NASA CONTRACTOR REPORT

NASA CR-437

SEQUENTIAL DECISION RULES FOR A MULTIPLE CHOICE PROBLEM

by Donald Roy Barr

Prepared under Grant No. NSG-569 *by*
COLORADO STATE UNIVERSITY
Fort Collins, Colo.
for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • APRIL 1966

NASA CR-437

SEQUENTIAL DECISION RULES FOR A MULTIPLE CHOICE PROBLEM

By Donald Roy Barr

Prepared under Grant No. NSG-569 by
COLORADO STATE UNIVERSITY
Fort Collins, Colo.

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – Price $1.00

ABSTRACT

The problem considered is that of finding a rule for deciding which of k known nonequivalent density functions $f_1, f_2, \ldots, f_k$ is the density of a random variable Y. It is well known that a Bayes solution to this "k-decision" problem for the case k=2 is given by the sequential probability ratio test. The "generalized probability ratio"

$$r(y)=(\sum f_i(y))^{-1}(f_1(y), f_2(y), \ldots, f_k(y))$$

is used in this paper to define the "generalized sequential probability ratio test" (GSPRT) for the case $k \geq 2$. The GSPRT is viewed as a random walk on a space $\mathcal{X}$ of k dimensional vectors $(x^1, x^2, \ldots, x^k)$ such that $\sum x^i=1$ and $x^i \geq 0$ for all i. The test terminates when the walk enters an absorbing barrier in $\mathcal{X}$. Some properties of this absorbing barrier are discussed for a class of GSPRT's which is essentially complete in the class of Bayes rules for the k-decision problem.

Integral equations are obtained for the operating characteristics of the GSPRT. Conditions are given under which the test almost surely terminates. Monotonicity

properties of the operating characteristics with respect to certain changes in the absorbing barrier are obtained. The distribution induced on the random variable r(Y) by the f_i's is discussed, and an identity is given which in some sense characterizes the distribution of probability ratios.

PREFACE

In order to minimize the burden of reading the large amount of specialized terminology and notation required in this paper, a table of symbols and terms is included as an appendix. The table can be used to find the page upon which each symbol and term is defined. Some of the notation is standard and is due mainly to Wald [8,9]. Considerable use has been made of notation introduced by Seo [6] and Skibinsky [7].

In order to simplify notation, the following conventions will be followed:

(i) All summation and union runs from 1 to k unless otherwise specifically stated. The index of summation or union may not be listed when there is no possibility of confusion.

(ii) Unless specifically denoted otherwise, all integration will be over the entire space under consideration.

(iii) The letters "a.e." may be omitted in statements when it is clear from the context that the statement holds only with probability unity.

(iv) Displayed equations are numbered only when they are referred to elsewhere in the paper.

The theorems are numbered consecutively throughout this paper in the order in which they are stated. Numbers in brackets following a reference refer to corresponding complete references in the bibliography. In some cases, the pagination of a reference is included by inserting "p. ---" in the brackets following the number of the reference.

Acknowledgment is due to the National Science Foundation and the National Aeronautics and Space Administration for financial assistance on work related to this investigation.

TABLE OF CONTENTS

Chapter I

INTRODUCTION

The problem we consider is that of making one of k decisions, $d_1, \ldots, d_k$ on the basis of observations on the components of a random vector $Y=(Y_1, Y_2, Y_3, \ldots)$ whose distribution F is known to belong to a set $\mathcal{F}$. Suppose $y=(y_1, y_2, \ldots)$ is a point of the sample space $\mathcal{Y}$, and D is the set of decisions d_i which can be made in the problem. We seek a decision function $\delta:\mathcal{Y}\rightarrow D$, so that the resulting procedure has certain "optimal" properties.

If the decision rule δ is adopted and $y\varepsilon\mathcal{Y}$ is observed, the smallest positive integer $n=n(y)$ with the property that $\delta(y)=\delta(y')$ for any $y'\varepsilon\mathcal{Y}$ for which $y_1=y_1'$, $y_2=y_2'$ $\ldots$, $y_n=y_n'$ is called the sample size of the rule δ, given the observation y. (n(y) may be identically zero under certain circumstances.) If n(y) is not necessarily constant, we say that δ is a sequential decision rule. Unless specifically stated otherwise, we shall confine ourselves to sequential decision rules based on a sequence $Y_1, Y_2, \ldots$ of independent, identically distributed random variables, whose distribution is specified under the various decisions. Thus the decision d_j indicates acceptance of the hypothesis that $F_j\varepsilon\mathcal{F}$ is the distribution of the components of Y.

A sequential decision rule δ for the problem considered here (hereafter called the k-decision problem) can be identified with it's sample size function $n: \mathcal{Y} \to \{o,1,2, \ldots\}$ and a terminal decision function $\Phi: \mathcal{Y} \to \mathcal{X}$, where $\mathcal{X}$ is a space of vectors $X=(x^1, x^2, \ldots, x^k)$ satisfying the conditions

$$\text{(i)} \quad x^i \geq o \text{ for all } i,$$

and

$$\text{(ii)} \quad \sum x^i = 1.$$

It is assumed that the components of Φ are measurable with respect to the smallest σ-field $\mathcal{Z}$ over $\mathcal{Y}$ containing all cylinder sets in $\mathcal{Y}$ with finite dimensional bases. We also assume that $\Phi(y)$ is dependent only on the first $n(y)$ components of y, and to emphasize this we write $\Phi_n(y)$ for $\Phi(y)$ in what follows.

The test (n, Φ_n) consists of taking one observation on each of the first $n(Y)$ random variables $Y_1, Y_2, \ldots, Y_n$, finding the corresponding value of Φ_n, and making the decisions d_i with respective probabilities Φ_n^i. If the probability that $n(Y)$ is greater than the integer m is zero or unity for each specific m, and the range of $\Phi_n(y)$ is restricted to the k vectors with one component equal to unity, the test (n, Φ_n) is called non-randomized. In what follows we shall consider only non-randomized decision functions, so that application of the decision rule (n, Φ_n) consists of observing each of the first $n(Y)$ components of Y, and then choosing the i-th distribution function F_i to be the true one if $\Phi^i_{n(y)}(y)=1$.

Let E_μ denote the expectation operator relative to a probability measure μ on $(\mathcal{Y}, \mathcal{B})$. When there is no possibility of confusion, we write E_i for the expectation operator relative to the measure induced on $(\mathcal{Y}, \mathcal{B})$ by F_i. Similarly, $P_i(A)$ will denote the probability of the event A given that F_i is the true distribution of Y_j. The operating characteristics (O.C.'s) of the test (n, Φ_n) are defined as follows:

$N_i(\delta)=E_i n(Y)$ is the expected sample size required by the decision rule δ.

$Q_{ij}(\delta)=E_i \Phi_n^j$ is the probability that the rule δ accepts the j-th distribution to be the true one, given that the i-th distribution is the true one. The Q_{ij}'s will be referred to as "error probabilities" of the rule δ.

Let w_{ij} denote the loss incurred by choosing the j-th distribution to be the true one when the i-th distribution is correct. We call $W=(w_{ij})$ the loss matrix, and assume that $w_{ij} \geq 0$ for all i and $j(i \neq j)$, and $w_{ii}=0$, $i=1,2, \ldots, k$. A criterion for judging the relative "goodness" of any rule δ is the risk of δ. If the cost of making observations on the components of Y is linearly related to the number of observations taken (as we shall assume it to be) the risk of the rule δ is defined to be

(1) $$R(y,W,\delta) = \sum_i y^i \, [cN_i(\delta) + \sum_j w_{ij} Q_{ij}(\delta)],$$

where c is the cost of a single observation and y is the vector of a priori probabilities that the corresponding distributions are the true distributions of the components of Y.

Remark. $R(y,W,\delta)$ represents the expected loss (to the experimenter) when the rule δ is used. Without loss of generality c could be taken to be unity, since this involves at most a scale change in the elements of W.

Definition. Let $\mathcal{T}$ be a class of decision rules for the k-decision problem. A Bayes rule in $\mathcal{T}$ relative to the vector of a priori probabilities y and loss matrix W (a Bayes y, W rule in $\mathcal{T}$) is a rule $\delta^* \varepsilon \mathcal{T}$ such that

$$R(y,W,\delta^*) \leq R(y,W,\delta) \text{ for all } \delta \varepsilon \mathcal{T}.$$

A. Wald [9, p.110] has given a characterization of Bayes rules which we include here for completeness and later reference. The notation required for a statement of Wald's theorem is the subject of the next paragraph.

We define three classes of sequential rules as follows:

$$\mathcal{J} = \left\{ \delta : P_i \left\{ n(\delta) < \infty \right\} = 1,\ i=1,2,\ \ldots,\ k \right\},$$

$$\mathcal{J}^* = \left\{ \delta \varepsilon \mathcal{J} : n(\delta) \geq 1 \right\}, \text{ and}$$

$$\mathcal{J}_m = \left\{ \delta : n \leq m \right\};\ m=0,1,2,\ \ldots .$$

Let

(2) $$\rho(y,W,\mathcal{T}) = \inf_{\delta \varepsilon \mathcal{T}} R(y,W,\delta).$$

For convenience of notation in what follows, we shall write

$$\rho(y,W) = \rho(y,W,\mathcal{J}),$$

$$\rho^*(y,W) = \rho(y,W,\mathcal{J}^*), \text{ and}$$

$$\rho_m(y,W) = \rho(y,W,\mathcal{J}_m).$$

Note that

(3) $$\rho_o(y,W) = \min_{j \varepsilon \{1,2,\ \ldots,\ k\}} \left\{ \sum_i y^i w_{ij} \right\},$$

and that

(4) $$\rho(y,W)=\min\left\{\rho_o(y,W),\ \rho^*(y,W)\right\}.$$

Wald's characterization of Bayes y,W rules in the class $\mathcal{S}$ of all decision rules for the k-decision problem (subject to the assumptions made above) is given as

Theorem 1. A necessary and sufficient condition for a decision rule $(n^*, \Phi^*_{n^*})$ to be a Bayes y_o, W rule in $\mathcal{S}$ is that the following four conditions be fulfilled for almost all (under y_o) points $y \varepsilon \mathcal{Y}$:

(i) For any integer $m < n^*(y)$ the a posteriori measure $\mu(y_o, y_1, y_2, \ldots, y_m)$ satisfies the inequality $\rho_o(\mu,W) \geq \rho^*(\mu,W)$.

(ii) If $\rho_o(\mu(y_o,y_1, \ldots, y_m),W) >$

$$\rho^*(\mu(y_o, y_1, \ldots, y_m),W),$$

then

$$n^*(y) > m.$$

(iii) $\rho_o(\mu(y_o, y_1, \ldots, y_{n^*}),W) \leq$

$$\rho^*(\mu(y_o, y_1, \ldots, y_{n^*}),W).$$

(iv) $R(\mu(y_o, y_1, \ldots, y_{n^*}), W, (n^*, \Phi^*_{n^*}))=$

$$\rho(\mu(y_o, y_1, \ldots, y_{n^*}),W).$$

Remark. The approach taken here for the class $\mathcal{S}$ can be used for any subclass $\mathcal{T}$ of $\mathcal{S}$. One could, for example, speak of a Bayes y,W rule in $\mathcal{T}$, the definitions of the functions ρ, ρ^*, and ρ_m being given with $\mathcal{S}$ replaced by $\mathcal{T}$. Theorem 1 would then characterize Bayes y,W rules in $\mathcal{T}$.

When k=2, any Bayes y_o, W rule in $\mathcal{S}$ is equivalent to a sequential probability ratio test (SPRT). In one sense, the SPRT provides the solution to the 2-decision problem, in that it has an even stronger property of optimality -- the so-called "optimal property". After a brief description of the SPRT, we shall discuss some of its' properties.

