MOST ECONOMICAL MULTIPLE-DECISION RULES[1]

by

Wm. Jackson Hall
Institute of Statistics
University of North Carolina

Institute of Statistics
Mimeograph Series No. 115
August, 1954

1. This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

ACKNOWLEDGMENT

The author wishes to express his sincere thanks to Professor Wassily Hoeffding whose stimulating lectures inspired this research, and whose criticisms and suggestions have been invaluable in accomplishing it.

Thanks are also due the Office of Naval Research and the United States Air Force for their generous financial assistance, and to the U. S. International Information Administration for the Fulbright Grant which enabled the author to study abroad and to incorporate into this research many valuable suggestions of Mr. D. V. Lindley, to whom the author is gratefully indebted.

TABLE OF CONTENTS

Table of Contents (Continued)

INTRODUCTION

Most economical multiple-decision theory is primarily an extension of Professor Hoeffding's formulation of most economical two-decision theory, as given in his course lectures at Chapel Hill. The concept is an application of Wald's formulation of sequential analysis to non-sequential decision theory in which the choice of a sample size is at the disposal of the experimenter - instead of minimizing the expected sample size of a sequential experiment subject to bounds on the probabilities of error, Hoeffding suggests minimizing the fixed sample size of a non-sequential experiment subject to these bounds. Assuming the cost of experimentation to be proportional to the sample size, such procedures are, in a sense, "most economical". In terms of the now classical Neyman-Pearson theory of testing hypotheses, such decision rules adjust the sample size so as to obtain the desired power against certain alternatives; such an approach is implicit in the writings of Neyman and Pearson. In the words of Ferris, Grubbs, and Weaver [5], "... if the sampler wants a given degree of assurance in rejecting the null hypothesis when a particular alternative is true, he would like to know the minimum sample size which would accomplish this when the probability of rejecting the null hypothesis when true is given." For this purpose, these authors have supplied graphs of the power functions for various sample sizes for some of the common statistical tests. Hoeffding's lectures constitute a formalization, and generalization as well, of this approach to testing hypotheses and give, in addition, a number of existence theorems.

The multiple-decision extensions treated here are, first, to the consideration of decision rules for deciding among m alternatives, rather than just two, which minimize the sample size subject to bounds on the m probabilities of choosing the alternative which is to be preferred, or which, in a sense, is the "correct" alternative, under the prevailing one of m possible situations (true distributions or classes of distributions). A second extension is to decision rules which minimize the sample size subject to bounds on each of the probabilities (less than $\ell \cdot m$ in number) of making "incorrect" decisions - the probabilities of choosing one of the m alternatives which is not to be preferred when one of the ℓ possible situations prevails. Solutions to the first problem turn out to be "likelihood ratio" decision rules, and to the latter, "unlikelihood ratio" decision rules, defined precisely in the text. They are obtained by an application of Wald's minimax theory for fixed sample sizes; however, the minimax theory is used solely as a tool and does not lend itself to any interpretation according to Wald's theories. But it is an effective tool in that it is proved that one has nothing to lose by restricting his consideration solely to minimax solutions with respect to certain artificial loss functions for various sample sizes.

Problems of both "simple" and "composite" discrimination are considered, and characterization theorems and existence theorems are given. Various properties of the decision rules are derived, as well as relationships with works of Wald, Wolfowitz, Lindley, Rao, and others. A number of examples are treated which are generalizations of some of the common statistical tests. The theory as given is applicable to multivariate problems as well as univariate ones, and this generality enables some k-population problems - such as deciding which of several populations has the largest mean - to be covered also by considering a set of k observations, one from each univariate population, as one observation from a k-variate population.

Finally, utilizing the analogies with Wald's statistical decision functions, a generalization to a most economical theory of decision functions is considered in which the maximum expected cost is minimized subject to bounds on the expected loss function. This approach to statistical decision functions overcomes one of the common criticisms of Wald's theory, that losses due to incorrect decisions and cost of experimentation are treated on an equal footing, simply by summing them. The relationship with analogous work of Blyth and Konijn is pointed out.

Some of the theorems and their proofs, as well as some of the concepts, are fairly straightforward generalizations and adaptations of the work of Hoeffding, Wald, Lindley, and others, and indication to that effect is given wherever appropriate.

In Chapter I, the basic assumptions and definitions are given and problems of "simple discrimination" are treated - that is, problems in which one of a finite number of possible distributions underlies the decision problem. Chapter II treats problems of composite discrimination - discrimination among possible classes of distribution functions. Chapter III gives some very general necessary and sufficient conditions for the existence of most economical decision rules for both simple and composite discrimination. And Chapter IV treats generalizations of the foregoing theory to statistical decision functions, giving existence theorems as well as indicating various applications. An appendix treats some particular examples of two- and three- decision rules; a nomograph is given for obtaining such rules explicitly, and some brief tables of most economical sample sizes are computed from it.

NOTATION

We use X's to denote random variables and x's to denote observations on the corresponding random variables. X or x with a subscript denotes one real- or vector-valued variable, and without a subscript it denotes a sequence of such variables, usually n in number.

For clarity, we use "//" to denote the completion of a proof. Numbers in square brackets refer to bibliography. "Decision rule" is abbreviated "d.r." and "most economical", "M.E."; "with respect to" is abbreviated "w.r.t.".

For conciseness, any symbol, say s, when underlined denotes m (or sometimes ℓ) of the same with subscripts running from 1 to m (or ℓ); thus, $\underline{s}$ denotes $s_1, s_2, \ldots, s_m$. Throughout, m is a positive integer greater than unity. All other notation is introduced as needed.

CHAPTER I

MOST ECONOMICAL MULTIPLE-DECISION RULES FOR SIMPLE DISCRIMINATION

1.1 Formulation of the Problem.

1.1.1 Basic Assumptions and Definitions. We are concerned with a sequence $X_1, X_2, \ldots,$ of real- or vector-valued, independent, and identically distributed random variables, each having a generalized density function $f(x)$ w.r.t. a measure $\mu(x)$. $f(x)$ is not completely known, but is assumed to belong to some specified class Ω of density functions w.r.t. the specified measure μ.

We suppose we are faced with a number m of alternative decisions, $A_1, \ldots, A_m$, one of which is to be chosen after having taken a sample of size n, that is, after having observed the first n random variables $X_1, \ldots, X_n$, denoted simply X, n being at our disposal. We denote the sample values by $x = (x_1, \ldots, x_n)$. n is to be fixed in advance of experimentation and is to be completely non-random, not depending on the observations nor on any chance mechanism. We assume further that the cost of experimentation is proportional to the sample size. It would be necessary to assume only that the cost be some specified increasing function of the sample size by making minor alterations throughout, but we make the former assumption for expediency.

The decision procedure consists first in choosing a non-negative integer n, and then, after taking an observation on $X = (X_1, \ldots, X_n)$, in choosing one of the alternative decisions $A_1, \ldots, A_m$. The decision problem is to formulate a rule for choosing n, and having taken a sample of size n, for choosing among $A_1, \ldots, A_m$. A multiple-decision rule D (hereafter abbreviated m-d.r., or simply d.r., m denoting the number of alternatives available) for choosing among $A_1, \ldots, A_m$ after an observation x on X has been taken is defined by an ordered set of non-negative, real-valued, and measurable functions $\underline{\phi}(x) = [\phi_1(x), \ldots, \phi_m(x)]$ on the space $\mathcal{X}$ of x having the property $\sum_{i=1}^{m} \phi_i(x) = 1$ identically in x (for n = 0, D is defined by a set of non-negative constants $\underline{\phi} = (\phi_1, \ldots, \phi_m)$ summing to unity). D is interpreted as follows: having taken a sample of size n and computed $\underline{\phi}(x)$, a chance mechanism is used to select one of the alternatives $A_1, \ldots, A_m$, the probability of choosing A_i being $\phi_i(x)$ $(i = 1, \ldots, m)$ when x is observed. Such a d.r. is said to be "randomized" d.r. A d.r. is said to be "non-randomized" if the values of the ϕ_i's are restricted to 0 and 1; then each ϕ_i

is a characteristic (or "indicator") function of a set, commonly called a region, say R_i, of the sample space $\mathcal{X}$ and the property $\Sigma\phi_i = 1$ implies that $R_1, \ldots, R_m$ are mutually exclusive and exhaustive. We call such regions "acceptance regions" since a non-randomized d.r. is of the type: if $x \varepsilon R_i$, accept A_i ($i = 1, \ldots, m$). For every d.r. D, a set of functions $\underline{\phi}$ is implicitly assumed, and we shall refer to a d.r. either by D or by the set of functions $\underline{\phi}$. A subscript n on D is sometimes used to denote the corresponding sample size; $\underline{\phi}^n$ denotes the set of functions defining D_n. Unless otherwise specified, we assume the class of d.r.'s at our disposal consists of all possible d.r.'s as defined above for every possible fixed sample size $n = 0, 1, 2, \ldots$.

1.1.2 Problems of Simple Discrimination. We suppose throughout this chapter that $\underline{\Omega}$ consists of a finite number, say ℓ, of elements $f_1, \ldots, f_\ell$, and we denote $\underline{\Omega} = \underline{f} = (f_1, \ldots, f_\ell)$; when this is the case, we say that the corresponding decision problem is one of "simple discrimination" and a d.r. is a d.r. for "simple discrimination" or for "discriminating among $\underline{f}$". We use the same notation to denote the joint density functions of the first n random variables in the sequence, sometimes using a superscript n to denote the sample size when necessary to avoid confusion; similarly, the measure μ may be used to denote a joint measure of n random variables -- that is, we denote $\prod_{i=1}^{n} f_j(x_i) = f_j^n(x) = f_j(x)$, where x without a subscript denotes an n-vector, and we say $f_j(x)$ is a density function w.r.t. $\mu(x)$ $(j = 1, \ldots, m)$.

A d.r. D will be characterized by the functions

$$(1.1) \quad p_{ij}(D) = \Pr(D \text{ chooses } A_j \mid f_i) = \int_{\mathcal{X}} \phi_j(x)\, f_i(x)\, d\mu = E_i \phi_j(X)$$

$(i = 1, \ldots, \ell;\ j = 1, \ldots, m)$ where the subscript on the expectation operator E denotes the corresponding density function. If D is a non-randomized d.r., (1.1) implies

$$p_{ij}(D) = \int_{R_j} f_i(x)\, d\mu \ .$$

We shall consider two different criteria for choosing a d.r. for simple discrimination. The first criterion is applicable only to the case when the number ℓ of elements in $\underline{\Omega}$ is equal to the number m of alternatives and the alternative decisions correspond to the possible true density functions in such a way that the decision A_i is to be preferred when f_i is true $(i = 1, \ldots, m)$;

A_i is said to be a "correct" decision if f_i is true and "incorrect" if f_j is true ($j \neq i$). We denote $p_{ii}(D) = p_i(D) = 1-q_i(D)$ ($i = 1, \ldots, m$), so that when the d.r. D is used and f_i is true, p_i is the probability of a correct decision and q_i the probability of an incorrect decision. We now formulate the first criterion:

DEFINITION 1.1: Let $\underline{\alpha} = (\alpha_1, \ldots, \alpha_m)$ be a given vector of positive[1] constants each less than one. A d.r. D_N, based on a sample of size N, is said to be a most economical m-decision rule relative to the vector $\underline{\alpha}$ for discriminating among $\underline{f} = (f_1, \ldots, f_m)$ if it satisfies

$$(1.2) \qquad p_i(D) \geq \alpha_i \qquad (i = 1, \ldots, m)$$

and if N is the least integer n for which (1.2) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the most economical sample size.

Thus, by this definition (and assuming the sampling cost to be proportional to the sample size), a most economical d.r. (hereafter abbreviated M.E. d.r.) is one with a minimum sampling cost subject

1. If any α_i were zero, we may just as well never choose A_i and consider an (m-1)-d.r. for discriminating among the m-1 f_j's ($j \neq i$).

to lower bounds on the m probabilities of correct decisions. We note that (1.2) implies lower bounds on the probabilities of an incorrect decision, vis., $q_i(D) \leq 1 - \alpha_i$ $(i = 1, \ldots, m)$.

A second criterion for choosing a d.r. will now be given. We no longer require that $\ell = m$, but shall suppose that corresponding to each f_i one or more of the alternatives A_j is preferable (or "correct") when f_i is true.

DEFINITION 1.2: Let $\beta = (\beta_{ij})$ be a given $\ell \times m$ matrix of positive constants such that for every i, j pair for which A_j is a correct decision when f_i is true $\beta_{ij} = 1$. A d.r. D_N, based on a sample of size N, is said to be a most economical m-decision rule relative to the matrix β for discriminating among $\underline{f} = (f_1, \ldots, f_\ell)$ if it satisfies

$$p_{ij}(D) \leq \beta_{ij} \quad (i = 1, \ldots, \ell;\ j = 1, \ldots, m) \tag{1.3}$$

and if N is the least integer n for which (1.3) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the most economical sample size.

Thus, by this definition, a M.E. d.r. is one with a minimum sample size subject to an upper bound on each of the probabilities of incorrect decisions. If we wish to relinquish control of the probability of any particular incorrect decision, we simply set the corresponding $\beta_{ij} = 1$. If $\ell = m$ and A_i is preferred when f_i is true ($i = 1, \ldots, m$), then a M.E. d.r. relative to β also controls the probabilities of correct decisions if $\sum_{j \neq i} \beta_{ij} < 1$ since (1.3) implies

$$p_{ii}(D) = 1 - \sum_{j \neq i} p_{ij}(D) \geq 1 - \sum_{j \neq i} \beta_{ij} \quad (i = 1, \ldots, m).$$

For some applications, one may wish to restrict the class of d.r.'s to the class of all non-randomized d.r.'s and define M.E. non-randomized d.r.'s relative to a vector or matrix by the same definitions, but this may require an increased sample size. Alternatively, one may wish to allow a random choice between two (or more) consecutive sample sizes, and this may decrease the average sample size over a number of experiments. (Hoeffding [8] has considered such procedures in the two-decision case.) However, we shall assume throughout that the sample size is non-random, but shall allow randomized d.r.'s unless otherwise specified.

In the case of 2-d.r.'s when $\ell = m = 2$, we have $p_{12} = 1-p_{11}$ and $p_{21} = 1-p_{22}$. Suppose A_1 is the decision to accept the hypothesis H_1: $f = f_1$ and A_2 is the decision to reject H_1 in favor of H_2: $f = f_2$. Then p_{12} and p_{21} are the classical first and second kinds of error, respectively, associated with a test of H_1 against H_2, and, denoting $\alpha = 1 - \alpha_1 = \beta_{12}$ and $\beta = 1 - \alpha_2 = \beta_{21}$, it follows that both (1.2) and (1.3) reduce to upper bounds (α, β) on the two kinds of error. Hence, Definitions 1.1 and 1.2 are equivalent (if $\ell = m = 2$), both defining a M.E. 2-d.r. as one with minimum sample size subject to these bounds. Consider for each sample size n, the most powerful test T_n of size α of H_1 against H_2. It may be shown that if for some n the power of T_n is at least $1 - \beta$, then the test T_N, where N is the least n for which this is true, is "most economical". Hoeffding [8] has considered such two-decision problems in some detail. Definitions 1.1 and 1.2 are two extensions of these concepts to multiple-decision problems.

In the following sections we shall be concerned with deriving both types of M.E. m-d.r.'s for simple discrimination, as defined by Definitions 1.1 and 1.2. We shall consider minimax d.r.'s w.r.t. certain weight functions relative to the class of all d.r.'s based on a sample of fixed size n and shall prove that in order to obtain M.E. d.r.'s we need only consider minimax d.r.'s for various values of n. The question of "admissibility" of M.E. d.r.'s obtained in this way is also considered. It should be emphasized that we are using minimax theory only as a tool; the loss functions introduced are somewhat artificial. In Section 1.3.3 we shall consider another approach to obtaining

M.E. d.r.'s, but shall show that it is virtually equivalent to the minimax approach. First, in the next section, we shall review Wald's theory of minimax d.r.'s for fixed sample sizes when the number of possible decisions and the number of possible distributions are both finite.

1.2 Minimax Decision Rules for Fixed Sample Sizes. We shall review briefly some of the concepts introduced by Wald in [22], Section 1.1, as applied in Section 5.1.1, and shall repeat some of his theorems of Sections 3.5 and 5.1.1, altering his notation slightly to conform to the notation introduced in the previous section. There are only minor differences between the "data of the decision problem" as assumed by Wald and here: he treats distribution functions rather than generalized density functions w.r.t. some measure, and he assumes univariate random variables rather than allowing the possibility of multivariate random variables, but these and other differences are inessential in what follows. We assume throughout that the sample size n has been fixed.

We assume a bounded weight function $W(f_i, A_j) = W_{ij}$ $(i = 1, \ldots, \ell;\ j = 1, \ldots, m)$, representing the 'loss' incurred by accepting A_j when f_i is true. The corresponding risk function when using a d.r. D is defined as the expected loss:

$$(1.4) \qquad r(f_i, D) = \sum_{j=1}^{m} W_{ij}p_{ij}(D).$$

We introduce an a priori distribution $\underline{\xi} = (\xi_1, \ldots, \xi_\ell)$ over $\underline{\Omega}$ where $\xi_i \geq 0$ and $\sum_{i=1}^{\ell} \xi_i = 1$; ξ_i represents the 'probability that f_i is true'. The average risk relative to $\underline{\xi}$ is defined:

$$(1.5) \qquad r(\underline{\xi}, D) = \sum_{i=1}^{m} \xi_i r(f_i, D) = \sum_{i,j} \xi_i W_{ij} p_{ij}(D).$$

A d.r. D^* is said to be a Bayes d.r. relative to $\underline{\xi}$ if it minimizes the average risk relative to $\underline{\xi}$; i.e., if $r(\underline{\xi}, D^*) = \inf_D r(\underline{\xi}, D)$. A d.r. D^0 is said to be a minimax d.r. if it minimizes the maximum risk over $\underline{\Omega}$; i.e., if $\max_{1 \leq i \leq \ell} r(f_i, D^0) = \inf_D \max_{1 \leq i \leq \ell} r(f_i, D)$. An a priori distribution $\underline{\xi}^0$ is said to be a least favorable distribution if it maximizes w.r.t. $\underline{\xi}$ the minimum of the average risk; i.e., if $\inf_D r(\underline{\xi}^0, D) = \sup_{\underline{\xi}} \inf_D r(\underline{\xi}, D)$.

The following two theorems are given by Wald ([22], parts of Theorems 5.1, 5.3, and 3.9):

THEOREM 1.1: A necessary and sufficient condition for a d.r. D to be a Bayes d.r. relative to a given a priori distribution $\underline{\xi}$ is that $\phi_j(x) = 0$ for any x (except perhaps in a set of $\underline{\xi}$-measure zero[2]) and for any j for which

$$\sum_{i=1}^{\ell} \xi_i W_{ij} f_i(x) > \min_{1 \leq k \leq m} \left[\sum_{i=1}^{\ell} \xi_i W_{ik} f_i(x) \right] .$$

THEOREM 1.2: (i) There exists a minimax d.r. D^0;

(ii) there exists a least favorable distribution $\underline{\xi}^0$;

(iii) any minimax d.r. is a Bayes d.r. relative to any least favorable distribution, and conversely;

(iv) for any i for which $\xi_i^0 > 0$, $r(f_i, D^0)$ $= \max_{i \leq j \leq \ell} r(f_j, D^0)$.

ASSUMPTION 1.1: The measure μ is non-atomic.[3]

---

2. By the ξ-measure of a subset R of $\mathcal{X}$ we mean $\sum_{i=1}^{\ell} \xi_i \int_R f_i(x)\, d\mu$.

3. A measure is non-atomic if every set of non-zero measure has a subset of different non-zero measure; e.g., Lebesgue measure is non-atomic.

Dvoretzky, Wald, and Wolfowitz [4] have proved under Assumption 1.1 that any randomized d.r. for simple discrimination may be replaced by an "equivalent" non-randomized d.r., "equivalent" implying equality of risk functions.

We conclude this section with a lemma which will be useful later on. Conditions under which r_n, defined below, tends to zero are given in Chapter III.

LEMMA 1.1: For every fixed sample size $n = 0, 1, 2, \ldots,$ let D_n^0 be a minimax d.r. and denote $r_n = \max_i r(f_i, D_n^0)$. Then the sequence $\{r_n\}$, $n = 0, 1, 2, \ldots,$ is a non-increasing sequence.

Proof: Let $\mathcal{D}_n$ denote the class of all m-d.r.'s based on a sample of fixed size n ($n = 0, 1, 2, \ldots$). Now $\mathcal{D}_n \subseteq \mathcal{D}_N$ for all $n \leq N$ ($N = 0, 1, 2, \ldots$) since we may take $\phi(x_1, \ldots, X_N)$ independent of $x_{n+1}, \ldots, x_N$, in which case ϕ defines a d.r. in $\mathcal{D}_n$ and in $\mathcal{D}_N$. Hence

$$r_n = \inf_{D \in \mathcal{D}_n} \max_i r(f_i, D) \geq \inf_{D \in \mathcal{D}_N} \max_i r(f_i, D) = r_N \quad (n \leq N). \ //$$

1.3 Most Economical Decision Rules Relative to a Vector $\underline{\alpha}$. We shall apply the theory of Section 1.2 to two specific weight functions W_{ij} and develop in each case a method of obtaining M.E. d.r.'s as defined by Definition 1.1. In both cases we assume $\ell = m$. Various properties of these decision rules are considered and some examples indicated.

1.3.1 First Minimax Approach. Let

$$W(f_i, A_j) = W_{ij} = -\delta_{ij}/\alpha_i \quad (i, j = 1, \ldots, m) \tag{1.6}$$

where δ_{ij} denotes the Kroneker δ-function.[4] Then, from (1.4),

---

4. The loss function (1.6) may be interpreted in terms of gains. One might develop a decision theory completely analogous to Wald's theory but with Wald's loss function replaced by a gain function which is positive when a "correct" decision is made and zero otherwise. A Bayes d.r. might be defined as one which maximizes the average expected gain, and a "maximin" d.r. as one which maximizes the minimum expected gain. But such a theory would be completely equivalent to the Wald theory if one defines loss as negative gain, or, if one prefers a non-negative loss function, if one defines loss as the difference between the maximum possible gain (maximum over all possible distributions as well as over all possible decisions) and the actual gain. (Savage [19] considered Wald's loss functions in terms of gains, defining loss as the difference between the maximum gain for the prevailing true distribution and the actual gain, so that his loss is always zero when a correct decision is made and positive otherwise.) Either of these loss functions satisfies Wald's requirements (if all gains are finite) though it is not necessarily zero when a correct decision is made nor necessarily positive otherwise, as suggested -- but never required mathematically -- by Wald. Thus, we may consider $G_{ij} = \delta_{ij}/\alpha_i$ as a gain function and define the loss function in (1.6) by $W_{ij} = -G_{ij}$, or, alternatively, we may replace (1.6) by $W_{ij} = (\max_{i,j} G_{ij}) - G_{ij}$.

the risk w.r.t. W_{ij} is

$$(1.7) \quad r(f_i, D) = \sum_{j=1}^{m} \delta_{ij} p_{ij}(D)/\alpha_i = -p_i(D)/\alpha_i \quad (i = 1, \ldots, m).$$

Theorem 1.2 asserts the existence of a minimax d.r. D^0 for any (fixed) sample size. We shall consider minimax d.r.'s for each value of the sample size $n = 0, 1, 2, \ldots$, and prove the following theorem:

THEOREM 1.3: For each $n = 0, 1, 2, \ldots$, let D_n^0 be a minimax d.r. w.r.t. the weight function (1.6) for samples of fixed size n. Suppose for some n,

$$(1.8) \qquad \max_i r(f_i, D_n^0) \leq -1$$

and let N be the least such integer. Then D_N^0 is a M.E. d.r. relative to the vector $\underline{\alpha}$ for discriminating among $\underline{f}$. Conversely, if there exists a M.E. d.r. relative to $\underline{\alpha}$ for discriminating among $\underline{f}$ and the M.E. sample size is N, then D_N^0 is a M.E. d.r.

