \mu)\right] \frac{1}{2} \text{Tr}\left(\{(n-1)D_n + mV\right\} \Sigma^{-1}) \right] \]

\[ \times [K_v(2\pi)^{-pn/2} |mV|^{-(m+p+1)/2}]. \]

\[ = g_\Theta(\bar{Y}_{n1}, W_n; \mu, \Sigma) h(y_{n1}, \ldots, y_{nn}; mV), \]

where \((m+n-1)W_n = [(n-1)D_n + mV]\) and where the function \(g_\Theta\) is a nonnegative \(\mathcal{S}\)-measurable function and the nonnegative function \(h\) depends only on \(\{y_{n1}, \ldots, y_{nn}; mV\}\)—in fact, only on \(mV\). Hence, by Theorem 2.4 (the Factorization Criterion), \(\hat{\hat{\Theta}}_n = (\bar{Y}_{n1}, W_n)\) is sufficient for \(\hat{\Theta}_{n_0}\) for each \(n \geq 1\). q.e.d.

It should be noted that, since \(V\) is positive-definite with probability one and \(D_n\) is positive-semidefinite, \(W_n\) is positive-definite with probability one for each \(n \geq 1\). Thus, since the group \(\hat{\mathcal{G}}_S\) induced by \(\hat{\mathcal{G}}\) on the space of \(\hat{\Theta}_n\) has elements \(\hat{g}_s\) defined by (3.12) with \(W_n\) replacing \(D_n\), we have that a
maximal invariant under $\hat{G}_s$ is

$$\hat{T}_n^2 = V_n(\hat{S}_n) = n^{-1}\frac{\bar{y}_n}{\hat{y}_n} W_n^{-1} \bar{y}_n, \quad n \geq 1, \quad (3.24)$$

the verification of which is completely analogous to that of $T_n^2$ of Section B, except that $D_n$ is replaced by $W_n$.

Hence, applying Theorem 2.6 as in Section B, we have that the sequence $\{\hat{T}_n^2\}$ is an invariantly sufficient sequence for the class $\hat{Y}_\Omega$ of modified sequential models. Further, it is shown in Appendix C, using the recursive relation in (3.16), that

$$W_{n+1} = \frac{1}{m+n} [(m+n-1)W_n + \frac{n}{n+1}(\bar{y}_{n+1} - \bar{y}_n)(\bar{y}_{n+1} - \bar{y}_n)^{-1}]. \quad (3.25)$$

Hence, by Theorem 2.10, the sequence $\{\hat{S}_n = (\bar{y}_n, \bar{y}_n)\}$ is transitive under $\hat{G}_s$, and therefore the sequence $\{\hat{T}_n^2\}$ is transitive under $\hat{G}_s$. It should here be noted that, although $\hat{S}_n$ and $\hat{T}_n^2$ differ from the $S_n$ and $T_n^2$ of Section B, the function $V_n$, defining the maximal invariant under $\hat{G}_s$ and $G_s$, respectively, remains the same.

**Corollary 3.2**: Let $\{y_1, \ldots, y_n\}, \ n \geq 1, \bar{y}_n, D_n, V,$ and $W_n$ be as defined in Lemma 3.2, and let $\hat{T}_n^2 = n \frac{\bar{y}_n}{\hat{y}_n} W_n^{-1} \bar{y}_n, \ n \geq 1$. Then, $\hat{T}_n^2$ is distributed according to $T_p^2 (\tau^2; m + n - 1)$, where $\tau^2 = n^{-1} \Sigma^{-1} \bar{y}_n = n\lambda^2$.

**Proof**: Since $(n-1)D_n$ is distributed according to $W_p (\Sigma, n-1)$ and $mV$ is distributed independently according to $W_p (\Sigma, m)$ (Anderson, 1958, p. 159), we have, by Theorem 2.13,
that

\[(m + n - 1)W_n = mV + (n - 1)D_n\]  

(3.26)

is distributed according to \(W_p(\Sigma, m+n-1)\). Since \(\sqrt{n} \frac{\bar{y}}{s_n}\) is distributed according to \(N_p(\sqrt{n} \frac{\mu}{s_n}, \Sigma)\) independently of \(W_n\), the required result follows immediately from Theorem 2.14 with the following correspondences

\[\gamma_n \equiv \sqrt{n} \frac{\mu}{s_n} ; \quad \gamma \equiv \sqrt{n} \frac{\bar{y}}{s_n} ; \quad \text{and} \quad S \equiv W_n\]

and with \(m\) replaced by \((m + n - 1)\). q.e.d.

Thus, by application of Theorem 2.6, the invariant SPRT of (3.3) is specified by decision rules of the form (1.4)

where \(\hat{L}_n\) is the ratio of the p.d.f. of \(\hat{T}^2_n\) at \(\tau_1 = n \lambda_1^2\) and the p.d.f. of \(\hat{T}^2_n\) at \(\tau_0 = n \lambda_0^2\):

\[L_n = \frac{f(\hat{T}^2_n | n \lambda_1^2)}{f(\hat{T}^2_n | n \lambda_0^2)}\]  

(3.27)

which, by Corollary 2.2, becomes

\[\hat{L}_n = \exp\left[-\frac{1}{2n}(\lambda_1^2 - \lambda_0^2)\right] \frac{F\left[\frac{m+n}{2}, \frac{p}{2}; \frac{n \lambda_1^2 \hat{T}^2_n}{2(m+n-1+\hat{T}^2_n)}\right]}{F\left[\frac{m+n}{2}, \frac{p}{2}; \frac{n \lambda_0^2 \hat{T}^2_n}{2(m+n-1+\hat{T}^2_n)}\right]}\]  

(3.28)

By Corollary 2.3, since \(\hat{L}_n = \hat{L}_n(\hat{T}^2_n)\) is an increasing function of \(\hat{T}^2_n\) on \((0, \infty)\) for every fixed set \(p \geq 1\),
0 \leq \lambda_0^2 < \lambda_1^2; \ n \geq 1, \ m \geq p} \ an \ equivalent \ form \ of \ the \ invariant
SPRT \ is:
accept \ or \ reject \ H_0 \ according \ as \ the \ lower \ or \ upper \ inequality \ in
\begin{equation}
\hat{T}_n^2 < T_n^2 < \tilde{T}_n^2
\end{equation}
is first violated, \ where \ \hat{T}_n^2 \ is \ defined \ in \ (3.24) \ and \ the
critical \ limits \ are \ the \ solution \ of
\begin{equation}
\hat{L}_n (\hat{T}_n^2) = B, \ \hat{L}_n (\tilde{T}_n^2) = A.
\end{equation}

As \ in \ section \ B, \ termination \ with \ probability \ one \ follows
from \ Jackson \ and \ Bradley's \ (1961a) \ results, \ while, \ additionally \ from \ the \ MLR \ property \ of \ the \ family \ of \ densities
\{f(t^n_1|\tau^2), \ \tau^2 > 0\}, \ it \ follows \ that \ the \ above \ SPRT \ effectively
tests \ the \ hypotheses \ in \ (3.21).

It might be noted that, with \ m = 0, \ the \ SPRT \ of \ this
section \ reduces \ to \ the \ SPRT \ of \ the \ previous \ section--with, \ of
course, \ appropriate \ restrictions \ for \ n \leq p. \ On \ strictly
intuitive \ grounds, \ one \ might \ expect \ that, \ for \ the \ case \ when
there \ exists \ an \ independent \ estimate \ V \ of \ \Sigma \ with \ m \ degrees \ of
freedom, \ the \ ASN \ function \ for \ the \ sequential \ test \ based \ on
\{\hat{T}_n^2\} \ would \ be \ less \ than \ that \ for \ the \ sequential \ \bar{T}_n^2-test \ of
Section \ B. \ Hopefully, \ this \ conjecture \ may \ eventually \ be
substantiated \ by \ means \ of \ a \ large-scale \ empirical \ investigation.
Further, \ the \ fact \ that \ the \ SPRT \ based \ on \ \{\hat{T}_n^2\} \ begins \ with \ n = 1,
whereas the SPRT based on $\{T_n^2\}$ begins with $n = p + 1$, makes the latter sequential test especially appealing when an early decision is crucial.
IV. SINGLE-SAMPLE SEQUENTIAL TESTS ABOUT MEAN VECTORS ADJUSTED FOR COVARIATES

A. Introduction

As described in Chapter I, statistical analyses utilizing concomitant information have been used in fixed-sample-size experimentation for the multivariate case (see, for example, Rao, 1966), as well as for the univariate case (see, for example, Cochran, 1957); and, as noted therein, at least one purpose for the utilization of concomitant information is to remove the effects of disturbing variables in observational studies. With this purpose in mind, Roseberry (1965), Cox and Roseberry (1966), and Sampson (1968) developed univariate sequential procedures for testing statistical hypotheses about a mean adjusted for covariate-effects.

In this chapter, we develop sequential tests for the comparison of two treatments, wherein both the response and the covariate metameters are vector-valued. In fact, in the framework of the paired-comparison experimental design, the response metameter is actually the vector difference of the within-pair response vectors and, similarly, the covariate metameter is the vector difference of the corresponding within-pair covariate vectors. Since this can be considered as an application of a single-sample procedure, we state the basic problem of this chapter as that of testing sequentially
hypotheses about a mean vector adjusted for covariates, where it is assumed that the response vectors are multivariate-normally distributed and that the covariate vectors are controlled.

B. Case (i): \( \Sigma \) Unknown

Let \( \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n, \ldots \) be mutually independent \((p \times 1)\) random vectors; each \( \mathbf{y}_i \) being distributed according to \( N_p(\mathbf{a} + \mathbf{B} \mathbf{x}_i, \Sigma) \), where \( \mathbf{a} (p \times 1) \), \( \mathbf{B} (p \times q) \), and \( \Sigma (p \times p) \) are unknown parameters and each \( \mathbf{x}_i \) is a \((q \times 1)\) vector of known covariates. Equivalently, we shall write:

\[
\mathbf{y}_i = \mathbf{a} + \mathbf{B} \mathbf{x}_i + \varepsilon_i, \quad i = 1, 2, \ldots
\]  

(4.1)

where the \( \varepsilon_i \)'s are \( i.i.d. \) \( N_p(0, \Sigma) \). The parameter space \( \Omega \) consists of elements \( \theta = (\mathbf{a}, \mathbf{B}, \Sigma) \) and

\[
\Omega = \{ \theta : -\infty < a_j < \infty, -\infty < b_{jk} < \infty, \\
k = 1, \ldots, q, j = 1, \ldots, p \Sigma > 0 \}
\]

(4.2)

where \( \Sigma > 0 \) denotes \( \Sigma \) positive definite.

Suppose that we wish to discriminate sequentially between the two composite hypotheses

\[
H_0 : (\mathbf{a} - \mathbf{a}_0)' \Sigma^{-1} (\mathbf{a} - \mathbf{a}_0) = \lambda_0^2
\]

\[
H_1 : (\mathbf{a} - \mathbf{a}_0)' \Sigma^{-1} (\mathbf{a} - \mathbf{a}_0) = \lambda_1^2, \lambda_0^2 < \lambda_1^2
\]

(4.3)
where $\mathbf{a}_0$ is a vector of specified constants and $\mathbf{g}_i$ represents the mean vector adjusted for covariates. For convenience, we shall consider an equivalent hypotheses-formulation:

$$H_0 : \lambda^2 = \lambda^2_0$$

$$H_1 : \lambda^2 = \lambda^2_1, \lambda^2_0 < \lambda^2_1$$

where $\lambda^2 = \mathbf{a}'_0 \Sigma^{-1} \mathbf{a}_0$.

In passing here, it might be noted that, for the rare situation in which the matrix of regression coefficients $\mathbf{B}$ is known, the test procedure of Chapter III is applicable to the adjusted observations $\{z^*_i = (\mathbf{y}_i - \mathbf{B} \mathbf{x}_i)\}$.

For every $n \geq 1$, let $G$ be a group of component-wise transformations $g$ defined by

$$g(y^*_i, x^*_i) = (C y^*_i, D x^*_i), i = 1, 2, \ldots \quad (4.5)$$

where $C$ ($p \times p$) and $D$ ($q \times q$) are arbitrary nonsingular matrices. Thus, since the random vectors $\{y^*_i = C y^*_i\}$ are independently distributed, by Theorem 2.11,

$$y^*_i \sim N_p(C \mathbf{a}_0 + \mathbf{C} \mathbf{B}^{-1} x^*_i, \Sigma \mathbf{C}' \Sigma')$$

where $x^*_i = D x_i$.

we have that $Y_\Omega$ remains invariant under $G$ with the induced group $\tilde{G}$ defined by elements
\( \bar{g}(\theta) = (\bar{C}_\theta, \bar{C}_D^{-1}, \bar{C}_E) \). \hspace{1cm} (4.6)

More strictly, the notation for the class of sequential probability models under consideration might be presented as \((Y|X)_\Omega\). Instead, the less cumbersome notation \(Y_\Omega\) will be used in what follows.

The invariance of \(\lambda^2 = \alpha_i^\prime \Sigma^{-1}_i \alpha_i\) on \(\Omega\) under \(\bar{G}\) follows immediately from the sufficient condition of Lemma 3.1. The maximality of \(\lambda^2\) under \(\bar{G}\) is obtained from the necessary condition of Lemma 3.1 as follows.

Let \(\lambda^2_1\) correspond to \(\theta_i = (\alpha_i, \beta_i, \Sigma_i)\) for \(i = 1, 2\); if \(\lambda^2_1 = \lambda^2_2\), then by the necessary condition of Lemma 3.1, there exists a nonsingular \((p \times p)\) matrix \(C\) such that \(\alpha_2 = C\alpha_1\) and \(\Sigma_2 = C\Sigma_1 C'\). Now, considering the two matrices of rank \(q\), \(\beta_2\) and \(\beta_1\), it is clear that there exists a nonsingular matrix \(D\) such that \(\beta_2 = C\beta_1 D^{-1}\). Hence, \(\lambda^2_1 = \lambda^2_2\) implies that there exists \(\bar{g} \in \bar{G}\) such that \(\theta_2 = \bar{g}(\theta_1)\) so that \(\lambda^2\) is a maximal invariant on \(\Omega\) under \(G\) (Definition 2.8).

In developing the invariant SPRT of the hypotheses-formulation (4.4) using Theorem 2.6, we need to assume that, for \(n > q + 1\), the \((q \times n)\) matrix \(X_n\) has rank \(q\), where

\[
X_n = [x_{11} - \bar{x}_n, \ldots, x_{n1} - \bar{x}_n]. \hspace{1cm} (4.7)
\]
We now define the sample estimates of the components of \( \theta = (\alpha, \beta, \Sigma) \) at the \( n \)th stage, \( n > q + 1 \), as follows. The estimate of \( \beta \) is

\[
B_n = Y_n X_n' A_n^{-1},
\]

where

\[
Y_n = [y_{n1} - \bar{y}_n, \ldots, y_{nN} - \bar{y}_n] \tag{4.9}
\]

and

\[
A_n = X_n X_n'. \tag{4.10}
\]

The estimate of \( \alpha \) is

\[
\bar{\alpha}_n = \bar{y}_n - B_n \bar{x}_n. \tag{4.11}
\]

The estimate of \( \Sigma \) is

\[
E_n = \left( \frac{1}{n-q-1} \right) [Y_n Y_n' - Y_n X_n' B_n]. \tag{4.12}
\]

Thus, the \( n \)th stage sample estimate of \( \theta = (\alpha, \beta, \Sigma) \) will be denoted by \( S_n = (\bar{\alpha}_n, B_n, E_n) \). The relevant properties of \( S_n \) are given in the following results from multivariate linear model theory (see, for example, Anderson, 1958, p. 183).