Suppose that F_1 and F_2 are absolutely continuous with corresponding density functions f_1 and f_2. The SPRT is defined in terms of the "probability ratio"

$$\frac{f_{2m}}{f_{1m}} = \frac{f_2(y_1)\cdot f_2(y_2)\ \ldots\ f_2(y_m)}{f_1(y_1)\cdot f_1(y_2)\ \ldots\ f_1(y_m)},$$

and two positive constants A and B, $A \leq B$. The sample size and terminal decision functions are determined as follows: Before an observation is taken, decide whether $n > 0$.* If n=0, make the terminal decision minimizing the expected loss. If $n > 0$, observe the value y_1 of Y_1 and compute f_{21}/f_{11}. If this ratio is greater than or equal to B, accept f_2 as the true density with one observation. If $f_{21}/f_{11} \leq A$, accept f_1 with one observation. If $A<f_{21}/f_{11}<B$, observe y_2 and compute f_{22}/f_{12}. Continue sampling or terminate with the appropriate decision according to whether f_{22}/f_{12} is in (A,B) or not, respectively. In general, continue sampling as long as f_{2m}/f_{1m} is between A and B, and terminate as soon as this condition is violated. Accept f_1 if $\frac{f_{2n}}{f_{1n}} \leq A$, and accept f_2 if $f_{2n}/f_{1n} \geq B$.

---

*One can, for example, define f_{20}/f_{10} to be unity, so that n=0 if and only if $A \geq 1$ or $B \leq 1$.

The optimal property of the SPRT is that it requires, on the average, under both hypothesis, fewer (or at most not more) observations than any other test with the same or smaller error probabilities. More precisely, the optimal property of SPRT is stated as

Theorem 2. Let Q_{ij} and N_i denote the O.C.'s of a SPRT defined by two fixed numbers A and B, $0 < A < 1 < B$, and let Q^*_{ij} and N^*_i denote the O.C.'s of any other test (n, ϕ_n) in $\mathcal{S}$ for the 2-decision problem. Then

$$Q^*_{ij} \leq Q_{ij} \text{ for all } i,j, i \neq j$$

implies that

$$N^*_i \geq N_i \text{ for all } i.$$

Remark. A proof of Theorem 2 is given in [10].

Associated with the optimum property of the SPRT are the following two properties:

Uniqueness. Two SPRT's with the same error probabilities are equivalent.

Monotonicity. If a SPRT with stopping bounds A and B is changed by decreasing A and increasing B, and if the new test is not equivalent to the old one, then at least one of the error probabilities is decreased.

The connection between the optimum property, the monotonicity property, and the uniqueness property of the SPRT is discussed by Wijsman [11]. In particular, Wijsman shows that the monotonicity property can be proved independently of the optimal property, and that it implies the uniqueness property and the optimal property of the SPRT within the class of SPRT's.

Chapter II

THE GENERALIZED SEQUENTIAL PROBABILITY RATIO TEST

It is the purpose of the present chapter to introduce the "generalized sequential probability ratio test" (GSPRT), and of the following chapter to investigate some of its properties.

Let f_i be a probability density function with respect to a σ-finite measure μ on $(\mathcal{Y}, \mathcal{B})$, $i=1,2, \ldots, k$. In order to avoid trivialities, we shall assume that the f_i's are pairwise not equivalent. Let F_i denote the probability distribution function corresponding to f_i, i.e.,

$$F_i(B)= \int_B f_i(y)\, d\mu \text{ for all } B\epsilon\mathcal{B}.$$

Let $\mathcal{X}$ be the k-1 dimensional simplex defined in Chapter I (see page 2). Define a mapping r from $\mathcal{Y}$ to $\mathcal{X}$ by the equation

$$(5) \quad r(y)=(\textstyle\sum f_i(y))^{-1}\, (f_1(y),\ f_2(y),\ \ldots,\ f_k(y))$$

for all $y\epsilon\mathcal{Y}$ for which $\sum f_i(y)\neq 0$. We arbitrarily define $r(y)=(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$ if $\sum f_i(y)=0$. If $k=2$,

$$r=(1+f_2/f_1)^{-1}(1,\ f_2/f_1),$$

so the mapping is equivalent to the ordinary probability ratio in this case. Thus we will call r(y) the "generalized probability ratio" for $k \geq 2$.

Let $\mathcal{F}$ denote the σ-field over $\mathcal{X}$ induced by r, i.e.,

$$\mathcal{F} = \left\{A : r^{-1}(A) \in \mathcal{B}\right\}.$$

It is easily seen that the k-1 dimensional Borel sets are in $\mathcal{F}$.

We shall define the Euclidean metric m on the space $\mathcal{X}$. That is,

$$m(x,y) = \left(\sum (x^i - y^i)^2\right)^{\frac{1}{2}}.$$

Let P_i denote the distribution on $(\mathcal{X}, \mathcal{F})$ defined by

$$P_i(A) = \int_{r^{-1}(A)} f_i \, d\mu, \text{ for } A \in \mathcal{F}.$$

Let ν denote the measure $\sum P_i$ on $(\mathcal{X}, \mathcal{F})$. Note that $P_i \ll \nu$ and ν is totally finite. Then by the Radon-Nikodym theorem, densities $p_1, p_2, \ldots, p_k$ exist such that

$$P_j(A) = \int_A p_j(x) \, d\mu \text{ for all } A \in \mathcal{F}.$$

If we restrict our attention to decision rules depending on Y_j only through $r(Y_j)$, we can state the decision rule for deciding the true density of Y_j in terms of a sequence X of random variables

$$X = (X_1, X_2, \ldots) = (r(Y_1), r(Y_2), \ldots)$$

in the space $\mathcal{X}$, whose density is known to be one of the p_i's. The k-decision problem that we are considering can be stated in this context as the problem of finding a rule for deciding which of the p_i's is the true density of $r(Y_j)$, where the p_i's are the densities induced on the generalized probability ratio by the distributions which determine the alternatives to be considered. We shall see below that this restriction on the class of rules to be considered does not result in an increase in the minimum attainable risk.

We now show that the p_i's satisfy the following identity:

Theorem 3. $(p_1(x), p_2(x), \ldots, p_k(x)) = x$ a.e.ν.

Proof. For any integrable function $u(x)$ we have*

$$\int_{\mathcal{X}} u(x) p_i(x) d\nu(x) = \int_{\mathcal{Y}} u\left\{ r(y) \right\} f_i(y) d\mu(y)$$
$$= \int_{\mathcal{Y}} u\left\{ r(y) \right\} r^i(y) (\sum f_j(y) d\mu(y))$$
$$= \int_{\mathcal{X}} u(x) x^i d\nu(x).$$

Therefore,

$$p_i(x) = x^i \quad \text{a.e.}\nu.$$

Remark. This identity was originally stated by Seo [6]. The present proof is different from the one given by Seo, however.

Definition. We define a binary operation "o" on the elements of $\mathcal{X}$ as follows:

$$x o y = (\sum x^i y^i)^{-1} (x^1 y^1, x^2 y^2, \ldots, x^k y^k).$$

It is easily seen that the interior $\mathcal{X}^o$ of $\mathcal{X}$ forms an abelian group relative to this operation. (The identity element $(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$ will be denoted by e in that which follows.)

Remark. It is also true that the identity of

---

*For proof of the first step see, for example, Halmos [5, p.163].

Theorem 3 holds for the "cummulative sums" $s_n = x_o o x_1 o \ldots o x_n$. That is,

$$\left(\sum_i \prod_{j=o}^{n} p_i(s_j)\right)^{-1}\left(\prod_{j=o}^{n} (p_1(s_j)), \prod_{j=o}^{n}(p_2(s_j)), \ldots, \prod_{j=o}^{n} (p_k(s_n))\right) = s_n \quad \text{a.e.}\nu.$$

An argument similar to the proof of Theorem 3 can be used to prove this assertion.

Theorem 3 has several interesting consequences concerning properties of the random variable r(Y). Some of these are given in the following theorem. For simplicity of notation, we shall use the letter X to denote a component of the vector X in what follows, provided that the context makes confusion impossible.

Theorem 4. (i) $E_i(X^j)=E_j(X^i)$; $i,j=1,2, \ldots, k$, where X^j denotes the j-th component of a member of the vector X of random variables.

(ii) $E_i(X^i) > \frac{1}{k}$; $i=1,2, \ldots, k$.

(iii) $\sum_j E_j(X^i)=1$; $i=1,2, \ldots, k$.

(iv) $E_i(XoX)^i > \dfrac{1}{E_i\left(\sum_j \frac{X^i}{X^j}\right)}$

(v) $P_i(X=e)=\frac{1}{k}\,\nu(e) < 1$; $i=1,2, \ldots, k$.

Proof. (i) $E_i(X^j)=\int x^j x^i d\nu$ by Theorem 3. But this is equivalent to

$$\int x^i x^j d\nu = E_j(X^i).$$

(ii) By Schwarz's inequality,

$$\int (x^i)^2 d\nu \cdot \int 1^2 d\nu > \left[\int x^i \, 1 d\nu \right]^2 = 1,$$

so that

$$E_i(X^i) = \int (x^i)^2 d\nu > \frac{1}{\int d\nu} = \frac{1}{k} .$$

The strict inequality holds, since under the assumed nonequivalence of the P_i's, x^i cannot be equivalent (with respect to the measure ν) to a constant.

(iii) and (iv) are trivial consequences of Theorem 3 and part (i) of the present theorem.

(v) Since the p_j's are assumed to be distinct, they cannot all assign probability 1 to the set $\{ e \}$, that is,

$$\int_{\{e\}} x^i d\nu = \int_{\{e\}} \frac{1}{k} d\nu = \frac{1}{k}\nu(e) \neq 1.$$

But this, together with the fact that

$$p_1(e) = p_2(e) = \ldots = p_k(e) = \frac{1}{k},$$

implies that none of the densities can assign probability 1 to $\{ e \}$.

The sequential rules we shall consider will be given in terms of random walks on $\mathcal{X}$. For each random walk a starting point x_o and a sequence of measurable absorbing barriers $\{ \cup_i A_{in} \}$ are specified. (The components of x_o may be considered to be a priori probabilities of the corresponding p_i's.) Such rules will be identified by the set $\{ x_o, \cup_i A_{in} \}$.

A test of the k hypotheses

H_i: The density of X_j is p_i; $i=1,2, \ldots, k$

using the rule $\{ x_o, \cup_i A_{in} \}$ operates as follows: The set $\{ A_{in}: i=1,2, \ldots, k; n=0,1,2, \ldots \}$ and the vector x_o are given. (We shall see below that in order to avoid trivialities we may assume that $x_o \varepsilon \mathcal{X}^o$ and that the sets A_{in} are mutually disjoint for each n.) If $x_o \varepsilon A_{i0}$ for some i, the test accepts the i-th density to be the true one without taking an observation. If it is not, a value x_1 of X_1 is observed. If the vector of a posteriori probabilities $s_1 = x_0 \circ x_1$ is a point in A_{i1}, the test accepts p_i with one observation. If $s_1 \not\varepsilon \cup A_{i1}$, observe x_2, compute $s_2 = s_1 \circ x_2$, and determine whether $s_2 \varepsilon \cup A_{i2}$. If so the test terminates; if not it continues, and so on.

The remainder of this chapter is devoted to an investigation of the tests $\{ x_o, \cup A_{in} \}$. We shall show that a slightly narrower class of such tests contains tests attaining risks as small as any Bayes x_o,W tests in $\mathcal{S}$. Several of the theorems we give were originally proved by Wald [9], and rely heavily on the characterization of Bayes rules given in Theorem 1. These theorems are stated here in the framework of random walks on $\mathcal{X}$.