Proof: From (1.7) and (1.8), it follows that D_N^0 satisfies (1.2). Now suppose for some $n < N$, there exists a d.r. D_n satisfying (1.2). But D_n^0 is a minimax d.r. so that

$$(1.9) \qquad \max_i r(f_i, D_n^0) \leq \max_i r(f_i, D_n) = \max_i \left[-p_i(D_n)/\alpha_i \right] .$$

Since D_n satisfies (1.2), we have from (1.9) that $\max_i r(f_i, D_n^0) \leq -1$, in contradiction to the fact that N is the least integer n for which this is true. Hence, D_N^0 is a M.E. d.r.

To prove the converse, suppose D_N is a M.E. d.r. Then $-1 \geq \max_i \left[-p_i(D_N)/\alpha_i \right] = \max_i r(f_i, D_N) \geq \max_i r(f_i, D_N^0)$ since D_N^0 is a minimax d.r. Hence, (1.8) is satisfied for n = N, and since N is the M.E. sample size, D_N^0 is a M.E. d.r. //

Hence, to obtain M.E. d.r.'s relative to $\underline{\alpha}$, we need consider only minimax d.r.'s for various values of n. Lemma 1.1 should prove helpful in finding the M.E. sample size since it assures us that any n for which (1.8) is violated is too small. Now let us consider the structure of minimax d.r.'s for a fixed sample size n.

Let $\underline{\xi}$ be an a priori distribution. Now

$$\sum_{i=1}^{m} \xi_i W_{ij} f_i(x) = -\xi_j f_j(x)/\alpha_j \qquad (j = 1, \ldots, m).$$

Hence, by Theorem 1.1, a necessary and sufficient condition for D to be a Bayes d.r. relative to $\underline{\xi}$ is that for any x (except perhaps in a set of $\underline{\xi}$-measure zero) and any j for which $\phi_j(x) > 0$, we have

$$\xi_j f_j/\alpha_j = \max_{1\le i\le m} \xi_i f_i/\alpha_i \;. \tag{1.10}$$

DEFINITION 1.3: A d.r. D defined by $\underline{\phi}(x)$ is said to be a likelihood ratio d.r. if there exist positive constants $a_1, \ldots, a_m$ such that for any j and any x for which $\phi_j(x) > 0$, $a_j f_j(x) \geq a_i f_i(x)$ for all $i \neq j$.

(Note that $a_1, \ldots, a_m$ determine $\underline{\phi}$ completely except in sets of x for which $a_i f_i(x) = \max_j a_j f_j(x)$ for more than one value of i.) Setting $a_j = \xi_j/\alpha_j$, it follows from (1.10) that a Bayes d.r. relative to any $\underline{\xi}$ for which all $\xi_i > 0$ is a likelihood ratio d.r., and conversely.

Moreover, it follows from (1.7) and Theorem 1.2 (iv) that if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$p_1(D^0)/\alpha_1 = \ldots = p_m(D^0)/\alpha_m \;. \tag{1.11}$$

We shall give sufficient conditions for this to be true.

ASSUMPTION 1.2: If R is a subset of $\mathcal{X}$ for which $\int_R f_i(x)\, d\mu = 0$

for some i, then $\int_R f_i(x)\, d\mu = 0$ for all values of i. (Whenever this assumption is made, we shall tacitly assume that $\mathcal{X}$ is redefined so as to exclude all such sets R; this implies that $f_i(x) > 0$ for all i and for all $x \in \mathcal{X}$.)

We prove a theorem analogous to Wald's Theorem 5.4 [22][5]; the proof is also analogous.

THEOREM 1.4: If Assumption 1.2 holds, all components of a least favorable distribution ξ^0 w.r.t. the weight function (1.6) are positive.

Proof: Consider the d.r. D defined by $\phi_i(x) = \alpha_i / \sum_{j=1}^{m} \alpha_j$ identically in x (i = 1, ..., m); and let D^0 be a minimax d.r. Then

$$(1.12)\quad \max_j r(f_j, D^0) = \inf_D \max_j r(f_j, D) \leq \max_j r(f_j, D) = \max_j(-\alpha_j/\Sigma_i \alpha_i)$$

$$< 0.$$

5. It might be noted that Wald's condition (iii) of Theorem 5.4 is superfluous since it is always fulfilled; e.g., in Wald's notation, let $\delta_i = 1/u$ (i = 1, ..., u) identically in x, and then $r(F_j, \delta) = (u - 1)/u < 1$ for j = 1, ..., k.

Now suppose for some j, $\xi_j^0 = 0$. Then, under Assumption 1.2,

$$(1.13) \qquad 0 = \xi_j^0 f_j(x)/\alpha_j < \max_i \xi_i^0 f_i(x)/\alpha_i \qquad \text{for all } x.$$

But D^0 is a Bayes d.r. relative to $\underline{\xi}^0$ (by Theorem 1.2) so that (1.10) and (1.13) imply $\phi_j^0(x) = 0$ identically in x. Hence, $p_j(D^0) = 0$ and $r(f_j, D^0) = 0$, in contradiction to (1.12). Hence, $\xi_j^0 > 0$ for all j. //

1.3.2 Second Minimax Approach. We consider a second weight function W_{ij}:

$$(1.14) \qquad W(f_i, A_j) = W_{ij} = (1 - \delta_{ij})/(1 - \alpha_j) \qquad (i,j = 1,\ldots,m).$$

Let $1 - \alpha_j = \beta_j$. Then

$$(1.15) \qquad r(f_i, D) = q_i(D)/\beta_i \qquad (i = 1, \ldots, m).$$

Theorem 1.2 asserts the existence of a minimax d.r. Proceeding as with the first weight function, we state the following theorem, the proof of which is not given since it is analogous to the proof of Theorem 1.3.

THEOREM 1.5: For each $n = 0, 1, 2, \ldots$, let D_n^0 be a minimax d.r. w.r.t. the weight function (1.14) for samples of fixed size n. Suppose for some n, $\max_i r(f_i, D_n^0) \leq 1$, and let N be the least such integer. Then D_N^0 is a M.E. d.r. relative to the vector $\underline{\alpha}$ for discriminating among $\underline{f}$. Conversely, if there exists a M.E. d.r. relative to $\underline{\alpha}$ and N is the M.E. sample size, then D_N^0 is a M.E. d.r.

Thus, we have a second method of obtaining M.E. d.r.'s (if existent) from minimax d.r.'s.

Now let $\underline{\xi}$ be an a priori distribution. Then

$$\sum_{i=1}^{m} \xi_i W_{ij} f_i(x) = \sum_{i=1}^{m} \xi_i f_i(x)/\beta_i - \xi_j f_j(x)/\beta_j \quad (j = 1, \ldots, m).$$

By an argument analogous to the one used with the first weight function, we have: a necessary and sufficient condition for D to be a Bayes d.r. relative to a given a priori distribution $\underline{\xi}$ is that for any x (except perhaps in a set of $\underline{\xi}$-measure zero) and for any j for which $\phi_j(x) > 0$, $b_j f_j(x) \geq b_i f_i(x)$ $(i = 1, \ldots, m)$ where $b_j = \xi_j/\beta_j$ $(j = 1, \ldots, m)$. Hence, Bayes d.r.'s w.r.t. the weight function (1.14) relative to any $\underline{\xi}$ for which all $\xi_i > 0$ are also likelihood ratio d.r.'s, and conversely.

Moreover, it follows from (1.15) and Theorem 1.2 (iv) that if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$(1.16) \qquad q_1(D^0)/\beta_1 = \ldots = q_m(D^0)/\beta_m \; .$$

We shall give sufficient conditions for this to be true.

LEMMA 1.2: If Assumption 1.2 holds, and if there exists some d.r. D for which $r(f_i, D) < 1/\max_{1\le j\le m} \beta_j$ $(i = 1, \ldots, m)$, then all components of a least favorable distribution are positive.

Proof: Let D^0 be a minimax d.r.; then $\max_j r(f_j, D^0) \le \max_j r(f_j, D) < 1/\max_j \beta_j$ so that

$$(1.17) \qquad r(f_j, D^0) < 1/\beta_j \qquad \text{for all } j.$$

Now suppose for some j, $\xi_j^0 = 0$. By an argument analogous to the one in the proof of Theorem 1.4, this implies $\phi_j^0(x) = 0$ identically in x. Hence, $q_j(D^0) = 1$ and $r(f_j, D^0) = 1/\beta_j$, in contradiction to (1.17). //

LEMMA 1.3: If $\beta_i < \frac{1}{m-1} \sum_{j=1}^{m} \beta_j$ (i.e., $\alpha_i > \frac{\Sigma\alpha_j - 1}{m-1}$) for all i, then there exists a d.r. D for which $r(f_i, D) < 1/\max_j \beta_j$ for all i.

Proof: Let D be defined by $\phi_i(x) = 1-(m-1)\beta_i/\Sigma\beta_j > 0$ identically in x ($i = 1, \ldots, m$). Then $r(f_i, D) = q_i(D)/\beta_i = (m-1)/\Sigma\beta_j < 1/\beta_j$ for all j ($i = 1, \ldots, m$). //

LEMMA 1.4: Suppose Assumption 1.2 holds. If $\xi_j^0 > 0$ for some j for which $\beta_j > \beta_i$, then $\xi_i^0 > 0$.

Proof: In the proof of Lemma 1.2, we found that $\xi_i^0 = 0$ implies $r(f_i, D^0) = 1/\beta_i$; hence, we need only prove that $r(f_i, D^0) < 1/\beta_i$. By Theorem 1.2 (iv), we have $\max r(f_i, D^0) = r(f_j, D^0) = q_j(D^0)/\beta_j \leq 1/\beta_j < 1/\beta_i$. //

THEOREM 1.6: Suppose Assumption 1.2 holds. For any sample size greater than or equal to the M.E. sample size, all components of a least favorable distribution are positive.

Proof: Suppose $n \geq N$, the M.E. sample size, and that D_n^0 is a minimax d.r., for samples of size n; then, using Lemma 1.1 and Theorem 1.5, D_n^0 satisfies (1.2). This implies $q_i(D_n^0) \leq \beta_i$ and $r(f_i, D_n^0) = q_i(D_n^0)/\beta_i \leq 1 < 1/\max_j \beta_j$. Lemma 1.2 completes the proof of the theorem. //

Hence, under the conditions of the theorem, D_N^0 is a likelihood ratio d.r. A possible advantage in using this second weight function in order to obtain M.E. d.r.'s is that if for some n one of the components of a least favorable distribution is zero, we know immediately that n is less than the M.E. sample size. (It cannot be greater by Lemma 1.1.)

1.3.3 An Alternative Approach. It was shown in the previous sections that, under Assumption 1.2, a M.E. d.r. relative to $\underline{\alpha}$ obtained by one of the minimax methods satisfies (1.11) or (1.16). This suggests an alternative approach for obtaining these d.r.'s which is described below, the proofs of the various statements being indicated but not always given in detail. This method gives some geometric insight into the properties of the d.r.'s thus obtained.

Assume the sample size n is fixed, and, given $\underline{\alpha}$, define two classes of d.r.'s $C = C(\underline{\alpha})$ and $I = I(\underline{\beta})$ as follows:

$$C = \{ D : p_1(D)/\alpha_1 = \dots = p_m(D)/\alpha_m \}$$

$$I = \{ D : q_1(D)/\beta_1 = \dots = q_m(D)/\beta_m \}$$

where $\beta_i = 1 - \alpha_i$ $(i = 1, \dots, m)$ as before; that is, C is the class of d.r.'s for which the m probabilities of a correct decision are in assigned ratios and I is the class of d.r.'s for which the m probabilities of an incorrect decision are in assigned ratios. A d.r. is said to be optimum in C (or I) if it maximizes over all d.r.'s in C (or I) all probabilities of correct decisions; that is, if it maximizes the common ratio $p_i(D)/\alpha_i$ (or minimizes the common ratio $q_i(D)/\beta_i$). Clearly, optimum d.r.'s in C and I, when existent, are minimax d.r.'s w.r.t. the weight functions (1.6) and (1.14), respectively.

Rao ([17], page 311) has considered the class I of d.r.'s for problems of classification in multivariate analysis, in which case the sample consists of one observation from a multivariate population. This is no restriction, of course, since a sample of size n from a p-variate population may be treated as a sample of size 1 from an np-variate population. Rao gives sufficient conditions for a d.r. to be optimum in I; his conditions are that if there exists a likelihood ratio d.r. in I, then it is optimum in I. However, his conditions cannot always be fulfilled since I may be a null class (if $m > 2$). Heuristically, this may be seen as follows: suppose one β_i, say β_1, is very close to unity and all others are very close to zero; then for any d.r. in I, the ratio $(1 - p_i)/(1 - p_1) = \beta_i/\beta_1$ must be very small for all $i > 1$; but even for a d.r. for which $p_1 = 0$ it may not be possible to make all the other p_i's sufficiently large for this to hold. If this is the case, the obvious thing to do is to always reject A_1 and consider $(m - 1)$-d.r.'s in the corresponding class I.

A theorem completely analogous to Rao's theorem may be proved for the class C, giving the same condition as sufficient characterization of an optimum d.r. in C; and we shall see that C is never null. It will be shown by a geometric argument that similar conditions are also necessary for a d.r. to be optimum. Much of this development was suggested by similar arguments of Lindley [16].

Consider the m-dimensional Euclidean space of points $\underline{p} = (p_1, \ldots, p_m)$; every d.r. D has a corresponding point $\underline{p}$ with $p_i = p_i(D)$. Let P denote the set of points corresponding to all possible d.r.'s (for a fixed sample size), and let C' and I' denote sub-sets of P corresponding to C and I respectively. It may be proved (e.g., as a corollary to Lindley's Theorem 1.1 [16] since P is a subspace of the space considered by him, or as a special case of a theorem of Dvoretsky, Wald, and Wolfowitz (see footnote 2 in [24]) for the weight function (1.6) or (1.14)) that P is a convex body; i.e., P is closed, bounded, and convex. Clearly, P is enclosed in the unit cube U, and all points with one coordinate unity and the other m - 1 coordinates zero are points in P. Since P is convex, the intersection of U with the hyperplane $\Sigma p_i = 1$ determined by these m points is contained in P; denote this "flat" sub-space of P by P_0 [6].

The equations $p_1/\alpha_1 = \ldots = p_m/\alpha_m$ determine a line L_C passing through the interior of U (since all $\alpha_i > 0$) and passing through the origin, and C' is the segment of this line intersecting P; such a segment exists since L_C must intersect P_0. Since P is a convex body, C' is closed and the endpoint furthest from the origin, say $\underline{p}^C$, corresponds to any optimum d.r. in C. Let $a_1, \ldots, a_m$ be constants (at least one of which is positive) determining a supporting hyperplane to P at $\underline{p}^C$; i.e., $\Sigma a_i p_i = \Sigma a_i p_i^C$, and for any $\underline{p} \in P$, $\Sigma a_i p_i \leq \Sigma a_i p_i^C$, or

$$(1.18) \qquad \sum_{i=1}^{m} a_i \int_{\mathcal{X}} \phi_i(x) f_i(x) \, d\mu \leq \sum_{i=1}^{m} a_i \int_{\mathcal{X}} \phi_i^C(x) f_i(x) \, d\mu$$

6. It may be pointed out that all points on and below P_0 (that is, on the same side of P_0 as the origin) correspond to "trivial" d.r.'s and all points above P_0 to "non-trivial" ones, as defined in Chapter III.

where $\underline{\phi}^C$ defines an optimum d.r. in C and $\underline{\phi}$ defines an arbitrary d.r. Let Δ be a "small" subset of $\mathfrak{X}$ such that for a "small" positive ε, $\phi_1^C(x) > \varepsilon$ for all $x \in \Delta$ and $\int_\Delta f_1(x)\, d\mu > 0$. That such a set exists, at least for small ε, is guaranteed since $p_1^C > 0$. For some $j > 1$, define $\underline{\phi}(x)$ as follows:

$$\phi_i(x) = \begin{cases} \phi_1^C(x) - \varepsilon & \text{if } x \in \Delta \text{ for } i = 1 \\ \phi_j^C(x) + \varepsilon & \text{if } x \in \Delta \text{ for } i = j \\ \phi_i^C(x) & \text{if } x \notin \Delta \text{ for } i = 1,\ j \text{ and for all } x \text{ for} \\ & i \neq 1,\ j. \end{cases}$$

Then (1.18) implies

$$a_1 \int_\Delta f_1(x)\, d\mu \geq a_j \int_\Delta f_j(x)\, d\mu ,$$

which is true for any j, so that by taking Δ and ε sufficiently small we have that $a_1 f_1(x) \geq a_j f_j(x)$ $(j = 1, \ldots, m)$ for all x for which $\phi_1^C(x) > 0$ and $f_1(x) > 0$. Moreover, since at least one a_i is positive and $p_1^C > 0$, a_1 must be non-negative. A similar relation holds for $i > 1$ as well as for $i = 1$, so that any d.r. D for which $\underline{p}(D) = \underline{p}^C$ is of the likelihood ratio form with non-negative constants $a_1, \ldots, a_m$, which are not in general necessarily all positive. But clearly, since all $p_i^C > 0$, under Assumption 1.2 all a_i's must be positive, so that in this case an optimum d.r. in C is a likelihood ratio d.r.

Now consider the line L_I determined by the equations $(1 - p_1)/\beta_1 = \ldots = (1 - p_m)/\beta_m$. This line passes through the interior of U (since all $\beta_i < 1$) and through the point $(1,1,\ldots,1)$, but does not necessarily intersect P_0 and may not intersect P (see the remarks in the third paragraph of this section). In fact, it may be shown that L_I intersects P_0 if and only if every $\beta_i \leq \sum_{j=1}^{m} \beta_j/(m - 1)$.[7] If L_I intersects P, I' is the segment of L_I intersecting P, and $\underline{p}^I$, the endpoint of I' closest to the point $(1, 1, \ldots, 1)$, corresponds to any optimum d.r. in I. If L_I intersects P, it may be shown, as in the previous case, that

7. The sufficiency is proved by constructing a d.r. which is in both P_0 and I as in Lemma 1.3. The necessity is proved by showing that if $\beta_1 > \Sigma\beta_j/(m-1)$, say, then L_I intersects the hyperplane $\Sigma p_i = 1$ in a point with a negative first coordinate, and hence outside of U.

any d.r. D for which $\underline{p}(D) = \underline{p}^I$ is of the likelihood ratio form defined by the non-negative constants $b_1, \ldots, b_m$ which determine a supporting hyperplane of P at $\underline{p}^I$, and if Assumption 1.2 holds, all b_i are positive. If a M.E. d.r. relative to $\underline{\alpha}$ exists, then it may be proved under Assumption 1.2 that L_I intersects P for the M.E. sample size; i.e., that I is not null. A direct proof has been established but will not be given since Theorem 1.6 proves the same thing. The fact that I may be null for smaller sample sizes is essentially equivalent to the fact that some of the components of a least favorable distribution may be zero for such sample sizes.

Now let P_n, C_n, and I_n denote the corresponding classes P, C, and I for samples of size n, n = 0, 1, 2, ... For any n, the two lines L_{C_n} and L_{I_n} intersect in the point $\underline{p} = \underline{\alpha}$ which is interior to U. Intuitively, the least n (if any) for which this point is contained in P_n is the M.E. sample size; formally, we have:

<u>THEOREM</u> 1.7: Suppose for some n the point $\underline{p} = \underline{\alpha}$ is in P_n and let N be the least such integer. Then N is the M.E. sample size. Moreover, there exist optimum d.r.'s in C_N and I_N and any such d.r. is a M.E. d.r.

<u>Proof</u>: We assume $\Sigma\alpha_i > 1$ since otherwise 0 is the M.E. sample size and the theorem is trivial. Clearly, any d.r. D for which $\underline{p}(D) = \underline{\alpha}$ satisfies (1.2), and any such D is in C and in I so that d.r.'s corresponding to $\underline{p}^C$ and $\underline{p}^I$ exist and satisfy (1.2). We need only prove that N is the M.E. sample size. Suppose it is not and that D_n is a M.E. d.r. for $n < N$. Let D_{C_n} be an optimum d.r. in C_n, and suppose $a_1, \ldots, a_m$ are non-negative constants defining a supporting hyperplane at $\underline{p}^{C_n}$. Then $\Sigma a_i p_i^{C_n} \geq \Sigma a_i p_i(D_n) \geq \Sigma a_i \alpha_i$ since D_n satisfies (1.2); but since the $p_i^{C_n}$'s are proportional to the α_i's, this implies $p_i^{C_n} \geq \alpha_i$ for all i so that the point $\underline{p} = \underline{\alpha}$, being between $\underline{p}^{C_n}$ and P_0, is in P_n. This is a contradiction, so that N is the M.E. sample size. //

Clearly, for $n = 0$, $P_n = P_0$ defined previously, and similarly to Lemma 1.1, we have that $P_{n_1} \subseteq P_{n_2}$ if $n_1 \leq n_2$. Under "favorable" conditions, we might expect P_n to be a proper subset of P_{n+1} for every n and P_n to tend in the limit to U, thus guaranteeing the existence of a M.E. d.r. relative to any $\underline{\alpha}$. Such "favorable" conditions will be given in Chapter III. (Though stated in different terms, it is easy to verify that Theorem 3.6 does just this.)

1.3.4 Admissibility. When considering M.E. d.r.'s relative to a vector $\underline{\alpha}$, we shall define admissibility as follows: a d.r. D is said to be admissible if there does not exist a d.r. D' for the same sample size for which

$$(1.19) \qquad p_i(D') \geq p_i(D) \qquad (i = 1, \ldots, m)$$

with strict inequality for at least one i. For either of the weight functions introduced in Sections 1.3.1 and 1.3.2, this is equivalent to Wald's definition (see [22] or [24]).