**Theorem 4.1:** For each \( n > q + 1 \), \( S_n = (\bar{\alpha}_n, B_n, E_n) \), as defined above, is sufficient for \( Y_{(n)}^\prime \); \( (\bar{\alpha}_n, B_n) \) is normally distributed with mean \( (\alpha, \beta) \) and the covariance matrix of the \( i \)th and \( j \)th rows of \( (\bar{\alpha}_n, B_n) \) is \( \sigma_{ij} H_n^{-1} \), where
further, \((n - q - 1) E_n\) is independently distributed according to \(W_p(\Sigma, n - q - 1)\).

It can be easily shown that the inverse of the matrix \(H_n\) is given by

\[
H_n^{-1} = \begin{pmatrix}
\frac{1}{n} + \bar{x}_n' A_n^{-1} \bar{x}_n & -\bar{x}_n' A_n^{-1} \\
- A_n^{-1} \bar{x}_n & A_n^{-1}
\end{pmatrix}.
\]  

(4.14)

Now, restricting attention to the distribution of \(a_{\nu n}\), we state the following result from Theorem 4.1.

**Corollary 4.1:** For \(n > q + 1\), \(S_n = (a_{\nu n}, B_n, E_n)\) is sufficient for \(Y_{\nu n}\); \(a_{\nu n}\) is distributed according to \(N_p(a, c_{\nu n}^{-1}\Sigma)\) where

\[
c_{\nu n}^{-1} = \left(\frac{1}{n} + \bar{x}_n' A_n^{-1} \bar{x}_n\right),
\]

and \((n-q-1)E_n\) is distributed independently according to \(W_p(\Sigma, n-q-1)\).
Analogous to the group \( \bar{G} \) induced on \( \Omega \) by \( G \), the group \( G^s \) induced by \( G \) on the space of \( S_n \) has elements \( g^s \) defined by

\[
g^s(S_n) = (C \alpha^s_n, C B_n D^{-1}, C E_n C').
\]

(4.15)

Further, for \( n > p + q \), a maximal invariant under \( G^s \) is

\[
T_n^2 = V_n(S_n) = c_n \alpha^s_n E^{-1} \beta_n \alpha_n.
\]

(4.16)

where \( c_n \) is defined in Corollary 4.1. The invariance and maximality of \( T_n^2 \) under \( G^s \) may be verified by steps exactly analogous to those used already in showing that \( \lambda^2 \) is maximal invariant with respect to \( \bar{G} \). It should be noted however that, for \( n \leq p + q \), any constant is a maximal invariant under \( G^s \) since \( n \) must exceed \( (p + q) \) for \( E_n \) to be positive definite with probability one.

As in Chapter III, since the parameter space \( \Omega \) contains an open set in \( \mathbb{R}^k \), \( k = p + pq + p(p+1)/2 \), the sufficient statistic \( S_n \) for \( Y_{(n)} \) is complete, and therefore by Theorem 2.7, Assumption 2.1 obtains. Hence, by Theorem 2.6 under Assumption 2.1, the sequence \( \{T_n^2\} \) is an invariantly sufficient sequence for the class of sequential models \( Y_{(n)} \) under \( G \).

The transitivity of the sequence \( \{S_n = (\alpha_n, B_n, E_n)\} \) under \( G^s \) follows, via Corollary 2.1, from the recurrence relationships given below (see Appendix C for verification):
\[ B_{n+1} = B_n (A_n A_n^{-1}) \]

\[ + \frac{n}{n+1} (y_{n+1} - a_n - B_n \bar{x}_n) (x_{n+1} - \bar{x}_n)' A_n^{-1} \]  \hspace{1cm} (4.17)

\[ a_{n+1} = \left[ \frac{n}{n+1} \right] a_n + \left[ \frac{n}{n+1} \right] B_n \bar{x}_n \]

\[ - B_{n+1} \bar{x}_{n+1} + \left( \frac{1}{n+1} \right) y_{n+1} \]  \hspace{1cm} (4.18)

\[ E_{n+1} = \left( \frac{n-q-1}{n-q} \right) E_n + \left( \frac{1}{n-q} \right) B_n A_n B_n \]

\[ - \left( \frac{1}{n-q} \right) B_{n+1} A_{n+1} B_{n+1} \]

\[ + \frac{n}{(n+1)(n-q)} (y_{n+1} - a_n - B_n \bar{x}_n) (y_{n+1} - a_n - B_n \bar{x}_n)' . \]  \hspace{1cm} (4.19)

The following corollary to Theorem 2.14 establishes the distribution of \( T_n^2 \) for \( n > p + q \).

**Corollary 4.2:** Let \( y_{\chi_{1}}, \ldots, y_{\chi_{n}} \) be mutually independent, each \( y_{\chi_{i}} \) being distributed according to \( N_p (\chi + B X_{1}, \Sigma) \), and let \( T_n^2 \) be as defined in (4.16), \( n > p + q \). Then, \( T_n^2 \) is distributed according to \( T_p^2 (\tau^2; n - q - 1) \), where \( \tau^2 = c_n \chi^2 = c_n a_n \Sigma a_n^{-1} \chi \) and \( c_n \) is as defined in Corollary 4.1.

**Proof:** From Corollary 4.1, we have that \( \sqrt{c_n} a_n \) is distributed \( N_p (\sqrt{c_n} a_n \chi, \Sigma) \) and that \( (n - q - 1) E_n \) is distributed independently according to \( W_p (\Sigma, n - q - 1) \). Thus, by Theorem 2.14, \( T_n^2 = c_n a_n E_n^{-1} a_n \) is distributed according to \( T_p^2 (c_n \chi^2; n - q - 1) \). q.e.d.
It follows that the invariant SPRT of (4.4) is specified by decision rules of the form (1.4) where \( L_n \) is given as

\[
L_n = \frac{f(t_n^2 | c_n \lambda_1^2)}{f(t_n^2 | c_n \lambda_0^2)}
\]  

(4.20)

which, by Corollary 2.2, becomes

\[
L_n = \exp\left[ -\frac{1}{2} c_n (\lambda_1^2 - \lambda_0^2) \right] \frac{F\left( \frac{n-q}{2}, \frac{c_n \lambda_1^2 - T_n^2}{2(n-q-1+T_n^2)} \right)}{F\left( \frac{n-q}{2}, \frac{c_n \lambda_0^2 - T_n^2}{2(n-q-1+T_n^2)} \right)}
\]  

(4.21)

where \( F(\cdot,\cdot;\cdot) \) is defined in (2.21).

By Corollary 2.3, since \( L_n = L_n(T_n^2) \) is an increasing function of \( T_n^2 \) on \((0,\infty)\) for every fixed set \( \{p \geq 1, q \geq 1, 0 \leq \lambda_0^2 < \lambda_1^2; n > p + q\} \), an equivalent form of the invariant SPRT is:

accept or reject \( H_0 \) according as the lower or upper inequality in

\[
T_n^2 < T_n^2 < T_n^2, \quad n > p + q,
\]  

(4.22)

is first violated, where \( T_n^2 \) is defined in (4.16) and the critical limits are the solutions of equations of the form (3.20). Further, by the monotonicity of \( L_n \) in \( T_n^2 \), we have, by Theorem 2.9, that the SPRT given above effectively tests the hypotheses-formulation given in (3.21), where \( \lambda_i^2 = \sum_{i=1}^{\infty} \frac{1}{n_i} \frac{1}{\lambda_i} \).
Since the test described above is a sequential $T^2$-test, Jackson and Bradley's (1961a) result proves termination with probability one, thereby allowing one to use Wald boundaries with the procedure achieving approximately the specified Type I and Type II probabilities of error. As in the case of the sequential test procedures of Chapter III, neither ASN nor OC formulae are available in this case.

C. Case (ii): $\Sigma$ Unknown, but Estimated Independently

As in Section B of this chapter, let us assume that the random vectors $Y_1, \ldots, Y_n, \ldots$ satisfy model (4.1). Further, let us assume that there exists a random matrix $V$ such that $mV$ is distributed independently of the $Y_i$'s according to $W_p(\Sigma, m), m \geq p$. As in Chapter III, the class of modified sequential probability models is denoted by $\hat{Y}_\Omega = \{\hat{Y}_\theta; \hat{Y}_\theta = Y_\theta \times V_\theta, \theta \in \Omega\}$, where $Y_\theta$ represents the sequential probability model of Section B of this chapter and $V_\theta$ represents the probability model corresponding to the random matrix $V$. The parameter space $\Omega$ is again given by (4.2). Suppose that we wish to test sequentially $H_0$ against $H_1$, given by (4.4), utilizing the additional information on the covariance matrix $\Sigma$.

For every $n \geq 1$, let $\hat{G}$ be a group of component-wise transformations $\hat{g}$ defined by
\( g(y_i, x_i, v) = (C y_i, D x_i, C V C'), i = 1, 2, \ldots \) (4.23)

where \( C \) (\( p \times p \)) and \( D \) (\( q \times q \)) are arbitrary nonsingular matrices. By the invariance of the sequential model \( \hat{Y}_\Omega \) of Section B under the group \( G \) and by the result that \( mV^* = (mCV C') \) is independently distributed according to \( W_p(C \Sigma C', m) \), we have that the class of modified sequential models \( \hat{Y}_\Omega \) is invariant under the group \( G \) with the induced group \( \hat{G} \) equal to \( \bar{G} \), the elements of which are defined in (4.6). Hence, as in Section B of this Chapter, the maximal invariant under \( \hat{G} = \bar{G} \) is \( \lambda^2 = \alpha \Sigma^{-1} \beta \).

As before, we shall assume that, for \( n > q + 1 \), the \( (q \times n) \) matrix \( X_n \), defined in (4.7), has rank \( q \). The following lemma establishes the sufficient sequence for the class of modified sequential models \( \hat{Y}_\Omega \).

**Lemma 4.1:** Let \( y_1, \ldots, y_n \), \( n > q + 1 \), be \((p \times 1)\) independent random vectors, each \( y_i \) distributed according to \( N_p(\alpha + B x_i, \Sigma) \), and let \((a_n, B_n, E_n)\) be as defined in (4.8) through (4.12). Let \( mV \) be distributed independently of \( y_1, \ldots, y_n \) according to \( W_p(\Sigma, m) \), \( m \geq p \); and let

\[
W_n = \frac{1}{(m+n-q-1)} [mV + (n-q-1)E_n].
\] (4.24)

Then, \( \hat{S}_n = (a_n, B_n, W_n) \) is sufficient for \( \hat{Y}_n(\Omega) \) for each \( n > q + 1 \).
Proof: Since $mV$ and $y_1, \ldots, y_n$ are independent and since $mV$ is distributed $W_p(\Sigma, m)$ we can write the joint p.d.f. as:

$$f_1(y_1, \ldots, y_n, mV|\theta) = f_1(y_1, \ldots, y_n|\theta) f_2(mV|\Sigma)$$

where, as shown in (Anderson, 1958, p. 183),

$$f_1(y_1, \ldots, y_n|\theta) = (2\pi)^{-n/2} |\Sigma|^{-n/2} \times \exp\left(-\frac{1}{2} \text{Tr}( [(a_n, B_n) - (\alpha, \beta)] H_n [(a_n, B_n) - (\alpha, \beta)]')^{-1} \right)$$

$$\times \exp\left(-\frac{1}{2} \text{Tr}( [(n-q-1)E_n]^{-1}) \right)$$

and, as given in the proof of Lemma 3.2,

$$f_2(mV|\Sigma) = K^{-1}_V |mV|^{(m-p+1)/2} |\Sigma|^{-m/2} \times \exp\left(-\frac{1}{2} \text{Tr}(mV\Sigma^{-1}) \right).$$

Therefore, using the distributivity property of the trace-operator, we can write the product of the last terms of (4.25) and (4.26), respectively, as

$$\exp\left(-\frac{1}{2} \text{Tr}( [(n-q-1)E_n + mV]\Sigma^{-1}) \right)$$

$$= \exp\left(-\frac{1}{2} \text{Tr}( [(m+n-q-1)W_n]\Sigma^{-1}) \right).$$

Hence, by the Factorization Criterion for sufficient statistics (Theorem 2.4), we have that $\hat{S}_n = (a_n, B_n, W_n)$ is sufficient for $\hat{y}(n)|\Omega$, for each $n > q + 1$. q.e.d.
Since $V$ is positive-definite with probability one and $E_n$ is positive-semidefinite, $W_n$ is positive-definite for $n > q + 1$. Thus, since the group $\hat{G}_S$ induced by $\hat{G}$ on the space of $S_n$ has elements defined by (4.15) with $W_n$ replacing $E_n$, we have that for $n > q + 1$, a maximal invariant under $\hat{G}_S$ is

$$\hat{T}_n^2 = V_n(\hat{S}_n) = c_n^{1/2} W_n^{-1/2} \alpha_n,$$  \hspace{1cm} (4.27)$$

where $c_n$ is defined in Corollary 4.1.

The verification that $\hat{T}_n^2$ is a maximal invariant under $\hat{G}_S$ is completely analogous to the verification for $T_n^2$ of Section B of this chapter, except that $E_n$ is replaced by $W_n$. Therefore, applying Theorem 2.6 as in Section B, we have that the sequence $\{\hat{T}_n^2\}$ is an invariantly sufficient sequence for the class of modified sequential models $\hat{\mathcal{Y}}_\Omega$ under $\hat{G}$.

As verified in Appendix C, the transitivity of the sequence $\{\hat{S}_n = (\alpha_n, B_n, W_n)\}$ under $\hat{G}_S$ follows, via Corollary 2.1, from the recurrence relationships given in (4.17) through (4.19) and from the fact that

$$W_{n+1} = \frac{1}{(m+n-q)[mV + (n-q)E_{n+1}]}.$$  \hspace{1cm} (4.28)$$

Therefore, as noted in Chapter II, the transitivity of the invariantly sufficient sequence $\{\hat{T}_n^2\}$ follows from the transitivity of the sufficient sequence $\{\hat{S}_n\}$.

The following corollary to Theorem 2.14 establishes the distribution of $\hat{T}_n^2$ for $n > q + 1$. 
Corollary 4.3: Let \( Y_1, \ldots, Y_n \), \( n > q + 1 \), \( (a_n, B_n, E_n) \), \( V \), and \( W_n \) be as defined in Lemma 4.1, and let \( \hat{T}_n^2 \) be as defined in (4.27). Then, \( \hat{T}_n^2 \) is distributed according to \( T_p^2(\tau^2; m+n-q-1) \), where \( \tau^2 = c_n \lambda^2 \) and \( c_n \) is as defined in Corollary 4.1.

Proof: From Corollary 4.1, we have that \( \sqrt{c_n} a_n \) is distributed \( N(\sqrt{c_n} a, \Sigma) \) and that \( (n - q - 1)E_n \) is distributed independently according to \( W_p(\Sigma, n - q - 1) \). Since \( mV \) is distributed independently according to \( W_p(\Sigma, m) \), we have, by Theorem 2.13, that \( (m + n - q - 1) W_n \) is distributed according to \( W_p(\Sigma, m + m - q - 1) \) and independently of \( \sqrt{c_n} a_n \). Thus, by Theorem 2.14, \( \hat{T}_n^2 = \sqrt{c_n} a_n' W_n^{-1} a_n \) is distributed according to \( T_p^2(c_n \lambda^2; m + n - q + 1) \). q.e.d.