It follows from Theorem 1 that a Bayes x_o,W rule in $\mathcal{S}$ can be given in terms the sequential rules $\{ x_o, \cup_i A_{in} \}$ defined above. Thus it is true that a Bayes x_o,W rule in $\mathcal{S}$ for the k-decision problem can be associated with the random walk starting at x_o and stopping in the absorbing barrier $\bigcup_i A_{in}$. In fact, the sequence

$\{ \cup_i A_{in} \}$ is constant over n for a Bayes x_o,W rule in $\mathcal{S}$, a fact which is stated as

Theorem 5. For a Bayes x_o,W rule in $\mathcal{S}$, the stopping region $\bigcup_i A_{in}$ is independent of n.

Proof. From Theorem 1, we see that the stopping region can be defined as the set

$$\bigcup_i A_{in} = \left\{ x \varepsilon \mathcal{X} : \rho_o(x,W) \leq \rho^*(x,W) \right\} . \tag{6}$$

It follows from conditions (ii) and (iii) of Theorem 1 that for each x_o and W there is a Bayes rule in $\mathcal{S}$ which terminates when the walk enters $\bigcup_i A_{in}$. Since the inequality in the right hand side of eq. 6 does not depend on n, neither does $\bigcup_i A_{in}$.

Remark. In view of Theorem 5, we shall henceforth designate the rule $\left\{ x_o, \bigcup_i A_{in} \right\}$ by the set $\left\{ x_o, \cup A_i \right\}$, since by restricting our attention to rules of the latter type we do not increase the minimal risk attainable.

In what follows we assume that the starting point x_o is in $\mathcal{X}^o$. If x_o is a point of the boundary of $\mathcal{X}$, that is, if at least one of the components x_o^i is zero, the resulting procedure may be replaced by a test of correspondingly fewer hypotheses without increasing the risk. In particular, if one of the components of x_o is unity, a Bayes x_o,W rule in $\mathcal{S}$ accepts (with zero risk) the corresponding hypothesis without taking an observation, since the off-diagonal elements of W are assumed to be non-negative, and the diagonal elements are assumed to be zero. This observation is equivalent to stating that

A_j contains the point $(\delta_{1j}, \delta_{2j}, \ldots, \delta_{kj})$, the "j-th vertex of the space $\mathcal{X}$."

Remark. We could assume, without loss of generality, that $x_o = e$, since the test $\{x_o, \cup A_i\}$ is equivalent to the test $\{x_o, \cup (x_o^{-1} o A_i)\}$. We shall not do this, however, since we shall have occasion to consider the O.C.'s as functions of x_o for fixed stopping region.

We next show that under certain circumstances other points (if they exist) should be included in A_j. Suppose x is a point in $\mathcal{X}$ such that all random walks containing x at some stage almost surely eventually terminate in A_j. Then it seems natural that this point should also be in A_j, a fact which is stated as

Theorem 6. Suppose $U \subset \mathcal{X}$ is a set defined by

$$U = \left\{ x : P_i \{ s_N \varepsilon A_j \mid x_o = x \} = 1 \text{ for all } i \right\},$$

where N denotes the number of observations required by the test $\{x_o, \cup A_i\}$. Then we may consider only tests for which $U \subset A_j$, inasmuch as any test for which this is not true can be replaced by one for which it is and the latter test has risk at most as small as the former test.

Proof: Let $A_1, A_2, \ldots, A_k, Q_{ij}$, and N_i denote stopping regions and O.C's of a test S. Suppose S' is a new test in which $A_j' = A_j \cup U$, with O.C.'s denoted by Q_{ij}' and N_i'. Assume some a priori point x_o is given. Now $Q_{ij}'(x_o) = Q_{ij}(x_o)$ for all i, since

$$Q_{ij}(x_o) = P_i \{ s_N \varepsilon A_j \mid x_o \} = P_i \{ s_N \varepsilon A_j \text{ or } s_n \varepsilon U,$$

$$\text{for some } n \le N \mid x_o \} = P_i \{ s_{N'} \varepsilon A'_j \mid x_o \} = Q'_{ij}(x_o).$$

Also, $Q_{im}(x_o)=Q'_{im}(x_o)$ for $m \neq j$, since

$$P_i \{ s_N \varepsilon A_m \mid x_o \} = P_i \{ s_N \varepsilon A_m \text{ and } s_n \notin U, \text{ all } n < N \mid x_o \}$$

$$= P_i \{ s_n \notin U \text{ all } n < N \mid s_N \varepsilon A_m, x_o \} \cdot P_i \{ s_N \varepsilon A_m \mid x_o \}$$

$$= P_i \{ s_{N'} \varepsilon A_m \mid x_o \} .$$

Since $\bigcup A_i \subset \bigcup A'_i$, it follows that

$$N_j(x_o) - N'_j(x_o) \ge 0 \text{ for } j=1,2, \ldots, k,$$

a fact which proved below (Theorem 18). In conclusion we observe that the test S has risk

$$R(x_o, W, S) = \sum_i x_o^i \, [cN_i + \sum_j w_{ij} Q_{ij}(x_o)]$$

which is greater than or equal to the risk

$$R(x_o, W, S') = \sum_i x_o^i \, [cN'_i + \sum_j w_{ij} Q'_{ij}(x_o)]$$

of the test S'.

Our next aim is to show that specifying the loss matrix W determines the stopping region $\bigcup A_i$ for a Bayes x_o,W rule in $\mathcal{L}$. In order to prove this, we make use of two theorems due to Wald [9, p.105]. We let p_{x_o} denote the probability density defined by

$$p_{x_o}(x) = \sum x_o^j x^i .$$

Theorem 7. $\rho_{m+1}(x_o,W)=\min\left\{\rho_o(x_o,W),\right.$

$$\left. 1+\int \rho_m(s_1,W)p_{x_o}(x_1)d\nu\right\}; \; m=0,1,2,\ \ldots\ .$$

Theorem 8. The function $\rho(x_o,W)$ satisfies the equation $\rho(x_o,W)=\min\left\{\rho_o(x_o,W),\ 1+\int\rho(s_1,W)p_{x_o}(x_1)d\nu\right\}$.

Theorem 9. For a Bayes x_o,W rule in $\mathcal{S}$, the components of W uniquely determine the A_i's.

Proof. For each x_o and W, $\rho_o(x_o,W)$ and $\rho(x_o,W)$ are uniquely determined by Theorems 7 and 8. Thus for each x, $\rho_o(x,W)$ and $\rho(x,W)$ are uniquely determined by W. The theorem follows by recalling that

$$\bigcup A_i=\left\{x:\rho_o(x,W)\le\rho(x,W)\right\},$$

and by the fact that the A_i's have at most boundary points in common. That the A_i's have at most boundary points in common is seen as follows:

$$A_j=\left\{x:\rho_o(x,W)\le\rho(x,W)\text{ and the test accepts }H_j\right\}$$

$$\subset\left\{x:\sum_i x^i w_{ij}\le\rho(x,W)\right\}$$

$$\subset\left\{x:\sum_i x^i w_{ij}=\rho_o(x,W)\right\}=A_j^*\text{ (say)}.$$

But the A_j^*'s have only boundary points in common.

Theorem 10. If $\left\{x_o,\bigcup A_i\right\}$ is a Bayes x_o,W rule in $\mathcal{S}$, the components A_i of the stopping region are convex.

Proof. For each δ, the risk $R(x,W,\delta)$ defined by equation 1 is linear in x, so that $\inf_\delta R(x,W,\delta)$ is a concave function of x.

Another property of the A_j's is that they can be considered to be closed. In order to prove this, we first need to show that the function $\rho(x,W)$ is continuous in x. This is the aim of the following two theorems.

Theorem 11. $\rho_m(x,W)$ converges to $\rho(x,W)$ (as $m\to\infty$) uniformly in x. The proof of this fact is given by Wald [9, p.106].

Theorem 12. Let $\{x_n\}$, n=1,2,3, ..., be a sequence of points converging in the metric m to a point x_o (also in $\mathcal{X}$). Then

$$\lim_{n\to\infty} \rho(x_n,W) = \rho(x_o,W).$$

Proof. $\rho_o(x,W) = \min_j \sum_i x^i w_{ij}$ is continuous in x, and by Theorem 7, $\rho_m(x,W)$ is continuous in x. Since by Theorem 11 $\rho_m(x,W)$ converges to $\rho(x,W)$ uniformly in x, it follows that $\rho(x,W)$ is continuous in x.

Remark. This implies that for a Bayes x,W rule in $\mathcal{S}$,

$$N_i + \sum_j Q_{ij} w_{ij}$$

is continuous in x for each i such that $x^i > 0$.

Theorem 13. For a Bayes x_o,W rule in $\mathcal{S}$, the A_j's may be considered to be closed.

Proof. For the convergent sequence $\{x_n\}$ of the preceding theorem we have, by the continuity of $\rho(x,W)$ and $\rho_o(x,W)$ that

$$\rho(x_n,W) \to \rho(x,W),$$

and

$$\rho_o(x_n,W) \to \rho_o(x,W)$$

as $n\to\infty$. Then $x_n \varepsilon A_j$ implies that

$$\rho_o(x_n,W) \le \rho^*(x_n,W) \text{ for all } n,$$

and thus

$$\rho_o(x,W) \le \rho^*(x,W),$$

so that $x\varepsilon A_j$.

Our next task is to state conditions sufficient to insure that the test $\{x_o, \cup A_i\}$ eventually almost surely terminates. For convenience of notation, we shall restrict ourselves in the remainder of the paper to the case k=3. The arguments used can readily be adapted to any other finite $k>1$, however.

Theorem 14. Suppose that each component A_i of the stopping region of the test $\{e,\cup A_i\}$ contains a spherical neighborhood (with respect to the topology induced on $\mathcal{X}$ by the metric m) centered at its corresponding extreme point $(\delta_{i1},\delta_{i2},\delta_{i3})$. Then the test terminates with probability 1 under all hypotheses.

Proof. We shall actually prove a slightly stronger statement: Under H_i the "cumulative sum" s_n converges almost surely to the vertex $(\delta_{i1},\delta_{i2},\delta_{i3})$. Suppose we consider the case in which H_1 is true. Maintaining our convention of using upper-case letters to designate random variables, we write $S_n=X_1 o X_2 o \ldots o X_n$, so that

$$S_n = \left(\frac{\prod_{j=1}^{n} X_j^1}{\sum_i \prod_j X_j^i}, \frac{\prod X_j^2}{\sum_i \prod_j X_j^i}, \frac{\prod X_j^3}{\sum_i \prod_j X_j^i} \right)$$

$$= \left(\frac{1}{1 + \frac{\prod X_j^2}{\prod X_j^1} + \frac{\prod X_j^3}{\prod X_j^1}}, \frac{\frac{\prod X_j^2}{\prod X_j^1}}{1 + \frac{\prod X_j^2}{\prod X_j^1} + \frac{\prod X_j^3}{\prod X_j^1}}, \frac{\frac{\prod X_j^3}{\prod X_j^1}}{1 + \frac{\prod X_j^2}{\prod X_j^1} + \frac{\prod X_j^3}{\prod X_j^1}} \right).$$

The latter equality holds a.s., since under H_1, $\prod_{j=1}^{n} X_j^1 > 0$ a.s. for any integer n. In view of this expression, we need only show that

(7) $$\frac{\prod X_j^2}{\prod X_j^1} \xrightarrow{a.s.} 0 \quad \text{and} \quad \frac{\prod X_j^3}{\prod X_j^1} \xrightarrow{a.s.} 0$$

under H_1.

Let

(8) $$Z_j = \ln \frac{X_j^2}{X_j^1} .$$

(Z_j is defined to be $-\infty$ if $\frac{X_j^2}{X_j^1}$ is zero.) We shall show that

(9) $$\sum_{j=1}^{n} Z_j \xrightarrow{a.s.} -\infty,$$

which implies the first part of expression (7).