A Bayes d.r. D^* relative to the sequence of a priori distributions $(\underline{\xi}^1, \ldots, \underline{\xi}^h)$ is defined inductively as follows: for $h = 1$, D^* is a Bayes d.r. relative to $\underline{\xi}^1$; for $h > 1$, D^* minimizes the average risk relative to $\underline{\xi}^h$ w.r.t. all d.r.'s which are Bayes d.r.'s relative to the sequence $(\underline{\xi}^1, \ldots, \underline{\xi}^{h-1})$. This definition is due to Wald and Wolfowitz [24]; they proved the following theorem:

<u>THEOREM</u> 1.8: A necessary and sufficient condition for a d.r. to be admissible is that it is a Bayes d.r. relative to a sequence of h ($\leq$ m) a priori distributions ($\underline{\xi}^1, \ldots, \underline{\xi}^h$) such that

(i) for any j, $\xi_j^i > 0$ for some i, and

(ii) the sequence ($\underline{\xi}^1, \ldots, \underline{\xi}^{h-1}$) does not have property (i).

It follows that a Bayes d.r. relative to any $\underline{\xi}$ of which all components are positive is admissible. Hence, any likelihood ratio d.r. is admissible, and M.E. d.r.'s obtained by either of the minimax approaches of Sections 1.3.1 and 1.3.2 are admissible if all components of a least favorable distribution are positive; under Assumption 1.2, this latter condition holds (by Theorems 1.4 and 1.6).

But suppose Assumption 1.2 does not hold. We shall now show, for any sample size, how a minimax d.r. w.r.t. the weight function (1.6) may be altered to obtain an admissible d.r. The same method may be used for the M.E. sample size for the weight function (1.14). This will enable us to obtain M.E. d.r.'s which are admissible.

Geometrically, the situation is this: the line L_C intersects the surface of P in a point $\underline{p}^C$; if this point is at a "flat" part of the surface of P, then we may be able to obtain a "better" point by choosing a boundary point of this flat subspace which is further from P_0; that is, without changing any of the coordinates of $\underline{p}^C$ determining the sub-space, we may increase some of the coordinates of $\underline{p}^C$ within the subspace.

Let $\underline{\xi}^0$ be a least favorable distribution and D^0 a minimax d.r. w.r.t. the weight function (1.6), and suppose k $(0 < k < m)$ of the components, say $\xi_1^0, \ldots, \xi_k^0$, are zero. We shall introduce below a distribution $\underline{\xi}^1$ with the first k components positive and re-define D^0, without increasing the maximum risk, in such a way that it is a Bayes d.r. relative to the sequence $(\underline{\xi}^0, \underline{\xi}^1)$. Using Theorem 1.8, this d.r. will be an admissible minimax d.r.

Define the class $\mathcal{D}^0 = \{ D : D$ is a Bayes d.r. relative to $\underline{\xi}^0 \}$; i.e., $\mathcal{D}^0$ is the set of all minimax d.r.'s. The d.r. D^0, defined by $\phi^0(x)$, is a particular member of this class. Let $S = \{ x : f_{k+1}(x) = \ldots = f_m(x) = 0 \}$; we shall verify presently that S is not empty. Now for all $x \in \mathcal{X} - S$, $\max\limits_{k<i\leq m} \xi_i^0 f_i(x)/\alpha_i > 0$ so that by Theorem 1.1 (see (1.10)), $\phi_1(x) = \ldots = \phi_k(x) = 0$ for all $x \in \mathcal{X} - S$ for any $\phi(x)$ defining a d.r. $D \in \mathcal{D}^0$. Hence, using (1.1), for any such d.r.,

$$(1.20) \quad p_i(D) = \begin{cases} \int_S \phi_i(x) f_i(x)\, d\mu & \text{for } i = 1, \ldots, k \\ \int_{\mathcal{X} - S} \phi_i(x) f_i(x)\, d\mu & \text{for } i = k+1, \ldots, m. \end{cases}$$

Then, if we re-define $\underline{\phi}^0(x)$ in S, we will not affect $p_{k+1}(D^0)$, $\ldots$, $p_m(D^0)$.

Let $s_i = \int_S f_i(x)\, d\mu$ $(i = 1, \ldots, k)$. Suppose $s_i = 0$ for some i, say i = 1; then from (1.20), $p_1(D^0) = 0$, implying $r(f_1, D^0) = 0$, in contradiction to (1.12) (which does not depend on Assumption 1.2). Therefore, all $s_i > 0$, implying that S is not empty as well.

Let $f_i^*(x)$ $(i = 1, \ldots, k)$ be the conditional density function of x w.r.t. μ given $x \in S$; i.e., $f_i^*(x) = f_i(x)/s_i$. Let $\underline{\xi}^* = (\xi_1^*, \ldots, \xi_k^*)$ be some a priori distribution over $(f_1^*, \ldots, f_k^*)$ with $\xi_i^* > 0$ for all i. Let $\underline{\phi}^*(x) = [\phi_1^*(x), \ldots, \phi_k^*(x)]$ define a k-d.r. D^*, say, which is a Bayes d.r. relative to $\underline{\xi}^*$ for discriminating among $f_1^*, \ldots, f_k^*$; i.e., D^* satisfies

$$(1.21)\quad r(\underline{\xi}^*,D^*) = \inf_D r(\underline{\xi}^*,D) = -\sup_{\underline{\phi}} \sum_{i=1}^{k} \frac{\xi_i^*}{\alpha_i} \int_S \phi_i(x) f_i^*(x)\, d\mu$$

$$= -c \sup_{\underline{\phi}} \sum_{i=1}^{k} \frac{\xi_i^1}{\alpha_i} \int_S \phi_i(x) f_i(x)\, d\mu$$

where $\xi_i^1 = \xi_i^*/s_i c$ and $c = \sum_{i=1}^{k} \xi_i^*/s_i$. (If $k = 1$, $\underline{\phi}^*(x)$ has the single component "1".) By Theorem 1.1, for any $x \in S$ and for any j for which $\phi_j^*(x) > 0$, we have $\xi_j^* f_j^*(x)/\alpha_j = \max_{1\leq i\leq k} \xi_i^* f_i^*(x)/\alpha_i$ or $\xi_j^1 f_j(x)/\alpha_j = \max_{1\leq i\leq k} \xi_i^1 f_i(x)/\alpha_i$.

Let $\underline{\xi}^1 = (\xi_1^1, \ldots, \xi_m^1)$ where $\xi_{k+1}^1 = \ldots = \xi_m^1 = 0$ so that $\Sigma\xi_i^1 = 1$, and let D^1 be an m-d.r. defined by

$$(1.22)\quad \underline{\phi}^1(x) = \begin{cases} \underline{\phi}^0(x) & \text{if } x \in \mathcal{X} - S \\ [\phi_1^*(x), \ldots, \phi_k^*(x), 0, \ldots, 0] & \text{if } x \in S. \end{cases}$$

Now, by definition, D^1 is a Bayes d.r. relative to the sequence $(\underline{\xi}^0, \underline{\xi}^1)$ if $D^1 \in \mathcal{D}^0$ and if

$$(1.23)\qquad r(\underline{\xi}^1, D^1) = \inf_{D\varepsilon\mathcal{D}^0} r(\underline{\xi}^1, D) = -\sup_{D\varepsilon\mathcal{D}^0} \sum_{i=1}^{k} \frac{\xi_i^1}{\alpha_i} \int_S \phi_i(x) f_i(x)\, d\mu.$$

Now D^0 and D^1 are equal except in S so that by (1.20), $r(\underline{\xi}^0, D^1) = r(\underline{\xi}^0, D^0)$; hence, $D^1 \varepsilon \mathcal{D}^0$. Moreover, any $D \varepsilon \mathcal{D}^0$ for which the supremum in (1.23) is attained must have $\phi_{k+1}(x) = \ldots = \phi_m(x) = 0$ for $x \varepsilon S$ since $\Sigma\phi_i(x) = 1$. Clearly, then, from (1.21), (1.22), and (1.23), D^1 is a Bayes d.r. relative to $(\underline{\xi}^0, \underline{\xi}^1)$.

1.3.5 Likelihood Ratio Decision Rules. It was shown in the previous sections that to obtain M.E. d.r.'s relative to a vector $\underline{\alpha}$ under Assumption 1.2 we need consider only likelihood ratio d.r.'s, and that any likelihood ratio d.r. is admissible.

Suppose the density functions in $\underline{\Gamma}$ are completely specified except for a parameter θ, taking on one of the values $\theta_1, \ldots, \theta_m$ (assuming $\ell = m$). Clearly, then, to obtain a likelihood ratio d.r. we need consider only sufficient statistics, if existent. The following theorem is of a related nature.

THEOREM 1.9: Suppose there exists a statistic $t = t(x)$ which is a monotone increasing function of $f_j(x)/f_i(x)$ for every j and for every $i < j$ (for some ordering of the subscripts). Then there exist constants $c_1, \ldots, c_{m-1}$ such that if $\underline{\phi}(x)$ defines a likelihood ratio d.r., we have

$$\phi_j(x) > 0 \quad \text{implies} \quad \begin{cases} t \leq c_1 & \text{for } j = 1 \\ c_{j-1} \leq t \leq c_j & \text{for } j = 2,\ldots,m-1 \\ t \geq c_{m-1} & \text{for } j = m \end{cases}$$

for any j and any $x \in \mathcal{X}$.

<u>Proof</u>: By Definition 1.3, there exist positive constants $a_1, \ldots, a_m$ such that $\phi_j(x) > 0$ implies $f_j(x)/f_i(x) \geq a_i/a_j$ for all $i \neq j$. Now $f_j/f_i \geq a_i/a_j$ for all $i < j$ implies $t \geq b_{ij}$ for all $i < j$ for some constants b_{ij} $(i < j)$, and $f_j/f_i \geq a_i/a_j$ for all $i > j$ (or $f_i/f_j \leq a_j/a_i$) implies $t \leq b_{ji}$ for all $i > j$. Hence

$$\phi_j(x) > 0 \text{ implies } \begin{cases} t \leq \min\limits_{i>1} b_{1i} & \text{for } j = 1 \\ \max\limits_{i<j} b_{ij} \leq t \leq \min\limits_{i>j} b_{ji} & \text{for } j=2,\ldots,m-1 \\ t \geq \max\limits_{i<m} b_{im} & \text{for } j = m \text{ ;} \end{cases}$$

or, in particular,

$$\phi_j(x) > 0 \text{ implies } \begin{cases} t \leq b_{12} & \text{for } j = 1 \\ b_{j-1,j} \leq t \leq b_{j,j+1} & \text{for } j = 2, \ldots, m-1 \\ t \geq b_{m-1,m} & \text{for } j = m. \end{cases}$$

Set $c_j = b_{j,j+1}$ $(j = 1, \ldots, m-1)$, and the proof is completed. //

This theorem characterizes M.E. d.r.'s obtained by either of the minimax methods of Sections 1.3.1 and 1.3.2. If all probabilities of correct decisions are to be positive, then $-\infty < c_1 \leq c_2 \leq \cdots \leq c_{m-1} < \infty$ and

$$(1.24) \qquad \underline{\phi}(x) = \begin{cases} (1, 0, 0, \ldots, 0) & \text{if } t < c_1 \\ (a_1, 1-a_1, 0, \ldots, 0) & \text{if } t = c_1 \\ (0, 1, 0, \ldots, 0) & \text{if } c_1 < t < c_2 \\ (0, a_2, 1-a_2, \ldots, 0) & \text{if } t = c_2 \\ (0, 0, 1, \ldots, 0) & \text{if } c_2 < t < c_3 \\ \cdots\cdots\cdots\cdots & \\ (0, 0, 0, \ldots, 1) & \text{if } t > c_{m-1} \end{cases}$$

where $a_1, \ldots, a_{m-1}$ are appropriately chosen constants between zero and one. If Assumption 1.1 is satisfied, the a_i's may be set equal to 0 or 1 arbitrarily and ϕ then defines a non-randomized d.r.

Note that if each $f_i(x)$ is of the form

$$(1.25) \qquad f_i(x) = A(x) \exp [\rho(\theta_i)\, t(x) + \tau(\theta_i)]$$

where $\theta_1, \ldots, \theta_m$ are so ordered that $\rho(\theta_1) < \ldots < \rho(\theta_m)$, then the conditions of Theorem 1.9 are satisfied.

1.3.6 Four Examples. In this section we give four examples of M.E. m-d.r.'s relative to a vector $\underline{\alpha}$ for simple discrimination, each of which is obtained by the characterization theorem of the previous section. For $m = 2$, they reduce to the standard "one-sided" tests of hypotheses.

EXAMPLE 1: Mean of Normal Distribution. Each $f_i(x)$ is a normal density function with variance σ^2 (known) and with mean θ_i $(-\infty < \theta_1 < \ldots < \theta_m < \infty)$. The m alternatives $A_1, \ldots, A_m$ correspond to the true densities $f_1, \ldots, f_m$. Now $f_i(x)$ is of the form (1.25) with $t(x) = \bar{x}$, the sample mean, so that, by the results of the previous section, a non-randomized d.r. D_n with acceptance regions

$$(1.26)\quad \begin{cases} R_1^n = \{x : \bar{x} \leq c_1^n\} \\ R_j^n = \{x : c_{j-1}^n < \bar{x} \leq c_j^n\},\ 1 < j < m \\ R_m^n = \{x : c_{m-1}^n < \bar{x}\} \end{cases}$$

is a likelihood ratio d.r. Since Assumption 1.2 is satisfied, a likelihood ratio d.r. satisfying (1.11) (or (1.16)) with n chosen so that the ratio in (1.11) is approximately, but not less than, unity is a M.E. d.r. Hence, we may obtain a M.E. d.r. by first solving the following equations for n, $c_1^n, \ldots, c_{m-1}^n$, with $\rho_n = 1$:

$$(1.27)\quad \begin{cases} p_1(D_n) = \Phi\left(\sqrt{n}\,\dfrac{c_1^n - \theta_1}{\sigma}\right) = \alpha_1 \rho_n \\ p_j(D_n) = \Phi\left(\sqrt{n}\,\dfrac{c_j^n - \theta_j}{\sigma}\right) - \Phi\left(\sqrt{n}\,\dfrac{c_{j-1}^n - \theta_j}{\sigma}\right) = \alpha_j \rho_n,\ 1<j<m \\ p_m(D_n) = 1 - \Phi\left(\sqrt{n}\,\dfrac{c_{m-1}^n - \theta_m}{\sigma}\right) = \alpha_m \rho_n \end{cases}$$

where Φ denotes the standard normal distribution function, and then, choosing N to be the least integer $\geq$ n and re-solving (1.27) for $c_1^N, \ldots, c_{m-1}^N, \rho_N$, with n = N, the acceptance regions (1.26) with n = N will define a M.E. d.r. relative to $\underline{\alpha}$ which is a minimax d.r. w.r.t. (1.6) for samples of size N. Alternatively, we may replace (1.27) by equations of the form $1 - p_i(D_n) = (1-\alpha_i)\rho_n'$ and proceed as before, obtaining a M.E. d.r. which is a minimax d.r. w.r.t. (1.14). Or, solving for N as above, we may choose any c_i's for which a likelihood ratio d.r. D_N satisfies (1.2), and this d.r. will be a M.E. d.r.

Equations (1.27) with $\rho_n = 1$ may be solved iteratively by choosing a trial value of n, solving the first equation for c_1^n, the second for c_2^n, and so on until the j^{th} equation is unsolvable, and then choosing another trial value of n, and so on; at each stage it will be obvious from the j^{th} equation whether to try a larger or smaller n.

EXAMPLE 2: Variance of Normal Distribution. Each $f_i(x)$ is a normal density function with mean μ (known) and variance σ_i^2 $(0 < \sigma_1 < \ldots < \sigma_m < \infty)$. $f_i(x)$ is again of the form (1.25) but now with $t(x) = \sum_k (x_k - \mu)^2/n$. Hence, the results are similar to Example 1 with $\bar{x}$ replaced by $\sum_k (x_k - \mu)^2/n$ in (1.26), and the normal distribution functions replaced by the corresponding χ_n^2 distribution functions in (1.27).

EXAMPLE 3: Binomial Parameter. Each $f_i(x)$ is a point binomial probability function with parameter θ_i $(0 < \theta_1 < \ldots < \theta_m < 1)$, and hence is of the form (1.25) with $t = \Sigma x_k$. The results are similar to Example 1 except that here a randomized d.r. will usually be required. The necessary modifications to the methods of Example 1 are obvious.

EXAMPLE 4: Poisson Parameter. Each $f_i(x)$ is a Poisson probability function with parameter λ_i $(0 < \lambda_1 < \ldots < \lambda_m < \infty)$, and hence is of the form (1.25) with $t = \bar{x}$. The results are similar to those of Example 3.

1.4 Most Economical Decision Rules Relative to a Matrix β. We shall derive a method of obtaining M.E. d.r.'s as defined by Definition 1.2, using minimax theory. In this development we replace each β_{ij} which is equal to unity by $+\infty$ which, of course, makes no effective change in Definition 1.2. This has the same effect as omitting completely all corresponding terms throughout, but because it is complicated notationally to do so we use the above device.

Suppose n fixed, and let $\underline{\Omega}'$ be a set of density functions g_{ij} w.r.t. μ $(i = 1, \ldots, \ell;\ j = 1, \ldots, m)$ where $g_{ij} = f_i$ identically in x; thus, there are $\ell' = \ell m$ elements in $\underline{\Omega}'$. Define a weight function $W(g_{ij}, A_k) = W_{ijk}$ where

$$(1.28) \qquad W_{ijk} = \begin{cases} 1/\beta_{ij} & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1,\ldots,\ell; j,k = 1, \ldots, m).$$

We shall consider m-d.r.'s D for choosing among $A_1, \ldots, A_m$ when one of the ℓ' density functions g_{ij} is "true", and where the "loss" incurred by choosing A_k when g_{ij} is "true" is $W(g_{ij}, A_k)$. This interpretation is meaningless, but it serves to clarify the approach. The risk function when using a d.r. D and g_{ij} is "true" is: $r(g_{ij}, D) = \sum_k W_{ijk} p'_{ijk}(D)$ where $p'_{ijk}(D) = \Pr(D \text{ chooses } A_k \mid g_{ij}) = p_{ik}(D)$. From (1.28), then, we have

$$(1.29) \qquad r(g_{ij},D) = p_{ij}(D)/\beta_{ij} \quad (i = 1,\ldots,\ell;\ j = 1,\ldots,m).$$

Theorem 1.2 when applied to the class $\underline{(\overline{\ })}'$ and the weight function (1.28) asserts the existence of a minimax d.r. D^0; i.e., a d.r. D^0 such that $\max_{i,j} [p_{ij}(D^0)/\beta_{ij}] = \inf_D [\max_{i,j} p_{ij}(D)/\beta_{ij}]$.

We have the theorem:

THEOREM 1.10: For each $n = 0, 1, 2, \ldots$, let D_n^0 be a minimax d.r. w.r.t. the weight function (1.28) for discriminating among $(g_{11}, g_{12}, \ldots, g_{\ell m})$ for samples of fixed size n. Suppose for some n,

$$\max_{i,j} r(g_{ij}, D_n^0) \leq 1 \tag{1.30}$$

and let N be the least such integer. Then D_N^0 is a M.E. d.r. relative to the matrix β for discriminating among $(f_1, \ldots, f_\ell)$. Conversely, if there exists a M.E. d.r. relative to β and N is the M.E. sample size, then D_N^0 is a M.E. d.r.

Proof: From (1.29) and (1.30), it follows that D_N^0 satisfies (1.3). Now suppose for some $n < N$, there exists a d.r. D_n satisfying (1.3). But D_n^0 is a minimax d.r. so that $\max_{i,j}[p_{ij}(D_n^0)/\beta_{ij}]$ $= \max_{i,j} r(g_{ij}, D_n^0) \leq \max_{i,j} r(g_{ij}, D_n) = \max_{i,j}[p_{ij}(D_n)/\beta_{ij}] \leq 1$ by (1.3), in contradiction to the fact that N is the least integer n for which this is true. Hence, D_N^0 is a M.E. d.r.

To prove the converse, suppose D_N is a M.E. d.r. Then $1 \geq \max_{i,j}[p_{ij}(D_N)/\beta_{ij}] \geq \max_{i,j} r(g_{ij}, D_N^0)$. Hence, (1.30) is satisfied for $n = N$, and since N is the M.E. sample size, D_N^0 is a M.E. d.r. //

Thus, to obtain M.E. d.r.'s relative to β, we need only consider minimax d.r.'s relative to the weight function (1.28) for various values of n. Lemma 1.1 may be helpful in finding the M.E. sample size. Now let us consider the structure of minimax solutions.

Let $\xi = (\xi_{11}, \xi_{12}, \ldots, \xi_{\ell m})$ where $\xi_{ij} \geq 0$ and $\sum_{i,j} \xi_{ij} = 1$ denote an a priori distribution over $\underline{(\bar{\ })}'$. By Theorem 1.1, a necessary and sufficient condition for a d.r. D to be a Bayes d.r. relative to ξ is that for any x (except perhaps in a set of ξ-measure zero) and any k for which $\phi_k(x) > 0$, we have

$$\sum_{i,j} \xi_{ij} W_{ijk} g_{ij}(x) = \min_{1 \leq k \leq m} \sum_{i,j} \xi_{ij} W_{ijk} g_{ij}(x)$$

or

$$(1.31) \qquad \sum_{i=1}^{\ell} \xi_{ik} f_i(x)/\beta_{ik} \leq \sum_{i=1}^{\ell} \xi_{ij} f_i(x)/\beta_{ij} \qquad (j = 1, \ldots, m).$$

Since for any i,j pair corresponding to a correct decision, $\beta_{ij} = \infty$, the sums in (1.31) may be replaced by sums over all i for which i,k on the left or i,j on the right correspond to incorrect decisions. Setting $a_{ij} = \xi_{ij}/\beta_{ij}$, we thus have from (1.31) that any Bayes d.r. relative to ξ is a "minimum unlikelihood" d.r., as defined by Lindley [16]. Hereafter, we shall suppose $\xi_{ij} = 0$ for every i,j for which $\beta_{ij} = \infty$ without loss of generality.

Theorem 1.2 asserts the existence of a least favorable distribution ξ^0, and that any Bayes d.r. relative to ξ^0 is a minimax d.r. and conversely; moreover,

$$(1.32) \qquad p_{ij}(D^0)/\beta_{ij} = \max_{i,j}[p_{ij}(D^0)/\beta_{ij}] \text{ for any } i,j \text{ for which } \xi^0_{ij} > 0.$$

Apparently, however, there are no general conditions under which all ξ^0_{ij} are positive, and consequently we have no proof of the admissibility of a minimax d.r. In fact, supposing $\ell = m$ and the β_{ij}'s satisfy $\sum_{j=1}^{m} \beta_{ij} = 1$ for every i, if $\xi^0_{ij} > 0$ for all i, j then $p_{ij} = \beta_{ij}$, regardless of the sample size. This, of course, is too much to expect! Geometrically, the convex body in the $\ell \cdot m$-dimensional space with coordinate axes p_{ij}, corresponding to all possible d.r.'s for a fixed sample size, is not necessarily intersected by the line determined by $p_{ij}/\beta_{ij} = p_{i'j'}/\beta_{i'j'}$ for all pairs of subscripts corresponding to incorrect decisions.