Hence, by application of Theorem 2.6, the invariant SPRT of (4.4) consists of decision rules of the form (1.4) with \( L_n \) replaced by \( \hat{L}_n \), given below:

\[
\hat{L}_n = \frac{f(\hat{T}_n^2 | c_n \lambda_1^2)}{f(\hat{T}_n^2 | c_n \lambda_0^2)}, \quad n > q + 1.
\] (4.29)

By Corollary 2.2, we can write \( \hat{L}_n \) as

\[
L_n = \exp\left[ -\frac{1}{2} c_n (\lambda_1^2 - \lambda_0^2) \right] \left( \begin{array}{c} \frac{m+n-q}{2} \frac{p}{2}; \frac{c_n \lambda_1^2 \hat{T}_n^2}{2(m+n-q-1+\hat{T}_n^2)} \end{array} \right). \] (4.30)
As in Section B, termination with probability one follows from Jackson and Bradley's (1961a) result. Further, by the monotonicity of \( \hat{L}_n \) as a function of \( \hat{T}_n^2 \) (Corollary 2.3), an equivalent form of the invariant SPRT is given by decision rules of the form (3.29) for \( n > q + 1 \) with critical limits as defined in (3.30). Also, by the MLR property of the family of densities for \( \hat{T}_n^2 \), the above SPRT effectively tests the one-sided hypotheses

\[
H_0 : \lambda^2 \leq \lambda^2_0
\]

\[
H_1 : \lambda^2 \geq \lambda^2_1,
\]

where \( \lambda^2 = \alpha' \Sigma^{-1} \alpha \).
V. TWO-SAMPLE SEQUENTIAL TESTS
ABOUT MEAN VECTORS

A. Introduction

In Chapters III and IV we developed multivariate sequential test procedures for the comparison of two treatments for a model based on within-pair differences of observations. Thus in effect, as noted in those chapters, we actually developed one-sample sequential tests with the above as the principal application. In Chapters V and VI, we will develop two-sample sequential tests which do not require the pairing restriction.

In the univariate case, Hajnal (1961) developed a two-sample sequential $t^2$-test by application of Cox's Theorem (Theorem 2.7). Later, Sampson (1968) showed that an almost identical test could be developed through the Waldian weight function approach.

With respect to the multivariate case, it was noted in Chapter I that, although Jackson and Bradley (1961a) considered two-sample situations, their suggestion was, in fact, to reduce the two-sample problem to an application of the one-sample sequential $T^2$-test via sample paired-vector differences. On strictly intuitive grounds, one can conjecture that the advantages and disadvantages of paired as compared with independent samples in fixed-sample-size analysis also
generally obtain in sequential procedures. In particular, when there is zero correlation between observations within pairs, one might suspect that the loss of degrees of freedom in estimating the covariance matrix using paired samples would, on the average, result in sample sizes greater than those required by an analysis using two independent samples.

HWG (1965, p. 591) have suggested a two-sample sequential $T^2$-test but, unfortunately, did not give a complete verification of its development. Further, unlike Hajnal's (1961) and Sampson's (1968) univariate procedures, the two-sample sequential $T^2$-test suggested by HWG (1965) requires equal sample sizes for the two independent samples, a restriction which may not always be desirable or convenient in practice. For example, in a clinical trial comparing a drug to a placebo, the clinical investigator may prefer to take more observations on the drug than on the placebo.

Accordingly, instead of restricting ourselves to sampling only one observation from each multivariate-normal population at each stage, we shall assume that, at each stage, we may sample $r_1$ and $r_2$ observations from the first and second populations, respectively, where $r_1$ and $r_2$ are integers greater than or equal to unity. For example, at each stage, we may sample one observation from the first population and four from the second population so that at the $n^{th}$ stage we have accumulated $n$ and $4n$ observations, respectively.
Wald (1947, pp. 102, 103) discussed the effects of such grouping; his general conclusions were that: (i) the realized values of the probabilities of Type I and Type II errors cannot exceed the intended values \( \alpha \) and \( \beta \) except by an exceedingly small quantity (which, he states, may be ignored for all practical purposes), (ii) the expected number of observations required to decision will be increased from that of sampling single observations at each stage, and (iii) the realized values of the probabilities of error may be substantially smaller than the intended values \( \alpha \) and \( \beta \) (a feature of grouping that Wald regarded as compensation for the increase in the number of observations).

Further, it might be noted that, with \( r_1 = r_2 = 1 \), the two-sample sequential test, developed in this chapter, reduces to the two-sample sequential \( T^2 \)-test suggested by HWG (1965).

B. Case (i): \( \Sigma \) Unknown

Let \( \gamma_{11}, \ldots, \gamma_{1n_1}, \ldots \) and \( \gamma_{21}, \ldots, \gamma_{2n_2}, \ldots \) be mutually independent \((p \times 1)\) random vectors; for \( i = 1, 2 \), \( \gamma_{ij} \) being distributed according to \( N_p(\mu_i, \Sigma) \), \( j = 1, \ldots, n_i, \ldots \), where \( \mu_i \) \((p \times 1)\) and \( \Sigma \) \((p \times p)\) are unknown parameters. Thus, the parameter space \( \Omega \) is given by

\[
\Omega = \{ \theta = (\mu_1, \mu_2, \Sigma); -\infty < \mu_{ik} < \infty, k = 1, \ldots, p, \quad \Sigma > 0 \}
\]
where, as before, $\Sigma > 0$ denotes $\Sigma$ positive-definite. For convenience, we will write this probability model in the form:

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i, \ldots, \quad i = 1, 2$$

(5.2)

where the $\epsilon_{ij}$'s are i.i.d. $N_p(0, \Sigma)$.

As described in Section A, we shall assume that, at each stage, we sample $r_1$ and $r_2$ observations from the first and second populations, respectively, so that $n_1 = nr_1$ and $n_2 = nr_2$ observations have been accumulated from the respective populations at the $n^{th}$ stage. Thus, at the $n^{th}$ stage of sampling, $nr$ observations have been accumulated in all where $r = r_1 + r_2$.

Suppose that we wish to test sequentially hypotheses about the difference $\delta = (\mu_1 - \mu_2)$ between the two mean vectors. We may consider the following hypotheses-formulation:

$$H_0 : (\delta - \delta_0)' \Sigma^{-1} (\delta - \delta_0) = \lambda_0^2$$

$$H_1 : (\delta - \delta_0)' \Sigma^{-1} (\delta - \delta_0) = \lambda_1^2, \quad \lambda_0^2 < \lambda_1^2,$$

(5.3)

where $\delta_0$ is a $(p \times 1)$ vector of specified constants. Without loss of generality, we assume that we wish to discriminate sequentially between the two composite hypotheses

$$H_0 : \lambda^2 = \lambda_0^2$$

$$H_1 : \lambda^2 = \lambda_1^2, \quad \lambda_0^2 < \lambda_1^2,$$

(5.4)

where $\lambda^2 = \delta' \Sigma^{-1} \delta = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$. 
For every $n \geq 1$, let $G$ be a group of component-wise
transformations $g$ defined by
\begin{equation}
g(y_{ij}) = C(y_{ij} + b),
\end{equation}
for $j = 1, \ldots, n$, and $i = 1, 2,$
where $b$ is an arbitrary $(p \times 1)$ vector and $C$ is an arbitrary
$(p \times p)$ nonsingular matrix. For $i = 1, 2$, it follows from
Theorem 2.11 that the random vectors $\{y_{ij}^* = C(y_{ij} + b)\}$ are
independent identically distributed according to $N_p(C(y_{ij} + b),$
$CEC')$. Thus the class of sequential probability models $Y_\Omega$
remains invariant under $G$ with the induced group $\bar{G}$ consisting
of elements $\bar{g}$ of the form
\begin{equation}
\bar{g}(0) = (C(y_{ij} + b), C(y_{ij} + b), CEC').
\end{equation}

By means of the step-wise procedure given in Theorem 2.1,
we will now show that $\chi^2 = \delta' \Sigma^{-1} \delta$ is a maximal invariant with
respect to $\bar{G}$ on $\Omega$.

We first note that the group $G$ is clearly generated by
the two subgroups $H$ and $K$, the elements of which are
respectively
\begin{equation}
h(y_{ij}) = (y_{ij} + b) \text{ and }
\end{equation}
\begin{equation}
k(y_{ij}) = Cy_{ij},
\end{equation}
where \( b \) and \( C \) are as defined in (5.5). Again from Theorem 2.11, it follows that the random vectors \( \{\gamma_i + b\} \) are i.i.d. \( N_p(\mu_i + b, \Sigma) \) and that the random vectors \( \{C\gamma_i\} \) are i.i.d. \( N_p(C\mu_i, C\Sigma C') \), for \( i = 1, 2 \).

Thus, the subgroups \( H \) and \( K \) induce \( \bar{H} \) and \( \bar{K} \), respectively, on the parameter space \( \Omega \), where the elements of \( \bar{H} \) and \( \bar{K} \) are respectively

\[
\bar{H}(\theta) = ((\mu_1 + b), (\mu_2 + b), \Sigma) \quad \text{and} \quad (5.9)
\]

\[
\bar{K}(\theta) = (C\mu_1, C\mu_2, C\Sigma C') \quad (5.10)
\]

Hence, \( \bar{G} \) is clearly the group generated by the subgroups \( \bar{H} \) and \( \bar{K} \) on the parameter space \( \Omega \).

As the first step in the step-wise procedure given in Theorem 2.1, we wish to show that, with respect to \( \bar{H} \) on \( \Omega \), a maximal invariant is

\[
\gamma(\theta) = (\delta, \Sigma) = (\mu_1 - \mu_2, \Sigma). \quad (5.11)
\]

Note first that \( \Sigma \) is unchanged by the subgroup \( \bar{H} \) on \( \Omega \). Since

\[
(\mu_1 + b) - (\mu_2 + b) = (\mu_1 - \mu_2) = \delta,
\]

\( \delta \) is invariant under \( \bar{H} \). Suppose \( \delta_1(1) = \delta_2(2) \); then

\[
(\mu_1(1) - \mu_2(1)) = (\mu_1(2) - \mu_2(2)), \quad \text{which implies that}
\]

\[
(\mu_1(2), \mu_2(2)) = (\mu_1(1) + b, \mu_2(1) + b) \quad \text{for some} \quad (p \times 1)
vector $\mathbf{b}$—in particular, $\mathbf{b} = (\mathbf{h}_2(2) - \mathbf{h}_2(1))$—thereby satisfying the maximality condition (Definition 2.8).

Let $\mathbf{C}$ be an arbitrary $(p \times p)$ nonsingular matrix corresponding to some $\mathbf{k} \in \bar{K}$. Now, since

$$Y(k(\theta)) = Y(C_{\mathbf{h}_1}, C_{\mathbf{h}_2}, CEC')$$

$$= \begin{pmatrix} C_{\mathbf{h}_1} - C_{\mathbf{h}_2} \\ C_\gamma \end{pmatrix},$$

it follows that, for $k \in K$,

$$\gamma(\theta_1) = \gamma(\theta_2) \text{ implies } \gamma(k(\theta_1)) = \gamma(k(\theta_2)).$$

(5.13)

Let $\bar{K}^*$ be the group of transformations $\bar{k}^*$ defined by

$$\bar{k}^*(\delta, \Sigma) = \gamma(k(\theta)) \text{ when } (\delta, \Sigma) = \gamma(\theta).$$

(5.14)

Thus, combining (5.12) and (5.14), we can write

$$\bar{k}^*(\delta, \Sigma) = (C_{\delta}, CEC').$$

(5.15)

From (5.15), we see that the group $\bar{K}^*$ is identical to the group $\bar{G}$ of Chapter III (except that $\delta$ replaces $\mu$). Hence, it follows from the arguments of Chapter III that $\lambda^2 = \delta_\gamma^\prime \Sigma^{-1} \delta_\gamma$ is a maximal invariant under the group $\bar{K}^*$ of transformations on $\Omega$. Finally, by Theorem 2.1—with (5.13) corresponding to condition (2.2), we have that $\lambda^2 = (\mu_{\mathbf{h}_1} - \mu_{\mathbf{h}_2})^\prime \Sigma^{-1}(\mu_{\mathbf{h}_1} - \mu_{\mathbf{h}_2})$ is maximal invariant with respect to the group $\bar{G}$. 
For \( n \geq 1 \) and for each of the two samples \((i = 1, 2)\), we denote the sample mean vector and the sample covariance matrix at the \( n^{th} \) stage by \( \bar{y}_{\infty} \) and \( D_{\infty} \), respectively, where

\[
\bar{y}_{\infty} = \frac{1}{nr_i} \sum_{j=1}^{nr_i} y_{ij}
\]

(5.16)

\[
D_{\infty} = \frac{1}{(nr_i-1)} \sum_{j=1}^{nr_i} (y_{ij} - \bar{y}_{\infty})(y_{ij} - \bar{y}_{\infty})'.
\]

(5.17)

Further, we denote the pooled sample covariance matrix at the \( n^{th} \) stage by \( D_n \), where

\[
D_n = \frac{1}{nr - 2} \left[ (nr_1-1)D_{1n} + (nr_2-1)D_{2n} \right]
\]

(5.18)

\[
= \frac{1}{nr - 2} \left[ \sum_{i=1}^{nr_1} \sum_{j=1}^{nr_i} (y_{ij} - \bar{y}_{\infty})(y_{ij} - \bar{y}_{\infty})' \right].
\]

Clearly, \( D_n \) is a positive-semidefinite matrix; and since \((nr - 2)D_n\) is distributed according to \( W_p(\Sigma, nr - 2) \) (Anderson, 1958, p. 109), we have that \( D_n \) is positive definite for \((nr - 2) \geq p\). Further, it is a standard result (see, for example, HWG, 1965, p. 591) that \( S_n = (\bar{y}_{\infty}^{1n}, \bar{y}_{\infty}^{2n}, D_n) \) is sufficient for \( Y_{(n)}^{(n)} \), for each \( n \geq 1 \).

Analogous to the group \( \tilde{G} \) induced by \( G \) on \( \Omega \), the group \( G_1 \) induced by \( G \) on the space of \( S_n \) has elements \( g_1 \) defined by

\[
g_1(S_n) = (C(\bar{y}_{\infty}^{1n} + b), C(\bar{y}_{\infty}^{2n} + b), CD_nC').
\]

(5.19)
Further, for \( nr > (p + 1) \), a maximal invariant under \( G_s \) is

\[
T_n^2 = nc (\bar{\gamma}_{1n} - \bar{\gamma}_{2n})' D_n^{-1} (\bar{\gamma}_{1n} - \bar{\gamma}_{2n}),
\]

where \( c = r_1 r_2 / r \). The invariance and maximality of \( T_n^2 \) under the group \( G_s \) can be easily verified by steps completely analogous to those used in showing that \( \lambda^2 \) is maximal invariant with respect to \( \tilde{G} \). As in previous examples, however, any constant is a maximal invariant under \( G_s \) until the sample estimate of \( \Sigma \) is positive-definite with probability one—in this case, until \( nr \) exceeds \( p + 1 \).

As in Chapters III and IV, since the parameter space \( \Omega \) contains an open set in \( \mathbb{R}^k \), \( k = 2p + p(p+1)/2 \), the sufficient statistic \( S_n \) for \( Y(n) \Omega \) is complete, and therefore by Theorem 2.7, Assumption 2.1 is satisfied. It thus follows from Theorem 2.6 (under Assumption 2.1) that \( \{T_n^2\} \) is an invariantly sufficient sequence for the class of sequential models \( Y_\Omega \) under \( G \).

The transitivity of the sequence \( \{S_n = (\bar{\gamma}_{1n}, \bar{\gamma}_{2n}, D_n)\} \) under \( G_s \)—and consequently, the transitivity of the sequence \( \{T_n^2\} \)—follows, via Corollary 2.1, from (5.18) and from the following recurrence relationships (see Appendix C for derivation): for \( i = 1, 2 \).
The distribution of $T_n^2$ has been derived, for example, by Anderson (1958, p. 109) and is presented here as another corollary to Theorem 2.14.