Since the logarithm function is strictly concave, and since $\frac{X_j^2}{X_j^1}$ is assumed to be not identically 1, it follows by Jensen's inequality that

$$E_1 Z_j < \ln E_1 \left[\frac{X_j^2}{X_j^1}\right] = 0.$$

Also, we have that $E_1 \left|Z_j\right|$ is finite or $E_1 Z_j = -\infty$. To see this, it suffices to note that

$$E_1 \left|Z_j\right| = -\int_{x^2<x^1} \ln\left(\frac{x^2}{x^1}\right) x^1 d\nu + \int_{x^2\geq x^1} \ln\left(\frac{x^2}{x^1}\right) x^1 d\nu$$

$$\leq -\int_{x^2<x^1} \ln\left(\frac{x^2}{x^1}\right) x^1 d\nu + 1 .$$

If $E_1 \left|Z_j\right| < \infty$, the Kolmogorov strong law of large numbers ensures that

$$\frac{1}{n} \sum_{j=1}^{n} Z_j \xrightarrow{a.s.} E_1 Z_j < 0$$

so that expression (9) holds.

If $E_1 Z_j = -\infty$, consider the "truncated" random variable $Z_j^c = \begin{cases} Z_j, & \text{if } Z_j > c \\ c & \text{otherwise} \end{cases}$, where c is a constant. The constant c may be chosen sufficiently small so that $E_1 Z_j^c < 0$, and since $E_1 \left|Z_j^c\right| < \infty$ we again apply the strong law of large numbers (to the sequence Z_j^c) to obtain

$$\sum Z_j \leq \sum Z_j^c \xrightarrow{a.s.} -\infty.$$

A similar argument can be used to establish the second part of expression (7).

This argument may be repeated under the assumption that H_2 or H_3 holds, so that we have

$$S_n \xrightarrow{a.s.} (\delta_{1i}, \delta_{2i}, \delta_{3i})$$

under H_i. By the assumption that A_i contains a neighborhood of $(\delta_{1i}, \delta_{2i}, \delta_{3i})$, it follows that the walk S_n enters the stopping region $\cup A_i$ of the test with probability 1 under H_i.

Remark. Theorem 14 holds if the test under consideration is the test $\left\{x_o, \cup A_i\right\}$ for any $x_o \varepsilon \chi^o$. This follows by the fact that if the test $\left\{x_o, \cup A_i\right\}$ satisfies the conditions of the theorem, then the equivalent test $\left\{e, \cup(x_o^{-1} o A_i)\right\}$ will also satisfy these conditions, since the transformation "x_o^{-1}" preserves the existence of the neighborhoods required in the theorem.

In the remainder of this paper we consider the class of tests $\left\{x_o, \cup A_i\right\}$ such that

(i) $x_o \varepsilon \chi^o$,

(ii) the A_i's are convex and closed,

(iii) the A_i's have at most boundary points in common, and

(iv) the test terminates as soon as the walk enters $\cup A_i$, that is, randomization on the boundary of $\cup A_i$ is not considered.

In view of Theorems 5,9,10, and 13, the class of such tests is essentially complete in the class $\mathcal{S}$ of sequential rules for the k-decision problem, so that this class is sufficiently wide from the standpoint of minimizing risk.

Chapter III

MONOTONICITY PROPERTIES OF THE TEST $\{x_o, \cup A_i\}$

In this chapter we show a connection between the stopping region and the O.C.'s. In particular, we show that if the stopping region $\cup A_i$ of a test is made larger in certain ways to form a new test, certain combinations of the error probabilities for the new test will be larger than the corresponding ones of the old test, and the expected sample sizes of the new test will be smaller than the corresponding ones for the old test. The results given here are of the same type as some of those given by Wijsman [11, p.680], although he considered the special case k=2.

Define $\pi_i(x)$ to be the set characteristic function of A_i, i=0,1,2,3, that is,

$$\pi_i(x) = \begin{cases} 1, & \text{if } x \in A_i \\ 0, & \text{if } x \notin A_i, \end{cases}$$

where A_o is defined to be $\mathcal{X} - \cup A_i$. If the stopping region $\cup A_i$ of the test $\{x_o, \cup A_i\}$ is held fixed, the O.C.'s of the test are functions of x_o only, and will be denoted by $Q_{ij}(x_o)$ and $N_i(x_o)$.

<u>Theorem 15</u>. The O.C.'s of the rules $\{x_o, \cup A_i\}$ (with stopping region held fixed) satisfy the integral equations

(10) $$Q_{ij}(x_o) = \pi_j(x_o) + \pi_o(x_o)\int Q_{ij}(s_1)x^i d\nu,$$

and

(11) $$N_i(x_o) = \pi_o(x_o) + \pi_o(x_o)\int N_i(s_1)x^i d\nu,$$

for i,j = 1,2,3.

Proof. Since the argument used to establish eq. (11) is essentially the same as that used for eq. (10), we shall give only the latter. (This argument is similar to one used by Albert [1].) In order to establish eq. (10), let

(12) $$q_{kij}(x_o) = P_i\left\{s_k \varepsilon A_j \text{ and } s_m \varepsilon A_o \text{ for } m < k \;\middle|\; x_o\right\}.$$

Then (by Theorem 14)

(13) $$\sum_{k=o}^{\infty} q_{kij}(x_o) = P_i\left\{s_n \varepsilon A_j \;\middle|\; x_o\right\} = Q_{ij}(x_o).$$

The $q_{kij}(x_o)$'s satisfy the following relations:

$$q_{oij}(x_o) = \pi_j(x_o), \text{ and}$$

$$q_{(k+1)ij}(x_o) = \int\cdots\int \left[\prod_{t=o}^{k} \pi_o(s_t)\right]\pi_j(s_{k+1})$$
$$dP_i(s_{k+1} \mid s_k)\ \ldots\ dP_i(s_1 \mid x_o)$$

$$= \pi_o(x_o)\int q_{kij}(s_1)dP_i(s_1 \mid x_o)$$

$$= \pi_o(x_o)\int q_{kij}(s_1)x_1^i d\nu,$$

where P_i is the distribution of s when p_i is the density of X. The fact that $dP_i(s_1 \mid x_o)$ can be replaced by $x_1^i d\nu$ in the above argument follows from Theorem 3. Now

$$(14)\qquad \sum_{k=o}^{\infty} q_{kij}(x_o) = \pi_j(x_o) + \pi_o(x_o) \sum_{k=o}^{\infty} \int q_{kij}(s_1)x^i d\nu,$$

which (by the use of Lebesque's Monotone Convergence Theorem and the eq. (13)) may be written

$$Q_{ij}(x_o) = \pi_j(x_o) + \pi_o(x_o) \int Q_{ij}(s_1)x^i d\nu(x),$$

so that eq. (10) is obtained.

In what follows, we shall make use of equations of the type seen in eq. (10), so we next define the notion of lower and upper functions for the solutions of such equations, and examine the uniqueness of their solutions.

Consider the integral equation

$$(15)\qquad Q_{ij}(x) = \pi_j(x) + \pi_o(x) \int Q_{ij}(x o x_1)x_1^i d\nu(x_1).$$

A nonnegative function $h(x)$ is an upper function for the solution $Q_{ij}(x)$ of eq. (15) if its iterate $h_1(x)$, defined by

$$h_1(x) = \pi_j(x) + \pi_o(x) \int h(x o x_1)x_1^i d\nu(x_1)$$

satisfies the inequality

$$(16)\qquad h_1(x) \leq h(x) \text{ for all } x \varepsilon \mathcal{X}.$$

Similarly, $h(x)$ is called a lower function of $Q_{ij}(x)$ if

$$h_1(x) \geq h(x) \text{ for all } x \varepsilon \mathcal{X}.$$

The usefulness of upper functions for the solution of eq. (15) follows from

Theorem 16. An upper function $h(x)$ for the solution $Q_{ij}(x)$ of eq. (15) is an upper bound for $Q_{ij}(x)$ on $\mathcal{X}$.

Proof. Let $h(x)$ be an upper function for $Q_{ij}(x)$, and assume that $h(x)$ is not an upper bound for $Q_{ij}(x)$ on $\mathcal{X}$.

Then

$$U = \operatorname*{lub}_{x \in \mathcal{X}} \left\{ Q_{ij}(x) - h(x) \right\} > 0.$$

By iteration of the integration in eq. (15), we see that

$$Q_{ij}(x) = \sum_{m=o}^{n-1} I_m(i,j,x) + \int \cdots \int \left[\prod_{m=o}^{n-1} \pi_o(xou_m) \right] Q_{ij}(xou_n) \prod_{k=1}^{n} (x_k^i d\nu),$$

where

$$u_m = \begin{cases} e \text{ if } m=o \\ x_1 o x_2 o \ldots o x_m \text{ if } m > o \end{cases},$$

$$I_o(i,j,x) = \pi_j(x),$$

and

$$I_m(i,j,x) = \int \cdots \int \left[\prod_{k=o}^{m-1} \pi_o(xou_k) \right] \pi_j(xou_m) \prod_{t=1}^{m} (x_t^i d\nu).$$

Also, iteration in the upper function inequality (16) can be used to show that

$$h_n(x) \leq h_{n-1}(x) \leq \ldots \leq h_1(x) \leq h(x),$$

where

$$h_n(x) = \sum_{m=o}^{n-1} I_m(i,j,x) + \int \cdots \int \left[\prod_{m=o}^{n-1} \pi_o(xou_m) \right] h(xou_n) \prod_{k=1}^{n} (x_k^i d\nu),$$

with the same I_m's as above. Thus it follows that

$$Q_{ij}(x) - h(x) \leq Q_{ij}(x) - h_n(x)$$

$$= \int \cdots \int \left[\prod_{m=o}^{n-1} \pi_o(xou_m) \right] [Q_{ij}(xou_n) - h(xou_n)] \prod_{k=1}^{n} (x_k^i d\nu)$$

$$\leq U \int \cdots \int \prod_{m=o}^{n-1} (\pi_o(xou_m) x_m^i d\nu).$$

But by Theorem 14 the latter integral tends to zero with increasing n, so that $Q_{ij}(x) - h(x) \leq 0$, a contradiction.

Remark. A similar argument can be used to show that a lower function for the solution $Q_{ij}(x)$ is a lower bound for $Q_{ij}(x)$ for all $x \epsilon \mathcal{X}$.

<u>Theorem 17</u>. The solution $Q_{ij}(x)$ of eq. (15) is unique.

Proof. Assume that there are two solutions $Q_{ij}(x)$ and $Q'_{ij}(x)$. Let

$\Delta Q_{ij}(x) = Q_{ij}(x) - Q'_{ij}(x)$. Then

$$\Delta Q_{ij}(x) = \pi_o(x) \int \Delta Q_{ij}(xox_1) x_1^i d\nu.$$

We wish to show that $\Delta Q_{ij}(x) = 0$ for all $x \epsilon \mathcal{X}$. By iteration of the above integral, we obtain

$$\Delta Q_{ij}(x) = \int \cdots \int \left[\prod_{m=o}^{n-1} \pi_o(xou_m) \right] \Delta Q_{ij}(xou_n) \prod_{t=1}^{n} (x_t^i d\nu).$$

Let $\underset{x \epsilon \mathcal{X}}{\text{lub}} \; \Delta Q_{ij}(x) = \alpha \leq 1$. Then

$$\Delta Q_{ij}(x) \leq \alpha \int \cdots \int \prod_{m=o}^{n-1} (\pi_o(xou_m) x_m^i d\nu).$$

It was observed in the proof of Theorem 16 that the right hand side of this inequality tends to zero with increasing n, so that $\Delta Q_{ij}(x) \leq 0$. Since the choice of $Q_{ij}(x) - Q'_{ij}(x)$ for $\Delta Q_{ij}(x)$ was arbitrary, it also follows that $-\Delta Q_{ij}(x) \leq 0$, and the proof of the theorem is complete.