(See [16] and Section 1.3.3.) However, we do have the following theorem in this regard, assuming $\ell = m$ and A_i is "correct" when f_i is true ($i = 1, \ldots, m$).

<u>THEOREM</u> 1.11: Suppose Assumption 1.2 holds and that $\sum_{j \neq i} \beta_{ij} < 1$ for every i. For any sample size greater than or equal to the M.E. sample size, a least favorable distribution ξ^0 has the property $\sum_{i=1}^{m} \xi^0_{ij} > 0$ for every j.

<u>Proof</u>: Suppose the theorem false; i.e., for some j, $\xi^0_{ij} = 0$ for every i. Then, for some $k \neq j$, $\xi^0_{ik} > 0$ for at least one i. Hence, for all x (using Assumption 1.2), $\sum_i \xi^0_{ij} f_i(x)/\beta_{ij} = 0 < \sum_i \xi^0_{ik} f_i(x)/\beta_{ik}$. Therefore, a Bayes d.r. relative to ξ^0 must have $\phi_k(x) = 0$ identically in x. Hence, denoting a minimax d.r. by D^0, $p_{ik}(D^0) = 0$ for all i; in particular, $p_{kk}(D^0) = 0$. Now, since the sample size is $\geq N$ (the M.E. sample size) and using Lemma 1.1, D^0 satisfies (1.30) -- i.e., it satisfies (1.3) -- so that

$$0 = \sum_{j=1}^{m} \xi_{kj}(D^0) - 1 = \sum_{j\neq k} p_{kj}(D^0) + p_{kk}(D^0) - 1$$

$$\leq \sum_{j\neq k} \beta_{kj} + p_{kk}(D^0) - 1 < p_{kk}(D^0) ,$$

a contradiction. //

According to this theorem and (1.32), it follows that $p_{ij}(D_N^0)/\beta_{ij}$ attains its maximum for at least one value of i for every j, where D_N^0 is a minimax d.r. for samples of the M.E. size. This may be useful for obtaining M.E. d.r.'s.

EXAMPLE: Mean of a Normal Distribution, Variance Known. We shall consider briefly minimum unlikelihood d.r.'s for samples of size n for discriminating among m normal density functions $f_1, \ldots, f_m$ with means θ_i $(-\infty < \theta_1 < \ldots < \theta_m < \infty)$ and common variance σ^2. For simplicity, suppose $\sigma^2 = 1$ and $\ell = m = 3$, the alternatives A_1, A_2, A_3 corresponding respectively to the densities f_1, f_2, f_3. Suppose further, without loss of generality, that $\theta_2 = 0$.

We need consider only non-randomized d.r.'s. A d.r. with acceptance regions

$$(1.33)\quad \begin{cases} R_1^n = \{ x : h_1^n(x) \le h_2^n(x),\ h_1^n(x) \le h_3^n(x) \} \\ R_2^n = \{ x : h_2^n(x) < h_1^n(x),\ h_2^n(x) \le h_3^n(x) \} \\ R_3^n = \{ x : h_3^n(x) < h_1^n(x),\ h_3^n(x) < h_2^n(x) \} \end{cases}$$

where $h_i^n(x) = a_{ji}f_j^n(x) + a_{ki}f_k^n(x)$ and (i, j, k) is a permutation of $(1, 2, 3)$, is a minimum unlikelihood d.r. for the weights (a_{ij}). Denoting the sample mean by $\bar{x}$, we have

$$h_i^n(x) = C(x)[\, a_{ji} \exp(n\theta_j\bar{x} - n\theta_j^2/2) + a_{ki} \exp(n\theta_k\bar{x} - n\theta_k^2/2)\,]$$

where $C(x) = (2\pi)^{-n/2}\exp(-\Sigma x^2/2)$ (> 0). Let $g_i^n(x) = h_i^n(x)/C(x)$. Clearly, we may replace $h_i^n(x)$ by $g_i^n(x)$ throughout (1.33), and thus the acceptance regions depend on x only through $\bar{x}$. It may easily be verified that g_1^n is an increasing function of $\bar{x}$ and g_3^n is a decreasing function of $\bar{x}$. Setting $dg_2^n/d\bar{x} = 0$ and solving for $\bar{x}$, we obtain the single stationary point of g_2^n, $\bar{x} = [\, n(\theta_3 - \theta_1)\,]^{-1}\log(-a_{12}\theta_1/a_{32}\theta_3)$, which is a minimum of g_2^n since the second derivative is everywhere positive. Hence, by sketching the three g_i^n functions, it is clear that if none of the acceptance regions is to be empty, one of three possibilities must obtain:

the acceptance regions are of the form

$$
(1.34) \qquad \begin{cases} R_1 = \{ x : \bar{x} \leq c_1 \text{ or } c_3 \leq \bar{x} \leq c_4 \} \\ R_2 = \{ x : c_2 \leq \bar{x} \leq c_3 \} \\ R_3 = \{ x : c_1 \leq \bar{x} \leq c_2 \text{ or } \bar{x} \geq c_4 \} \end{cases}
$$

where either $c_1 = c_2$, or $c_3 = c_4$, or both. (Equality signs have been assigned everywhere in (1.34) for simplicity.) Let c (= 2 or 3) denote the number of c_i's to be determined. The c_i's are obtained by solving $c + 1$ of the six equations $p_{ij} = \rho\, \beta_{ij}$ $(i, j = 1, 2, 3; i \neq j)$ for the c_i's and ρ, the choice of the equations to be solved being such that $p_{ij} \leq \rho \beta_{ij}$ for all six pairs of subscripts. Theorem 1.11 may be helpful in this choice of equations to be solved.

To obtain a M.E. d.r., the sample size n is to be minimized subject to $\rho = \rho^n \leq 1$.

Similar methods may be applied to simple discrimination problems concerning any distribution of the form (1.25).

1.5 A Generalization of Most Economical Decision Rules Relative to a Vector.

DEFINITION 1.4: Given an $m \times m$ matrix $W = (w_{ij})$ of non-negative elements and a vector $\underline{\beta} = (\beta_1, \ldots, \beta_m)$ of positive constants, a d.r. D_N, based on a sample of size N, is said to be a M.E. m-d.r. relative to $\underline{\beta}$ w.r.t. the matrix W for discriminating among $\underline{f} = (f_1, \ldots, f_m)$ if it satisfies

$$(1.35) \qquad \sum_{j=1}^{m} w_{ij} p_{ij}(D_n) \leq \beta_i \quad (i = 1, \ldots, m)$$

and if N is the least integer n for which (1.35) may be satisfied by some m-d.r. D_n based on a sample of size n.

Letting $w_{ij} = 1 - \delta_{ij}$ (Kronecker δ) and $\beta_i = 1 - \alpha_i$, this definition reduces to Definition 1.1.

Suppose $f_i(x) = f(x, \theta_i)$ $(i = 1, \ldots, m)$ and $w_{ij} = (\theta_i - \theta_j)^2$. If we interpret the alternative A_j as the decision to estimate the true θ by θ_j, then θ_j (the estimate) is a random variable with probability function p_{ij} if θ_i is the true θ, and w_{ij} is the squared error. Hence, a M.E. d.r. relative to $\underline{\beta}$ w.r.t. W is a d.r. with minimum sample size subject to bounds ($\underline{\beta}$) on the m mean-squared deviations corresponding to the m possible true values of θ.

Minimax theory for fixed sample sizes may be applied as before to obtain M.E. d.r.'s. Let

$$W(f_i, \Lambda_j) = w_{ij}/\beta_i \quad (i, j = 1, \ldots, m). \tag{1.36}$$

Then $r(f_i, D) = \frac{1}{\beta_i} \sum_{j=1}^{m} w_{ij} p_{ij}(D)$. By Theorem 1.2, for any fixed sample size, there exists a minimax d.r. D^0 for the weight function (1.36). By considering minimax d.r.'s D_n^0 for each sample size n, we may prove in a manner completely analogous to Theorem 1.5 that D_N^0 is a M.E. d.r. if N is the least n for which $\max_i r(f_i, D_n^0) \leq 1$.

In Chapter IV, we shall consider further generalizations.

CHAPTER II

MOST ECONOMICAL MULTIPLE-DECISION RULES FOR COMPOSITE DISCRIMINATION

2.1 The Problem of Composite Discrimination.

In this chapter we no longer require that Ω be a finite class of density functions but allow a denumerable or non-denumerable infinity of elements as well, all of which are density functions w.r.t. a specified measure μ. We shall assume that Ω is a parametric class of generalized density functions, completely specified except for some unknown real- or vector-valued parameter θ; this is somewhat restrictive, but with only minor changes the whole of this chapter may be extended to more general classes of distributions. We denote the parameter space by Ω and the corresponding density function by $f(x,\theta)$. All other assumptions and definitions of Section 1.1.1 carry over.

We suppose, moreover, that some finite number, say ℓ, of disjoint subsets $\omega_1, \ldots, \omega_\ell$ of the space Ω are specified such that for every pair i, j ($i = 1, \ldots, m$; $j = 1, \ldots, \ell$), there is a definite preference for or against the decision A_j if the true $\theta \epsilon \omega_i$; we suppose that none of the decisions is definitely preferred if $\theta \epsilon (\Omega - \bigcup_{i=1}^{\ell} \omega_i)$, the "indifference region", and therefore we shall hereafter denote for simplicity $\Omega = \bigcup_{i=1}^{\ell} \omega_i$, which we often denote $\omega = (\omega_1, \ldots, \omega_\ell)$. Under these assumptions, we say that the corresponding decision problem is one of "composite discrimination" and a d.r. is a d.r. for "composite discrimination" or for "discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_\ell)$". As before, f or f^n denotes the joint density function w.r.t. $\mu = \mu(x)$ of the first n random variables.

A d.r. D will be characterized by the functions $p_j(\theta,D)$, $j = 1, \ldots, m$, defined for all $\theta \in \Omega$, where

$$p_j(\theta,D) = \Pr(D \text{ chooses } A_j|\theta) \tag{2.1}$$

$$= \int_{\mathcal{X}} \phi_j(x)\ f(x,\theta)\ d\mu = E_\theta \phi_j(X)$$

where the subscript on the expectation operator E denotes the corresponding parameter point. If D is a non-randomized d.r. with acceptance regions $R_1, \ldots, R_m$,

$$p_j(\theta,D) = \int_{R_j} f(x,\theta)\ d\mu\ .$$

As in the case of simple discrimination, we shall consider two different criteria for choosing a d.r. for composite discrimination. The first criterion is applicable to the case when the number of subsets, ℓ, of Ω is equal to the number, m, of alternatives and the alternative decisions correspond to the possible true parameter points in such a way that A_i is preferred when $\Theta \epsilon \omega_i$ $(i = 1, \ldots, m)$; A_i is said to be a correct decision if $\Theta \epsilon \omega_i$ and incorrect if $\Theta \epsilon \omega_j$ $(j \neq i)$.

<u>DEFINITION</u> 2.1: Let $\underline{\alpha} = (\alpha_1, \ldots, \alpha_m)$ be a given vector of positive constants each less than one. A d.r. D_N, based on a sample of size N, is said to be a most economical m-decision rule relative to the vector $\underline{\alpha}$ for discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_m)$ if it satisfies

$$(2.2) \qquad p_i(\Theta, D) \geq \alpha_i \quad \text{for all } \Theta \epsilon \omega_i \qquad (i = 1, \ldots, m)$$

and if N is the least integer n for which (2.2) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the most economical sample size.

Thus, assuming the sampling cost to be proportional to the sample size, a M.E. d.r. is one with minimum sampling cost subject to lower bounds on the probabilities of correct decisions. (2.2) may be written

$$\inf_{\theta\epsilon\omega_i} p_i(\theta,D) \geq \alpha_i \qquad (i = 1, \ldots, m)$$

or

$$\sup_{\theta\epsilon\omega_i} q_i(\theta,D) \leq 1 - \alpha_i \qquad (i = 1, \ldots, m)$$

where $q_i(\theta,D) = 1 - p_i(\theta,D)$, so that lower bounds on the probabilities of an incorrect decision are implicit in the definition.

A second criterion for choosing a d.r. will now be given. We suppose that corresponding to each ω_i one or more alternatives A_j is preferable, or correct, when $\theta\epsilon\omega_i$.

<u>DEFINITION</u> 2.2: Let $\beta = (\beta_{ij})$ be a given ℓxm matrix of positive constants where for every i, j pair for which A_j is a correct decision when $\theta\epsilon\omega_i$, $\beta_{ij} = 1$. A d.r. D_N, based on a sample of size N, is said to be a M.E. d.r. relative to the matrix β for

discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_m)$ if it satisfies

$$(2.3) \qquad p_j(\theta,D) \leq \beta_{ij} \quad \text{for all } \theta\epsilon\omega_i \quad (i=1,\ldots,\ell;j=1,\ldots,m)$$

and if N is the least integer n for which (2.3) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the M.E. sample size.

The remarks following Definition 1.2, with minor modification, are applicable here as well. In particular, when $\ell = m = 2$, the two definitions are equivalent. Considering a 2-d.r. as a test of the hypothesis that $\theta\epsilon\omega_1$ against the class of alternatives $\theta\epsilon\omega_2$, both (2.2) and (2.3) specify bounds on the two kinds of error; Hoeffding [8] has proved that a M.E. 2-d.r. may be obtained by considering, for each n, tests of size $\alpha(= 1-\alpha_1)$ w.r.t. ω_1 which maximize the minimum power w.r.t. ω_2 and choosing that test for which n is a minimum subject to the minimum power being at least $1 - \beta$ $(= \alpha_2)$. Definitions 2.1 and 2.2 are two extensions of these concepts to multiple decision problems.

In the following sections, we shall derive both types of M.E. d.r.'s for composite discrimination from minimax d.r.'s w.r.t. certain weight functions relative to the class of all d.r.'s based on a fixed sample size. First, we shall review parts of Wald's theory with some additional theorems to be applied to M.E. theory afterwards.

2.2 Minimax Decision Rules for Fixed Sample Sizes.

We shall review briefly some of Wald's concepts of Sections 1.1 and 5.14 [22], with minor changes as indicated in Section 1.2 above. We assume throughout that the sample size n has been fixed.

We assume a bounded[1] weight function $W(\Theta,A_j) = W_j(\Theta)$ $(j = 1, \ldots, m)$, defined for all $\Theta \epsilon \underline{\overline{(\ )}}$ representing the "loss" incurred by accepting A_j when Θ is true. The corresponding risk function when using a d.r. D is:

$$(2.4) \qquad r(\Theta,D) = \sum_{j=1}^{m} W_j(\Theta)\, p_j(\Theta,D) \quad .$$

We introduce an a priori distribution, denoted $\xi = (\underline{\xi},\lambda)$, over the Borel subsets $\{\omega\}$ of $\underline{\overline{(\ )}}$:

1. The boundedness is not always required (see Ghosh [6]) but is not restrictive here.

$$\xi(\omega) = \Pr(\Theta\epsilon\omega) = \sum_{i=1}^{\ell} \xi_i \lambda_i(\omega)$$

where

$$\xi_i = \xi(\omega_i),\ \lambda_i(\omega) = \Pr(\Theta\epsilon\omega \mid \Theta\epsilon\omega_i) \quad (i=1,\ldots,\ell) \quad .$$

The average risk relative to ξ is:

$$(2.5) \quad r(\xi,D) = \int_{\Omega} r(\Theta,D)\, d\xi = \sum_{i=1}^{\ell} \xi_i \int_{\omega_i} r(\Theta,D)\, d\lambda_i \ .$$

A d.r. D^* is said to be a Bayes d.r. relative to ξ if it minimizes the average risk relative to ξ; i.e., if $r(\xi,D^*) = \inf_D r(\xi,D)$. A d.r. D^* is said to be a Bayes d.r. relative to the infinite sequence $\{\xi^\nu\}$ of a priori distributions if $\lim_{\nu=\infty} [\inf_D r(\xi^\nu,D) - r(\xi^\nu,D^*)] = 0$. D^* is said to be a Bayes d.r. in the strict sense if there exists a ξ such that D^* is a Bayes d.r. relative to it, and a Bayes d.r. in the wide sense if there exists an infinite sequence of distributions such that D^* is a Bayes d.r. relative to the sequence. A d.r. D^0 is said to be a minimax d.r. if it minimizes the maximum risk over Ω; i.e., if $\sup_{\Omega} r(\Theta,D^0) = \inf_D \sup_{\Omega} r(\Theta,D)$. An a priori distribution ξ^0 is said to be least favorable if it maximizes w.r.t. ξ the minimum of the average risk; i.e., if for any other ξ

$$\inf_D r(\xi^0,D) \geq \inf_D r(\xi,D) \quad . \tag{2.6}$$

Wald has proved the following two theorems ([22] remarks on page 148 and parts of Theorems 3.8, 3.9, 3.10, and 5.12):

THEOREM 2.1: A necessary and sufficient condition for a d.r. D to be a Bayes d.r. relative to a given a priori distribution ξ is that $\phi_j(x) = 0$ for any x (except perhaps on a set of ξ-measure zero) and for any j for which

$$\int_{\Omega} W_j(\Theta)\, f(x,\Theta)\, d\xi > \min_{1\leq k\leq m} \int_{\Omega} W_k(\Theta)\, f(x,\Theta)\, d\xi \quad . \tag{2.7}$$

THEOREM 2.2: (i) There exists a minimax d.r. D^0;

(ii) any minimax d.r. is a Bayes d.r. in the wide sense.

(iii) suppose there exists a least favorable distribution ξ^0; then any minimax d.r. D^0 is a Bayes d.r. relative to it and conversely; and $\sup_{\Omega} r(\theta,D^0) = r(\xi^0,D^0)$;

(iv) if ξ^0 is a least favorable distribution, D^0 a minimax d.r., and ω the set of all $\theta \epsilon \Omega$ for which $r(\theta,D^0) < \sup_{\Omega} r(\theta,D^0)$, then $\xi^0(\omega) = 0$.

We formulate two assumptions ([22] Assumptions 5.1 and 5.6) and state a lemma and a theorem proved by Wald ([22] Lemma 5.1 and Theorem 5.11).

<u>ASSUMPTION</u> 2.1: Ω is closed and bounded and $W_j(\theta)$ $(j=1,\ldots,m)$ is a continuous function of θ.

<u>ASSUMPTION</u> 2.2: If $\{\theta^\nu\}$ is a sequence of parameter points such that $\lim_{\nu=\infty} \theta^\nu = \theta^0$, then

$$\lim_{\nu=\infty} \int_R f(x,\theta^\nu)\, d\mu = \int_R f(x,\theta^0)\, d\mu$$

uniformly in all subsets R of $\mathcal{X}$.

LEMMA 2.1: If $f(x,\theta)$ is continuous in θ, Assumption 2.2 holds.

THEOREM 2.3: If Assumptions 2.1 and 2.2 hold, then there exists a least favorable distribution.

Another existence theorem for least favorable distributions has been given by Lehmann [14].

We consider another assumption and prove three theorems which may be helpful in finding minimax d.r.'s. Sverdrup [20] gives some other theorems which may prove useful in this same regard.

ASSUMPTION 2.3: For each i, j pair ($i = 1, \ldots, \ell$; $j = 1, \ldots, m$), $W_j(\theta)$ equals a constant, say W_{ij}, for all $\theta \epsilon \omega_i$.

This assumption is that for each alternative the loss varies only from subset to subset among $\omega_1, \ldots, \omega_\ell$ and not within any subset. For a given set of conditional distributions $\lambda = (\lambda_1, \ldots, \lambda_\ell)$, we denote

$$f_i^\lambda(x) = \int_{\omega_i} f(x,\theta)\, d\lambda_i \qquad (i = 1, \ldots, \ell). \tag{2.8}$$

THEOREM 2.4: If Assumption 2.3 holds, a necessary and sufficient condition for a d.r. D^* to be a Bayes d.r. relative to $\xi = (\underline{\xi}, \lambda)$ for discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_\ell)$ is that D^* be a Bayes d.r. relative to $\underline{\xi}$ for discriminating among $\underline{f}^\lambda = (f_1^\lambda, \ldots, f_\ell^\lambda)$ w.r.t. the weight function W_{ij}. The average risks in the two cases are equal.

Proof: Using Assumption 2.3 and (2.8), we have

$$(2.9) \qquad \int_{\Omega} W_j(\Theta)\, f(x,\Theta)\, d\xi = \sum_{i=1}^{\ell} \xi_i \int_{\omega_i} W_{ij}\, f(x,\Theta)\, d\lambda_i$$

$$= \sum_{i=1}^{\ell} \xi_i\, W_{ij}\, f_i^\lambda(x) \quad .$$

The first part of the theorem follows immediately from (2.9) and Theorems 1.1 and 2.1.

We now show that for any d.r. D, the two risks are equal. We have, using (2.4), Assumption 2.3, (2.1), Fubini's Theorem, (2.8), (1.1), and (1.4):

$$(2.10) \quad \int_{\omega_i} r(\Theta,D)\, d\lambda_i = \sum_{j=1}^{m} \int_{\omega_i} W_j(\Theta)\, p_j(\Theta,D)\, d\lambda_i$$

$$= \sum_j W_{ij} \int_{\omega_i} \int_{\mathcal{X}} \phi_j(x)\, f(x,\Theta)\, d\mu\, d\lambda_i$$

$$= \sum_j W_{ij} \int_{\mathcal{X}} \phi_j(x)\, f_i^{\lambda}(x)\, d\mu$$

$$= \sum_j W_{ij}\, p_{ij}(\lambda,D) = r(f_i^{\lambda},D)$$

where we have denoted $p_{ij}(\lambda,D) = Pr(D \text{ chooses } A_j \mid f_i^{\lambda})$. Hence, by (2.5) and (1.5),

$$(2.11) \qquad r(\xi,D) = r_{\lambda}(\underline{\xi},D)$$

where $r_{\lambda}(\underline{\xi},D)$ is the average risk relative to $\underline{\xi}$ when discriminating among $\underline{f}^{\lambda}$. //

THEOREM 2.5: Suppose Assumption 2.3 holds. Necessary and sufficient conditions that $\xi^0 = (\underline{\xi}^0, \lambda^0)$ be a least favorable distribution and D^0 a minimax d.r. for discriminating among $\underline{\omega}$ are that

(i) $\underline{\xi}^0$ is a least favorable distribution and D^0 is a minimax d.r. w.r.t. W_{ij} for discriminating among $\underline{f}^{\lambda^0}$; and

(ii) for any i for which $\xi_i^0 > 0$,

$$\int_{\omega_i} r(\theta, D^0)\, d\lambda_i^0 = \sup_{\theta \in \omega_i} r(\theta, D^0) \quad .$$

Moreover, the maximum risks in the two cases are equal; i.e.,

$$\text{(2.12)} \qquad \sup_{(\bar{\ })} r(\theta, D^0) = \max_{1 \le i \le \ell} r(f_i^{\lambda^0}, D^0) \quad .$$

Proof: We first prove the necessity. Since (2.6) holds for any $\xi = (\underline{\xi}, \lambda)$, we have in particular $\inf_D r(\xi^0, D) \ge \inf_D r((\underline{\xi}, \lambda^0), D)$ for any $\underline{\xi}$, so that, using (2.11), $\inf_D r_{\lambda^0}(\underline{\xi}^0, D) \ge \inf_D r_{\lambda^0}(\underline{\xi}, D)$;

that is, $\underline{\xi}^0$ is a least favorable distribution for discriminating among $\underline{f}^{\lambda^0}$. By Theorem 2.2 (iii), D^0 is a Bayes d.r. relative to ξ^0, so that, by Theorems 2.4 and 1.2 (iii), D^0 is a minimax d.r. for discriminating among $\underline{f}^{\lambda^0}$.