**Corollary 5.1:** For $i = 1, 2$, let $y_{i1}, \ldots, y_{in_i} \sim \text{i.i.d. } N(\mu_i, \Sigma)$, and let $T_n^2$ be as defined in (5.20), $nr \equiv n(r_1 + r_2) > (p + 1)$. Then, $T_n^2$ is distributed according to $T_p^2(\tau; nr - 2)$, where $\tau^2 = nc\lambda^2$, $c = r_1r_2/r$, and

$$\lambda^2 = \delta\Sigma^{-1}\delta = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2).$$

Since $\{T_n^2\}$ is an invariantly sufficient and transitive sequence for $Y_\Omega$, it follows from Theorem 2.6 and the remarks in Section C of Chapter II that the invariant SPRT of (5.4) is given by decision rules of the form (1.4) where

$$L_n = \frac{f(t_n^2 \mid nc\lambda^2_i)}{f(t_n^2 \mid nc\lambda^2_o)}.$$  (5.23)
which, by Corollary 2.2, can be written as

\[
L_n = \exp\left[-\frac{1}{2}nc(\lambda_1^2 - \lambda_0^2)\right] \frac{F\left(\frac{nr-1}{2}, \frac{nc}{2}, \frac{\lambda_1^2}{2(nr-2+T_n^2)}\right)}{F\left(\frac{nr-1}{2}, \frac{nc}{2}, \frac{\lambda_0^2}{2(nr-2+T_n^2)}\right)}
\]  
(5.24)

where \(F(\cdot, \cdot; \cdot)\) is defined in (2.21), \(r = r_1 + r_2\), and \(c = r_1 r_2 / r\).

By the monotonicity of \(L_n\) in \(T_n^2\) (Corollary 2.3) on \((0, \infty)\) for every fixed set

\[
\{p \geq 1, r_1 \geq 1, r_2 \geq 1, 0 \leq \lambda_0^2 < \lambda_1^2; nr > (p + 1)\},
\]

an equivalent form of the invariant SPRT is:

accept or reject \(H_0\) as the lower or upper inequality in

\[
T_n^2 < T_n^2 < T_n^2, \quad nr > p + 1,
\]  
(5.25)

is first violated, where \(T_n^2\) is defined in (5.20) and the critical limits are the solutions of equations of the form (3.20). Also, by Corollary 2.3, the SPRT given above effectively tests the hypotheses-formulation

\[
H_0 : \lambda^2 \leq \lambda_0^2
\]

\[
H_1 : \lambda^2 \geq \lambda_1^2, \quad \lambda_0^2 < \lambda_1^2
\]  
(5.26)

where \(\lambda^2 = \delta^5 \Sigma_{\infty}^{-1} \delta\).
Termination with probability one follows from Jackson and Bradley's (1961a) result for sequential $T^2$-tests so that Wald boundaries can be used with this procedure to achieve approximately the specified Type I and Type II probabilities of error.

C. Case (ii): $\Sigma$ Unknown, but Estimated Independently

As in Section B of this chapter, we shall assume that the random vectors $\{y_{ij}, j = 1, \ldots, n_i, \ldots, i = 1, 2\}$ satisfy model (5.2). Also, we shall assume that there exists a random matrix $V$ such that $mV$ is distributed independently of the $y_{ij}$'s according to $W_p(\Sigma, m)$, $m \geq p$.

As in the previous examples involving a preliminary estimate of $\Sigma$, we shall denote the class of modified sequential probability models by $\hat{Y}_\Omega$, which we defined in Chapter III. The parameter space $\Omega$ is again given by (5.1). Suppose that now we wish to discriminate sequentially between the two hypotheses in (5.4) utilizing all available information.

For every $n \geq 1$, let $\hat{G}$ be a group of component-wise transformations $\hat{g}$ defined by

$$\hat{g}(y_{ij}, V) = (C(y_{ij} + b), CV')$$

for $j = 1, \ldots, nr_i$ and $i = 1, 2$,

where $b$ is an arbitrary $(p \times 1)$ vector and $C$ is a $(p \times p)$ nonsingular matrix. Since $mV^* = mCV'$ is distributed according
to \( W_p(CEC^', m) \), the invariance of the class of modified sequential models \( \hat{\gamma}_n \) under \( \hat{G} \) follows immediately from the invariance of \( \hat{\gamma}_n \) under the group \( G \) of Section B. Further, since the induced group \( \bar{G} \) on \( \Omega \) is equal to \( \bar{G} \), as defined in (5.6), the maximal invariant with respect to \( \bar{G} \) is

\[
\lambda^2 = \delta^\Omega (\Sigma^{-1} \delta) = (\mu_1 - \mu_2) (\Sigma^{-1} (\mu_1 - \mu_2)).
\]

The following lemma establishes the sufficient sequence for the class of modified sequential models \( \hat{\gamma}_n \).

**Lemma 5.1:** For \( i = 1, 2 \), let \( y_{i1}, \ldots, y_{in_i} \) (\( n_i = nr_i \)) be i.i.d. \( N(\mu_i, \Sigma) \); and let \( (\bar{V}_1, \bar{V}_2, D_n) \) be as defined in (5.16) through (5.18). Let \( mV \) be distributed independently of the \( y_{ij} \)'s according to \( W_p(\Sigma, m) \), \( m > p \); and let

\[
\lambda_n = \frac{1}{(m + nr - 2)} [mV + (nr-2)D_n],
\]

where \( r = r_1 + r_2 \). Then, \( \hat{S}_n = (\bar{V}_1, \bar{V}_2, W_n) \) is sufficient for \( \hat{\gamma}_{(n)\Omega} \) for each \( n \geq 1 \).

Clearly, Lemma 5.1 can be easily proved by arguments analogous to those of the proof of Lemma 3.2.

Since \( D_n \) is positive-semidefinite and \( V \) is positive-definite with probability one, \( W_n \) is positive-definite with probability one. Further, since the group \( \hat{G}_s \) induced by \( \hat{G} \) on the space of \( \hat{S}_n \) has elements of the form (5.19) with \( D_n \) replaced by \( W_n \), a maximal invariant with respect to \( \hat{G}_s \), for \( n \geq 1 \), is
\[ \hat{T}_n^2 = V_n(S_n) = nc(\bar{\alpha}_{1n} - \bar{\alpha}_{2n})'W_n^{-1}(\bar{\alpha}_{1n} - \bar{\alpha}_{2n}) \] (5.29)

where \( c = r_1r_2/r \). Therefore, applying Theorem 2.6 as in Section B, we have that \( \{\hat{T}_n^2\} \) is an invariantly sufficient sequence for the class \( \hat{\gamma}_\Omega \) of modified sequential models under \( \hat{\gamma} \).

Moreover, since
\[ W_{n+1} = \frac{1}{(m+(n-1)r-2)} [mV + ((n+1)r - 2)D_{n+1}] , \] (5.30)
the transitivity of the sequence \( \{\hat{S}_n = (\bar{\alpha}_{1n}, \bar{\alpha}_{2n}, W_n)\} \) under \( \hat{\gamma}_s \) follows from Corollary 2.1 and from the transitivity of \( \{(\bar{\alpha}_{1n}, \bar{\alpha}_{2n}, D_n)\} \) under the group \( \gamma_s \), which was established by the recurrence relations (5.21) and (5.22). Hence, \( \{\hat{T}_n^2\} \) is an invariantly sufficient and transitive sequence for \( \hat{\gamma}_\Omega \). The distribution of \( \hat{T}_n^2 \) is now derived in the following corollary to Theorem 2.14.

**Corollary 5.2:** Let \( \{\gamma_{ij}, j = 1, \ldots, nr_i, i = 1, 2\} \), \( \{\bar{\alpha}_{1n}, \bar{\alpha}_{2n}, D_n\} \), \( V \), and \( W_n \) be as defined in Lemma 5.1, and let \( \hat{T}_n^2 \) be as defined in (5.29). Then, \( \hat{T}_n^2 \) is distributed according to \( T_p(\tau^2; m+nr-2) \), where \( \tau^2 = nc\lambda^2 \).

**Proof:** Since \( \sqrt{nc}(\bar{\alpha}_{1n} - \bar{\alpha}_{2n}) \) is distributed according to \( N_p(\sqrt{nc}(\mu_{1n} - \mu_{2n}), \Sigma) \), where \( c = \left( \frac{1}{r_1} + \frac{1}{r_2} \right)^{-1} = \frac{r_1r_2}{r} \) , and \( (nr - 2)D_n \) is distributed independently according to
\( W_p(\Sigma, nr-2) \) (Anderson, 1958, p. 109) and since \( mV \) is distributed independently according to \( W_p(\Sigma, m) \), we have the following results:

(i) by Theorem 2.13,

\[
(m+nr-2)W_n \sim W_p(\Sigma, m+nr-2),
\]

independently of \( \sqrt{nC} (\bar{y} - \bar{y}_{2n}) \); and

(ii) by Theorem 2.14,

\[
\frac{T^2_n}{\lambda^2} \sim T^2_p(\lambda^2; m+nr-2),
\]

where \( \lambda^2 = (\mu_{11} - \mu_{21})' \Sigma^{-1} (\mu_{11} - \mu_{21}) \). q.e.d.

Hence, application of Theorem 2.6 yields an invariant SPRT of hypotheses (5.4) with decision rules of the form (1.4), except that \( L_n \) is replaced by \( \hat{L}_n \), where

\[
\hat{L}_n = \frac{f(t^2_n \mid \nu_1 \lambda^2)}{f(t^2_n \mid \nu_2 \lambda^2)}, \quad n \geq 1. \tag{5.31}
\]

By Corollary 2.2,

\[
\hat{L}_n = \exp\left\{-\frac{1}{2}nc(\lambda^2_1 - \lambda^2_0)\right\} \frac{F\left[\frac{m+nr-1}{2}, \frac{p}{2}; \frac{nc \lambda^2_1 T^2_n}{2(m+nr-2+T^2_n)}\right]}{F\left[\frac{m+nr-1}{2}, \frac{p}{2}; \frac{nc \lambda^2_0 T^2_n}{2(m+nr-2+T^2_n)}\right]}. \tag{5.32}
\]

As in Section B of this chapter, Jackson and Bradley's (1961a) result proves termination with probability one.
Further, by the monotonicity of \( \hat{I}_n \) as a function of \( \hat{I}_n^2 \) (Corollary 2.3), we have that: (i) an equivalent form of the invariant SPRT given above consists of decision rules of the form (5.25) with critical limits as defined in (3.30); and (ii) the above SPRT effectively tests hypotheses of the form (5.26).
VI. TWO-SAMPLE SEQUENTIAL TESTS ABOUT MEAN VECTORS ADJUSTED FOR COVARIATES

A. Introduction

In Chapter IV, we developed single-sample sequential tests for statistical hypotheses about mean vectors adjusted for covariates; in Chapter V, we developed two-sample sequential tests for hypotheses about the difference of two mean vectors. In this chapter, we develop what might be considered as an extension of the procedures of Chapters IV and V—that is, two-sample sequential tests for statistical hypotheses about the difference of two mean vectors adjusted for covariates.

Using Wald's method of weight functions, Sampson (1968) developed a univariate two-sample sequential \( t^2 \)-test for testing the difference of two means adjusted for covariate-effects. With \( p = 1 \), the sequential test in Section B of this chapter reduces to the sequential test derived by Sampson (1968) via weight functions except that the first argument of the confluent hypergeometric function differs by the constant \( 1/2 \).

As in Chapter V, we will assume that, at each stage of sampling, we may sample \( r_1 \geq 1 \) and \( r_2 \geq 1 \) observations from the first and second multivariate-normal populations, respectively, so that \( n_1 = nr_1 \) and \( n_2 = nr_2 \) observations have been accumulated at the \( n^{th} \) stage.
B. Case (i): $\Sigma$ Known

Let $\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1}, \ldots$ and $\mathbf{y}_{21}, \ldots, \mathbf{y}_{2n_2}, \ldots$ be mutually independent $(p \times 1)$ random vectors, each $\mathbf{y}_{ij}$ being distributed according to $N_p(\alpha_i + \mathbf{B} \mathbf{x}_{ij}, \Sigma)$ for $i = 1, 2$, where $\alpha_i (p \times 1)$, $\mathbf{B} (p \times q)$, and $\Sigma (p \times p)$ are unknown parameters and each $\mathbf{x}_{ij}$ is a $(q \times 1)$ vector of known covariates. Thus, the parameter space $\Omega$ is given by

$$\Omega = \{\theta = (\alpha_1, \alpha_2, \mathbf{B}, \Sigma) : -\infty < \alpha_{ij} < \infty, -\infty < \beta_{jk} < \infty, k = 1, \ldots, q, j = 1, \ldots, p, i = 1, 2; \Sigma > 0\} \quad (6.1)$$

where, as before, $\Sigma > 0$ denotes $\Sigma$ positive-definite.

Equivalently, we shall write:

$$\mathbf{y}_{ij} = \alpha_i + \mathbf{B} \mathbf{x}_{ij} + \xi_{ij} \quad (6.2)$$

for $j = 1, \ldots, n_i (= n r_1), \ldots$ and $i = 1, 2$

where the $\xi_{ij}$'s are i.i.d. $N_p(0, \Sigma)$. It should be noted that, besides assuming equal covariance matrices for the two populations, we are also assuming that the matrix of regression coefficients $\mathbf{B}$ is the same for both populations.

In the univariate case ($p = 1$), the parameter $(\alpha_1 - \alpha_2)$ is the distance (parallel to the $y$-axis) between the two parallel hyperplanes defined in (6.2) and represents the difference of the two population means adjusted for covariates. Correspondingly, we take the vector difference $(\alpha_1 - \alpha_2)$ to
represent the difference of the population mean vectors adjusted for covariates.

We shall assume that, without loss of generality, the problem is to discriminate sequentially between the two composite hypotheses

\[ H_0 : \lambda^2 = \lambda_0^2 \]

\[ H_1 : \lambda^2 = \lambda_1^2, \lambda_0^2 < \lambda_1^2 \]

where \( \lambda^2 = (\alpha_1 - \alpha_2)'\Sigma^{-1}(\alpha_1 - \alpha_2) \). If \( \beta \) were known, one could simply form adjusted observations \( \{z_{ij} = (y_{ij} - \beta x_{ij}) \} \) and apply the sequential test procedure of Chapter V. In our context, however, \( \beta \) is an unknown nuisance parameter.

For every \( n \geq 1 \), let \( G \) be a group of component-wise transformations \( g \) defined by

\[ g(y_{ij}, x_{ij}) = (C(y_{ij} + b), D(x_{ij} + e)) \]

(6.4)

for \( j = 1, \ldots, nr_i \) and \( i = 1, 2 \)

where \( b(p \times 1) \) and \( e(q \times 1) \) are arbitrary vectors and \( C(p \times p) \) and \( D(q \times q) \) are arbitrary nonsingular matrices. By Theorem 2.11, the random vectors \( \{y_{ij}^* = C(y_{ij} + b)\} \) are independently distributed,

\[ y_{ij}^* \sim N_p(\tilde{\alpha}_{ij} + \beta^* x_{ij}^*, I^*) \]

where \( x_{ij}^* = D(x_{ij} + e) \).
Thus, the class of sequential probability models \( Y_\omega \) remains invariant under the group of transformations \( G \) with the induced group \( \bar{G} \) on \( \Omega \) defined by

\[
g(\theta) = (C(\alpha_1 + b - \beta e), C(\alpha_2 - \beta e), C\beta D^{-1}, CE^\prime). \tag{6.5}
\]

Using the step-wise procedure given in Theorem 2.1, we now wish to show that a maximal invariant with respect to the group \( \bar{G} \) on \( \Omega \) is

\[
\lambda^2 = (\alpha_1 - \alpha_2)'\Sigma^{-1}(\alpha_1 - \alpha_2).
\]

We first note that the group \( G \) is clearly generated by the two subgroups \( H \) and \( K \), defined respectively by

\[
h(y, x) = (y + b, x + e) \tag{6.6}
\]

\[
k(y, x) = (Cy, Dx). \tag{6.7}
\]

By Theorem 2.11, the random vectors \( \{y_i + b\} \) are independently distributed,

\[
(y_i + b) \sim N_p ((\alpha_i + b - \beta e) + \beta(x_i + e), \Sigma),
\]
so that the subgroup $H$ induces the subgroup $\tilde{H}$ on $\Omega$ with elements

$$\tilde{h}(\theta) = (\tilde{\alpha}_1 + b - \tilde{\beta} e, \tilde{\alpha}_2 + b - \tilde{\beta} e, \tilde{\beta}, \Sigma). \quad (6.8)$$

As the first step in verifying the invariance and maximality of $\lambda^2$ under the group $G$ on $\Omega$, we will show that a maximal invariant under the subgroup $\tilde{H}$ is

$$\gamma(\theta) = (\tilde{\alpha}_1 - \tilde{\alpha}_2, \tilde{\beta}, \Sigma). \quad (6.9)$$

It should be noted that $\tilde{\beta}$ and $\Sigma$ are unchanged by the subgroup of transformations $\tilde{H}$ on $\Omega$. The invariance of $(\tilde{\alpha}_1 - \tilde{\alpha}_2)$ under $\tilde{H}$ on $\Omega$ is accordingly clear. The maximality of $\gamma(\theta)$ is obtained as follows. Suppose that $\gamma(\theta_1) = \gamma(\theta_2)$ so that $(\tilde{\alpha}_1(1) - \tilde{\alpha}_2(1)) = (\tilde{\alpha}_1(2) - \tilde{\alpha}_2(2))$. Then, with $b = (\tilde{\alpha}_2(1) - \tilde{\alpha}_2(2))$ and $e = 0$, it follows that

$$\tilde{\alpha}_i(2) = \tilde{\alpha}_i(1) + b - \tilde{\beta} e \quad (i = 1, 2)$$

thereby satisfying the maximality condition (Definition 2.8).