Remark. Theorem 17 might also be proved by observing that if $\pi_j(x) = 0$ for all $x \varepsilon \mathcal{X}$, then $Q_{ij}(x) = 0$. The uniqueness of the solutions to eqs. (10) and (11) is important in our case, since in order to prove certain properties of the O.C.'s we shall argue in terms of their integral representations. Thus we can ensure that in dealing with a solution of an equation of these types, we are dealing with the corresponding O.C.'s.

It has been observed [11] that if the upper stopping bound of a SPRT is increased and the lower one decreased, and if the new test is not equivalent to the old one, then at least one of the error probabilities is decreased. In the remainder of the present chapter, the analogous properties of the GSPRT will be investigated. Some implications of these "monotonicity properties" of the O.C.'s are explored below.

In order to prove that increasing the size of the stopping region strictly changes certain of the O.C.'s, we make the following

Definition. Let $\triangle$ be the symmetric difference of the stopping regions for the two GSPRT's $\left\{ x_o, \cup A_i \right\}$ and $\left\{ x_o, \cup A_i' \right\}$. These tests are said to be not equivalent if, for some integers n and i,

$$P_i \left\{ s_n \varepsilon \triangle, \; s_m \varepsilon A_o' \cap A_o \text{ for all } m < n \right\} > 0.$$

Remark. Roughly speaking, the definition states that two GSPRT's are equivalent if they have the same O.C.'s.

We are now in a position to prove

<u>Theorem 18</u>. If two nonequivalent GSPRT's $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ (with expected sample sizes $N_i(x_o)$ and $N_i'(x_o)$, respectively) are such that $\cup A_i \subset \cup A_i'$, then

$$\Delta N_i(x_o) = N_i'(x_o) - N_i(x_o) \leq 0$$

for all i, with strict inequality for at least one i.

Proof. From Theorem 15 we see that $\Delta N_i(x_o)$ satisfies the equation

$$(17)\quad \Delta N_i(x_o) = -\delta(x_o)(1 + \int N_i(x_o o x)x^i d\nu) + \pi_o'(x_o) \int \Delta N_i(x_o o x)x^i d\nu,$$

where δ is the characteristic function of $\triangle$. Since zero is an upper function for $\Delta N_i(x_o)$ we have (by Theorem 16) that $\Delta N_i(x_o) \leq 0$ for all i. Iterating the integration in (17) n times with zero as a first approximation, we obtain the expression

$$I_n = \sum_{j=o}^{n-1} I_j - \int \cdots \int \Big[\prod_{j=o}^{n-1} \pi_o'(s_j)\Big]\delta(s_n)(1+N_i(s_{n+1})) \prod_{k=1}^{n+1}(x_k^i d\nu),$$

where I_j denotes the j-th iterate and

$$I_o = -\delta(x_o)(1+\int N_i(s_1)x_1^i d\nu).$$

Since $\Delta N_i(x_o) \leq I_n \leq \ldots \leq I_1 \leq 0$, it suffices to show that for some i and some n, $I_n < 0$, that is, that

$$\int \cdots \int \Big[\prod_{j=o}^{n-1} \pi_o'(s_j)\Big]\delta(s_n)(1+N_i(s_{n+1})) \prod_{k=1}^{n+1}(x_k^i d\nu) > 0$$

for some i and n. But this is guaranteed by the condition of nonequivalence of the tests, since nonequivalence of $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ implies that

$$\int\cdots\int\Big[\prod_{j=o}^{n-1}\pi_o'(s_j)\Big]\delta(s_n)\prod_{k=1}^{n}(x_k^i d\nu) > 0$$

for some i and n.

Monotonicity theorems on the error probabilities are our next consideration. For the first two theorems, we consider increasing only certain parts of the stopping region $\cup A_i$. The first result is given as

Theorem 19. Given a test $\{x_o, \cup A_i\}$, let a new stopping region $\cup A_i'$ be defined so that $A_1 \subset A_1'$, $A_2 = A_2'$, and $A_3 = A_3'$. Denote the error probabilities for the new rule by Q_{ij}'. Then $Q_{i1}' \geq Q_{i1}$ for i = 1,2,3, and $Q_{ij}' \leq Q_{ij}$ for all i and $j \neq 1$.

Proof: Let

(18) $$\delta_1(x_o) = \pi_1'(x_o) - \pi_1(x_o),$$

where π_1' is the characteristic function of the new region A_1'. Define

(19) $$\Delta Q_{ij}(x_o) = Q_{ij}'(x_o) - Q_{ij}(x_o).$$

Then using equations (10), (18), and (19) it follows that

(20) $$\Delta Q_{i1}(x_o) = t(x_o) - \pi_o'(x_o)\int \Delta Q_{i1}(s_1)x_1^i d\nu,$$

where

$$t(x_o) = \delta_1(x_o)[1 - \int Q_{i1}(s_1)x_1^i d\nu] \geq 0.$$

It follows that zero is a lower function, and hence a lower bound, for the solution to eq. (20). Thus the first part of the theorem follows.

To complete the proof of Theorem 19, we consider Q_{ij} and Q'_{ij} for $j \neq 1$. An argument analogous to that given for eq. (20) yields

$$- \Delta Q_{ij}(x_o) = \delta_1(x_o) \int Q_{ij}(s_1) x_1^i d\nu + \pi_o'(x_o) \int - \Delta Q_{ij}(s_1) x_1^i d\nu.$$

Since $\delta_1(x_o) \int Q_{ij}(s_1) x_1^i d\nu \geq 0$, the second part of the theorem follows.

Remark. The inequalities in Theorem 19 can be strengthened if we make the additional assumptions that:

(i) the two tests $\left\{ x_o, \cup A_i \right\}$ and $\left\{ x_o, \cup A_i' \right\}$ are not equivalent,

and

(ii) $Q_{i1}(x) < 1$ on some subset of $\triangle$ with positive measure under H_i.

In particular, if

$$Q_{i1}(x) = 1 \text{ for all } i \text{ implies that } x \epsilon A_1,$$

then the assumption of nonequivalence of the tests is sufficient to guarantee that $\Delta Q_{i1}(x) > 0$ for some i. In view of Theorem 6, condition (ii) above can be considered to be a condition on the shape of the stopping region $\cup A_i$. If this condition is not satisfied, the performance of the test can be uniformly improved by changing the configuration of $\cup A_i$. A proof of strict inequality under conditions (i) and (ii) above can be obtained by iterating the integration in eq. (20) and arguing as in the proof of Theorem 18.

Theorem 20. If in the test $\{x_o, \cup A_i\}$ a new stopping region $\cup A_i'$ is defined so that $A_1 \subset A_1'$ and $A_2 \subset A_2'$, with $A_3 = A_3'$, then

$$\Delta Q_{i1} + \Delta Q_{i2} \geq 0 \text{ for } i = 1,2,3.$$

Proof. Let $\delta_i(x_o) = \pi_i'(x_o) - \pi_i(x_o)$ for $i = 1,2$. We have

$$(21) \quad \Delta Q_{i1}(x_o) = \delta_1(x_o) - (\delta_1(x_o) + \delta_2(x_o)) \int Q_{i1}(s_1) x_1^i d\nu$$
$$+ \pi_o'(x_o) \int \Delta Q_{i1}(s_1) x_1^i d\nu$$

with a similar expression for ΔQ_{i2}. Thus

$$(22) \quad \Delta Q_{i1} + \Delta Q_{i2} = (\delta_1 + \delta_2)(1 - \int (Q_{i1} + Q_{i2}) dP_i)$$
$$- \pi_o' \int (\Delta Q_{i1} + \Delta Q_{i2}) dP_i.$$

Since $\sum_j Q_{ij} = 1$, it follows that $1 - \int (Q_{i1} + Q_{i2}) dP_i \geq 0$, and zero is a lower function for the solution to eq. (22), which completes the proof.

Remark. The inequality of Theorem 20 may be strengthened under the condition of nonequivalence of the tests. The argument is outlined in the remark following Theorem 19. It should also be noted that Theorem 20 remains true if the assumption that $A_3 = A_3'$ is replaced by the assumption that $A_3 \supset A_3'$. A similar remark holds for Theorem 19.

The derivation of the integral equation for $Q_{ij}(x_o)$ involves the assumption of independent identically distributed elements in $X = (X_1, X_2, \ldots)$, and from the equation certain monotonicity properties have been obtained (Theorems 18, 19, and 20). These results can also be obtained without assumptions on the distribution properties of X. Instead of explicitly using the error probabilities $Q_{ij}(x_o)$, one can argue in terms of the performance of the test under various sample sequences.

As an example of this type of argument, consider the following situation:

Given two GSPRT's $T = \{x_o, \cup A_i\}$ and $T^* = \{x_o, \cup A_i'\}$, where $A_1 \subset A_1'$, $A_2' \subset A_2$, $A_3' \subset A_3$. Then $Q_{i1} \leq Q_{i1}'$, $i = 1,2,3$, and $Q_{ij}' \leq Q_{ij}$ for $i = 1,2,3$, $j \neq 1$, where

$$Q_{i1}' = P_i \{T^* \text{ accepts } H_1\}, \text{ and}$$

$$Q_{i1} = P_i \{T \text{ accepts } H_1\}.$$

Every sample sequence $y_1, y_2, \ldots y_N$ leading to the acceptance of H_1 under T will also lead to its acceptance under T*. Thus the event $\{y{:}T \text{ accepts } H_1\}$ is contained in the event $\{y{:}T^* \text{ accepts } H_1\}$, so that

$$Q_{i1} = P_1 \{y{:}T \text{ accepts } H_1\} \leq Q_{i1}'.$$

Also, it is possible that for some sequences such that T accepts H_2 or H_3 T* will accept H_1, so that the inequality is strengthened.

The situation for $j \neq 1$ (say $j=2$) can be argued as follows: A sequence y resulting in acceptance of H_2 by T may no longer result in its acceptance by T*. There are two reasons:

(i) entrance of the walk s_m into A_2 is no longer sufficient (but is necessary) for acceptance of H_2 under T*, and

(ii) T* may accept H_1 before the walk enters A_2.

Thus we have:

$$\{y: \text{T accepts } H_2\}$$

contains points y not in

$$\{y: \text{T* accepts } H_2\} .$$

However, we do not have the strict inclusion between these events, as was the case in the first part of this argument. The difficulty is that $\{y: \text{T* accepts } H_2\}$ may contain points not in $\{y: \text{T accepts } H_2\}$. Thus we must consider events of the type $\{y: \text{T accepts } H_2 \text{ or } H_3\}$, and obtain an inequality involving $Q_{i2}+Q_{i3}$ and $Q'_{i2}+Q'_{i3}$ as was the case in Theorem 20.

It is interesting to note that Theorems 18, 19, and 20 hold for any GSPRT. These theorems are quite general in that the measure space $(\mathcal{X}, \mathfrak{F}, \nu)$ may be considered to be the image (through r) or a quite general measure space $(\mathcal{Y}, \mathcal{B}, \mu)$, so that the sequential test defined in terms of a random walk on $\mathcal{X}$ may be applied to a wide class of problems. Unfortunately, there does not seem to be a monotonicity theorem as general as Theorems 19 and 20 for the case in which all components A_i of the stopping region are simultaneously enlarged. Under certain restrictions on the shape of the A_i's, however, such a monotonicity theorem can be proved. We devote the next several paragraphs to a discussion of two such shapes, and their application to a more general monotonicity theorem.