We shall now verify (2.12). By Theorem 1.2 (iv),

$$\max_{1\leq i\leq \ell} r(f_i^{\lambda^0},D^0) = \sum_{i=1}^{\ell} \xi_i^0\, r(f_i^{\lambda^0},D^0) = r_{\lambda^0}(\underline{\xi}^0,D^0)$$

so that together with (2.11) and Theorem 2.2 (iii), we have

$$\max_{1\leq i\leq \ell} r(f_i^{\lambda^0},D^0) = r(\xi^0,D^0) = \sup_{\Omega} r(\theta,D^0) \quad .$$

For any i for which $\xi_i^0 > 0$, we have $r(f_i^{\lambda^0},D^0) = \max_i r(f_i^{\lambda^0},D^0)$ by Theorem 1.2 (iv) and $\sup_{\omega_i} r(\theta,D^0) = \sup_{\Omega} r(\theta,D^0)$ by Theorem 2.2 (iv), which, together with (2.12) and (2.10), prove (ii).

We now prove the sufficiency. By Theorems 1.2 (iii) and 2.4, D^0 is a Bayes d.r. relative to $\xi^0 = (\underline{\xi}^0,\lambda^0)$; i.e.,

$$r(\xi^0, D^0) = \inf_D r(\xi^0, D) \quad . \tag{2.13}$$

Suppose ξ^0 is not a least favorable distribution; then there exists a $\xi = (\underline{\xi}, \lambda)$ such that

$$\inf_D r(\xi^0, D) < \inf_D r(\xi, D) \quad . \tag{2.14}$$

But

$$\inf_D r(\xi, D) \leq r(\xi, D^0) = \sum_{i=1}^{\ell} \xi_i \int_{\omega_i} r(\theta, D^0)\, d\lambda_i \tag{2.15}$$

$$\leq \sum_{i=1}^{\ell} \xi_i \sup_{\omega_i} r(\theta, D^0) \leq \max_{1 \leq i \leq \ell} \sup_{\omega_i} r(\theta, D^0)$$

$$= \sup_{\Omega} r(\theta, D^0) \ .$$

By Theorem 1.2 (iv), for any i for which $\xi_i^0 > 0$, $r(f_i^{\lambda^0}, D^0) = \max_i r(f_i^{\lambda^0}, D^0)$, which, together with (2.10) and (ii) implies $\sup_{\omega_i} r(\Theta, D^0) = \max_i \sup_{\omega_i} r(\Theta, D^0) = \sup_{\Omega} r(\Theta, D^0)$. Hence, from (ii),

$$r(\xi^0, D^0) = \sum_i \xi_i^0 \int_{\omega_i} r(\Theta, D^0)\, d\lambda_i^0 = \sup_{\Omega} r(\Theta, D^0) ,$$

in contradiction to (2.13), (2.14), and (2.15). Hence, ξ^0 is a least favorable distribution, and, by Theorem 2.2 (iii), D^0 is a minimax d.r. for discriminating among $\underline{\omega}$. //

<u>THEOREM</u> 2.6: Suppose Assumption 2.3 holds, and suppose $\{\lambda^\nu\}$ is a sequence of sets of conditional a priori distributions and D^0 a d.r. such that

$$(2.16) \quad \lim_{\nu=\infty} \int_{\omega_i} r(\Theta, D^\nu)\, d\lambda_i^\nu = \sup_{\omega_i} r(\Theta, D^0) \quad (i = 1, \ldots, \ell)$$

where for each $\nu = 1, 2, \ldots, D^{\nu}$ is a minimax d.r. for discriminating among $\underline{f}^{\lambda^{\nu}}$. Then D^{0} is a minimax d.r. for discriminating among $\underline{\omega}$.

Proof: By Theorem 1.2 (ii) and (iii), for each ν there exists a least favorable distribution $\underline{\xi}^{\nu}$ and D^{ν} is a Bayes d.r. relative to $\underline{\xi}^{\nu}$ for discriminating among $\underline{f}^{\lambda^{\nu}}$; i.e., for any d.r. D, $r_{\lambda^{\nu}}(\underline{\xi}^{\nu},D^{\nu}) \leq r_{\lambda^{\nu}}(\underline{\xi}^{\nu},D)$. Hence, using (2.10),

$$(2.17) \qquad \sum_{i=1}^{\ell} \xi_i^{\nu} \int_{\omega_i} r(\theta,D^{\nu})\, d\lambda_i^{\nu} \leq \sum_{i=1}^{\ell} \xi_i^{\nu} \int_{\omega_i} r(\theta,D)\, d\lambda_i^{\nu}$$

$$\leq \sum_i \xi_i^{\nu} \sup_{\omega_i} r(\theta,D)$$

$$\leq \sup_{\underline{\Omega}} r(\theta,D) \quad .$$

Now each sequence $\{\xi_i^{\nu}\}$ has at least one limit point; let $\{\xi^{\nu_j}\}$,

$j = 1, 2, \ldots$, be a sub-sequence of $\{\xi^{\nu} = (\underline{\xi}^{\nu}, \lambda^{\nu})\}$ for which each $\xi_i^{\nu_j}$ converges to a limit, say ξ_i^0; then $\sum_i \xi_i^0 = 1$. By Theorem 1.2 (iv) and (2.10), for each i for which $\xi_i^{\nu} > 0$,

$$\int_{\omega_i} r(\Theta, D^{\nu})\, d\lambda_i^{\nu} = \max_i \int_{\omega_i} r(\Theta, D^{\nu})\, d\lambda_i^{\nu}$$

so that, from (2.16), for each i for which $\xi_i^0 > 0$, $\sup_{\omega_i} r(\Theta, D^0) = \max_i \sup_{\omega_i} r(\Theta, D^0) = \sup_{\Omega} r(\Theta, D^0)$. Hence, from (2.16),

$$\lim_{j=\infty} \sum_{i=1}^{\ell} \xi_i^{\nu_j} \int_{\omega_i} r(\Theta, D^{\nu_j})\, d\lambda_i^{\nu_j} = \sum_{i=1}^{\ell} \xi_i^0 \sup_{\omega_i} r(\Theta, D^0)$$

$$= \sup_{\Omega} r(\Theta, D^0) ,$$

which, together with (2.17), asserts $\sup_{\Omega} r(\Theta, D^0) \leq \sup_{\Omega} r(\Theta, D)$

for any D; i.e., D^0 is a minimax d.r. for discriminating among $\underline{\omega}$. //

If a least favorable distribution exists, the problem reduces to one of simple discrimination, so that if μ is non-atomic only non-randomized d.r.'s need be considered. We state a lemma for the case of composite discrimination analogous to Lemma 1.1; the proof (not given) is also analogous.

LEMMA 2.1: For every fixed sample size n = 0, 1, 2, ..., let D_n^0 be a minimax d.r. and denote $r_n = \sup_{\underline{\Omega}} r(\Theta, D_n^0)$. Then the sequence $\{r_n\}$, n = 0, 1, 2, ..., is a non-increasing sequence.

2.3 Most Economical Decision Rules Relative to a Vector $\underline{\alpha}$.

As in Sections 1.3.1 and 1.3.2, we shall apply the theory of Section 2.2 to two specific weight functions $W_j(\Theta)$ and develop in each case a method of obtaining M.E. d.r.'s as defined by Definition 2.1. In both cases, we assume $\ell = m$. First, let

$$(2.18)\quad W(\Theta,A_j)=W_j(\Theta) = \begin{cases} -\,1/\alpha_j & \text{if } \Theta \in \omega_j \\ 0 & \text{otherwise}^2. \end{cases}$$

From (2.4), the risk w.r.t. $W_j(\Theta)$ is

$$r(\Theta,D) = -p_i(\Theta,D)/\alpha_i \quad \text{if } \Theta \in \omega_i \quad (i = 1, \ldots, m),$$

and

$$(2.19)\qquad \sup_{\omega_i} r(\Theta,D) = -\inf_{\omega_i} p_i(\Theta,D)/\alpha_i \qquad (i = 1, \ldots, m)\ .$$

By Theorem 2.2, there exists a minimax d.r. D^0. We consider such d.r.'s for each sample size n and prove:

<u>THEOREM</u> 2.7: For each n = 0, 1, 2, ..., let D_n^0 be a minimax d.r. w.r.t. the weight function (2.18) for samples of fixed

---

2. The remarks in footnote 4, Chapter I, are applicable here as well.

size n. Suppose for some n, $\sup_{\Omega} r(\theta,D_n^0) \leq l$ and let N be the least such integer. Then D_N^0 is a M.E. d.r. relative to the vector $\underline{\alpha}$ for discriminating among $\underline{\omega}$. Conversely, if there exists a M.E. d.r. relative to $\underline{\alpha}$ for discriminating among $\underline{\omega}$ and the M.E. sample size is N, then D_N^0 is a M.E. d.r.

<u>Proof:</u> The proof is exactly like the proof of Theorem 1.3, replacing $p_i(D_n)$ by $\inf_{\omega_i} p_i(\theta,D_n)$. //

Note that the weight function (2.18) satisfies Assumption 2.3 with W_{ij} given by (1.6). Hence, if a least favorable distribution $\xi^0 = (\underline{\xi}^0,\lambda^0)$ exists, Theorems 2.4 and 2.5 imply that the composite discrimination problem may be treated as a simple discrimination problem with $f_i(x)=f_i^{\lambda^0}(x)=\int_{\omega_i} f(x,\theta)d\lambda_i^0$, and the theory of Chapter I will be applicable. If a least favorable distribution does not exist, Theorem 2.6 asserts that by a similar treatment for a sequence of a priori distributions having certain properties in the limit, it may be possible to solve the composite discrimination problem. Now suppose a least favorable distribution $\xi^0 = (\underline{\xi}^0,\lambda^0)$ exists.

Then, by Theorem 2.6,

$$(2.20) \quad \int_{\omega_i} p_i(\Theta, D^0)\, d\lambda_i^0 = \inf_{\Theta \epsilon \omega_i} p_i(\Theta, D^0) \text{ for any } i \text{ for which } \xi_i^0 > 0.$$

<u>ASSUMPTION</u> 2.4: If R is a subset of $\mathcal{X}$ for which $\int_R f(x,\Theta)d\mu = 0$ for some $\Theta \epsilon \Omega$, then $\int_R f(x,\Theta)\, d\mu = 0$ for all $\Theta \epsilon \Omega$.

This assumption implies Assumption 1.2 for the density functions $f_1^\lambda, \ldots, f_m^\lambda$, defined by (2.8), for any set of conditional distributions λ. If Assumption 2.4 holds, and if a least favorable distribution exists, it follows from Theorems 1.4 and 1.2 (iv) and (2.20) that

$$(2.21) \quad \frac{1}{\alpha_1} \inf_{\Theta \epsilon \omega_1} p_1(\Theta, D^0) = \ldots = \frac{1}{\alpha_m} \inf_{\Theta \epsilon \omega_m} p_m(\Theta, D^0)$$

where D^0 is a minimax d.r.

We now consider a second weight function $W_j(\Theta)$:

$$(2.22)\quad W(\Theta,A_j) = W_j(\Theta) = \begin{cases} 1/\beta_i & \text{if } \Theta\epsilon\omega_i, \quad i\neq j \\ 0 & \text{otherwise} \end{cases}$$

where $\beta_i = 1 - \alpha_i$ as before. From (2.4), the risk w.r.t. $W_j(\Theta)$ is

$$(2.23)\qquad r(\Theta,D) = q_i(\Theta,D)/\beta_i \quad \text{if } \Theta\epsilon\omega_i \quad (i = 1, \ldots, m),$$

and

$$\sup_{\omega_i} r(\Theta,D) = \sup_{\omega_i} q_i(\Theta,D)/\beta_i \quad (i = 1, \ldots, m).$$

By Theorem 2.2, there exists a minimax d.r. D^0. The proof of the following theorem is analogous to the proof of Theorem 2.7 and therefore will not be given.

THEOREM 2.8: For each $n = 0, 1, 2, \ldots,$ let D_n^0 be a minimax d.r. w.r.t. the weight function (2.22) for samples of fixed size n. Suppose for some n, $\sup_{\Omega} r(\Theta,D_n^0) \leq 1$ and let N be the least such integer. Then D_N^0 is a M.E. d.r. relative to the vector $\underline{\alpha}$ for discriminating among $\underline{\omega}$. Conversely, if there exists a M.E. d.r. relative to $\underline{\alpha}$ for discriminating among $\underline{\omega}$ and the M.E. sample size is N, then D_N^0 is a M.E. d.r.

Note that the weight function (2.22) satisfies Assumption 2.3 with W_{ij} given by (1.14). Hence, Theorems 2.4 to 2.6 and the theory of Chapter I may be applied to obtain M.E. d.r.'s.

Suppose a least favorable distribution $\xi^0 = (\underline{\xi}^0, \lambda^0)$ does exist. Then, by Theorem 2.5 and (2.23)

$$(2.24) \quad \int_{\omega_i} q_i(\theta, D^0) d\lambda_i^0 = \sup_{\omega_i} q_i(\theta, D^0) \text{ for any } i \text{ for which } \xi_i^0 > 0.$$

If Assumption 2.4 holds, if a least favorable distribution exists, and if

$$(2.25) \qquad \sup_{\underline{\Omega}} r(\theta, D^0) < 1 / \max_{1 \leq j \leq m} \beta_j$$

where D^0 is a minimax d.r., then by Lemma 1.2 and (2.24)

$$(2.26) \quad \frac{1}{\beta_1} \sup_{\Theta\epsilon\omega_1} q_1(\Theta,D^0) = \ldots = \frac{1}{\beta_m} \sup_{\Theta\epsilon\omega_m} q_m(\Theta,D^0) \quad .$$

The following two lemmas give sufficient conditions for (2.25) to hold.

<u>LEMMA</u> 2.2: Suppose Assumption 2.4 holds. For any sample size greater than or equal to the M.E. sample size, (2.25) holds.

<u>Proof:</u> Suppose $n \geq N$, the M.E. sample size, and that D_n^0 is a minimax d.r. for samples of size n; then, using Lemma 2.1 and Theorem 2.8, D_n^0 satisfies (2.2). This implies $q_i(\Theta,D_n^0) \leq \beta_i$ for all $\Theta\epsilon\omega_i$ and $\sup_{\Omega} r(\Theta,D_n^0) = \max_i \sup_{\omega_i} q_i(\Theta,D_n^0)/\beta_i \leq 1$ $< 1/\max_j \beta_j$. //

<u>LEMMA</u> 2.3: If $\beta_i < \sum_{j=1}^{m} \beta_j/(m-1)$ (i.e., $\alpha_i > (\Sigma\alpha_j - 1)/(m-1)$) for all i, then (2.25) holds.

Proof: The lemma may be proved analogously to Lemma 1.3. //

Thus, we have given two methods for obtaining M.E. d.r.'s relative to $\underline{\alpha}$ for composite discrimination. If Assumption 2.4 holds and if a least favorable distribution exists, the first method leads to M.E. d.r.'s for which the minimum probabilities of a correct decision are in the ratios $\alpha_1 : \alpha_2 : \ldots : \alpha_m$ (see (2.21)), and the second method leads to M.E. d.r.'s for which the maximum probabilities of an incorrect decision are in the ratios $\beta_1 : \beta_2 : \ldots : \beta_m$ (see (2.26)). This suggests an alternative approach: we may consider for each n the class of d.r.'s satisfying (2.21) (or (2.26)) and define an "optimum" d.r. from among this class as a d.r. which maximizes (or minimizes) the common ratio. By considering these "optimum" d.r.'s for n = 0, 1, 2, ..., and choosing the least n for which the ratio is $\leq$ (or $\geq$) 1, we can obtain M.E. d.r.'s. Results similar to Theorems 2.5 and 2.6, stated in terms of the weight function (2.18) (or (2.22)), may be derived as sufficient conditions for a d.r. to be optimum in the corresponding class. This is clearly equivalent to the minimax approach.

Thus, to obtain a M.E. d.r., we look for a set of "least favorable conditional distributions", or a sequence of distributions which is "least favorable in the limit", and then determine n and constants defining a likelihood ratio d.r. for discriminating among these "average" density functions for which (2.21) (or (2.26)) are satisfied.

Wald's definition of admissibility for either of the weight functions considered here is: a d.r. D is said to be admissible if there does not exist a d.r. D' for the same sample size for which $p_i(\Theta,D') \geq p_i(\Theta,D)$ for all $\Theta\epsilon\omega_i (i=1,\ldots,m)$ with strict inequality for at least one $\Theta\epsilon\Omega$. No proof of admissibility of the M.E. d.r.'s derived in this section has been obtained. However, if Assumption 2.4 holds and there exists a least favorable distribution, it can easily be verified that there does not exist a d.r. D_N' based on a sample of size N for which $\inf_{\omega_i} p_i(\Theta,D_N') \geq \inf_{\omega_i} p_i(\Theta,D_N)$ $(i = 1, \ldots, m)$ with strict inequality for at least one i, where D_N is a M.E. d.r. obtained by either of the minimax methods of this section.

2.4 Most Economical Decision Rules Relative to a Matrix β.

We now consider M.E. d.r.'s for composite discrimination as defined by Definition 2.2. Just as the approach of Section 1.3.2 was extended in Section 1.4, we shall extend the approach of Section 2.3 in this section. The argument is very brief because of this analogy. We replace each β_{ij} which is equal to unity by $+\infty$.

Suppose n fixed, and consider parameter spaces $\Theta_1, \ldots, \Theta_m$, each Θ_j being identical to Θ, and denote $\Theta' = \bigcup_j \Theta_j$. For each j, denote the corresponding subsets by $\omega_{1j}, \ldots, \omega_{\ell j}$. Define a weight function $W_k(\theta)$ for $k = 1, \ldots, m$, by

$$(2.27)\quad W(\theta, A_k) = W_k(\theta) = \begin{cases} 1/\beta_{ij} \text{ if } \theta \epsilon \omega_{ij} \text{ and } j=k (i=1,\ldots,\ell; j=1,\ldots,m) \\ 0 \quad \text{otherwise.} \end{cases}$$

We obtain

$$r(\theta, D) = \frac{1}{\beta_{ij}} p_j(\theta, D) \quad \text{if } \theta \epsilon \omega_{ij}.$$

Let ξ be an a priori distribution over Θ' and for any subset ω in Θ' denote

$$\xi(\omega) = \sum_{i,j} \xi_{ij} \lambda_{ij}(\omega)$$

where

$\xi_{ij} = \xi(\omega_{ij})$ and $\lambda_{ij}(\omega) = \Pr(\Theta\epsilon\omega|\Theta\epsilon\omega_{ij})$ $(i=1,\ldots,\ell;\ j=1,\ldots,m)$.

For a given set of conditional distributions $\lambda=(\lambda_{11},\lambda_{12},\ldots,\lambda_{\ell m})$, denote

$$(2.28) \qquad g_{ij}^{\lambda}(x) = \int_{\omega_{ij}} f(x,\Theta)\, d\lambda_{ij} \quad .$$

In a manner completely analogous to Sections 2.3 and 1.4, we have:

<u>THEOREM</u> 2.9: For each n = 0, 1, 2, ..., let D_n^0 be a minimax d.r. w.r.t the weight function (2.27) for samples of fixed size n. Suppose for some n, $\sup_{\underline{\Omega}'} r(\Theta,D_n^0) \leq 1$ and let N be the least such integer. Then D_N^0 is a M.E. d.r. relative to the matrix β for discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_\ell)$. Conversely, if there exists a M.E. d.r. relative to β for discriminating among $\underline{\omega}$ and the M.E. sample size is N, then D_N^0 is a M.E. d.r.

The theorems of Section 2.2 may be applied to obtain minimax d.r.'s for composite discrimination w.r.t. the weight function (2.27) by replacing ℓ in the theroems by $\ell' = \ell \cdot m$ and replacing single subscripts i by ij and f_i^λ by g_{ij}^λ. If a least favorable distribution exists, then the composite discrimination problem reduces to a problem of simple discrimination among the "average" density functions g_{ij}^λ defined by (2.28) w.r.t. a set of "least favorable conditional distributions" λ, and Theorem 1.11 and the remarks of Section 1.4 are applicable. Thus, this method of solution gives minimum unlikelihood d.r.'s as M.E. d.r.'s. If a least favorable distribution does not exist, then a minimax d.r. will be a Bayes d.r. in the wide sense and Theorem 2.6 may be applicable.

2.5 Some Parametric Examples of Three-Decision Rules.

EXAMPLE 1: Normal Mean, Variance Known. Suppose $f(x,\theta)$ is a normal density function with variance σ^2 (known) and mean θ, and $\omega_1 = \{\theta : \theta \leq \theta_1\}$, $\omega_2 = \{\theta : \theta_2' \leq \theta \leq \theta_2''\}$, $\omega_3 = \{\theta : \theta \geq \theta_3\}$, where $\theta_1 < \theta_2' \leq \theta_2'' < \theta_3$. Define a set of conditional distributions $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ over $\omega_1, \omega_2, \omega_3$, respectively, where $\lambda_i(\theta_i) = 1$ $(i = 1, 2, 3)$ and $\theta_2 = \theta_2'$ or θ_2'', to be determined later. For fixed n, we shall show that such a set of λ's is "least favorable" in the sense of Theorem 2.5 (ii), w.r.t. the weight function (2.18), and hence, this composite discrimination problem reduces to the simple discrimination problem considered in Example 1, Section 1.3.6, with θ_2 determined as follows:

$$(2.29) \qquad \begin{aligned} \theta_2 &= \theta_2' \text{ if } p_2(\theta_2', D') \leq p_2(\theta_2'', D') \\ \theta_2 &= \theta_2'' \text{ if } p_2(\theta_2'', D'') < p_2(\theta_2', D'') \end{aligned}$$

where D' and D'' are the solutions to the corresponding simple discrimination problems (with $\theta_2 = \theta_2'$ or θ_2'') for fixed n; we shall show below that such a determination of θ_2 is complete and consistent. We shall only treat this extension of Example 1 of Section 1.3.6, but the other three examples of that section may be extended in a completely analogous manner. Also, we may use the weight function (2.22) instead of (2.18) by making minor changes throughout.