Let us now consider the subgroup $K$. As shown in Chapter IV, the random vectors $\{\tilde{C}_{\nu ij}\}$ ($i = 1, 2$) are independently distributed,

$$\tilde{C}_{\nu ij} \sim N_p (\tilde{C}_{\nu i} + \tilde{C} \tilde{D}^{-1}(\tilde{D}_{\nu ij}), \Sigma \Sigma')$$
Thus, the subgroup $\mathcal{K}$ induced by the subgroup $K$ consists of elements of the form

$$\mathcal{K}(\theta) = (C_{a_1}, C_{a_2}, C_{B^{-1}}, C_{E'}) \tag{6.10}$$

Since, for every $k \in \mathcal{K}$,

$$\mathcal{K}(\theta) = (C_{a_1} - C_{a_2}, C_{B^{-1}}, C_{E'})$$

$$= (C(a_1 - a_2), C_{B^{-1}}, C_{E'}), \tag{6.11}$$

it follows that

$$\gamma(\theta_1) = \gamma(\theta_2) \text{ implies } \gamma(\mathcal{K}(\theta_1)) = \gamma(\mathcal{K}(\theta_2)), \tag{6.12}$$

thereby satisfying condition (2.2) of Theorem 2.1.

Let $\mathcal{K}^*$ be the group of transformations $\mathcal{K}^*$ defined by

$$\mathcal{K}^*(a_1 - a_2, \beta, \Sigma) = \gamma(\mathcal{K}(\theta)) \tag{6.13}$$

when

$$\gamma(\theta) = (a_1 - a_2, \beta, \Sigma).$$

Therefore, combining (6.11) and (6.13), we can write

$$\mathcal{K}^*(a_1 - a_2, \beta, \Sigma) = (C(a_1 - a_2), C_{B^{-1}}, C_{E'}).$$

Hence, it follows from the results of Chapter IV—with $a$ replaced by $(a_1 - a_2)$—that a maximal invariant under $\mathcal{K}^*$ is

$$\lambda^2 = (a_1 - a_2)' \Sigma^{-1} (a_1 - a_2).$$

Finally, by Theorem 2.1, it follows that $\lambda^2$ is maximal invariant with respect to the group $\mathcal{G}$ on $\Omega$. 
For purposes of estimation, we need to assume that, for \( nr > q + 1 \ (r = r_{1} + r_{2}) \), the \((q \times q)\) matrix \( A_{n} \) has rank \( q \), where \( A_{n} \) is defined as follows:

\[
A_{n} = A_{1n} + A_{2n}, \quad \text{where for } i = 1, 2
\]

\[
A_{in} = x_{in} x_{in}'
\]

\[
x_{in} = \begin{bmatrix} x_{i1} - \bar{x}_{in}, \ldots, x_{i(nr_{i})} - \bar{x}_{in} \end{bmatrix}
\]

\[
\bar{x}_{in} = \frac{1}{nr_{i}} \sum_{j=1}^{nr_{i}} x_{ij}
\]

(6.14) (6.15) (6.16) (6.17)

Denote the sample mean vectors by \( \bar{y}_{in} \) \((i = 1, 2)\), as defined in (5.16); and for \( i = 1, 2 \), let

\[
y_{in} = \begin{bmatrix} y_{i1} - \bar{y}_{in}, \ldots, y_{i(nr_{i})} - \bar{y}_{in} \end{bmatrix}.
\]

(6.18)

It follows from (5.17) and (5.18) of Chapter V that

\[
(nr_{i} - 1)D_{in} = y_{in} y_{in}'
\]

(6.19)

and

\[
(nr - 2)D_{n} = y_{in} y_{in}' + y_{2n} y_{2n}'.
\]

(6.20)

The sufficient statistic for \( Y_{(n)_{0}} \) is given by the following standard result from multivariate linear model theory (see, for example, Anderson, 1958, p. 183).
Theorem 6.1: Let \( \{y_{ij}, j = 1, \ldots, nr, i = 1, 2\} \) be independently distributed, each \( y_{ij} \) according to \( N_p(\alpha_i + Bx_{ij}, \Sigma) \). For each \( n \) such that \( nr \equiv n(r_1 + r_2) > q + 1 \), 
\[
S_n = (a_{1n}, a_{2n}, B_n, E_n) \text{ is sufficient for } Y(n) \Omega', \text{ where}
\]
\[
B_n = (Y_{1n}X_{1n}' + Y_{2n}X_{2n}') A_n^{-1}; \quad (6.21)
\]
\[
a_{in} = y_{in} - B_n x_{in} \quad (i = 1, 2); \quad (6.22)
\]
\[
E_n = \left( \frac{1}{nr-q-2} \right) [(nr-2)D_n - B_n A_n B_n^t]. \quad (6.23)
\]

Further, \( (a_{1n}, a_{2n}, B_n) \) is normally distributed with mean \( (\alpha_1, \alpha_2, B) \) and the covariance matrix of the \( i^{th} \) and \( j^{th} \) rows of \( (a_{1n}, a_{2n}, B_n) \) is \( \sigma_{ij} H_n^{-1} \), where
\[
H_n = \begin{bmatrix}
    nr_1 & 0 & nr_1 x_{1n}' \\
    0 & nr_2 & nr_2 x_{2n}' \\
    nr_1 x_{1n} & nr_2 x_{2n} & \Sigma \sum_{i=1}^q x_{1i}x_{2j}' \\
\end{bmatrix} \quad (6.24)
\]

and \( (nr - q - 2)E_n \) is independently distributed according to \( W_p(\Sigma, nr - q - 2) \).

In the following corollary to Theorem 6.1, we restrict our attention to the distribution of \( (a_{1n} - a_{2n}) \), the sample estimate of \( (\alpha_1 - \alpha_2) \) at the \( n^{th} \) stage.
Corollary 6.1: For \( nr > q + 1 \), \( S_n = (\alpha_{1n}, \alpha_{2n}, B_n, E_n) \), as defined in Theorem 6.1, is sufficient for \( y(n) \in \Omega \) and further, \( (\alpha_{1n} - \alpha_{2n}) \) is distributed according to \( N_p(\alpha_1 - \alpha_2, z_n^{-1} \Sigma) \), where

\[
z_n^{-1} = \left( \frac{1}{nc} + \frac{x_{1n} - \bar{x}_{2n}}{z_n^{-1}} \right) \times A_n^{-1}(\alpha_{1n} - \alpha_{2n}) \tag{6.25}
\]

and \( c = \left( \frac{1}{r_1} + \frac{1}{r_2} \right)^{-1} \), and \( (nr - q - 2)E_n \) is distributed independently according to \( W_p(\Sigma, nr - q - 2) \).

**Proof:** As can be easily verified, the inverse of \( H_n \) is

\[
H_n^{-1} = \begin{bmatrix}
    u_{11n} & u_{12n} & -\bar{x}_{1n}A_n^{-1} \\
u_{21n} & u_{22n} & -\bar{x}_{2n}A_n^{-1} \\

    -A_n^{-1}x_{1n} & -A_n^{-1}x_{2n} & A_n^{-1}
\end{bmatrix} \tag{6.26}
\]

where \( A_n \) is defined in (6.14) and where

\[
u_{ijn} = \begin{cases}
    \frac{1}{nr_i} + \frac{x_i'A_n^{-1}x_i}{\beta_{ijn}A_n'x_{jn}} & \text{for } i = j \\
    \frac{x_i'A_n^{-1}x_j}{\beta_{ijn}A_n'x_{jn}} & \text{for } i \neq j
\end{cases} \tag{6.27}
\]

Thus, it follows from Theorem 6.1 that the \((2p \times 1)\) vector

\[
\begin{bmatrix}
    \alpha_{1n} \\
    \alpha_{2n}
\end{bmatrix}
\]

is distributed \( N_{2p}\left(\begin{bmatrix}
    \alpha_1 \\
    \alpha_2
\end{bmatrix}, \begin{bmatrix}
    u_{11n}\Sigma & u_{12n}\Sigma \\
    u_{21n}\Sigma & u_{22n}\Sigma
\end{bmatrix}\right) \).
Then, since \((a_{1n} - a_{2n}) = [I_p, -I_p] \begin{pmatrix} \bar{a}_{1n} \\ \bar{a}_{2n} \end{pmatrix}\), we have from Theorem 2.11 that \((a_{1n} - a_{2n})\) has a \(p\)-dimensional normal distribution with mean vector \((a_{1n} - a_{2n})\) and covariance matrix

\[
\begin{aligned}
\Sigma_{1n} - \Sigma_{2n} - \Sigma_{12n} + \Sigma_{22n} \\
= (\Sigma_{1n} - \Sigma_{12n} - \Sigma_{21n} + \Sigma_{22n}) \\
= \left(\frac{1}{n_1} + \frac{1}{n_2} + \bar{x}_{1n} A_{1n}^{-1} \bar{x}_{1n} - \bar{x}_{2n} A_{2n}^{-1} \bar{x}_{2n}\right) - \bar{x}_{1n} A_{1n}^{-1} \bar{x}_{2n} + \bar{x}_{2n} A_{2n}^{-1} \bar{x}_{1n} \\
= z_n^{-1} \Sigma_{1n} - \Sigma_{2n} - \Sigma_{12n} + \Sigma_{22n} \\
&\quad = z_n^{-1} \Sigma_{1n}.
\end{aligned}
\]  

(6.28)

Finally here, since, by Theorem 6.1, \(E_n\) is distributed independently of \((a_{1n}, a_{2n})\), it follows that \((nr - q - 2)E_n\) is distributed independently of \((a_{1n} - a_{2n})\) according to \(W_p(\Sigma, nr - q - 1)\). q.e.d.

The group \(G_s\) induced by \(G\) on the space of \(S_n\) has elements \(g_s\) defined by

\[
g_s(S_n) = (C(a_{1n} + b - B_n e), C(a_{2n} + b - B_n e), C D_n^{-1}, CE_n C')
\]  

(6.29)

where, as before, \(b_{(p \times 1)}\) and \(e_{(q \times 1)}\) are arbitrary vectors and \(C_{(p \times p)}\) and \(D_{(q \times q)}\) are arbitrary nonsingular matrices. Further, for \(nr > p + q + 1\), a maximal invariant with respect to \(G_s\) is
where $z_n$ is given by (6.25). The verification that $T^2_n$ is maximal invariant with respect to $G_S$ is not given here since it would be completely analogous to the verification that $\lambda^2$ is maximal invariant with respect to $\overline{G}$. Again, because the distribution of $E_n^*$ is singular for $nr \geq p + q + 1$, any constant is a maximal invariant with respect to $G_S$ until $nr$ exceeds $(p + q + 1)$.

Since the parameter space $\Omega$ clearly contains an open set in $\mathbb{R}^k$, $k = 2p + pq + p(p+1)/2$, the sufficient statistic $S_n$ for $Y_{(n)}\Omega$ is complete so that, via Theorem 2.7, Assumption 2.1 obtains. Therefore by Theorem 2.6, the sequence $\{T^2_n\}$ is an invariantly sufficient sequence for the class $Y_{\Omega}$ of sequential models under $G$.

Moreover, since $S_{n+1}$ can be expressed as a function of $S_n$ and of the observations $\{y_{ij}, j = nr_{ij}, ..., (n + 1)r_{ij}\}$ at the $(n + 1)^{st}$ stage (see Appendix C for the pertinent recurrence relationships, as well as their derivation), it follows from Corollary 2.1 that the sequence $\{S_n = (a_{1n}, a_{2n}, B_n, E_n)\}$ is transitive under $G_S$.

Hence, the sequence $\{T^2_n\}$ is an invariantly sufficient and transitive sequence for $Y_{\Omega}$ so that, via Theorem 2.6, the invariant SPRT of (6.3) is given by decision rules of the form (1.4) where $L_n$ is the ratio of the p.d.f. of $T^2_n$ under $H_1$ and
the p.d.f. of $T_n^2$ under $H_0$.

The following corollary to Theorem 2.14 establishes the distribution of $T_n^2$.

**Corollary 6.2:** For $i = 1, 2$, let $Y_{ii}, \ldots, Y_{i_{n_i}}$ ($n_i = n_{r_i}$) be independent random vectors, each $Y_{ij}$ distributed according to $N_p(\alpha_i + \beta X_{ij}, \Sigma)$, and let $T_n^2$ be as defined in (6.30), $n_r > p + q + 1$. Then, $T_n^2$ is distributed according to $W_p(z_n^2; n_r - q - 2)$, where $z_n$ is defined in (6.25) and $\lambda^2 = (\alpha_{n_1} - \alpha_{n_2})'z_n^{-1}(\alpha_{n_1} - \alpha_{n_2})$.

**Proof:** From Corollary 6.1, we have that $\sqrt{n_r}(\alpha_{1n} - \alpha_{2n})$ is distributed $N_p(\sqrt{n_r}(\alpha_{n_1} - \alpha_{n_2}), \Sigma)$ and that $(n_r - q - 2)E_n$ is distributed independently according to $W_p(\Sigma, n_r - q - 2)$, which is nonsingular for $(n_r - q - 2) > p$. The required result therefore follows from Theorem 2.14. q.e.d.