Suppose in particular that A_o is convex as well as A_1, A_2, and A_3. Then the boundaries of the A_i's are straight lines in $\mathcal{X}$ characterized by their points of intersection with the boundary of $\mathcal{X}$. Suppose, for example, that

$$\max \left\{ x^1 : x \varepsilon A_1 \text{ and } x^2 = 0 \right\} = a_{31},$$

(23

$$\max \left\{ x^1 : x \varepsilon A_1 \text{ and } x^3 = 0 \right\} = a_{21}.$$

Then an equation of the line forming the boundary between A_1 and A_o (that is, the line containing the two points $(a_{21}, 1-a_{21}, 0)$ and $(a_{31}, 0, 1-a_{31})$) is given by

$$(24) \qquad x^1 = \frac{a_{21}}{1 - a_{21}} \; x^2 + \frac{a_{31}}{1 - a_{31}} \; x^3$$

assuming that $a_{i1} < 1$. Similarly, equations of the lines forming the boundaries between A_2 and A_o, and A_3 and A_o, are given by

$$x^2 = \frac{a_{12}}{1 - a_{12}} \; x^1 + \frac{a_{32}}{1 - a_{32}} \; x^3$$

$$x^3 = \frac{a_{13}}{1 - a_{13}} \; x^1 + \frac{a_{23}}{1 - a_{23}} \; x^2$$

where the a_{ij}'s are defined in a manner similar to that used in eqs. (23). Thus in this case,

(25) $A_i = \left\{ x\varepsilon\mathcal{X}:\ x^i \geq \frac{a_{ji}}{1-a_{ji}}\, x^j + \frac{a_{ki}}{1-a_{ki}}\, x^k;\ i,j,k \text{ distinct} \right\}.$

Remark. A_1 and A_3 are disjoint if and only if $a_{13} + a_{31} > 1$. A similar statement holds for the other combinations of A_i's. Also, $e\varepsilon A_o$ if and only if

$$2a_{ji}-3a_{ji}a_{ki}+2a_{ki} > 1$$

for all distinct i,j, and k.

Theorem 21. Suppose the components A_i of the stopping region of the test $\{x_o, \cup A_i\}$ are defined as in eq. (25). Then the probabilities $Q_{ij}(x_o)$ satisfy the inequalities

$$\begin{aligned} x_o^1 Q_{11}(x_o) &\geq \frac{a_{21}}{1-a_{21}}\, x_o^2 Q_{21}(x_0) + \frac{a_{31}}{1-a_{31}}\, x_o^3 Q_{31}(x_o), \\ x_o^2 Q_{22}(x_o) &\geq \frac{a_{12}}{1-a_{12}}\, x_o^1 Q_{12}(x_o) + \frac{a_{32}}{1-a_{32}}\, x_o^3 Q_{32}(x_o), \qquad (26) \\ x_o^3 Q_{33}(x_o) &\geq \frac{a_{13}}{1-a_{13}}\, x_o^1 Q_{13}(x_o) + \frac{a_{23}}{1-a_{23}}\, x_o^2 Q_{23}(x_o). \end{aligned}$$

Proof: If $x_o\varepsilon A_j$, then $Q_{ij}(x_o) = 1$ for $i = 1,2,3$, and $Q_{ik}(x_o) = 0$ for $k\neq j$. Thus the probabilities in two of the above inequalities are all zero, and in the remaining inequality the probabilities are all one, so that

$$x_o^j \geq \frac{a_{ij}}{1-a_{ij}}\, x_o^i + \frac{a_{kj}}{1-a_{kj}}\, x_o^k \ (i,j,k \text{ distinct})$$

which (in view of eq.(25)) is equivalent to the assumption that $x_o\varepsilon A_j$.

Suppose, then, that $x_o\varepsilon A_o$. We shall use an argument similar to that used by Wald [8, p.41] to establish similar inequalities for the SPRT. For any sample $x_1, x_2, \ldots, x_n$ such that the walk $s_n = x_o o x_1 o \ldots o x_n$ enters A_1 (say), it is true (by eq. (25)) that

$$s_n^1 \geq \frac{a_{21}}{1-a_{21}} s_n^2 + \frac{a_{31}}{1-a_{31}} s_n^3,$$

or equivalently,

$$(27) \quad x_o^1 \prod_{t=1}^{n} x_t^1 \geq \frac{a_{21}}{1-a_{21}} x_o^2 \prod_{t=1}^{n} x_t^2 + \frac{a_{31}}{1-a_{31}} x_o^3 \prod_{t=1}^{n} x_t^3.$$

For $q_{kij}(x_o)$ as defined in eq. (12), we have

$$(28) \quad q_{kij}(x_o) = \int \cdots \int \left(\prod_{v=o}^{k-1} \pi_o(s_v) \right) \pi_j(s_k) \prod_{t=1}^{k} (x_t^i d\nu).$$

Equations (27) and (28) imply that for each k,

$$(29) \quad x_o^1 q_{k11}(x_o) \geq \frac{a_{21}}{1-a_{21}} x_o^2 q_{k21}(x_o) + \frac{a_{31}}{1-a_{31}} x_o^3 q_{k31}(x_o).$$

Summing both sides of inequality (29) over all values of k we obtain the first of the inequalities (26) by eq. (13). The remaining inequalities in (26) are established in a similar manner.

Theorem 22. Suppose the components A_i of the stopping region of the test $\{x_o, \cup A_i\}$ are defined by eq. (25), where the a_{tm}'s are greater than 1/2. Then for any test $\{x_o, \cup A_i'\}$ such that $\cup A_i \subset \cup A_i'$,

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i \Delta Q_{ij}(x_o) \geq 0.$$

Proof. By equations (10) and (19) we have

$$(30) \quad \Delta Q_{ij}(x_o) = \delta_j(x_o) - \sum_{k=1}^{3} \delta_k(x_o) \int Q_{ij}(s_1) x_1^i d\nu + \pi_o'(x_o).$$

$$\cdot \int \Delta Q_{ij}(s_1)x_1^i d\nu,$$

where δ_j is the characteristic function of $A_j - A_j'$. By eq. (10) it also follows that for $x_o \varepsilon A_o$,

$$\int Q_{ij}(s_1)x_1^i d\nu = Q_{ij}(x_o),$$

so that eq. (30) may be written

$$(31)\quad \Delta Q_{ij}(x_o) = \delta_j(x_o) - \sum_{k=1}^{3} \delta_k(x_o)Q_{ij}(x_o) + \pi_o'(x_o)\int \Delta Q_{ij}(s_1)x_1^i d\nu.$$

Iterating the integration in (31) once, we obtain

$$\Delta Q_{ij}(x_o) = [\delta_j(x_o) - \sum_{k=1}^{3} \delta_k(x_o)Q_{ij}(x_o)]$$

$$+ \pi_o'(x_o)\int [\delta_j(s_1) - \sum_{k=1}^{3} \delta_k(s_1)Q_{ij}(s_1)]x_1^i d\nu$$

$$+ \pi_o'(x_o)\int \pi_o'(s_1)[\int \Delta Q_{ij}(s_2)x_2^i d\nu]x_1^i d\nu.$$

Similarly, by iterating the integration n-1 times, we obtain

$$(32)\quad \Delta Q_{ij}(x_o) = \sum_{m=o}^{n-1} I_m(i,j,x_o) + \int \cdots \int [\prod_{m=o}^{n-1} \pi_o'(s_m)]$$

$$\Delta Q_{ij}(s_n) \prod_{t=1}^{n} (x_t^i d\nu)$$

for all i and j, and for n=1,2,3, ..., where

$$I_m(i,j,x_o) = \begin{cases} \delta_j(x_o) - \sum_{k=1}^{3} \delta_k(x_o)Q_{ij}(x_o) & \text{for } m=0, \\ \int \cdots \int [\prod_{k=o}^{m-1} \pi_o'(s_k)][\delta_j(s_m) - \sum_{k=1}^{3} \delta_k(s_m) & \\ \qquad Q_{ij}(s_m)] \prod_{v=1}^{m} (x_v^i d\nu) & \text{for } m=1,2, \ldots . \end{cases}$$

In view of eq. (31), an integral equation for $\sum_{\substack{i,j\\ i\neq j}} x_o^i \Delta Q_{ij}(x_o)$ is

$$(33)\qquad \sum_{\substack{i,j\\ i\neq j}} x_o^i \Delta Q_{ij}(x_o) = \sum_{\substack{i,j\\ i\neq j}} x_o^i[\delta_j(x_o) - \sum_{k=1}^{3}\delta_k(x_o)Q_{ij}(x_o)] + \pi_o'(x_o)\int \sum_{\substack{i,j\\ i\neq j}} \Delta Q_{ij}(s_1)x_o^i\ x_1^i d\nu.$$

Iteration of the integration in eq. (33) leads to an equation of the form

$$(34)\qquad \sum_{\substack{i,j\\ i\neq j}} x_o^i \Delta Q_{ij}(x_o) = \sum_{m=o}^{n-1}\ \sum_{\substack{i,j\\ i\neq j}} x_o^i I_m(i,j,x_o) + \int\cdots\int [\prod_{m=o}^{n-1}\pi_o'(s_m)] \sum_{\substack{i,j\\ i\neq j}} (\Delta Q_{ij}(s_n)\prod_{t=0}^{n} x_t^i) d\nu \ldots d\nu,$$

where $I_m(i,j,x_o)$ is as defined in eq. (32).

In view of the remark following Theorem 16, it will suffice to show that for all m,

$$(35)\qquad \sum_{\substack{i,j\\ i\neq j}} x_o^i I_m(i,j,x_o) \geq 0.$$

For each m, this sum may be written as

$$(36)\qquad \int\cdots\int_{s_m \epsilon A_1' - A_1} [\prod_{k=o}^{m-1}\pi_o'(s_k)][(Q_{22}(s_m)\prod_{v=o}^{m} x_v^2 - Q_{12}(s_m)\prod_{v=o}^{m} x_v^1) + (Q_{33}(s_m)\prod_{v=o}^{m} x_v^3 - Q_{13}(s_m)\prod_{v=o}^{m} x_v^1)] d\nu \ldots d\nu \quad +$$

$$+ \int_{s_m \varepsilon A_2' - A_2} \cdots \int [\prod_{k=o}^{m-1} \pi_o'(s_k)][(Q_{11}(s_m) \prod_{v=o}^{m} x_v^1 - Q_{21}(s_m) \prod_{v=o}^{m} x_v^2)$$

$$+ (Q_{33}(s_m) \prod_{v=o}^{m} x_v^3 - Q_{23}(s_m) \prod_{v=o}^{m} x_v^2)]d\nu \ \ldots \ d\nu$$

$$+ \int_{s_m \varepsilon A_3' - A_3} \cdots \int [\prod_{k=o}^{m-1} \pi_o'(s_k)][(Q_{11}(s_m) \prod_{v=o}^{m} x_v^1 - Q_{31}(s_m) \prod_{v=o}^{m} x_v^3)$$

$$+ (Q_{22}(s_m) \prod_{v=o}^{m} x_v^2 - Q_{32}(s_m) \prod_{v=o}^{m} x_v^3)]d\nu \ \ldots \ d\nu.$$

Using the assumption that $a_{tm} > 1/2$ (so that $\frac{a_{tm}}{1-a_{tm}} > 1$), inequalities (26) imply that in particular

$$s_m^j Q_{jj}(s_m) \geq s_m^j Q_{ij}(s_m) \text{ for all } i,j, \text{ and } m,$$

or equivalently,

(37) $$Q_{jj}(s_m) \prod_{v=o}^{m} x_v^j \geq Q_{ij}(s_m) \prod_{v=o}^{m} x_v^i \text{ for all } i,j \text{ and } m.$$

Thus each integrand in (36) is nonnegative for all m, so that

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i I_m(i,j,x_o) \geq 0$$

for all m and x_o, which in turn implies that

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i \Delta Q_{ij}(x_o) \geq 0$$

for all x_o. This completes the proof of Theorem 22.

<u>Theorem 23</u>. The inequality of Theorem 22 is strict

if the tests $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ are nonequivalent.