We now show that λ satisfies Theorem 2.5 (ii) for fixed n. Let D, defined by c_1 and c_2, be a minimax d.r. for discriminating among θ_1, θ_2, θ_3, as given in Section 1.3.6; primes on any symbol refer to the corresponding value of Θ_2. Now since Φ is an increasing function and using (1.26),

$$\inf_{\omega_1} p_1(\Theta,D) = \inf_{\Theta \leq \Theta_1} \Phi\left(\sqrt{n}\,\frac{c_1-\Theta}{\sigma}\right) = \Phi\left(\sqrt{n}\,\frac{c_1-\Theta_1}{\sigma}\right)$$

$$= p_1(\Theta_1,D) = \int_{\omega_1} p_1(\Theta,D)\,d\lambda_1$$

and similarly

$$\inf_{\omega_3} p_3(\Theta,D) = 1 - \Phi\left(\sqrt{n}\,\frac{c_2-\Theta_3}{\sigma}\right) = \int_{\omega_3} p_3(\Theta,D)\,d\lambda_3 \quad .$$

Furthermore

$$\inf_{\omega_2} p_2(\Theta,D) = \inf_{\Theta_2' \leq \Theta \leq \Theta_2''} \left[\Phi\left(\sqrt{n}\,\frac{c_2-\Theta}{\sigma}\right) - \Phi\left(\sqrt{n}\,\frac{c_1-\Theta}{\sigma}\right) \right]$$

$$= \min \left[p_2(\Theta_2',D),\ p_2(\Theta_2'',D) \right]$$

since $p_2(\Theta,D)$ is an increasing function of Θ for $\Theta \leq (c_1+c_2)/2$ and a decreasing function of Θ for $\Theta \geq (c_1+c_2)/2$. Hence, if Θ_2 is chosen according to (2.29), Theorem 2.5 (ii) is satisfied.

We now show that such a choice of Θ_2 is possible by showing that if $p_2(\Theta_2'',D'') > p_2(\Theta_2',D'')$ then $p_2(\Theta_2'',D') > p_2(\Theta_2',D')$ and conversely. From (1.27), with either a prime or double-prime on ρ, D, c_1, and c_2, we have $\alpha_1\rho = \Phi(\sqrt{n}\,\frac{c_1-\Theta_1}{\sigma})$ and $\alpha_3\rho = 1 - \Phi(\sqrt{n}\,\frac{c_2-\Theta_3}{\sigma})$ so that

$$c_1 = \Theta_1 + \sigma\,\Phi^{-1}(\alpha_1\rho)/\sqrt{n}$$

$$c_2 = \Theta_3 + \sigma\,\Phi^{-1}(1-\alpha_3\rho)/\sqrt{n}$$

where $\Phi^{-1}(x) = t$ is defined by $\Phi(t) = x$. Hence

$$p_2(\Theta,D) = \Phi\left[\sqrt{n}\,\frac{\Theta_3-\Theta}{\sigma} + \Phi^{-1}(1-\alpha_3\rho)\right] - \Phi\left[\sqrt{n}\,\frac{\Theta_1-\Theta}{\sigma} + \Phi^{-1}(\alpha_1\rho)\right]$$

which is obviously a decreasing function of ρ for fixed Θ. Now $\alpha_2\rho' = p_2(\Theta',D')$ and $\alpha_2\rho'' = p_2(\Theta'',D'')$ so that, upon subtracting,

$$(2.30) \qquad \alpha_2(\rho'' - \rho') = p_2(\Theta_2'',D'') - p_2(\Theta_2',D').$$

Suppose $p_2(\Theta_2'',D'') > p_2(\Theta_2',D'')$ so that from (2.30), $\alpha_2(\rho'' - \rho') > p_2(\Theta_2',D'') - p_2(\Theta_2',D')$, implying that $\rho'' > \rho'$ since $p_2(\Theta_2',D)$ is a decreasing function of ρ. Then, since $p_2(\Theta_2'',D)$ is a decreasing function of ρ, we have from (2.30) that $0 < \alpha_2(\rho'' - \rho') < p_2(\Theta_2'',D') - p_2(\Theta_2', D')$. Conversely in the same manner, if $p_2(\Theta_2'',D') > p_2(\Theta_2', D')$, then from (2.30), $\alpha_2(\rho'' - \rho') > p_2(\Theta_2'',D'') - p_2(\Theta_2'',D')$, and ρ'' must be greater than ρ'; hence, from (2.30), $0 < \alpha_2(\rho'' - \rho') < p_2(\Theta_2'',D'') - p_2(\Theta_2',D'')$.

<u>EXAMPLE 2: Normal Mean, Variance Unknown.</u> Now suppose $f(x,\Theta)$, $\Theta = (\mu,\sigma)$, is a normal density function with both mean μ and variance σ^2 unknown, and suppose, given Θ_1, Θ_2', Θ_2'', Θ_3 ($\Theta_1 < \Theta_2' \leq \Theta_2'' < \Theta_3$), that $\omega_1 = \{\Theta\colon \mu/\sigma \leq \Theta_1,\ 0 \leq \sigma < \infty\}$, $\omega_2 = \{\Theta\colon \Theta_2' \leq \mu/\sigma \leq \Theta_2'',\ 0 \leq \sigma < \infty\}$, $\omega_3 = \{\Theta\colon \mu/\sigma \geq \Theta_3,\ 0 \leq \sigma < \infty\}$.

We shall show that, for a fixed sample size n, a non-randomized d.r. D^0 with acceptance regions of the form (1.26) with $\bar{x}$ replaced by Student's t-statistic is a minimax d.r. w.r.t. the weight function (2.18). c_1, c_2, and ρ are determined by equations of the form (1.27) with the normal distribution functions replaced by non-central t distribution functions. By considering such d.r.'s for various sample sizes, a M.E. d.r. may be obtained according to Theorem 2.7. (Alternatively, we may use the weight function (2.22) and Theorem 2.8.)

To prove that D^0 is minimax, we consider a sequence of distributions $\{\lambda^\nu\}$ and a corresponding sequence of minimax d.r.'s (D^ν) for discriminating among $\underline{f}^{\lambda^\nu}$, defined by (2.8), and apply Theorem 2.6. The methods of this example are adapted from Hoeffding's lecture notes [8] where he shows that a test based on Student's t maximizes the minimum power against a one-sided class of alternatives.

Let n be fixed throughout. For each $\nu = 1, 2, 3, \ldots$, consider a set of conditional distributions $(\lambda_1^\nu, \lambda_2^\nu, \lambda_3^\nu)$ where λ_i^ν assigns probability one to sets of Θ in which $\mu/\sigma = \theta_i$ $(i = 1, 2, 3)$, $\theta_2 = \theta_2'$ or θ_2'' to be determined later as in (2.29), and σ is distributed over $(0,\infty)$ according to the probability density

$$(2.31) \qquad \frac{c^m \tau^{m-1}}{\Gamma(m)} \; e^{-c\tau} \; d\tau, \quad 0 < \tau < \infty$$

where $\tau = 1/(2\sigma^2)$, $m = 1/\nu$, and c is a positive constant. For each ν(or m), a non-randomized likelihood ratio d.r. for discriminating among $(f_1^{\lambda^\nu}, f_2^{\lambda^\nu}, f_3^{\lambda^\nu})$ which satisfies (1.11) is a minimax d.r. w.r.t. the weight function (1.6) for simple discrimination. Denote such a d.r. by D^ν. Now D^ν is determined by the ratios $L_{ij}^{\lambda^\nu}(x) = f_j^{\lambda^\nu}(x)/f_i^{\lambda^\nu}(x)$ for various values of i, j (i < j). We have, using (2.8) and (2.31),

$$f_i^{\lambda^\nu}(x) = \pi^{-n/2} \, e^{-n\theta_i^2/2} \, \frac{c^m}{\Gamma(m)}$$

$$\cdot \int_0^\infty \tau^{n/2 + m-1} \exp\left[-\tau(c+\Sigma x^2)+\theta_i \Sigma x \sqrt{2\tau}\right] d\tau$$

$$(i = 1, 2, 3).$$

Let $u = \tau(c + \Sigma x^2)$ and denote $t_c(x) = \Sigma x / \sqrt{\Sigma x^2 + c}$; then

$$L_{ij}^{\lambda^\nu} = e^{n(\theta_i^2 - \theta_j^2)/2} \frac{\int_0^\infty u^{n/2 + m-1} \exp [-u + \theta_j t_c \sqrt{2u}\,] du}{\int_0^\infty u^{n/2 + m-1} \exp [-u + \theta_i t_c \sqrt{2u}\,] du} ,$$

a function of t_c only. By a theorem of Kruskal ([12], see his equation (4.2)), $L_{ij}^{\lambda^\nu}$ is an increasing function of t_c for $\theta_i < \theta_j$ $(i < j)$ so that Theorem 1.9 is applicable, and D^ν is of the form (1.26) with $\bar{x}$ replaced by t_c and the constants c_{1mc}, c_{2mc}, and ρ^ν are determined by the equations

$$(2.32) \qquad \alpha_i \rho^\nu = \int_{\omega_i} p_i(\theta, D^\nu)\, d\lambda_i^\nu \quad (i = 1, 2, 3).$$

When Θ_1 is true, $y_k = (x_k - \Theta_1\sigma)/\sigma$ is $N(0,1)$ so that

$$(2.33) \qquad p_1(\Theta_1, D^\nu) = \Pr [t_c(x) \leq c_{1mc} \mid \Theta_1]$$

$$= \Pr \left[\frac{\Sigma(y_k+\Theta_1)}{\sqrt{\Sigma(y_k+\Theta_1)^2+2c\tau}} \leq c_{1mc} \mid \Theta_1 \right]$$

$$= s_1(c\tau, c_{1mc}), \text{ say,}$$

which depends on σ through τ; and in a similar manner, using an obvious notation, $p_2(\Theta_2, D^\nu) = \Pr(c_{1mc} < t_c \leq c_{2mc} \mid \Theta_2)$ $= s_2(c\tau,\ c_{1mc},\ c_{2mc})$ and $p_3(\Theta_3,\ D^\nu) = \Pr(t_c > c_{2mc} \mid \Theta_3)$ $= s_3(c\tau,\ c_{2mc})$. From (2.31), (2.32), and (2.33),

$$(2.34) \quad \alpha_1\rho^\nu = \int_0^\infty s_1(c\tau, c_{1mc}) \frac{c^m \tau^{m-1}}{\Gamma(m)} e^{-c\tau} d\tau$$

$$= \int_0^\infty s_1(u, c_{1mc}) \frac{u^{m-1}}{\Gamma(m)} e^{-u} du ,$$

and similar equations may be obtained with $\alpha_2\rho^\nu$ and $\alpha_3\rho^\nu$ on the left. These are the equations determining c_{1mc}, c_{2mc}, and ρ^ν, and thus it is clear that c_{1mc} and c_{2mc} are independent of c, which therefore is arbitrary. Hereafter, we omit c as a subscript.

Now the distribution of t depends on θ only through μ/σ so that $p_i(\theta_i, D^0)$ is independent of σ $(i = 1, 2, 3)$. It may be verified as in Example 1 that

$$(2.35) \quad p_i(\theta_i, D^0) = \inf_{\omega_i} p_i(\theta, D^0) \qquad (i = 1, 2, 3)$$

where $\Theta_2 = \Theta_2'$ or Θ_2'', to be determined as in (2.29). (That such a choice of Θ_2 is possible may be proved as in the previous example.)

Student's t may be written $t(x) = \sqrt{(n-1)/n}\ \Sigma x/\sqrt{\Sigma(x-\bar{x})^2}$ $= \sqrt{(n-1)/n}\ t_0/\sqrt{1 - t_0^2/n}$ where $t_0 = \Sigma x/\sqrt{\Sigma x^2}$, and clearly t is an increasing function of t_0. Hence, the d. r. D^0 may be expressed in terms of t_0 rather than t with corresponding constants c_1' and c_2' defining the acceptance regions; that is, c_1', c_2', and ρ are determined by $p_i(\Theta_i, D^0)/\alpha_i = \rho$ $(i = 1, 2, 3)$; explicitly, using the notation introduced in (2.33) and the following equations,

$$(2.36)\quad \begin{cases} p_1(\Theta_1,D^0)=\Pr(t_0 \le c_1' \mid \Theta_1) = s_1(0,c_1') = \alpha_1\rho \\ \\ p_2(\Theta_2,D^0)=\Pr(c_1' < t_0 \le c_2' \mid \Theta_2) = s_2(0,c_1',c_2') = \alpha_2\rho \\ \\ p_3(\Theta_3,D^0)=\Pr(t_0 > c_2' \mid \Theta_3) = s_3(0,c_2') = \alpha_3\rho \quad . \end{cases}$$

We shall prove presently that

$$\lim_{\nu=\infty} \rho^{\nu} = \rho \quad . \tag{2.37}$$

Assuming it to be true temporarily, we have from (2.32) and (2.36)

$$\lim_{\nu=\infty} \int_{\omega_i} p_i(\Theta, D^{\nu})\, d\lambda_i^{\nu} = p_i(\Theta_i, D^0) \quad (i=1,2,3) \; . \tag{2.38}$$

From (2.38), (2.35), and (2.19), (2.16) is satisfied for the weight function (2.18) so that D^0 is a minimax d.r.

We need only prove that (2.37) holds. First, we prove

$$\lim_{m=\infty} c_{im} = c_i' \quad (i = 1, 2) \tag{2.39}$$

where c_1' and c_2' are defined by (2.36).

For any $v > 0$, we have

$$(2.40)\qquad \int_v^\infty \frac{u^{m-1}}{\Gamma(m)} e^{-u}\,du < \frac{1}{v}\int_v^\infty \frac{u^m}{\Gamma(m)} e^{-u}\,du$$

$$< \frac{1}{v}\int_0^\infty \frac{u^m}{\Gamma(m)} e^{-u}\,du = \frac{m}{v} .$$

Hence, remembering that $0 \le s_1 \le 1$, we have from (2.34)

$$(2.41)\quad \alpha_1\rho^\nu = \left(\int_0^v + \int_v^\infty\right) s_1(u,c_{1m}) \frac{u^{m-1}}{\Gamma(m)} e^{-u}\,du$$

$$= s_1(\delta_1 v, c_{1m}) \int_0^v \frac{u^{m-1}}{\Gamma(m)} e^{-u}du + \delta_1' \int_v^\infty \frac{u^{m-1}}{\Gamma(m)} e^{-u}\,du$$

$$(0 \le \delta_1,\ \delta_1' \le 1)$$

$$= s_1(\delta_1\ v, c_{1m})\ \left[1 - \int_v^\infty \frac{u^{m-1}}{\Gamma(m)}\ e^{-u}du\right] + \delta_1' \int_v^\infty \frac{u^{m-1}}{\Gamma(m)} e^{-u}du$$

$$= s_1(\delta_1 v, c_{1m}) + (\delta_1' - \delta_1'') \int_v^\infty \frac{u^{m-1}}{\Gamma(m)}\ e^{-u}du \quad (0 \leq \delta_1' \leq 1)$$

$$= s_1(\delta_1 v, c_{1m}) + \delta_1^*\ m/v \quad (|\delta_1^*| \leq 1) \text{ from (2.40).}$$

Let $v = \sqrt{m}$ so that, from (2.41),

$$\text{(2.42)} \qquad \rho^\nu = s_1(\delta_1 \sqrt{m},\ c_{1m})/\alpha_1 + \delta_1^* \sqrt{m}\ /\alpha_1 \quad ;$$

similarly, we may obtain

$$\text{(2.43)} \quad \rho^\nu = s_2(\delta_2 \sqrt{m}, c_{1m}, c_{2m})/\alpha_2 + \delta_2^* \sqrt{m}\ /\alpha_2 \ \ (0 \leq \delta_2 \leq 1,\ |\delta_2^*| \leq 1)$$

$$\text{(2.44)} \quad \rho^\nu = s_3(\delta_3 \sqrt{m}, c_{2m})/\alpha_3 + \delta_3^* \sqrt{m}\ /\alpha_3 \ \ (0 \leq \delta_3 \leq 1,\ |\delta_3^*| \leq 1).$$

Suppose, for at least one i (i = 1, 2), c_{im} does not tend to c_i'; e.g., suppose that $\limsup_{m=0} c_{1m} = c_1' + \epsilon$ ($\epsilon > 0$). Then there exists a sequence $\{m_j\}$, j = 1, 2, ..., such that $\lim_{j=\infty} m_j = 0$ and $\lim_{j=\infty} c_{1m_j} = c_1' + \epsilon$, and, since $s_1(u,c_1')$ is continuous in u and c_1', the right hand side of (2.42) tends to $s_1(0,c_1'+\epsilon)/\alpha_1$ and the left side tends to $\rho^* \equiv \limsup_{\nu=\infty} \rho^\nu$. Suppose $\limsup_{j=\infty} c_{2m_j} = c_2' + \delta$ ($\delta \gtreqless 0$). Then there exists a subsequence $\{m_{j_k}\}$, k = 1,2,..., redefined as $\{m_k\}$, such that $\lim_{k=\infty} c_{2m_k} = c_2' + \delta$ and the right sides of (2.43) and (2.44) tend to $s_2(0,c_1'+\epsilon,c_2'+\delta)/\alpha_2$ and $s_3(0,c_2'+\delta)/\alpha_3$, respectively, whereas the left sides both tend to ρ^*. But, from (2.36), it is seen that $s_1(0,c_1')$ is an increasing function of c_1', $s_2(0,c_1',c_2')$ is a decreasing function of c_1' and an increasing function of c_2', and $s_3(0,c_2')$ is a decreasing function of c_2'. Hence, $s_2(0,c_1',c_2')/\alpha_2 = s_1(0,c_1')/\alpha_1 < s_1(0,c_1'+\epsilon)/\alpha_1 = s_2(0,c_1'+\epsilon,c_2'+\delta)/\alpha_2$, implying $\delta > 0$, and $s_3(0,c_2')/\alpha_3 = s_1(0,c_1')/\alpha_1 < s_1(0,c_1'+\epsilon)/\alpha_1 = s_3(0,c_2'+\delta)/\alpha_3$, implying $\delta < 0$, a contradiction. Hence, $\limsup_{m=0} c_{1m} \leq c_1'$.

Similarly, we may show that $\limsup_{m=0} c_{2m} \leq c_2'$ and $\liminf_{m=0} c_{im} \geq c_i'$ (i=1,2). Consequently, (2.39) holds. Taking the limit in (2.42) and using (2.36), (2.37) is verified.

Note that in the sequence of distributions $\{\lambda^\nu\}$, the parameter c was left completely arbitrary (c > 0). Note also that the sequence of d.r.'s $\{D^\nu\}$ does not converge to D^0 if c is held fixed. This can be achieved, however, by letting c tend to zero as ν tends to infinity.

2.6 A Non-Parametric Example of a Three-Decision Rule.

We shall give an extension of the sign test for the median of an aribtrary distribution function by adapting an example given by Hoeffding in [7]. (See also Ruist [18] in this regard.) In an analogous manner, a M.E. d.r. concerning any quantile of an arbitrary distribution may be derived.

As was remarked in Section 2.1, the theory holds also for non parametric classes of density functions. Let $\underline{\Gamma}$ be the class of all density functions f w.r.t. a fixed measure μ on the real line such that $\mu\{x \leq 0\} > 0$, $\mu\{x > 0\} > 0$. Denote

$f^n(x) = \prod_{i=1}^{n} f(x_i)$ and $\Theta(f) = \int_{-\infty}^{0} f(x)\, d\mu$. Given

$\theta_1, \theta_2', \theta_2'', \theta_3$ $(0 < \theta_1 < \theta_2' \leq 1/2 \leq \theta_2'' < \theta_3 < 1)$, let $\omega_1 = \{f\colon \theta(f) \leq \theta_1\}$, $\omega = \{f\colon \theta_2' \leq \theta(f) \leq \theta_2''\}$, $\omega_3 = \{f\colon \theta(f) \geq \theta_3\}$. The alternatives, A_1, A_2, A_3, corresponding to ω_1, ω_2, ω_3, might be that the median of the unknown distribution is "significantly" less than zero, "close" to zero, "significantly" greater than zero, respectively.

Consider the density function

$$f(x,\theta) = \begin{cases} \theta^{b(x)}(1-\theta)^{1-b(x)}/A' & \text{if } |x| \leq A \\ 0 & \text{otherwise} \end{cases}$$

where $b(x) = 1$ if $x \leq 0$ and 0 otherwise, $A' = \mu\{(-A,0)\}$, and A is an arbitrary positive constant, and denote $f^n(x,\theta) = \prod_{i=1}^{n} f(x_i,\theta)$. Define a set of conditional distributions $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ over ω_1, ω_2, ω_3, respectively, where λ_i assigns probability 1 to $f(x,\theta_i)$ $(i = 1, 2, 3)$, hereafter denoted simply f_i, and where $\theta_2 = \theta_2'$ or θ_2'', to be determined later as in the parametric examples above. Note that $\Theta(f_i) = \int_{-\infty}^{0} f(x,\Theta_i)d\mu$ $= \Theta_i$ so that f_i is in ω_i $(i = 1, 2, 3)$. Consider the simple discrimination problem of discriminating among $f^n(x,\Theta_1)$, $f^n(x,\Theta_2)$, $f^n(x,\Theta_3)$, and let D_n be a minimax d.r. w.r.t. the weight function (1.6) for samples of fixed size n; clearly, D_n is a likelihood ratio d.r. Now $f^n(x,\Theta_i)$ is of the form (1.25) with $t(x) = \sum_k b(x_k)$ so that Theorem 1.9 is applicable, and D_n, defined by $\phi^n(x)$, is of the form (1.24) with $m = 3$.