Accordingly, by Corollary 2.2, $I_n$ can be written as

$$I_n = \exp \left[ -\frac{1}{2}z_n^2(\lambda_1^2 - \lambda_2^2) \right] \frac{\left[ \frac{nr-q-1}{2}, \frac{p}{2}; \frac{z_n^2}{2(nr-q-2+T_n^2)} \right]}{\left[ \frac{nr-q-1}{2}, \frac{p}{2}; \frac{T_n^2}{2(nr-q-2+T_n^2)} \right]}$$

(6.31)

where $F(\cdot, \cdot; \cdot)$ is the confluent hypergeometric function defined in (2.21).
Since \( I_n = I_n(T_n^2) \) is an increasing function of \( T_n^2 \) (Corollary 2.3) on \((0, \infty)\) for every fixed set
\[ \{p \geq 1, q \geq 1, r_1 \geq 1, r_2 \geq 1, 0 \leq \lambda_o^2 < \lambda_1^2; nr > p + q + 1 \}, \]
an equivalent form of the invariant SPRT is:
accept or reject \( H_0 \) according as the lower or upper inequality in
\[ T_n^2 < t^*_n < t^*_n, \quad nr > p + q + 1 \]  \tag{6.32}
is first violated, where \( T_n^2 \) is defined in (6.30) and the critical limits are the solutions of equations of the form (3.20). Also, by the monotonicity of \( I_n \) in \( T_n^2 \), it follows from Theorem 2.9 that the SPRT given above also, in effect, tests the hypotheses-formulation
\[ H_0 : \lambda^2 \leq \lambda_o^2 \]  \tag{6.33}
\[ H_1 : \lambda^2 \geq \lambda_1^2, \quad \lambda_o^2 < \lambda_1^2 \]
where \( \lambda^2 = (\gamma_1 - \gamma_2)\mu^{-1}(\gamma_1 - \gamma_2). \)

From Jackson and Bradley's (1961a) result for sequential \( T^2 \)-tests, termination with probability one follows so that Wald boundaries can be used with the sequential test procedure achieving approximately the Type I and Type II probabilities of error. As in the case of the sequential tests considered in previous chapters, however neither ASN nor OC formulae are available in this case.
C. Case (ii): Σ Unknown, but Estimated Independently

In correspondence with Section B of this chapter, we shall assume that the $(p \times 1)$ random vectors \( \{y_{ij}, j = 1, \ldots, n_i, \ldots, i = 1, 2 \} \) satisfy model (6.2). And as in previous chapters, we now wish to discriminate between the two composite hypotheses in (6.3) when an independent estimate, \( V \) say, is available. We assume that \( V \) is a random matrix such that \( mV \) is distributed independently of the \( \hat{\gamma}_{ij} \)'s according to \( W_p(\Sigma, m) \), \( m \geq p \). The parameter space \( \Omega \) remains the same as in (6.1).

For every \( n \geq 1 \), let \( \hat{G} \) be a group of component-wise transformations \( \hat{g} \) given by

\[
g(\hat{\gamma}_{ij}', \hat{x}_{ij}', V) = (C(\hat{\gamma}_{ij} + b), D(\hat{x}_{ij} + e), \text{CVC}') \tag{6.34}
\]

for \( j = 1, \ldots, n_i (= nr_i) \) and \( i = 1, 2 \)

where \( b, e, C, \) and \( D \) are as defined in (6.4). Let \( \hat{Y}_\Omega \) denote the class of modified sequential probability models, as defined in Chapter III. Clearly, \( \hat{Y}_\Omega \) is invariant under the group of transformations \( \hat{G} \) since \( Y_\Omega \) is invariant under the group \( G \) (of Section B) and \( m\text{CVC}' \) is independently distributed \( W_p(\Sigma\text{C}' \Sigma, m) \). Further, since the group \( \hat{G} \) induced by \( G \) on \( \Omega \) equals \( \tilde{G} \), it follows that the maximal invariant with respect to \( \tilde{G} \) is

\[
\chi^2 = (\hat{a}_1 - \hat{a}_2)' \Sigma^{-1} (\hat{a}_1 - \hat{a}_2).
\]
As in Section B of this chapter, we shall assume that the matrix \( A_n \), defined in (6.14), has rank \( q \) for \( nr > q + 1 \), where \( r = r_1 + r_2 \). In the following lemma, we establish the sequence of sufficient statistics for the class of modified sequential probability models \( \hat{Y}_n \).

**Lemma 6.1:** For \( i = 1, 2 \), let \( y_{ij}^{(1)}, \ldots, y_{in_i}^{(1)} (n_i = nr_1) \) be independently distributed, each \( y_{ij}^{(1)} \) according to \( N(a_i + Bx_i, \Sigma) \); let \( (a_{1n}, a_{2n}, B_n, E_n) \) be as defined in (6.21) - (6.23), and let \( nr \equiv n(r_1 + r_2) > q + 1 \). Further, let \( mV \) be distributed independently of the \( y_{ij}^{(1)} \)'s according to \( W_p(\Sigma, m) \), \( m > p \); and let

\[
W_n = \frac{1}{(m+nr-q-2)} [mV + (nr-q-2)E_n].
\]

Then, \( \hat{S}_n = (a_{1n}, a_{2n}, B_n, W_n) \) is sufficient for \( \hat{Y}(n) \), for each \( n \) such that \( nr > q + 1 \).

**Proof:** Lemma 6.1 may be proved exactly as was Lemma 4.1 since the joint p.d.f. of the \( y_{ij}^{(1)} \)'s is the same as that given in (4.25) except that \( n \) and \( (n-q-1) \) are replaced by \( nr \) and \( (nr-q-2) \), respectively, and \( (a_n, B_n) \) and \( (a, B) \) are replaced by \( (a_{1n}, a_{2n}, B_n) \) and \( (a_1, a_2, B) \), respectively.

Since \( V \) is positive-definite with probability one and \( E_n \) is positive-semidefinite, \( W_n \) is positive-definite with probability one for \( nr > q + 1 \). Therefore, since the group \( \hat{G}_s \) induced by \( \hat{G} \) on the space of \( \hat{S}_n \) has elements \( \hat{g}_s \) defined by (6.29) with \( W_n \) replacing \( E_n \), it follows that a maximal
invariant with respect to \( G_S \) is

\[
\hat{T}_n^2 = V_n(S_n) = z_n(a_{1n} - a_{2n})' W_n^{-1} (a_{1n} - a_{2n}),
\]

for \( nr > q + 1 \), where \( z_n \) is defined in (6.23). Accordingly, by application of Theorem 2.6 as in Section B, it follows that \( \{\hat{T}_n^2\} \) is an invariantly sufficient sequence for the class of modified sequential models \( \hat{Y}_\Omega \) under the group of transformations \( G \).

Moreover, since

\[
W_{n+1} = \frac{1}{(m+(n+1)r-q-2)} [mV + ((n+1)r-q-2)E_{n+1}],
\]

the transitivity of the sequence \( \{\hat{S}_n = (a_{1n}, a_{2n}, B_n, W_n)\} \) under \( G_S \) follows directly from the transitivity of the sequence \( \{(a_{1n}, a_{2n}, B_n, E_n)\} \) under the group \( G_S \). Hence, it follows that the sequence \( \{\hat{T}_n^2\} \) is transitive for \( \hat{Y}_\Omega \).

In the following corollary to Theorem 2.14, we next derive the distribution of \( \hat{T}_n^2 \) for each \( n \) such that \( nr > q + 1 \).

**Corollary 6.3:** Let \( \{v_{ij}, j = 1, \ldots, nr_i, i = 1, 2\}, (a_{1n}, a_{2n}, B_n, E_n), V, \) and \( W_n \) be as defined in Lemma 6.1, and let \( \hat{T}_n^2 \) be as defined in (6.36). Then, \( \hat{T}_n^2 \) is distributed according to \( T_p^2 (z_n \lambda^2; m + nr - q - 2) \), where \( z_n \) is defined in (6.25) and \( \lambda^2 = (a_{1} - a_{2})' E^{-1} (a_{1} - a_{2}) \).
Proof: Since \((nr - q - 2)E_n^*\) is distributed according to \(W_p(\Sigma, nr - q - 2)\) and \(mV\) is distributed independently according to \(W_p(\Sigma, m)\), it follows from Theorem 2.13 that
\((m + nr - q - 2)W_n^*\) is distributed according to \(W_p(\Sigma, m)\).

Further, since \(\sqrt{z_n}(a_{1n} - a_{2n})^2\) is distributed according to \(N_p(a_{1n} - a_{2n}, \Sigma)\) and independently of \((nr - q - 2)E_n^*\) (Corollary 6.1), it follows that \(\sqrt{z_n}(a_{1n} - a_{2n})^2\) is distributed independently of \(W_n^*\) so that, via theorem 2.14, \(\hat{T}_n^2 = z_n(a_{1n} - a_{2n})^2W_n^{-1}(a_{1n} - a_{2n})^2\) is distributed according to \(T_p^2(z_n \lambda^2; m + nr - q - 2)\). q.e.d.

Hence, by application of Theorem 2.6, the invariant SPRT of (6.3) is given by decision rules of the form (1.4) with \(\hat{L}_n\) replacing \(L_n\), where

\[
\hat{L}_n = \frac{f(t_n^2 \mid z_n \lambda_1^2)}{f(t_n^2 \mid z_n \lambda_0^2)}, \quad nr > q + 1. \tag{6.38}
\]

From Corollary 2.2, we have

\[
\hat{L}_n = \exp\left[-\frac{1}{2}z_n(\lambda_1^2 - \lambda_0^2)\right] \begin{pmatrix} \frac{m + nr - q - 1}{2} & \frac{z_n \lambda_1^2}{2} & \frac{\hat{T}_n^2}{2(m + nr - q - 2 + \hat{T}_n^2)} \\ \frac{m + nr - q - 1}{2} & \frac{z_n \lambda_0^2}{2} & \frac{\hat{T}_n^2}{2(m + nr - q - 2 + \hat{T}_n^2)} \end{pmatrix} \tag{6.39}
\]
where $F(\cdot,\cdot;\cdot)$ is the confluent hypergeometric function defined in (2.21) and, as noted previously, is equal to unity for the case $\lambda^2_0 = 0$.

Termination of this sequential $T^2$-test with probability one follows from Jackson and Bradley's (1961a) result for sequential $T^2$-tests. Further, by appealing once more to the MLR property of the family of densities of the $T^2_p(\tau^2; m)$ distribution, we have the following two results:

(i) the above invariant SPRT can be given equivalently in the form (3.29) with critical limits as defined in (3.30); and

(ii) via Theorem 2.9, the above invariant SPRT effectively tests one-sided hypotheses of the form (6.33).
VII. RELATED CONSIDERATIONS AND TOPICS FOR FURTHER RESEARCH

A. The ASN Function

As noted in Theorem 1.5, Wald (1947) gave approximate procedures for determining the average sample number (ASN) function when the sequential observations are independent. Little is known about the ASN function when the successive observations are not independent; in particular, no ASN formulae have been developed for invariant SPRT's of the type considered in this thesis. However, for purposes of planning sequential experiments, there may be information about the expected sample size from one or both of the following sources: (i) empirical investigations; and (ii) heuristic approximations, such as that proposed by Bhat in unpublished work (given, for example, in Jackson and Bradley, 1961a, pp. 1071-1073).

Sampson (1968) investigated empirically the univariate sequential t-test utilizing concomitant information, which he developed through Wald's method of weight functions. In comparing the sequential t-test in which a single covariate was used to the sequential t-test without covariates, he found empirically that a substantial saving in sample number was achieved when the covariate was used, provided the correlation coefficient ρ between the response and the covariate exceeded 0.6, and that a slight saving was achieved when the covariate
was used, if \( p \) was close to 0.6.

In an empirical evaluation of Jackson and Bradley's (1961a) sequential \( T^2 \)-test, Appleby and Freund (1962) found that the empirical error rates \( \alpha' \) and \( \beta' \) were slightly less than the nominal error rates \( \alpha \) and \( \beta \) and that the empirical ASN were appreciably smaller than the corresponding fixed sample sizes and approximate the ASN that Jackson and Bradley (1961a) obtained using Bhate's conjecture, which we now discuss.

Generally, heuristic approximations to the ASN function are based on the fact that, ignoring excesses over the boundaries at the termination of a sequential test, we have approximately (see, for example, Ghosh, 1970, p. 133):

\[
E[\ln L_n|H_i] = h_i(\alpha, \beta) \quad \text{when } H_i \text{ is true} \quad (i = 0, 1), \quad (7.1)
\]

where \( (\ln L_n) \) denotes the natural logarithm of the probability ratio \( L_n \) and where

\[
h_0(\alpha, \beta) = (1 - \alpha) \ln \left( \frac{\beta}{1 - \alpha} \right) + \ln \left( \frac{1 - \beta}{\alpha} \right)
\]

and

\[
h_1(\alpha, \beta) = \beta \ln \left( \frac{\beta}{1 - \alpha} \right) + (1 - \beta) \ln \left( \frac{1 - \beta}{\alpha} \right).
\]

In general, \( (\ln L_n) \) depends on the sample size \( n \) and a statistic \( Z_n \) based on the first \( n \) observations (\( Z_n \) will be defined for our particular examples later in this section). In order to solve (7.1) for the required ASN values under \( H_0 \) and \( H_1 \) respectively, one needs to express \( E(\ln L_n) \) as a function
of the parameters involved and of \( E(n) \), the ASN function.

Bhate's conjecture consists of approximating \( E(\ln L_n) \) by replacing \( Z_n \) and \( n \) in the expression for \( \ln L_n \) by \( E[Z_n | n = E(n)] \) and \( E(n) \) respectively, where the expectations \( E[Z_n | n = E(n)] \) are obtained under both \( H_0 \) and \( H_1 \). Jackson and Bradley (1961a) point out that this procedure is seen intuitively to give a central value for the distribution of \( \ln L_n \) and, upon appropriate substitutions in (7.1), to give equations in \( E_0(n) \) and \( E_1(n) \) for solution, where \( E_i(n) \) is the value of \( E(n) \) under the hypothesis \( H_i \) (\( i = 0, 1 \)).

This method of approximating the ASN has been used, for example, by Ray (1956) for sequential analysis of variance, by Hajnal (1961) for a two-sample sequential t-test, and, in particular, by Jackson and Bradley (1961a) for their one-sample sequential T^2-test.

A possible typographical error in the formula given by Jackson and Bradley (1961a) should, however, be noted, apart from which a considerably simpler form for \( E[Z_n | n = E(n)] \) can be obtained. The derivation is given below where it can be seen that the form is applicable to all sequential T^2-tests developed in this thesis.

Suppose that the random variable \( T^2 \) is distributed according to \( T_p^2(\tau^2; \nu) \). Let \( Z = \frac{T^2}{\nu + T^2} \). By Theorem 2.14, we
have that \( Z \) is distributed as \( \frac{p_{F'}}{(v-p+1)+p_{F'}} \) where \( F' \) has the noncentral \( F \)-distribution with \( p \) and \((v-p+1)\) degrees of freedom and noncentrality parameter \( \tau^2 \). Thus, it can be shown (see Graybill, 1961, p. 79) that \( Z \) has the noncentral Beta-distribution with p.d.f. given by

\[
f(z) = \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{v+1}{2} + i\right)}{\Gamma\left(\frac{v-p+1}{2}\right) \Gamma\left(\frac{p}{2} + i\right)} \frac{(\tau^2/2)^i}{i!} z^{(p/2)+i-1} (1-z)^{(v-p+1)/2-i-1}
\]

(7.3)

for \( 0 \leq z \leq 1 \). Following Wishart's (1932) derivation of the mean of the noncentral Beta distribution, we then have

\[
E(Z) = \left[ 1 - \left(\frac{v-p+1}{v+1}\right) F\left(1, \frac{v+3}{2}, -\frac{\tau^2}{2}\right) \right]
\]

(7.4)

where \( F(\cdot, \cdot; \cdot) \) is the confluent hypergeometric function defined in (2.21).

As summarized in Table 1, each sequential test developed in this thesis is based, at the \( n^{th} \) stage, on a \( T^2 \)-statistic for which the number of degrees of freedom \( \nu_n \) is a linear function of \( n \) while the noncentrality parameter \( \tau^2 \) is equal to \( c_n \lambda^2 \), where \( c_n \) is a function of \( n \) and, in fact, a linear function of \( n \) for the tests of Chapters III and V.