Proof. The nonequivalence of the tests implies that, for some i, j and m,

$$(38) \qquad \underset{s_m \varepsilon A_j' - A_j}{\int \cdots \int} \left(\prod_{k=o}^{m-1} \pi_o'(s_k) \right) \prod_{v=1}^{m} (x_v^i d\nu) > 0,$$

and since this integral is no larger than

$$\underset{s_m \varepsilon A_j' - A_j}{\int \cdots \int} \left(\prod_{k=o}^{m-1} \pi_o'(s_k) \right) (d\nu)^m \text{ (m-fold integral)},$$

one of the integrals (36) is strictly positive provided the integrand is positive on $A_j'-A_j$.

Suppose, for example, that the inequality (38) holds for j=2. We wish to show that the corresponding integrand

$$(39) \quad [Q_{11}(s_m) \prod_{v=0}^{m} x_v^1 - Q_{21}(s_m) \prod_{v=o}^{m} x_v^2] + [Q_{33}(s_m) \prod_{v=o}^{m} x_v^3 - Q_{23}(s_m) \prod_{v=o}^{m} x_v^2]$$

in expression (36) is strictly positive on $A_2'-A_2$. In view of inequalities (26) and the fact that $\frac{a_{tm}}{1-a_{tm}} > 1$, it follows that expression (39) is positive for $s_m \varepsilon A_2'-A_2$ if one of the error probabilities $Q_{21}(s_m)$, $Q_{31}(s_m)$, $Q_{23}(s_m)$,$Q_{13}(s_m)$ is positive or if one of the probabilities $Q_{11}(s_m)$, $Q_{33}(s_m)$ is positive. But this is always true, since a contradiction results from the assumption that these probabilities are all zero. This is seen as follows:

If $Q_{21}(s_m)$ and $Q_{23}(s_m)$ are both zero, then $Q_{22}(s_m) = 1$.

Similarly,

$$Q_{11}(s_m) = Q_{13}(s_m) = 0 \text{ implies } Q_{12}(s_m) = 1,$$

and

$$Q_{33}(s_m) = Q_{31}(s_m) = 0 \text{ implies } Q_{32}(s_m) = 1.$$

Now from the second inequality (26) we must have

$$s_m^2 \geq \frac{a_{12}}{1-a_{12}} s_m^1 + \frac{a_{32}}{1-a_{32}} s_m^3$$

which implies that $s_m \varepsilon A_2$, contradicting the assumption that $s_m \varepsilon A_2' - A_2$.

It follows that at least one of the integrals (36) is strictly positive for some m, so that

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i I_m(i,j,x_o) > 0$$

for some m, which completes the proof.

Remark. The assumption that the constants a_{tm} are greater than 1/2 is equivalent to the condition that

$$A_j \subset \left\{ x : x^j \geq x^k \text{ for all } k \right\}$$

in the present case. The latter condition would be satisfied for Bayes x_o,W rules in $\mathcal{S}$, for example, if the loss matrix W had equal off-diagonal elements.

<u>Theorem 24</u>. Given two tests $\left\{ x_o, \cup A_i \right\}$ and $\left\{ x_o, \cup A_i' \right\}$ such that $\cup A_i \subset \cup A_i'$ and A_i' is defined as in equation (25), where the a_{tm}'s are greater than 1/2. Then

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i \Delta Q_{ij}(x_o) \geq 0.$$

Proof. The proof is similar to that given for Theorem 22, except that the integral equation (30) for $\Delta Q_{ij}(x_o)$ is replaced by

$$(40) \quad \Delta Q_{ij}(x_o) = [\delta_j(x_o) - \sum_{k=1}^{3} \delta_k(x_o) \int Q'_{ij}(s_1) x_1^i d\nu] + \pi_o(x_o) \int \Delta Q_{ij}(s_1) x_1^i d\nu .$$

Iteration of the integration in eq. (40) yields

$$(41) \quad \Delta Q_{ij}(x_o) = \sum_{m=o}^{n-1} I'_m(i,j,x_o) + \int \cdots \int [\prod_{k=o}^{n-1} \pi_o(s_k)] \Delta Q_{ij}(s_n) \prod_{t=1}^{n} \left(x_t^i d\nu \right) ,$$

where

$$I'_m(i,j,x_o) = \begin{cases} \delta_j(x_o) - \sum_{k=1}^{3} \delta_k(x_o) \int Q'_{ij}(s_1) x_1^i d\nu & \text{for } m=0, \\ \int \cdots \int [\prod_{k=o}^{m-1} \pi_o(s_k)][\delta_j(s_m) \\ \quad - \sum_{t=1}^{3} \delta_t(s_m) \int Q'_{ij}(s_{m+1}) x_{m+1}^i d\nu] \prod_{v=1}^{m} (x_v^i d\nu) . \end{cases}$$

As was the case in the proof of Theorem 22, we have

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i I'_m(i,j,x_o) \geq 0$$

for all m, since

$$(42) \quad \sum_{\substack{i,j \\ i \neq j}} x_o^i I'_m(i,j,x_o) = \int \cdots \int_{s_m \epsilon A'_1 - A_1} [\prod_{k=o}^{m-1} \pi_o(s_k)][(Q'_{22}(s_{m+1}) \prod_{v=o}^{m+1} x_v^2 - Q'_{12}(s_{m+1}) \prod_{v=o}^{m+1} x_v^1) + (Q'_{33}(s_{m+1}) \prod_{v=o}^{m+1} x_v^3 - Q'_{13}(s_{m+1}) \prod_{v=o}^{m+1} x_v^1)](d\nu)^{m+1}$$

plus integrals over $A_2'-A_2$ and $A_3'-A_3$ corresponding to those in eq. (36), where the integrands in eq. (42) are all nonnegative by inequality (37).

Thus it follows that

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i \Delta Q_{ij}(x_o) \geq 0,$$

which completes the proof.

Remark. The preceding three theorems imply the following statement: If the stopping regions of the two tests $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ are such that for each i,

(i) $A_i \subset A_i'$,

and

(ii) there is a line of the type described by eq. (24) in the set $A_i'-A_i$, where the a_{ki}'s are greater than 1/2,

then

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i \Delta Q_{ij}(x_o) \geq 0,$$

with strict inequality if the region $\cup A_i$ and the region formed by the lines define nonequivalent tests.

The arguments in the preceding three theorems (Theorems 22, 23, and 24) in which a certain shape of stopping region is considered rely upon this shape only through use of the fundamental inequalities of Theorem 21. Thus, the arguments can be used for any shape for which such inequalities can be obtained. As an example, we consider next another particular shape of stopping region for which we can obtain such inequalities.

Suppose that A_i is defined by the intersection of two lines in $\mathcal{X}$, each containing an extreme point opposite the one in A_i. We shall assume that the i-th component of each non-vertex point of intersection of these lines with the boundary of $\mathcal{X}$ is greater than 1/2. Thus, for example, A_1 is given by

$$(43) \qquad A_1 = \left\{ x: \frac{x^1}{x^2} \geq c_{12} \right\} \cap \left\{ x: \frac{x^1}{x^3} \geq c_{13} \right\},$$

where c_{12} and c_{13} are constants greater than 1.

The shape of the stopping region mentioned above arises quite naturally from a consideration of the probability ratio test as follows: We wish to test the hypotheses

$$H_i : f_i \text{ is the density of } Y \;;\; i=1,2,3$$

against each other. Suppose we use the test defined by the probability ratios:

$$\text{Accept } H_i \text{ if } \min_{j \neq i} \left\{ f_{jn}/f_{in} \right\} \leq C_{in} \;;\; i=1,2,3,$$

continue otherwise.

It is easy to show that this test is equivalent to the GSPRT $\left\{ e, \bigcup A_i \right\}$, where the A_i's have the shape we are presently considering.

The monotonicity theorems stated above (Theorems 22 through 24) for the case in which the components of the stopping region are defined by single straight lines can be given also for stopping regions with components of the type defined by eq. (45). In this case, we make use of the inequalities

(44) $$x_o^i Q_{ii}(x_o) \geq c_{ij} x_o^j Q_{ji}(x_o); \; i,j=1,2,3,$$

where $c_{ij}=1$ if $i=j$. These inequalities are proved as follows:

If $x_o \varepsilon A_j$, the inequalities follow immediately as in the proof of Theorem 21. If $x_o \varepsilon A_o$, then for any sample $x_1, x_2, \ldots, x_n$ such that $s_n \varepsilon A_i$, we have (by eq. (43))

$$x_o^i \prod_{t=1}^{n} x_t^i \geq c_{ij} x_o^j \prod_{t=1}^{n} x_t^j,$$

so that by eq. (12),

$$x_o^i q_{kii}(x_o) \geq c_{ij} x_o^j q_{kji}(x_o).$$

But this implies eq. (44).

In this case, the analogue of Theorems 22 and 23 can be stated as follows: Given two tests $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ such that $\cup A_i \subset \cup A_i'$, where the A_i's are defined as in eq. (43). Then $\sum_{\substack{i,j \\ i \neq j}} x_o^i \Delta Q_{ij}(x_o) \geq 0$ with strict inequality if the tests are not equivalent.

In order to prove this, we argue as in the proofs of Theorems 22 and 23. Note that expression (36) does not depend upon the shape of $\cup A_i$, and that the developement of the proof of Theorem 22 through expression (36) holds also in the present case. By inequalities (44) it follows that the integrands in expression (36) are nonnegative for all m, so that

$$\sum_{\substack{i,j \\ i \neq j}} x_o^i I_m(i,j,x_o) \geq 0. \tag{45}$$

The inequality in (45) is strict under the assumption of nonequivalence of the tests by the argument of Theorem 23. For example, if expression (39) is zero on $A_2'-A_2$ we obtain the inequalities

$$s_m^2 \geq c_{21} s_m' \text{ and } s_m^2 \geq c_{23} s_m^3$$

which implies that $s_m \varepsilon A_2$, a contradiction.

The remainder of the present chapter is devoted to stating a characterization of the condition under which enlarging all components of the stopping region leads to an increase in at least one of the ΔQ_{ij}'s, and to giving some implications of such an increase. We shall show that the general monotonicity property mentioned above is equivalent to

Condition I. Given the two tests $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ such that $\cup A_i \subset \cup A_i'$, there exist integers i, j, and m such that

$$\int \cdots \int \left[\prod_{k=o}^{m-1} \pi_o'(s_k)\right]\left[\delta_j(s_m) - \sum_t \delta_t(s_m)\right.$$

$$\left. Q_{ij}(s_m)\right] \prod_{q=1}^{m} (x_q^i d\nu) \neq 0.$$

Remark. Note that if Condition I is satisfied, then the tests $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ are not equivalent, if it is assumed that $x_o \varepsilon A'_o$. This can be seen by observing that if the integral of Condition I is not zero, then

$$\int \cdots \int [\prod_{k=o}^{m-1} \pi'_o(s_k)] \sum_t \delta_t(s_m) \prod_{v=1}^{m} (x_v^i dv) > 0$$

for some i and m, which is the condition of nonequivalence. Condition I is a condition on the way in which the regions $\cup A_i$ and $\cup A'_i$ differ. In particular, if $A_i \subset A'_i$ for one or two i's, nonequivalence of the tests implies Condition I. (In fact, this is a result of Theorems 19 and 20.) If, however, $A_i \subset A'_i$ for all i, Condition I has the following interpretation in terms of sample sequences in the space $\mathcal{Y}$: For some i and j, the event

$$\left\{ s_k \varepsilon A'_o \text{ for } k < m,\ \delta_j(s_m)=1,\ s_N \varepsilon \bigcup_{t \neq j} A_t \right\}$$

has probability under the hypothesis H_i different from the event

$$\left\{ s_k \varepsilon A'_o \text{ for } k < m,\ \sum_{t \neq j} \delta_t(s_m)=1,\ s_N \varepsilon A_j \right\}$$

for some stage $m < N$. (N is the sample size function of the test $\left\{ x_o, \cup A_i \right\}$.) Thus Condition I is the condition that the increase from $\cup A_i$ to $\cup A'_i$ is not "symmetric" in the above sense.