Now t has a binomial distribution with parameter $\Theta = \Theta(f)$ and index n where f is the true density function. Clearly, then, for D_n defined above, $p_i(f,D_n)$ depends on f only through $\Theta = \Theta(f)$, so we denote $p_i(f,D_n) = p_i(\Theta,D_n)$ $(i = 1, 2, 3)$. Then, denoting the binomial distribution function and probability function by $B_{n,\Theta}$ and $b_{n,\Theta}$, respectively,

$$p_1(\Theta,D_n) = B_{n,\Theta}(c_1-1) + a_1\, b_{n,\Theta}(c_1)$$

$$p_2(\Theta,D_n) = B_{n,\Theta}(c_2-1)+a_2 b_{n,\Theta}(c_2)-B_{n,\Theta}(c_1)+(1-a_1)b_{n,\Theta}(c_1)$$

$$p_3(\Theta,D_n) = 1-B_{n,\Theta}(c_2)+(1-a_2)\, b_{n,\Theta}(c_2)$$

where a_1, a_2, c_1, c_2 depend on n and are determined so that $\inf_{f\epsilon\omega_i} p_i(f,D_n) = \alpha_i\rho$ ($i = 1, 2, 3$) for some ρ. Now $B_{n,\theta}(t)$ is a decreasing function of θ, and $b_{n,\theta}(t)$ is a decreasing or an increasing function of θ according as $t <$ or $> (n-1)$. Hence $p_1(\theta,D_n)$ is a decreasing function of θ using the fact that if $c_1 > (n-1)\ \theta$, p_1 may be written $B_{n,\theta}(c_1) - (1-a_1)\ b_{n,\theta}(c_1)$; similarly, $p_3(\theta,D_n)$ is an increasing function of θ. Since $\theta(f_i) = \theta_i$, clearly Theorem 2.5 (ii) for the weight function (2.18) is satisfied for $i = 1, 3$. Now $p_2(\theta,D_n)$ may be shown to have a maximum between $\theta = \theta_2'$ and $\theta = \theta_2''$, say at $\theta = \theta^0$, which is near $\theta = (c_1+c_2)/2(n-1)$ (if this point is between θ_2' and θ_2''; otherwise the maximum is at θ_2' or θ_2''), and to be an increasing or a decreasing function of θ according as $\theta <$ or $> \theta^0$. Hence, $\inf_{f\epsilon\omega_2} p_2(f,D_n) = \min [p_2(\theta_2',D_n),\ p_2(\theta_2'',D_n)]$, and since $\theta(f_2) = \theta_2$, by a proper choice of $\theta_2 = \theta_2'$ or θ_2'', Theorem 2.5 (ii) is satisfied for $i = 2$. That such a choice of θ_2 is possible may be proved as in Section 2.5.

Thus λ is "least favorable" in the sense of Theroem 2.5 (ii) and a M.E. d.r. may be obtained by solving $p_i(f_i, D_n) = \alpha_i \rho$ ($i = 1, 2, 3$) with $\rho = 1$ for n, c_1, c_2, a_1, a_2, and then with N equal to the least integer $\leq n$, re-solving for ρ, c_1, c_2, a_1, a_2; D_N is then a M.E. d.r. for discriminating among ω_1, ω_2, ω_3.

CHAPTER III

EXISTENCE THEOREMS FOR MOST ECONOMICAL MULTIPLE-DECISION RULES

3.1 Introduction.

In the two-decision case, it can easily be shown that the existence of a uniformly consistent sequence of d.r.'s (see Definition 3.3) implies the existence of M.E. d.r.'s. Berger [1] has given sufficient conditions for the existence of such sequences (in the two-decision case). Hoeffding [8], having defined non-trivial 2-d.r.'s for fixed sample sizes (see Definition 3.1) and given some sufficient conditions for their non-existence, has proved the existence of a uniformly consistent sequence from the existence of a non-trivial d.r. for some n by an adaptation of Berger's theorem. Berger and Wald [2] have given rather broad sufficient conditions for the existence of certain two-decision rules (what we shall define as strongly selective d.r.'s), the existence of which implies the existence of non-trivial 2-d.r.'s, as defined by Hoeffding, for any $n > 0$. Briefly, their conditions are that Ω_1 and Ω_2, defined below, be disjoint, assuming the existence of least favorable distributions[1]:

---

1. See Theorem 2.3 and the remarks following it.

$$\Omega_i = \{f(x) : f(x) = \int_{\omega_i} f(x;\theta)\, d\lambda_i\} \qquad (i = 1, 2)$$

where λ_i is any distribution function over ω_i. Thus, Berger and Wald's results supply sufficient conditions for the existence of a uniformly consistent sequence of 2-d.r.'s, and, therefore, for the existence of a M.E. 2-d.r.

We shall extend this work to the case of m-decision rules. The concept of non-triviality does not appear very fruitful in this case; instead, we define "strongly selective" m-d.r.'s and use them to prove that the existence of certain non-trivial 2-d.r.'s is both necessary and sufficient for the existence of a uniformly consistent sequence of m-d.r.'s. Finally, the existence of M.E. m-d.r.'s (relative to $\underline{\alpha}$ or β for $\ell = m$) is proved from the existence of a uniformly consistent sequence of m-d.r.'s. Thus, the existence of M.E. m-d.r.'s basically depends on the existence of certain non-trivial 2-d.r.'s, sufficient conditions for which are given by Berger and Wald, and some necessary conditions for which are given by Hoeffding's sufficient conditions for non-existence. Hoeffding's results are that if there exists a sequence $\{\lambda^\nu = (\lambda_1^\nu, \lambda_2^\nu)\}$, $\nu = 1, 2, \ldots$, of conditional distributions over ω_1 and ω_2, respectively, such that

$$\lim_{\nu=\infty} \int_{[f_1^{\lambda^\nu} > f_2^{\lambda^\nu}]} [f_1^{\lambda^\nu}(x) - f_2^{\lambda^\nu}(x)]\, d\mu = 0$$

where f_i^{λ} is defined by (2.8), then a non-trivial 2-d.r. for discriminating between ω_1 and ω_2 does not exist.

The results are derived for composite discrimination among parametric classes of density functions, $\omega_1, \ldots, \omega_m$, but they hold as well for more general classes of density functions and hence, in particular, for simple discrimination .

3.2 Non-Trivial, Selective, and Consistent Sequences of Decision Rules.

Many of the results of this section hold for sequences of random variables which are not necessarily identically distributed and some do not require independence, but since the results are derived primarily for application to most economical theory, we make both assumptions throughout.

We find it convenient throughout this section to specify a d.r. D by $\phi(x)$, sometimes adding a superscript n to denote the sample size.

DEFINITION 3.1: A d.r. $\phi(x) = [\phi_1(x), \ldots, \phi_m(x)]$ is said to be non-trivial for discriminating among $\omega_1, \ldots, \omega_m$ if the conditions

$$E_\Theta \phi_i(X) \geq \alpha_i \quad \text{for } \Theta \in \omega_i \qquad (i = 1, \ldots, m)$$

can be satisfied with some numbers $\alpha_1, \ldots, \alpha_m$ $(0 \leq \alpha_i \leq 1)$ such that $\sum_{i=1}^{m} \alpha_i > 1$; or, equivalently, if

$$\sum_{i=1}^{m} \inf_{\Theta \in \omega_i} E_\Theta \phi_i(X) > 1 \quad .$$

The term "non-trivial" is used since, if $\Sigma\, \alpha_i \leq 1$, the conditions can always be satisfied without taking any observatins. (See footnote 7, Chapter I.)

The existence of a non-trivial 2-d.r. for discriminating between ω_i and ω_j for some i, j $(i \neq j)$ implies the existence of a non-trivial m-d.r. for discriminating among $\omega_1, \ldots, \omega_m$ - no matter what the remaining ω_k's are; for, suppose $\phi_i(x) + \phi_j(x) = 1$ identically in x and $\inf_{\omega_i} E_\Theta \phi_i + \inf_{\omega_j} E_\Theta \phi_j > 1$; set $\phi_k(x) = 0$ identically in x for all $k \neq i$, j; then

$$\sum_{i=1}^{m} \phi_i(x) = 1 \quad \text{and} \quad \sum_{i=1}^{m} \inf_{\omega_i} E_\theta \phi_i(X) > 1 \; .$$

(A proof can also be given in which $\phi_k(x) > 0$ for $k \neq i$, j). Hence, the concept of non-trivial m-d.r.'s does not appear to be particularly useful except for $m = 2$. We shall introduce a slightly more restrictive class of d.r.'s:

DEFINITION 3.2: A d.r. $\underline{\phi}(x)$ is said to be selective (or weakly selective) for discriminating among $\underline{\omega}$ if

$$\inf_{\theta\epsilon\omega_i} E_\theta \phi_i(X) \geq \sup_{\theta\epsilon\omega_i} E_\theta \phi_j(X) \qquad \text{for all } j \neq i \qquad (i = 1, \ldots, m).$$

A d.r. $\underline{\phi}(x)$ is said to be strongly selective for discriminating among $\underline{\omega}$ if all the above inequalities hold strictly.

(We should be careful not to confuse selectivity with the similar concept of unbiasedness, which might be defined as implying

$$\inf_{\theta\epsilon\omega_i} E_\theta \phi_i(X) \geq \sup_{\theta\epsilon\omega_j} E_\theta \phi_i(X) \qquad \text{for all } j \neq i \qquad (i = 1, \ldots, m);$$

this definition reduces to what has been termed an unbiased test in the literature.) In the classical notation, strongly selective in the two-decision case means $\alpha < 1/2$, $\beta < 1/2$, whereas non-trivial means $\alpha + \beta < 1$.

It may easily be shown that selectivity does not imply non-triviality but strong selectivity does. We give some sufficient conditions for strong selectivity:

(1) $\sup_{\theta\epsilon\omega_i} E_\theta \phi_j(X) < 1/m$ for all i, j ($i \neq j$), since

$$\inf_{\omega_i} E_\theta \phi_i = \inf_{\omega_i} E_\theta (1 - \sum_{j\neq i} \phi_j) = 1 - \sup_{\omega_i} \sum_{j \neq i} E_\theta \phi_j$$

$$\geq 1 - \sum_{j\neq i} \sup_{\omega_i} E_\theta \phi_j > 1 - \frac{m-1}{m} = \frac{1}{m}$$

$$> \max_{j \neq i} \sup_{\omega_i} E_\theta \phi_j \qquad (i = 1, \ldots, m);$$

(2) $\inf_{\theta\epsilon\omega_i} E_\theta \phi_i(X) > 1/2$ for all i, since

$$\inf_{\omega_i} E_\theta \phi_i > 1/2 > 1 - \inf_{\omega_i} E_\theta \phi_i = \sup_{\omega_i} \sum_{j \neq i} E_\theta \phi_j$$

$$> \max_{j \neq i} \sup_{\omega_i} E_\theta \phi_j \qquad (i = 1, \ldots, m).$$

<u>DEFINITION</u> 3.3: The sequence $\{\underline{\phi}^n(x)\}$, n(sample size) = 0, 1, 2, ..., of d.r.'s is said to be uniformly consistent for discriminating among $\omega_1, \ldots, \omega_m$, if, for every $\alpha_1, \ldots, \alpha_m$ $(0 \leq \alpha_i < 1)$, there exists an $N = N(\underline{\alpha})$ such that for $n \geq N$, $E_\theta \phi_i^n(X) \geq \alpha_i$ for all $\theta \in \omega_i$ $(i = 1, \ldots, m)$; or, equivalently, if

$$\lim_{n = \infty} \inf_{\omega_i} E_\theta \phi_i^n(X) = 1 \qquad (i = 1, \ldots, m).$$

<u>THEOREM</u> 3.1: If there exists a strongly selective d.r. for discriminating among $\underline{\omega}$ for some n, then there exists a uniformly consistent sequence of d.r.'s for discriminating among $\underline{\omega}$.

<u>Proof</u>: Part 1: First, suppose for n = 1, a strongly selective d.r. $\underline{\phi}(x_1)$ exists; thus

$$(3.1) \qquad \inf_{\omega_i} E_\theta \phi_i(X_1) > \max_{j \neq i} \sup_{\omega_i} E_\theta \phi_j(X_1) \quad (i = 1, \ldots, m).$$

We define, for $n = 1, 2, \ldots$, the functions

$$\overline{\phi}_i^n(x) = \frac{1}{n} \sum_{j=1}^{n} \phi_i(x_j) \qquad (i = 1, \ldots, m),$$

and note the following properties of them (for any n, and $i, j = 1, \ldots, m$):

(i) $E_\theta \overline{\phi}_i^n(X) = E_\theta \phi_i(X_1)$

(ii) $\mathrm{Var}_\theta\, \overline{\phi}_i^n(X) = \frac{1}{n} \mathrm{Var}_\theta\, \phi_i(X_1) \leq \frac{1}{n}$

(iii) $\mathrm{Var}_\theta\, (\overline{\phi}_i^n - \overline{\phi}_j^n) \leq \mathrm{Var}_\theta\, \overline{\phi}_i^n + \mathrm{Var}_\theta\, \overline{\phi}_j^n + 2(\mathrm{Var}_\theta \overline{\phi}_i^n \cdot \mathrm{Var}_\theta \overline{\phi}_j^n)^{\frac{1}{2}}$

$$\leq 4/n \qquad (i \neq j).$$

We define, for each n, the d.r. $\underline{\psi}^n(x) = (\psi_1^n, \ldots, \psi_m^n)$:

$$\psi_i^n(x) = \begin{cases} \frac{1}{k+1} \text{ if } \overline{\emptyset}_i^n \geq \overline{\emptyset}_j^n & \text{for all } j \neq i \text{ with equality} \\ & \text{for } k \text{ values of } j \\ 0 \text{ otherwise} & (i = 1, \ldots, m). \end{cases}$$

Clearly, $\sum_{i=1}^{m} \psi_i^n(x) = 1$ identically in x. We have, for $i = 1,\ldots,m$,

$$E_\theta \psi_i^n(X) = \sum_{k=0}^{m-1} \frac{1}{k+1} P_\theta(\overline{\emptyset}_i^n \geq \overline{\emptyset}_j^n \text{ for all } j \neq i \text{ with equality for } k\ j\text{'s})$$

$$\geq P_\theta(\overline{\emptyset}_i^n > \overline{\emptyset}_j^n \text{ for all } j \neq i)$$

taking only the first term in the sum; hence, using (i),

$$E_\theta \psi_i^n(X) \geq 1 - P_\theta(\overline{\emptyset}_i^n \leq \overline{\emptyset}_j^n \quad \text{for at least one } j \neq i)$$

$$\geq 1 - \sum_{j \neq i} P_\theta(\overline{\emptyset}_i^n \leq \overline{\emptyset}_j^n)$$

$$= 1 - \sum_{j \neq i} P_\theta(\overline{\phi}_i^n - \overline{\phi}_j^n + E_\theta\overline{\phi}_j^n - E_\theta\overline{\phi}_i^n \leq$$

$$E_\theta\phi_j(X_1) - E_\theta\phi_i(X_1)\ ) .$$

Let $\delta = \min\limits_{1 \leq i \leq m} [\inf\limits_{\omega_i} E_\theta\phi_i(X_1) - \max\limits_{j \neq i} \sup\limits_{\omega_i} E_\theta\phi_j(X_1)]$. By (3.1), $\delta > 0$. Thus

$$\text{(3.2)} \quad \inf_{\omega_i} E_\theta\psi_i^n(X) \geq 1 - \sum_{j \neq i} P_\theta(-\overline{\phi}_i^n + \overline{\phi}_j^n - E_\theta\overline{\phi}_j^n + E_\theta\overline{\phi}_i^n \geq \delta)$$

$$\geq 1 - \sum_{j \neq i} P_\theta(|\ \overline{\phi}_i^n - \overline{\phi}_j^n - E_\theta(\overline{\phi}_i^n - \overline{\phi}_j^n)| \geq \delta)$$

$$\geq 1 - \sum_{j \neq i} \mathrm{Var}_\theta(\overline{\phi}_i^n - \overline{\phi}_j^n) \ \frac{1}{\delta^2}$$

by Tchebycheff's Inequality. By (3.2) and (iii), we have

$$\inf_{\omega_i} E_\theta \psi_i^n(X) \geq 1 - \frac{4(m-1)}{n\delta^2} \qquad (i = 1, \ldots, m) .$$

Hence, for $i = 1, \ldots, m$,

$$\lim_{n=\infty} \inf_{\omega_i} E_\theta \psi_i^n(X) = 1;$$

i.e., $\{\underline{\psi}^n(x)\}$ is a uniformly consistent sequence of d.r.'s for discriminating among $\underline{\omega}$.

Part 2: Now, suppose a strongly selective d.r., $\underline{\phi}(x_1, \ldots, x_\nu)$, exists for an integer ν. Take $n = \mu\nu$ for some integer μ, and define for $i = 1, \ldots, m$

$$\bar{\phi}_i^n(x_1, \ldots, x_n) = \frac{1}{\mu} [\phi_i(x_1, \ldots, x_\nu) + \phi_i(x_{\nu+1}, \ldots, x_{2\nu}) + \ldots +$$

$$\phi_i(x_{\mu\nu-\nu+1}, \ldots, x_{\mu\nu})] .$$

The remainder of the proof is analogous to that in Part 1. //

It may be shown that the above theorem does not hold for weakly selective d.r.'s. In the two-decision case, a more powerful theorem is possible:

<u>THEOREM</u> 3.2: If there exists a non-trivial 2-d.r. for discriminating between ω_i and ω_j $(i \neq j)$ for some n, then there exists a uniformly consistent sequence of 2-d.r.'s for discriminating between ω_i and ω_j.

This theorem is given in [8], being an adaptation of a theorem in [1], and will not be proved here. The proof of Theorem 3.1 given here is somewhat analogous to the proofs quoted above.

<u>THEOREM</u> 3.3: If there exists a non-trivial 2-d.r. for discriminating between ω_i and ω_j for some n_{ij} (sample size) for every i, j = 1, ..., m $(i \neq j)$, then, for some n, there exists a strongly selective m-d.r. for discriminating among $\omega_1, \ldots, \omega_m$.

<u>Proof</u>: Consider a particular pair i, j $(i \neq j)$. By Theorem 3.2, there exists a uniformly consistent sequence of d.r.'s for discriminating between ω_i and ω_j; denote such a sequence by $\overset{n}{\underline{\phi}}_{ij}(x) = (\overset{n}{\phi}_{i(j)}, \overset{n}{\phi}_{j(i)})$, n = 1, 2, ... This implies that there exists an N_{ij} such that for $n \geq N_{ij}$,

$$(3.3)\quad \inf_{\theta\in\omega_i} E_\theta \phi^n_{i(j)}(X) > \frac{m-1}{m}\,, \quad \inf_{\theta\in\omega_j} E_\theta \phi^n_{j(i)}(X) > \frac{m-1}{m}\,.$$

This is true for every pair i, j ($i \neq j$). Take $N = \max_{i,j} N_{ij}$. Then, for all $n \geq N$, (3.3) holds for every i, j = 1, ..., m ($i \neq j$). Consider some such n and define $\psi^n_i(x) = \frac{1}{\binom{m}{2}} \sum_{k\neq i} \phi^n_{i(k)}(x)$ ($i = 1, \ldots, m$). Now $\underline{\psi}^n(x) = (\psi^n_1, \ldots, \psi^n_m)$ is an m-d.r. since $0 \leq \psi_i \leq 1$ and

$$\sum_{i=1}^m \psi^n_i(x) = \frac{1}{\binom{m}{2}} \sum\sum_{\substack{i,k\\ i\neq k}} \phi^n_{i(k)}(x) = \frac{1}{\binom{m}{2}} \sum\sum_{k\neq i} \left[1 - \phi^n_{k(i)}(x)\right]$$

$$= 2 - \frac{1}{\binom{m}{2}} \sum\sum_{i\neq k} \phi^n_{i(k)}(x)$$

so that $\sum_{i=1}^m \psi^n_i(x) = 1$ identically in x. We have

$$(3.4) \qquad \inf_{\omega_i} E_\Theta \psi_i^n(X) = \frac{1}{\binom{m}{2}} \inf_{\omega_i} \sum_{k \neq i} E_\Theta \phi^n_{i(k)}(X)$$

$$\geq \frac{1}{\binom{m}{2}} \sum_{k \neq i} \inf_{\omega_i} E_\Theta \phi^n_{i(k)}(X)$$

$$> \frac{2}{m(m-1)} \quad (m-1) \quad \frac{m-1}{m} \quad \text{by } (3.3)$$

$$= \frac{2(m-1)}{m^2} \qquad (i = 1,\ldots,m);$$

furthermore, for $j \neq i$,

$$(3.5) \qquad \sup_{\omega_i} E_\Theta \psi_j^n(X) = \frac{1}{\binom{m}{2}} \sup_{\omega_i} \sum_{k \neq j} E_\Theta \phi^n_{j(k)}$$

$$\leq \frac{1}{\binom{m}{2}} \sum_{k \neq j} \sup_{\omega_i} E_\Theta \phi^n_{j(k)}$$

$$= \frac{1}{\binom{m}{2}} \left[\sum_{k\neq i,j} \sup_{\omega_i} E_\theta \phi^n_{j(k)} + \sup_{\omega_i} E_\theta (1-\phi^n_{i(j)}) \right]$$

$$\leq \frac{1}{\binom{m}{2}} \left[(m-2) + (1 - \inf_{\omega_i} E_\theta \phi^n_{i(j)}) \right]$$

$$< \frac{1}{\binom{m}{2}} (m - 1 - \frac{m-1}{m}) \quad \text{by (3.3)}$$

$$= \frac{2(m-1)}{m^2} \qquad (i = 1, \ldots, m) .$$

From (3.4) and (3.5) it follows that $\underline{\psi}^n(x)$ is a strongly selective d.r. for discriminating among $\underline{\omega}$. //

We shall now give a converse to Theorem 3.3:

<u>THEOREM</u> 3.4: If there exists a strongly selective m-d.r. for discriminating among $\omega_1, \ldots, \omega_m$, then there exists a non-trivial 2-d.r. for discriminating between ω_i and ω_j for every $i,j=1,\ldots,m$ $(i\neq j)$.

<u>Proof</u>: Suppose $\underline{\phi}(x) = (\phi_1, \ldots, \phi_m)$ is a strongly selective d.r. for discriminating among $\underline{\omega}$; i.e.,

$$(3.6) \qquad \inf_{\omega_i} E_\theta \phi_i > \sup_{\omega_i} E_\theta \phi_j \quad \text{for all } j \neq i \ (i = 1, \ldots, m) \ .$$

Consider some particular i, j ($i \neq j$), and suppose

$$(3.7) \qquad \sup_{\omega_i} E_\theta \phi_j \geq \sup_{\omega_j} E_\theta \phi_i \quad .$$

Now $\inf_{\omega_j} E_\theta(1-\phi_i) = 1 - \sup_{\omega_j} E_\theta \phi_i$, and upon adding this to (3.6), we obtain

$$\inf_{\omega_i} E_\theta \phi_i + \inf_{\omega_j} E_\theta(1-\phi_i) > 1 + \sup_{\omega_i} E_\theta \phi_j - \sup_{\omega_j} E_\theta \phi_i$$

$$\geq 1 \quad \text{by (3.7)} \ .$$

Hence, $(\phi_i, 1-\phi_j)$ is a non-trivial 2-d.r. for discriminating between ω_i and ω_j. If instead of (3.7), we have $\sup_{\omega_i} E_\theta \phi_j < \sup_{\omega_j} E_\theta \phi_i$, a similar argument will prove $(1-\phi_j, \phi_j)$ to be a non-trivial 2-d.r. for discriminating between ω_i and ω_j. This is true for every i, j $(i \neq j)$. //

<u>THEOREM</u> 3.5: A necessary and sufficient condition for the existence of a uniformly consistent sequence of m-d.r.'s for discriminating among $\omega_1, \ldots, \omega_m$ is that there exist non-trivial 2-d.r.'s for discriminating between ω_i and ω_j for some n_{ij} (sample size) for every $i, j = 1, \ldots, m (i \neq j)$.