The number of degrees of freedom for the \( \hat{T}^2 \)-statistic at the \( n^{th} \) stage is \((\nu_n + m)\) where \( \nu_n \) represents the number of degrees of freedom for the corresponding \( T^2 \)-statistic. The
Table 1. Degrees of freedom and noncentrality parameter for $T^2_n$-statistics

<table>
<thead>
<tr>
<th>Table 2 of Chapter:</th>
<th>Degrees of freedom</th>
<th>Noncentrality parameter $c_n$ ($\lambda^2 = 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>III</td>
<td>$n - 1$</td>
<td>$n$</td>
</tr>
<tr>
<td>IV</td>
<td>$n - q - 1$</td>
<td>$\left(\frac{1}{n} + \bar{x}<em>n \cdot A</em>{n}^{-1} \cdot \bar{x}_n\right)^{-1}$</td>
</tr>
<tr>
<td>V</td>
<td>$nr - 2$, where</td>
<td>$nc$, where</td>
</tr>
<tr>
<td></td>
<td>$r = r_1 + r_2$</td>
<td>$c = r_1 r_2 / r$</td>
</tr>
<tr>
<td>IV</td>
<td>$nr - q - 2$</td>
<td>$\left(\frac{1}{nc} + \frac{d}{\bar{x}<em>n} \cdot A</em>{n}^{-1} \cdot \bar{d}_n\right)^{-1}$, where</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{d}_n = (\bar{x}_n - \bar{x}_2 n)$</td>
</tr>
</tbody>
</table>

Noncentrality parameter $\tau^2$ for each $T^2_n$-statistic is the same as that for the corresponding $T^2$-statistic.

Thus, in general, with

$$z_n = \frac{T^2_n}{\nu_n + T^2_n} \quad \text{and} \quad \tau^2 = c_n \lambda^2,$$
equation (7.4) becomes

\[ E[Z_n|n = E(n)] = \left[ 1 - \frac{v_n}{v_n+p+1} \right] F\left( v_n+3, \frac{v_n}{2}; -\frac{c_n\lambda^2}{2} \right) \]  

(7.5)

Since the probability ratio for each of our sequential tests can be given in the general form

\[ I_n = \exp\left[ -\frac{1}{2} c_n (\lambda_i^2 - \lambda_0^2) \right] \frac{F\left( \frac{v_n+1}{2}, \frac{p_2}{2}; -\frac{c_n\lambda_i^2}{2}, \frac{z_n}{2} \right)}{F\left( \frac{v_n+1}{2}, \frac{p_2}{2}; -\frac{c_n\lambda_0^2}{2}, z_n \right)} \]  

(7.6)

the equation corresponding to (7.1) is given as

\[ \frac{1}{2} c_n (\lambda_i^2 - \lambda_0^2) + \ln F\left( \frac{v_n}{2}, \frac{p_2}{2}; \frac{c_n\lambda_i^2}{2}, E_i(Z_n) \right) \]

\[ - \ln F\left( \frac{v_n}{2}, \frac{p_2}{2}; \frac{c_n\lambda_0^2}{2}, E_i(Z_n) \right) = h_i(\alpha, \beta) \]  

(7.7)

where \( E_i(Z_n) = E[Z_n|n = E(n)] \) evaluated at \( \lambda_i^2 = \lambda_i^2 \) (\( i = 0, 1 \)).

Thus, solution of the two equations given in (7.7) for \( n \) yields \( E_0(n) \) and \( E_1(n) \), the required ASN values under \( H_0 \) and \( H_1 \) respectively.

As noted by Jackson and Bradley (1961a), solution of the two equations in (7.7) can be accomplished iteratively with a high-speed computer. However, it should be noted that for the tests of Chapter IV and VI, solution of the equations (7.7) requires knowledge of (or, at least, a good approximation for)
the quantities $x_n' \frac{A_n^{-1}}{\bar{x}_n} x_n$ and $(\bar{x}_{1n} - \bar{x}_{2n})' A_n^{-1} (\bar{x}_{1n} - \bar{x}_{2n})$, respectively.

As an example, let us consider the sequential $T^2$-test of Chapter III so that $Z_n = \frac{\frac{1}{n}}{n-1+T_n^2}$. Jackson and Bradley (1961a, p. 1072) give the formula

$$ E[Z_n|n = E(n)] = n\lambda^2 \left[ 1 - \left( \frac{n-\frac{p}{2}}{n} \right) F \left( \frac{n}{2}, \frac{n+2}{2}; \frac{n\lambda^2}{2} \right) e^{-n\lambda^2/2} \right], $$

(7.8)

whereas, from formula (7.5), we have

$$ E[Z_n|n = E(n)] = \left[ 1 - \left( \frac{n-\frac{p}{2}}{n} \right) F \left( 1, \frac{n+2}{2}; \frac{-n\lambda^2}{2} \right) \right] $$

(7.9)

Now, with $\lambda^2 = 0$, $Z_n$ has a central Beta distribution with parameters $\frac{p}{2}$ and $\frac{n-p}{2}$ so that equation (7.9) yields the correct mean value of $Z_n$, namely $E[Z_n]$, whereas equation (7.8) yields a mean value of zero, which is obviously incorrect. Ignoring the factor $(n\lambda^2)$ outside the brackets on the right-hand side of (7.8), one can show that (7.8) and (7.9) are equivalent by applying Kummer's identity (see Appendix A):

$$ F(a, b; x) = F(b - a, b; -x) e^x. $$

(7.10)

Clearly, equation (7.9)—as a function of $n$—is a much simpler form than the corrected version of (7.8).
With $\lambda_0^2 = 0$ in this example, we have that $E_0(Z_n) = \frac{D}{n}$ and

$$\ln F\left(\frac{n}{2}, \frac{D}{2}; \frac{n\lambda_0^2}{2} E_1(Z_n)\right) = 0$$

(for $i = 0, 1$) so that $E_0(n)$ and $E_1(n)$ can be found by solving equations (7.11) and (7.12), respectively:

$$-\frac{1}{2} n\lambda_1^2 + \ln F\left(\frac{n}{2}, \frac{D}{2}; \frac{\lambda_1^2}{2} p\right) = h_0(\alpha, \beta); \tag{7.11}$$

$$-\frac{1}{2} n\lambda_1^2 + \ln F\left(\frac{n}{2}, \frac{D}{2}; \frac{n\lambda_1^2}{2} E_1(Z_n)\right) = h_1(\alpha, \beta). \tag{7.12}$$

**B. Discussion**

We now discuss some problems that arise in using the sequential test procedures developed in this thesis.

1. **Tables**

   Direct applications of our sequential procedures involve comparison of the likelihood ratio $L_n$ (or $\hat{L}_n$) at each stage with prespecified boundaries $B$ and $A$, --typically, $\left[\frac{\beta}{1-\alpha}\right]$ and $\left[\frac{1-\beta}{\alpha}\right]$, respectively. This requires the evaluation of one or two confluent hypergeometric functions (depending on whether or not $\lambda_0^2$ equals zero) after each stage of sampling. As noted previously, tables of the confluent hypergeometric function are available, but it seems better to prepare tables of the boundary values $T_n^2$ and $\bar{T}_n^2$ (as defined in Chapters III through VI) so that only the test statistic need be computed in applications.
For the one-sample sequential $T^2$-test of Jackson and Bradley (1961a), tables have been given by Freund and Jackson (1960) for $\alpha = \beta = .05$ and $\lambda^2_0 = 0$; these tables show $T^2_n$ and $\overline{T}^2_n$ for $p$ from 2 to 9 and for several values of $\lambda^{2}_1$ between 0.5 and 10.0. In order to give tables that can be used for all sequential $T^2$-tests that we have developed, it seems easiest to compute boundaries $Z_n$ and $\overline{Z}_n$ for the statistic

$$Z_n = \frac{T^2_n}{(\nu_n + T^2_n)}$$

where $T^2_n$ is distributed according to $\chi^2_p(c_n, \lambda^2; \nu_n)$ and where the values of $\nu_n$ and $c_n$ are given in Table 1 for the various tests that we have developed. Obviously, such a project would require an appreciable amount of computer-time.

2. Determination of $H_0$ and $H_1$

As noted by Jackson and Bradley (1961a), specification of the noncentrality parameter $\lambda^2$ leads to difficult administrative decisions. In many applications, $\lambda^2_0$ is taken to be zero, but the determination of $\lambda^2_1$, however, is difficult because it depends on a $p$-dimensional ellipsoid related to the problem specifications. Jackson and Bradley (1961a, p. 1075) point out that no general rule for specifying $\lambda^2_1$ can be given so that each problem has to be handled individually.

Jackson and Bradley (1961b) do give a method for determining $\lambda^2_1$ in connection with the sampling inspection of ballistic missiles; the method consists essentially of
inscribing an ellipsoid inside the rectangular region bounded by the tolerance specifications. An alternative method for such a problem would be to circumscribe an ellipsoid around this rectangular region; as noted by Jackson and Bradley (1961b, p. 525), this second method could be used when the lot is considered passable even when all the characteristics are borderline.

3. The OC and ASN functions

As noted by Jackson and Bradley (1961a, p. 1075), no explicit or even approximate formulae yet exist for the OC and ASN functions when the hypotheses are composite. Although there exist some heuristic approximations for the ASN function (such as that described in Section A of this chapter), it appears that the statistician must, at present, rely on empirical investigations for a description of these properties.

4. Truncated and restricted schemes

Although the sequential test procedures developed in this thesis terminate with probability one, there still exists the possibility that in a particular case the sample number may become extremely large. As protection against such behavior, Wald (1947) discussed the truncation of the SPRT (for simple hypotheses) at some sample size N and gave a method of choosing N large enough to have a negligible effect on the OC and ASN functions. As noted by Jackson and Bradley (1961a), little
work has been done regarding truncation of sequential tests of composite hypotheses. A study of truncated versions of our sequential test procedures might prove valuable, but because of theoretical difficulties the investigation would undoubtedly have to be an empirical one.

Armitage (1957) and Armitage and Schneiderman (1962) presented some exact and approximate restricted (closed) sequential procedures for particular applications. One might consider applying these authors' ideas to the problems presented in this thesis. The development of such restricted multivariate sequential procedures certainly merits further study.

C. More Topics for Further Research

Much further research is needed in the field of multivariate sequential tests. Two topics from fixed-sample-size analysis that merit study in a sequential framework are the following: k-sample \((k > 2)\) analogues of the multivariate problems considered in this thesis; and the multivariate Behrens-Fisher problem (that is, the two-sample problem of Chapter V when the covariance matrices are unequal).

A problem related to the latter consideration is that of testing sequentially the equality of two covariance matrices. In this regard, Jackson and Bradley (1961a, pp. 1073-1074) discuss generalized \(\chi^2\)- and \(T^2\)-statistics for tests about covariances matrices. They developed a sequential generalized \(\chi^2\)-test for testing sequentially hypotheses about a single
covariance matrix. They did not, however, develop a sequential
generalized $T^2$-test for covariance matrices since, in their
opinion, such situations would rarely occur in sequential
experimentation. On the contrary, in the two-sample problem
of Chapter V, one might wish to test sequentially the equality
of the two covariances matrices subsequent to (or simultaneously
with) the sequential test about the difference of the means.
The development of such a sequential procedure, as well as a
study of its properties, seems worthy of attention.

In Chapter IV, we developed sequential tests about the
mean vector adjusted for covariates which were assumed to be
controlled. As an alternative specification, it might be
assumed that the $(p+q) \times 1$ vector

$$
\begin{pmatrix}
\chi \\
x
\end{pmatrix} \overset{\sim}{\sim} N_{p+q} \begin{pmatrix}
\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}, \\
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix}
$$

Suppose that one is interested in testing sequentially
hypotheses about the parameter $\begin{pmatrix}
\mu_1 \\
\mu_1
\end{pmatrix}$. If one ignores the
information contained in the covariates, the test procedures
of Chapter III are applicable. However, if there is high cor-
relation between $\chi$ and $x$, a reasonable conjecture is that
utilization of the concomitant information should thus lead to
a reduction in the ASN function.
Development of the appropriate sequential test procedure, as well as substantiation of this conjecture, is certainly worthy of further study. Similar considerations pertain to the two-sample problem. For a discussion of the appropriate fixed-sample-size analyses with respect to this problem, the reader is referred to (Subrahamiam, 1970) and (Rao, 1966).

In Chapter VI, it was assumed that the matrix of regression coefficients $\beta$ was the same for the two populations. Development of sequential test procedures analogous to those of Chapter VI, in which this restriction is relaxed to allow for the case of unequal matrices of regression coefficients $\beta_1$ and $\beta_2$, also merits further study.

As another suggestion for further research, we offer the following problem: any two-sample hypotheses-testing situation in which one sample is fixed and the other sample is sequential. As an example let $y_1$ be distributed $N_1(\mu_1, \sigma^2)$ and $y_2$ be distributed $N_1(\mu_2, \sigma^2)$. Suppose that there are available $m$ independent observations $y_{11}, \ldots, y_{1m}$ and suppose that we wish to discriminate between the two hypotheses:

$$H_0 : \frac{\mu_2 - \mu_1}{\sigma} = 0$$

$$H_1 : \frac{\mu_2 - \mu_1}{\sigma} = \delta_1.$$
According to available statistical test procedures, one has two choices: (i) sample \( n \) observations from the second normal population and apply the usual two-sample t-test from fixed-sample-size analysis; or (ii) ignore the available \( m \) observations from the first population and apply Hajnal's (1961) two-sample sequential t-test, sequentially sampling from both populations at each stage.

In many situations, neither choice may be entirely satisfactory. For example, if it is not possible to sample additional observations from the first population, one cannot choose a fully sequential sampling scheme. Further, it may happen that, due to cost considerations, choice (i) is not desirable. Development of a "semi-sequential" test procedure to handle such problems seems not to have been previously considered, nor does it appear that the development of such procedures would be a trivial task.
VIII. REFERENCES


Berk, R. H. 1972b. Private correspondence with the author.


IX. ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Professor C. Philip Cox for his many valuable suggestions throughout the research and preparation of this dissertation.

Thanks are also given to the National Science Foundation (Grant GZ-1596) for its financial support during the course of the author's graduate program at Iowa State University.

The author wishes especially to thank his wife, Diane, for her understanding and encouragement throughout his graduate work.
The Kummer confluent hypergeometric function is the infinite series

\[ F(a, c; x) = \sum_{i=0}^{\infty} \frac{(a)_i}{(c)_i} \cdot \frac{x^i}{i!} \quad (10.1) \]

where \( x \) is a real or complex variable, the parameters \( a \) and \( c \) are real or complex values (except that \( c \neq 0, -1, -2, \ldots \)), and

\[
(a)_i = a(a + 1)(a + 2)\ldots(a + i - 1) \\
(c)_i = c(c + 1)(c + 2)\ldots(c + i - 1) 
\]

As noted by Slater (1960, p. 2), this series is absolutely convergent for all finite values of \( a, c, \) and \( x \), real or complex, excluding \( c = 0, -1, -2, \ldots \). Further, it can be shown (Lebedev, 1965, p. 262) that the confluent hypergeometric function \( F(a,c;x) \) is a particular solution of the linear differential equation

\[
x \frac{d^2w}{dx^2} + (c - x) \frac{dw}{dx} - aw = 0 \quad (10.2)
\]

Commonly, equation (10.1) is written in the form given in Chapter II; that is,
\[ F(a, c; x) = \sum_{i=0}^{\infty} \frac{\Gamma(a+i)}{\Gamma(a)} \frac{\Gamma(c+i)}{\Gamma(c+i)} \frac{x^i}{i!} \]  

(10.3)

where \( \Gamma(\cdot) \) is the gamma function. It is particularly important to note here that \( c \) cannot equal zero or a negative integer.

A relation that is useful in the calculation of the confluent hypergeometric function is Kummer's identity:

\[ F(a, c; x) = e^x F(c - a, c; x). \]  

(10.4)

Thus, by Kummer's identity, \( F(a, c; x) \) can often be written as the product of the exponential function \( e^x \) and a series \( F(c-a, c; x) \), which is finite provided \( (c - a) \) is a nonpositive integer, (see, for example, Lebedev, 1965, p. 273). In the applications considered in this thesis, \( (c - a) \) will always be negative and will equal an integer or an integer plus 1/2.