<u>Theorem 25</u>. Given the tests $\left\{ x_o, \cup A_i \right\}$ and $\left\{ x_o, \cup A'_i \right\}$, where $\cup A_i \subset \cup A'_i$ and $x_o \varepsilon A'_o$. Then for some i and j, $\Delta Q_{ij} > 0$ if and only if Condition I is satisfied.

Proof. Note that by eq. (10) we have

$$\Delta Q_{ij}(x_o)=\sum_{m=o}^{n-1} I_m(i,j,x_o)+\int\cdots\int [\prod_{m=o}^{n-1} \pi_o^{\nu}(s_m)]$$

$$\Delta Q_{ij}(s_n)\prod_{t=1}^{n}(x_t^i d\nu),$$

for all i and j, and for n=1,2,3, ..., where

$$I_m(i,j,x_o) = \begin{cases} \delta_j(x_o) - \sum_k \delta_k(x_o)Q_{ij}(x_o) & \text{for } m=0 \\ \int\cdots\int [\prod_{k=o}^{m-1} \pi_o^{\nu}(s_k)][\delta_j(s_m)-\sum_t \delta_t(s_m)Q_{ij}(s_m)] & \\ \prod_{v=1}^{m}(x_v^i d\nu), & \text{for } m=1,2, \ldots . \end{cases}$$

Since by Theorem 14

$$\int\cdots\int [\prod_{m=o}^{n-1} \pi_o^{\nu}(s_m)]\Delta Q_{ij}(s_n)\prod_{t=1}^{n}(x_t^i d\nu)$$

tends to zero as $n\to\infty$,

$$\sum_{m=o}^{\infty} I_m(i,j,x_o)=\Delta Q_{ij}(x_o).$$

Thus it follows that if $\Delta Q_{ij}(x_o)>0$, then $I_m(i,j,x_o)\neq 0$ for some i,j and m, which is Condition I.

On the other hand, if Condition I holds, then at some stage $\sum_{m=o}^{n-1} I_m(i,j,x_o)$ is bounded away from zero which implies that $\Delta Q_{ij}(x_o)$ is bounded away from zero for some i,j. Since

$$\sum_j \Delta Q_{ij}(x_o)=0,$$

this implies that $\Delta Q_{ij}(x_o)>0$ for some i and j.

Suppose we consider a subclass $\mathcal{T}$ of $\mathcal{S}$ such that the stopping regions of any two tests in $\mathcal{T}$ are "similar" in the following sense: For any two tests $\{x_o, \cup A_i\}$ and $\{x_o, \cup A_i'\}$ in $\mathcal{T}$, at least one of the inclusions

$$A_i \subset A_i', \quad A_i \supset A_i'$$

holds for all i. A test $T \varepsilon \mathcal{T}$ will be said to have the "optimum property in $\mathcal{T}$" if, for any other test $T^* \varepsilon \mathcal{T}$ such that Condition I holds, and such that $Q_{ij}(T^*) \leq Q_{ij}(T)$ for $i \neq j$,

$$N_i(T^*) \geq N_i(T)$$

for all i, with strict inequality for at least one i. A test $T \varepsilon \mathcal{T}$ will be called "unique in $\mathcal{T}$" if, among all other tests in $\mathcal{T}$ such that Condition I holds, T is the unique test with error probabilities $Q_{ij}(T)$.

Theorem 26. A GSPRT $\{x_o, \cup A_i\}$ with error probabilities $Q_{ij}(x_o)$ has the optimum property in the class $\mathcal{T}$ of GSPRT's with similar stopping regions.

Proof. Consider any GSPRT $\{x_o, \cup A_i'\}$ such that Condition I is satisfied, and such that the stopping regions $\cup A_i$ and $\cup A_i'$ are similar. Then by Theorems 19, 20, and 25, the condition that

$$Q_{ij}'(x_o) \leq Q_{ij}(x_o)$$

for $i \neq j$ implies that $\cup A_i' \subset \cup A_i$. But then by Theorem 18, $N_i'(x_o) \geq N_i(x_o)$ for all i, with strict inequality for at least one i.

Remark. It similarly follows that a GSPRT $\{x_o, \cup A_i\}$ is unique in the class of GSPRT's with similar stopping regions.

Chapter IV

AN EXAMPLE

As was noted above, Theorems 18, 19, and 20 of the preceeding chapter hold for any GSPRT. However, the monotonicity property for the case in which all three components of the stopping region are simultaneously enlarged seems to require additional assumptions. In this chapter we give an example of a three choice problem for which such a theorem does hold, with strict inequalities, under only the assumption of nonequivalence.

We shall consider the problem of deciding among the three uniform densities

$$f_i(y) = \begin{cases} 1/i, 0 \leq y \leq i \\ 0 \text{ otherwise} \end{cases} ; \; i=1,2,3.$$

The testing procedure we use can be specified by two integers t and m. We assume, for simplicity, that $t > m$. The test operates as follows: The test accepts H_1 when the first t observations fall in the interval [0,1]. It accepts H_2 if an observation falls in the interval [1,2] before t observations have been taken, and the remaining observations (at least m-1, but not more than t-1 in number) fall in the interval [0,2]. The procedure accepts H_3 otherwise, that is, when an observation falls in the interval [2,3] before it accepts H_1 or H_2.

In terms of a random walk in the space $\mathcal{X}$, this test is equivalent to one which starts at $x_0=e=(1/3,1/3,1/3)$. At the first stage, the induced distribution of X is given by the table

x	p_1	p_2	p_3
(6/11,3/11,2/11)	1	1/2	1/3
(0,3/5,2/5)	0	1/2	1/3
(0,0,1)	0	0	1/3

In general, states which the walk may occupy are of the following three types:

(a) $(6^n+3^n+2^n)^{-1}(6^n,3^n,2^n)$; $n = 0, 1, 2, \ldots, t$,

(b) $(3^n+2^n)^{-1}(0, 3^n, 2^n)$; $n = 1, 2, \ldots, m$,

(c) $(0, 0, 1)$.

Since the states that the walk may occupy with positive probability under some hypothesis are of the form of discrete points lying on curved paths intersecting the components A_i of the stopping region $\cup A_i$, the number of points (t-1 and m-1) lying outside these components determine the test, that is, the shape of each component does not affect the test, as long as they meet certain mild conditions. The shape we shall use is that considered in Chapter III in which A_i is given by the intersection of two lines in $\mathcal{X}$, each containing one of the opposite extreme points of $\mathcal{X}$ (figure 1). (We drop the condition imposed on these lines in Chapter III.) Bayes rules for testing the three uniform densities can be given in this form.

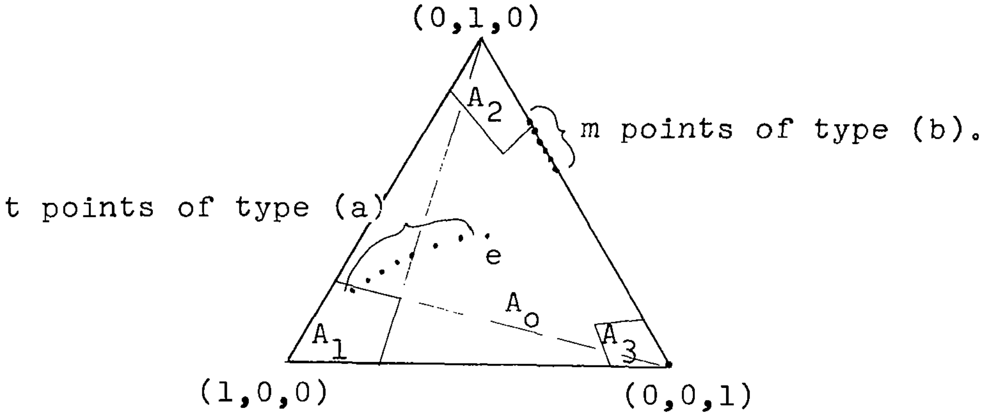

Figure 1

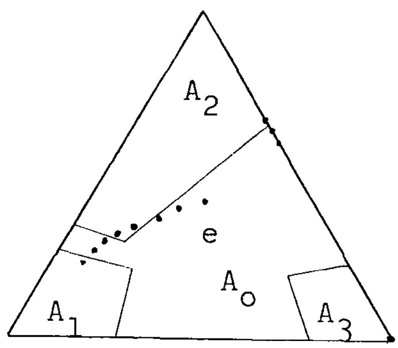

Figure 2

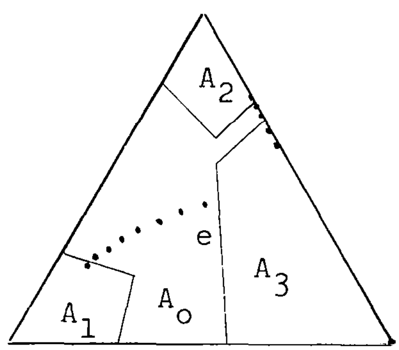

Figure 3

The components of the error probability matrix (Q_{ij}) for this test may be computed directly. They are:

$$Q_{11}=1,\ Q_{12}=Q_{13}=Q_{23}=0,$$

$$Q_{21}=(1/2)^t, Q_{22}=1-(1/2)^t, Q_{31}=(1/3)^t,$$

$$Q_{32}=(1/3)^m[2^m-1/2(1-(1/3)^{t-m})], Q_{33}=1-Q_{31}-Q_{32}.$$

Suppose we increase $\bigcup A_i$ to $\bigcup A_i'$, stopping region for a test defined by t' and m'. The new test may have a stopping region of the same type as the old (fig. 1), although it may not be true that $t'>m'$. There are also other possibilities, some of which are shown in figures 2 and 3. The error probability matrix (Q_{ij}') of the new test may also be computed directly for each of its possibilities. It is easily verified that in each case

$$tr(\Delta Q_{ij})=tr(Q_{ij}'-Q_{ij})<0,$$

provided that the tests are not equivalent, that is, provided that $t+m>t'+m'$.

Thus, in this example, any "reasonable" increase in $\bigcup A_i$ leads to a strict increase in at least one of the error probabilities in the set $\{Q_{21},Q_{31},Q_{32}\}$.

APPENDIX

Index of Terms and Notation

BIBLIOGRAPHY

1. Albert, G. E., On the computation of the sampling characteristics of a general class of sequential decision problems, Ann. Math. Stat. 25(1954), 340-356.

2. Armitage, P., Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis, Royal Stat. Soc. J. (ser. B) 12(1950), 137-144.

3. Arrow, K. J., Blackwell, D., and Girshick, M. A., Bayes and Minimax solutions of sequential decision problems, Econometrica 17 (1949), 213-244.

4. Blackwell, D., "Comparison of experiments", Proc. Second Berkeley Symp. Math. Stat. Probab., Univ. of California Press, Berkeley and Los Angeles (1951), 93-102.

5. Halmos, P. R., Measure theory, D. Van Nostrand Company, Inc., Princeton, 1950.

6. Seo, K., Invariant Bayes rules and sequential Bayes rules, Ph.D. Thesis (unpublished), Purdue University, 1962.

7. Skibinsky, M., Notes on the optimum character of the sequential probability ratio test, Special Report No. 8, Dept. of Stat., Univ. of Minn., Minneapolis, 1963.

8. Wald, A., Sequential analysis, John Wiley and Sons, Inc., New York, 1947.

9. Wald A., Statistical decision functions, John Wiley and Sons, Inc., New York, 1950.

10. Wald A., and Wolfowitz, J., Optimum character of the sequential probability ratio test, Ann Math. Stat. 19(1948), 326-339.

11. Wijsman, R. A., A monotonicity property of the sequential probability ratio test, Ann. Math. Stat. 31(1960), 677-684.