For evaluating the confluent hypergeometric function either by hand or electronic computer calculations, the following standard recurrence relations are most helpful.

\[ xF(a+1,c+1;x) = cF(a+1,c;x) - cF(a,c;x) \]
\[ aF(a+1,c+1;x) = (a-c)F(a,c+1;x) + cF(a,c;x) \]  

(10.5)
\[ aF(a+1,c;x) = (x+2a-c)F(a,c;x) + (c-a)F(a-1,c;x) \]
\[ (c-a)xF(a,c+1;x) = c(x+c-1)F(a,c;x) + c(1-c)F(a,c-1;x). \]

For further details about the confluent hypergeometric function, the reader is referred to Rushton (1954), Slater (1960), Lebedev (1965), and Abramowitz and Stegun (1964).
Let us consider the computation of the one-sample $T_n^2$ statistic defined in Section B of Chapter III as

$$T_n^2 = n\overline{\gamma}_n \cdot D_n^{-1} \overline{\gamma}_n$$  \hspace{1cm} (10.6)$$

where $\overline{\gamma}_n$ and $D_n$ denote the $n$th stage sample mean vector and sample covariance matrix, respectively. To compute $T_n^2$ directly as in (10.6), it is necessary to first compute $D_n^{-1}$. Instead, Anderson (1958, p. 107) shows that $T_n^2$ can be computed easily by the following procedure:

(i) Solve the linear system of equations

$$D_n b_n = \overline{\gamma}_n$$  \hspace{1cm} (10.7)$$

for the $(p \times 1)$ vector $b_n$. Anderson (1958), for example, shows that this can be done by the forward Doolittle method.

(ii) Then, since (10.7) implies that

$$b_n = D_n^{-1} \overline{\gamma}_n,$$  \hspace{1cm} (10.8)$$

it follows that $T_n^2$ can be computed according to

$$T_n^2 = n\overline{\gamma}_n \cdot b_n.$$  \hspace{1cm} (10.9)$$

Thus, this procedure avoids the explicit computation of $D_n^{-1}$.

Further, it is clear that this computational procedure can be applied to each $T^2$-statistic considered in this thesis.
since $T_n^2$, in each case, is of the form

$$T_n^2 = c_n w_n' V_n^{-1} w_n$$  \(10.10\)

where $w_n$ is a $(p \times 1)$ vector, $V_n$ is a $(p \times p)$ nonsingular matrix, and $c_n$ is a scalar quantity (given, for each case, in Table 1 of Chapter VII).

**C. Derivation of Recurrence Relationships for the Sufficient Sequences $\{S_n\}$**

We shall now derive the several recurrence relationships used in the preceding chapters to establish the transitivity of the sequence of sufficient statistics. Further, it might be noted here that, for any situation in which an additional observation is available, these recurrence relationships can be implemented in the computations of the updated value of the sufficient statistics using only the previous value of the sufficient statistics and the additional observation. With respect to calculations on a high-speed computer, since the observations from the previous stages need not be retained for later use, utilization of these recurrence relationships should lead to savings, not only in computing-time, but also in computer storage.

1. **Recurrence relationships of Chapter III**

   From the definition of the sample mean vector $\bar{y}_{n+1}$ at the $(n+1)^{st}$ stage, we have that
Now, from the definition of \( \bar{y}_n \) given in (3.10), it follows that

\[
(n+1)\bar{y}_{n+1} = \sum_{i=1}^{n+1} y_i
\]

\[
= \sum_{i=1}^{n} y_i + y_{n+1}.
\]

so that division of both sides of equation (10.11) by \((n+1)\) yields recurrence relationship (3.15).

The following expressions for the sample covariance matrix \( D_{n+1} \) at the \((n+1)\)st stage are a consequence first of the definition of \( D_{n+1} \) and secondly of the machine-formula representation of \( D_{n+1} \):

\[
nD_{n+1} = \sum_{i=1}^{n+1} (y_i - \bar{y}_{n+1})(y_i - \bar{y}_{n+1})'
\]

\[
= \sum_{i=1}^{n+1} y_i y'_i - (n+1)\bar{y}_{n+1}\bar{y}'_{n+1}.
\]

By adding and subtracting \( ny_i \bar{y}_n' \) in the right-hand side of the above expression and by substituting for \( \bar{y}_{n+1} \) from (3.15), we then have that
Finally, by collection of the last three terms in the above expression, as well as by the machine-formula representation of $D_n$, it follows that

$$nD_{n+1} = (n-1)D_n + \frac{n}{n+1} (y_{n+1} - \bar{y}_n) (y_{n+1} - \bar{y}_n)' + mV,$$

(10.12)

which is clearly equivalent to recurrence relationship (3.16) for $D_{n+1}$.

For the case of an independent estimate $V$ of $\Sigma$ with $m$ degrees of freedom, the recurrence relationship (3.25) for the pooled estimate $W_n$ of $\Sigma$, as defined in (3.23), can be easily derived via expression (10.12) as follows. By the definition of $W_{n+1}$ and then by substitution for $nD_{n+1}$ from (10.12), we have that

$$(m+n)W_{n+1} = nD_{n+1} + mV$$

$$= (n-1)D_n + \frac{n}{n+1} (y_{n+1} - \bar{y}_n) (y_{n+1} - \bar{y}_n)' + mV.$$ 

Thus, it follows from the definition of $W_n$ that

$$(m+n)W_{n+1} = (m+n-1)W_n + \frac{n}{n+1} (y_{n+1} - \bar{y}_n) (y_{n+1} - \bar{y}_n)' + mV.$$ 

(10.13)
Division of both sides of equation (10.13) by \((m+n)\) yields therefore the recurrence relationship (3.25).

2. Recurrence relationships of Chapter IV

From the definition of \(B_n\) given in (4.8), it follows that

\[ B_{n+1}A_{n+1} = Y_{n+1}X_{n+1} \]

\[ = \sum_{i=1}^{n+1} y_i \left( x_i - \bar{x}_{n+1} \right) \]

\[ = \sum_{i=1}^{n} y_i \left( x_i - \bar{x}_n \right) + Y_{n+1} \left( x_{n+1} - \bar{x}_{n+1} \right) \]

By adding and subtracting \(\bar{x}_n\) within \((x_i - \bar{x}_{n+1})\) in the first term of the above expression, we can write

\[ = \sum_{i=1}^{n} y_i \left( x_i - \bar{x}_n \right) + n\bar{y}_n \left( \bar{x}_n - \bar{x}_{n+1} \right) \]

Since, by (4.8), \(Y_n = B_n A_n\) and since, as is easily verified,

\[ \left( \bar{x}_n - \bar{x}_{n+1} \right) = \left( \frac{-1}{n+1} \right) \left( x_{n+1} - \bar{x}_n \right) \]

and

\[ \left( x_{n+1} - \bar{x}_{n+1} \right) = \left( \frac{n}{n+1} \right) \left( x_{n+1} - \bar{x}_n \right), \]

we have that

\[ B_{n+1}A_{n+1} = B_n A_n + \left( \frac{n}{n+1} \right) \left( Y_{n+1} - \bar{y}_n \right) \left( x_{n+1} - \bar{x}_n \right) \]

(10.14)
By the definition of \( \gamma_n \), given in (4.11), it then follows that

\[
\gamma_n = a_n + B_n \bar{x}_n
\]  

(10.15)

so that, by substitution in (10.14) for \( \gamma_n \), we finally have

\[
B_{n+1} \gamma_{n+1} = B_n \gamma_n + \frac{n}{n+1} (y_n + a_n - B_n \bar{x}_n)(x_{n+1} - \bar{x}_n)'.
\]  

(10.16)

Hence, the recurrence relationship (4.17) is obtained by simply multiplying (on the right) both sides of (10.16) by \( A_n^{-1} \).

From the definition of \( \gamma_n \), given in (4.11), and subsequently from recurrence relationship (3.15) for the sample mean vector, it follows that

\[
a_{n+1} = \gamma_{n+1} - B_{n+1} \bar{x}_{n+1}
\]

Thus, by substitution for \( \gamma_n \) from (10.15) and then by collection of terms, we have the recurrence relationship (4.18) as follows.

\[
a_{n+1} = \frac{1}{n+1} [n \gamma_n + B_n \bar{x}_n] + y_{n+1} - B_{n+1} \bar{x}_{n+1}
\]

Since \( y_n y_n' = (n-1)D_n \), where \( D_n \) is defined in (3.11), and since \( y_n x_n' = B_n A_n \), an expression for \( E_n \) equivalent to (4.12) is given by
Thus, it follows that

\[(n-q)E_{n+1} = nD_{n+1} - B_{n+1}A_{n+1}B_{n+1}'\]

so that, by (10.12) and by adding and subtracting \(B_{n}A_{n}B_{n}'\) in the right-hand side of the above expression, we have

\[(n-q)E_{n+1} = (n-1)D_{n} - B_{n}A_{n}B_{n}'\]

\[+ B_{n}A_{n}B_{n}' - B_{n+1}A_{n+1}B_{n+1}'\]

\[+ \frac{n}{n+1} \left( \bar{y}_{n+1} - \bar{y}_{n} \right) \left( \bar{y}_{n+1} - \bar{y}_{n} \right)'.\]

Finally, by (10.17) and by substitution for \(\bar{y}_{n}\) from (10.15), it follows that

\[(n-q)E_{n+1} = (n-q-1)E_{n} + B_{n}A_{n}B_{n}' - B_{n+1}A_{n+1}B_{n+1}'\]

\[+ \frac{n}{n+1} \left( \bar{y}_{n+1} - a_{n} - B_{n}x_{n} \right) \left( \bar{y}_{n+1} - a_{n} - B_{n}x_{n} \right)'. \quad (10.18)\]

Therefore, the recurrence relationship (4.19) is given by dividing both sides of (10.18) by \((n-q)\).

For the case of the pooled estimate \(\Sigma_{n}\) of \(\Sigma_{n}\), as defined in (4.24), it follows from (10.18) that \((n-q)E_{n+1}\) can be written as the sum of \((n-q-1)E_{n}\) and a function of \(a_{n}, B_{n}, B_{n+1}\) and \(\bar{y}_{n+1}, r(a_{n}, B_{n}, B_{n+1}, \bar{y}_{n+1})\) say. Thus, by expression (4.28) for \(\Sigma_{n+1}\), we have
3. Recurrence relationships of Chapter V

By the definition of the sample mean vector \( \bar{\gamma}_{(i)}^{(n+1)} \) at the \((n+1)\)th stage \((i = 1, 2)\), we have that

\[
s_{i}^{(n+1)} = \frac{r_{i}^{(n+1)}}{S_{j}^{(n+1)}} \sum_{j=1}^{r_{i}^{(n+1)}} \bar{\gamma}_{ij}^{(n+1)}
\]

Since, by the definition of \( \bar{\gamma}_{(i)}^{(n+1)} \) given in (5.16),

\[
r_{i}^{n} \bar{\gamma}_{(i)}^{(n+1)} = \sum_{j=1}^{r_{i}^{n}} \bar{\gamma}_{ij}^{(n+1)}
\]

it follows that division of both sides of equation (10.20) by \( r_{i}^{(n+1)} \) yields recurrence relationship (5.21).

By the definition of the sample covariance matrix \( D_{i}^{(n+1)} \) at the \((n+1)\)th stage and then by the machine-formula representation of \( D_{i}^{(n+1)} \), we have that

\[
[r_{i}^{(n+1)} - 1]D_{i}^{(n+1)} = \sum_{j=1}^{r_{i}^{(n+1)}} (\bar{\gamma}_{ij}^{(n+1)} - \bar{\gamma}_{i}^{(n+1)}) (\bar{\gamma}_{ij}^{(n+1)} - \bar{\gamma}_{i}^{(n+1)})'
\]

\[
= \sum_{j=1}^{r_{i}^{(n+1)}} \gamma_{ij}^{(n+1)}\gamma_{ij}^{(n+1)'} - r_{i}^{(n+1)} \bar{\gamma}_{i}^{(n+1)} \bar{\gamma}_{i}^{(n+1)'}
\]
By adding and subtracting \( r_i n^2 \), \( \bar{y}_{in} \) is the right-hand side of the above expression, we can write

\[
[r_i(n+1) - 1]D_i(n+1) = (r_i n - 1)D_i n + \sum_{j=r_i n + 1}^{r_i(n+1)} \bar{y}_{ij} \bar{y}_{ij}'
\]

\[
+ r_i [\bar{y}_{in} \bar{y}_{in}' - (n+1)\bar{y}_{in} \bar{y}_{in}(n+1)']
\]

Clearly, the above expression is equivalent to the recurrence relationship (5.22).

4. Recurrence relationships of Chapter VI

From the definition of \( B_n \) given in (6.21), it follows that

\[
B_{n+1} = Y_1(n+1)X_1(n+1)' + Y_2(n+1)X_2(n+1)',
\]

(10.21)

where \( Y_{in} \) and \( X_{in} \) are defined by (6.18) and (6.16), respectively, for \( i = 1, 2 \). Now, by arguments analogous to those of Subsection 2 of this section, it can be easily shown that

\[
Y_i(n+1)X_i(n+1)' = Y_i X_i' + nr_i (a_n + \bar{X}_n) (\bar{X}_i - \bar{X}_i(n+1))'
\]

Upon substitution in (10.21) and matrix multiplication (on the right) of both sides of equation (10.21), the following recurrence relationship for \( B_n \) obtains, that is,
\[ B_{n+1} = B_n (A_n^{-1} \cdot A_{n+1}^{-1}) + \left[ 2 \sum_{i=1}^{n} n r_i (a_i + B_{n i n} \cdot \bar{x}_i (n+1)) \right] A_{n+1}^{-1} \]

\[ + \left[ 2 \sum_{i=1}^{n} r_i (n+1) \sum_{j=r_i n+1}^{n+1} Y_{ij} (x_{ij} - \bar{x}_i (n+1)) \right] A_{n+1}^{-1}. \]  

(10.22)

From the definition of \( a_{in} \), given in (6.22), and then from recurrence relationship (5.21) for \( \bar{Y}_{in} \), it follows that

\[ a_{i(n+1)} = \bar{Y}_i (n+1) - B_{n+1} \bar{x}_i (n+1) \]

\[ = \left( \frac{n}{n+1} \right) \bar{Y}_{in} + \left( \frac{1}{n+1} \right) \sum_{j=r_i n+1}^{n+1} Y_{ij} - B_{n+1} \bar{x}_i (n+1). \]

Finally, by adding and subtracting \( \left( \frac{n}{n+1} \right) B_{n+1} \bar{x}_{in} \) in the right-hand side of the above expression, we have the following recurrence relationship for \( a_{in} \) (i = 1, 2):

\[ a_{i(n+1)} = \left( \frac{n}{n+1} \right) \bar{Y}_{in} + \left( \frac{1}{n+1} \right) \sum_{j=r_i n+1}^{n+1} Y_{ij} \]

\[ + \left( \frac{n}{n+1} \right) B_{n+1} \bar{x}_{in} - B_{n+1} \bar{x}_i (n+1). \]  

(10.23)

By the definition of \( E_{n} \), given in (6.23), it follows that

\[ [r(n+1) - q - 2] E_{n+1} = [r(n+1) - 2] D_{n+1} - B_{n+1} A_{n+1} B_{n+1}'. \]

Thus, by arguments analogous to those of Subsection 2 of this section, the following recurrence relationship for \( E_n \) can be easily derived:
\[ E_{n+1} = \frac{r^n - q - 2}{r^{(n+1)} - q - 2} E_n \]

\[ + \frac{1}{r^{(n+1)} - q - 2} \left[ B_n A_n B_n' - B_{n+1} A_{n+1} B_{n+1}' \right] \]

\[ + \frac{1}{r^{(n+1)} - q - 2} \left[ \sum_{i=1}^{2} \left( \frac{r_i^{(n+1)}}{\Sigma} \varphi_i j \varphi_i j' \right) \right] \]

\[ + r_i \left[ n \varphi_i \varphi_i' - (n+1) \varphi_i (n+1) \varphi_i (n+1)' \right] \]

where, by the definition of \( \bar{\varphi}_{in} \) in (6.22),

\[ \bar{\varphi}_{in} = \varphi_{in} + B_n \bar{\varphi}_{in} \]