<u>Proof:</u> The sufficiency follows directly from Theorems 3.3 and 3.1. To prove the necessity, suppose $\underline{\phi}^n = (\phi_1^n, \ldots, \phi_m^n)$, $n = 1, 2, \ldots,$ is a uniformly consistent sequence of d.r.'s for discriminating among $\omega_1, \ldots, \omega_m$. Then let N be an integer such that for $n \geq N$, $\inf_{\omega_i} E_\theta \phi_i^n > 1/2$ for $i = 1, \ldots, m$. By the second sufficient condition for strong selectivity given early in this section, and Theorem 3.4, the proof is completed. //

For some sufficient conditions for the existence, and for the non-existence, of non-trivial 2-d.r.'s, see the remarks in Section 3.1.

We shall now consider the case of simple discrimination. We say that two density functions $f(x)$ and $g(x)$ w.r.t. a measure μ are "distinct" if the set of all x for which $f(x) \neq g(x)$ has positive μ-measure; and a set of density functions is said to be "distinct" if every pair in the set is distinct. Let $f_1, \ldots, f_m$ be density functions w.r.t. a measure μ; we have the theorem:

THEOREM 3.6: A necessary and sufficient condition for the existence of a uniformly consistent sequence of m-d.r.'s for discriminating among $f_1, \ldots, f_m$ is that $f_1, \ldots, f_m$ be distinct.

Proof: The distinctness of any pair f_i, f_j $(i \neq j)$ implies the existence of a non-trivial 2-d.r. for discriminating between f_i and f_j for some n_{ij} (sample size) by Wald and Berger's theorem quoted in Section 3.1. The sufficiency part of this theorem follows, then, as a special case of Theorem 3.5 which is true for general classes of density functions as well as for parametric classes.

The necessity is proved as follows: choose an $\epsilon (0 < \epsilon < 1/2)$. Then, for some $n = n_\epsilon$, there exists a d.r. $\phi(x)$ such that

$$(3.8) \qquad E_i\phi_i(X) \geq 1 - \epsilon > 1/2 \quad \text{for all } i,$$

where the subscript on the expectation operator refers to the corresponding density function as in Chapter I. Now suppose for some i, j $(i \neq j)$ f_i and f_j are not distinct. Then

$$E_i\phi_i(X) + E_j\phi_j(X) = E_i\phi_i(X) + E_i\phi_j(X) \leq \sum_{k=1}^{m} E_i\phi_k(X) = 1,$$

in contradiction to (3.8). Hence, f_i and f_j are distinct for all i,j $(i \neq j)$. //

3.3 Existence Theorems for Most Economical Decision Rules.

We assume throughout this section that $\ell = m$ and that a pair i, j corresponds to a correct decision if $i = j$ and an incorrect decision if $i \neq j$.

THEOREM 3.7: A necessary and sufficient condition for the existence of a M.E. d.r. relative to any vector $\underline{\alpha}$ for discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_m)$ is that there exists a uniformly consistent sequence of m-d.r.'s for discriminating among $\underline{\omega}$.

Proof: The theorem is obvious from the definitions involved; however, a formal proof may be given analogous to the proof of Theorem 3.8. //

THEOREM 3.8: A necessary and sufficient condition for the existence of a M.E. m-d.r. relative to any matrix β for discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_m)$ is that there exists a uniformly consistent sequence of m-d.r.'s for discriminating among $\underline{\omega}$.

Proof: Part I: Sufficiency: Suppose there exists a uniformly consistent sequence of d.r.'s $\{D_n\}$. Then, for any positive $\epsilon \leq 1$, there exists an $N = N_\epsilon$ such that for $n \geq N_\epsilon$ $\inf\limits_{\omega_i} p_i(\Theta, D_n) \geq 1 - \epsilon$ $(i = 1, \ldots, m)$. Hence, for all i, j $(i \neq j)$,

$$\sup_{\omega_i} p_j(\Theta, D_n) \leq \sup_{\omega_i} \sum_{j \neq i} p_j(\Theta, D_n) = 1 - \inf_{\omega_i} p_i(\Theta, D_n) \leq \epsilon \quad .$$

Given any $\beta = (\beta_{ij})$, let $\Sigma = \min\limits_{\substack{i,j \\ i \neq j}} \beta_{ij}$. Then, for $n \geq N_\epsilon$, D_n satisfies (2.3). Since (2.3) can be satisfied for some n, there exists a least n for which it may be satisfied; i.e., there exists a M.E. d.r. relative to β.

Part 2: Necessity: Suppose, given any $\beta(0 < \beta_{ij} \leq 1)$, there exists a M.E. d.r. relative to β. Given $\epsilon \leq 1$, let

$$\beta_{ij\epsilon} = \begin{cases} 1 & \text{if } i = j \\ \epsilon/(m-1) & \text{if } i \neq j \quad (i, j = 1, \ldots, m), \end{cases}$$

and let D^{ϵ} be a M.E. d.r. relative to β_{ϵ}. Then, since $\sup\limits_{\omega_i} p_j(\theta,D^{\epsilon}) \leq \epsilon/(m-1)$ for all i, j ($i \neq j$), we have

$$(3.9) \quad \inf_{\omega_i} p_i(\theta,D^{\epsilon}) = 1 - \sup_{\omega_i} \sum_{j\neq i} p_j(\theta,D^{\epsilon}) \geq 1 - \sum_{j\neq i} \sup_{\omega_i} p_j(\theta,D^{\epsilon})$$

$$\geq 1-\epsilon \quad (i = 1, \ldots, m).$$

Let $\{\epsilon_{\nu}\}$, $\nu = 1, 2, \ldots$, be a decreasing sequence of positive constants converging to zero. Let $D^{\epsilon_{\nu}}$ be a M.E. d.r. relative to $(\beta_{ij\epsilon_{\nu}})$ defined above and let N_{ν} be the corresponding M.E. sample size. Now $\{N_{\nu}\}$ is a non-decreasing sequence since, for

$\mu < \nu$, N_μ is the least integer for which $\sup\limits_{\omega_i} p_j(\theta,D) \leq \beta_{ij\epsilon_\mu}$ $= \epsilon_\mu/(m-1)$ $(i, j = 1, \ldots, m;\ i \neq j)$ can be satisfied by some D, and N_ν is the least integer for which

$$\sup_{\omega_i} p_j(\theta,D) \leq \epsilon_\nu/(m-1) < \epsilon_\mu/(m-1) \quad (i, j = 1, \ldots, m; i \neq j)$$

can be satisfied by some D, and hence $N_\mu \leq N_\nu$. We shall suppose the sequence $\{N_\nu\}$ does not contain any integer more than once, for if it does we may delete some terms from $\{\epsilon_\nu\}$ and re-number the subscripts so that it will not. Consider the sequence $\{D_n\}$, $n = n_0, n_0 + 1, \ldots$, where $n_0 = \min\limits_\nu \{N_\nu\}$ and where $D_n = D^{\epsilon_\nu}$ if $N_\nu \leq n < N_{\nu+1}$. Hence, using (3.9), we have

$$\lim_{n=\infty} \inf_{\omega_i} p_i(\theta,D_n) = \lim_{\nu=\infty} \inf_{\omega_i} p_i(\theta,D^{\epsilon_\nu})$$

$$\geq \lim_{\nu=\infty} (1-\epsilon_\nu) = 1 \ (i = 1, \ldots, m);$$

i.e., $\{D_n\}$ is a uniformly consistent sequence of d.r.'s. //

Theroems 3.7 and 3.8 hold, of course, for the analogous cases of simple discrimination as well.

According to these two theorems, the necessary and sufficient conditions for the existence of a uniformly consistent sequence of d.r.'s given by Theorems 3.5 and 3.6 provide sufficient conditions for the existence of a M.E. d.r. relative to any specific $\underline{\alpha}$ or β and necessary conditions for the existence of M.E. d.r.'s relative to every $\underline{\alpha}$ or β. Thus we have existence theorems for M.E. d.r.'s defined by Definitions 1.1, 1.2, 2.1, and 2.2 (for $\ell = m$), and hence sufficient conditions for the assumptions of Theorems 1.3, 1.5, 1.7, 1.10, 2.7, 2.8, and 2.9, deriving M.E. d.r.'s from sequences of minimax d.r.'s for fixed sample sizes, to be satisfied.

CHAPTER IV

MOST ECONOMICAL DECISION FUNCTIONS

4.1 Introduction. In this chapter we extend the concept of M.E. d.r.'s to more general decision problems. We assume the formulation of the statistical decision problem as given by Wald ([22], Chapter 1) in its complete generality, except for one modification: we shall be concerned with two pairs of loss and cost functions. For conciseness, we denote the sum of one pair (one loss function and one cost function) by W_1 and the sum of the other pair by W_2. (In applying the results, we shall suppose that one of the W_i's is simply a loss function and the other a cost function, but we state the problem in this manner for symmetry and generality.) W_1 and W_2 are referred to as "weight functions". In considering risk functions, we denote

$$r^i(F, \delta) = r_1^i(F, \delta) + r_2^i(F, \delta) \quad (i = 1, 2)$$

where r_1^i and r_2^i correspond to Wald's r_1 and r_2, the i designating that r^i is the risk w.r.t. the weight function W_i. We assume Wald's subsequent definitions, notations, and theorems throughout. All references to Wald refer to [22] unless otherwise specified.

Let $\mathcal{D}$ denote the class of all decision functions at the disposal of the experimenter, and define the following subclasses:

$$\mathcal{D}^* = \{ \delta \in \mathcal{D} : \sup_{F \in (\ )} r^1(F, \delta) \leq 1 \}$$

$$\mathcal{D}_r = \{ \delta \in \mathcal{D} : \sup_{F \in (\ )} r^2(F, \delta) \leq r \}$$

for any non-negative real r.

We shall consider the following problem: to find a minimax solution w.r.t. W_2 relative to the class $\mathcal{D}^*$; that is, to find a decision function δ which minimizes the maximum risk w.r.t. the weight function W_2 subject to the condition that the risk w.r.t. the weight function W_1 is nowhere greater than unity.

Blyth [3] considered this problem (with minor modifications) and proved, under suitable conditions, that a minimax solution, δ^0, w.r.t. $cW_1 + W_2$ (relative to $\mathcal{D}$) where c is chosen so that $\sup_{F \in (\ )} r^1(F, \delta^0) = 1$ is a solution to the problem. His conditions are that there exists a class C of minimax solutions δ w.r.t. $cW_1 + W_2$ (for some c) for which

$$\sup_{F \in (\ )} [r^1(F,\delta) + r^2(F,\delta)] = \sup_{F \in (\ )} r^1(F,\delta) + \sup_{F \in (\ )} r^2(F,\delta)$$

for all $\delta \in C$, and that for every value L between the minimum and maximum of W_1 (over $\mathcal{X}$, $\underline{(\,)}$, D^t, s), there exists a $\delta_L \in C$ for which $\sup_{F \in \underline{(\,)}} r^1(F, \delta_L) = L$.

We shall consider a different approach. We prove under very general assumptions that a minimax solution w.r.t. W_1 relative to $\mathcal{D}_{r_0}$ is such a decision function, where r_0 is the minimum r for which a minimax solution w.r.t. W_1 relative to $\mathcal{D}_r$ is in $\mathcal{D}^*$, and shall also give sufficient conditions for the existence of r_0 and of such minimax solutions.

4.2 Preliminary Theory. Let δ_r denote any minimax solution (if existent) w.r.t. W_1 relative to $\mathcal{D}_r$, and define

$$\mathcal{D}^0 = \{\delta_r \ (r \geq 0) : \delta_r \in \mathcal{D}^*\}$$

and

$$\mathcal{D}^{0'} = \{\delta_r \in \mathcal{D}^0 : \sup_{\underline{(\,)}} r^2(F, \delta_r) = r\} .$$

LEMMA 4.1: Suppose there exists a minimax solution w.r.t. W_1 relative to $\mathcal{D}_r$ for every r for which $\mathcal{D}_r$ is non-null. Then $\mathcal{D}^{0'} = \mathcal{D}^0$.

<u>Proof</u>: Every $\delta \in \mathcal{D}^{0'}$ is also in $\mathcal{D}^0$, so that we need only show that every $\delta \in \mathcal{D}^0$ is also in $\mathcal{D}^{0'}$.

Consider an arbitrary element δ' of $\mathcal{D}^0$; then δ' must be a minimax solution w.r.t. W_1 relative to some $\mathcal{D}_r$, say $\mathcal{D}_{r'}$. If $\sup\limits_{\Omega} r^2(F, \delta') = r'$, then $\delta' \in \mathcal{D}^{0'}$.

Now suppose $\sup\limits_{\Omega} r^2(F, \delta') = r'' < r'$. Then $\delta' \in \mathcal{D}_{r''}$ so that

$$(4.1) \qquad \inf_{\mathcal{D}_{r'}} \sup_{\Omega} r^1(F, \delta) = \sup_{\Omega} r^1(F, \delta') \quad \inf_{\mathcal{D}_{r''}} \sup_{\Omega} r^1(F, \delta).$$

But $\mathcal{D}_{r''} \subsetneq \mathcal{D}_{r'}$ since $r'' < r'$ so that

$$(4.2) \qquad \inf_{\mathcal{D}_{r'}} \sup_{\Omega} r^1(F, \delta) \le \inf_{\mathcal{D}_{r''}} \sup_{\Omega} r^1(F, \delta).$$

From (4.1) and (4.2) we have that δ' is a minimax solution w.r.t. W_1 relative to $\mathcal{D}_{r''}$; but $\sup\limits_{\Omega} r^2(F, \delta') = r''$ so that $\delta' \in \mathcal{D}^{0'}$. //

According to Lemma 4.1, under certain conditions we may consider δ_r as a minimax solution w.r.t. W_1 relative to the class of all decision functions in $\mathcal{D}$ for which $\sup_{\underline{\Gamma}} r^2(F, \delta) = r$.

LEMMA 4.2: Suppose there exists a minimax solution w.r.t. W_2 relative to $\mathcal{D}^*$ and a minimax solution w.r.t. W_1 relative to $\mathcal{D}_r$ for every r for which $\mathcal{D}_r$ is non-null, and denote

$$(4.3) \qquad r^* = \inf_{\mathcal{D}^*} \sup_{\underline{\Gamma}} r^2(F, \delta).$$

Then (i) any minimax solution w.r.t. W_1 relative to $\mathcal{D}_{r^*}$ is a minimax solution w.r.t. W_2 relative to $\mathcal{D}^0$, and conversely; and (ii) any minimax solution w.r.t. W_2 relative to $\mathcal{D}^0$ is a minimax solution w.r.t. W_2 relative to $\mathcal{D}^*$.

Proof: Let δ^* be a minimax solution w.r.t. W_2 relative to $\mathcal{D}^*$. Now $\mathcal{D}^*_r$ is non-null since $\delta^* \varepsilon \mathcal{D}_{r^*}$. Let δ_{r^*} be a minimax solution w.r.t. W_1 relative to $\mathcal{D}_{r^*}$. Since $\mathcal{D}^0 \subseteq \mathcal{D}^*$ and $\delta_{r^*} \varepsilon \mathcal{D}_{r^*}$,

$$(4.4) \quad \inf_{\mathcal{D}^0} \sup_{\underline{\Gamma}} r^2(F, \delta) \geq \inf_{\mathcal{D}^*} \sup_{\underline{\Gamma}} r^2(F, \delta) = r^* \geq \sup_{\underline{\Gamma}} r^2(F, \delta_{r^*}).$$

Now $\delta^* \varepsilon \mathcal{D}^* \cap \mathcal{D}_{r^*}$ so that

$$1 \geq \sup_{\Omega} r^1(F, \delta^*) \geq \inf_{\mathcal{D}^*} \sup_{\Omega} r^1(F, \delta) = \sup_{\Omega} r^1(F, \delta_{r^*}).$$

Therefore, $\delta_{r^*} \varepsilon \mathcal{D}^*$ and hence $\delta_{r^*} \varepsilon \mathcal{D}^0$. Hence, the equality signs must hold in (4.4), and δ_{r^*} is a minimax solution w.r.t. W_2 relative to $\mathcal{D}^0$, proving the first part of (i).

Since equality must hold in (4.4), we have $\inf_{\mathcal{D}^0} \sup_{\Omega} r^2(F,\delta)$ $= \inf_{\mathcal{D}^*} \sup_{\Omega} r^2(F, \delta)$, and therefore, since $\mathcal{D}^0 \subseteq \mathcal{D}^*$, (ii) is proved.

By Lemma 4.1, we may replace $\mathcal{D}^0$ by $\mathcal{D}^{0'}$. Let δ' be a minimax solution w.r.t. W_2 relative to $\mathcal{D}^{0'}$. Since $\delta' \varepsilon \mathcal{D}^{0'}$, it must be a minimax solution w.r.t. W_1 relative to some $\mathcal{D}_r$, say $\mathcal{D}_{r'}$, and $\sup_{\Omega} r^2(F, \delta') = r'$. By (ii), $\sup_{\Omega} r^2(F, \delta')$ $= \inf_{\mathcal{D}^*} \sup_{\Omega} r^2(F, \delta)$; that is, $r' = r^*$ and $\mathcal{D}_{r'} = \mathcal{D}_{r^*}$. Thus, δ' is a minimax solution w.r.t. W_1 relative to $\mathcal{D}_{r^*}$, proving the converse of (i). //

<u>THEOREM</u> 4.1: Suppose there exists a minimax solution w.r.t. W_2 relative to $\mathcal{D}^*$ and a minimax solution δ_r w.r.t. W_1 relative to $\mathcal{D}_r$ for every r for which $\mathcal{D}_r$ is non-null. Then $\min\{r : \delta_r \in \mathcal{D}^*\} = r_0$, say, exists and δ_{r_0} is a minimax solution w.r.t. W_2 relative to $\mathcal{D}^*$, i.e.,

$$(4.5) \qquad \sup_{\Omega} r^2(F, \delta_{r_0}) = \inf_{\mathcal{D}^*} \sup_{\Omega} r^2(F, \delta).$$

Moreover, if δ^* is any other minimax solution w.r.t. W_2 relative to $\mathcal{D}^*$, then

$$(4.6) \qquad \sup_{\Omega} r^1(F, \delta_{r_0}) \leq \sup_{\Omega} r^1(F, \delta^*).$$

<u>Proof</u>: Now $\inf\limits_{\mathcal{D}^{0'}} \sup\limits_{\Omega} r^2(F, \delta) = \inf\{r : \delta_r \in \mathcal{D}^*\}$. By Lemma 4.2 (i) and Lemma 4.1, there exists a minimax solution w.r.t. W_2 relative to $\mathcal{D}^0$ or $\mathcal{D}^{0'}$. Hence, $\min\{r : \delta_r \in \mathcal{D}^*\}$ exists and δ_{r_0} is a minimax solution w.r.t. W_2 relative to $\mathcal{D}^0$. Lemma 4.2 (ii) completes the proof of (4.5).

By Lemma 4.2 (i), $\sup_{\Omega} r^1(F, \delta_{r_0}) = \inf_{\mathcal{D}_{r^*}} \sup_{\Omega} r^1(F, \delta)$

where r^* is defined by (4.3), and since $\delta^* \in \mathcal{D}_{r^*}$,

$\inf_{\mathcal{D}_{r^*}} \sup_{\Omega} r^1(F, \delta) \leq \sup_{\Omega} r^1(F, \delta^*)$, thus proving (4.6). //

Clearly, if $\mathcal{D}$ satisfies Wald's Assumptions 3.1 to 3.5 then any subset of $\mathcal{D}$ also satisfies them. Konijn [11] has proved that if $\mathcal{D}$ satisfies his formulation of Wald's Assumption 3.6, then $\mathcal{D}^*$ and $\mathcal{D}_r$ (for any non-negative real r) satisfy this assumption. Then, by Wald's Theorem 3.7, asserting the existence of a minimax solution under his Assumptions 3.1 to 3.6, we have Theorem 4.2 stated below.

<u>ASSUMPTION</u> 4.1: The stochastic process $X = (X_1, X_2, \ldots)$ underlying the decision problem, the class Ω of possible distribution functions of X, the space D^t of possible terminal decisions, both weight functions W_1 and W_2, and the class $\mathcal{D}$ of decision functions at the disposal of the experimenter satisfy Wald's Assumptions 3.1 to 3.5 and Konijn's formulation of Wald's Assumption 3.6.[1]

1. Some of these assumptions may be relaxed somewhat; e.g., see Ghosh [6] and Lehmann [14].

THEOREM 4.2: Suppose Assumption 4.1 holds. Then

(i) if $\mathcal{D}^*$ is non-null, there exists a minimax solution w.r.t. W_2 relative to the class $\mathcal{D}^*$; and

(ii) for any r for which $\mathcal{D}_r$ is non-null, there exists a minimax solution w.r.t. W_1 relative to $\mathcal{D}_r$.

Moreover, minimax solutions relative to $\mathcal{D}^*$ and $\mathcal{D}_r$ are characterized according to Wald's general theory as given in [22], Chapter 3.

4.3 Most Economical Decision Functions. Let $c(x; s)$ be a cost function as defined by Wald and denote the expected cost function when using δ by $r_2(F, \delta)$. Let $w(F, d^t)$ be a loss function as defined by Wald and denote the expected loss function when using δ by $r_1(F, \delta)$. Let $\beta(F)$ be a given positive-valued function defined for all $F \in \Omega$.

DEFINITION 4.1: A decision function δ^0 is said to be most economical (relative to the class $\mathcal{D}$) if it satisfies

$$(4.7) \qquad r_1(F, \delta) \leq \beta(F) \qquad \text{for all } F \in \Omega$$

and if, for any other δ satisfying (4.8), $\sup_{F\in\Omega} r_2(F,\delta^0) \leq \sup_{F\in\Omega} r_2(F,\delta)$.

Thus, a most economical (M.E.) decision function minimizes the maximum expected cost subject to upper bounds on the expected loss.

Define two weight functions:

$$(4.8) \qquad W_1 = w(F, d^t)/\beta(F), \quad W_2 = c(x; s).$$

Clearly, the problem of finding a most economical decision function is simply the problem introduced in Section 4.1 with W_1 and W_2 defined by (4.8). Theorem 4.1 gives a method of obtaining such decision functions: Define the class $\mathcal{D}_r$ of decision functions for which the expected cost is nowhere greater than r and obtain a minimax solution δ_r w.r.t. W_1 relative to $\mathcal{D}_r$. Letting r_0 be the minimum r for which δ_r satisfies (4.7), δ_{r_0} is a M.E. decision function. And Theorem 4.2 gives sufficient conditions for the existence of a solution by this method; explicitly, if Assumption 4.1 is satisfied and if there exists some decision function satisfying (4.7), then there exists a M.E. decision function.

Wald's sequential probability ratio test (see [21]) is an example of a M.E. decision function. Suppose X_1, X_2, ..., are independent and identically distributed and that Ω consists of but two elements, F_0 and F_1. Suppose the cost depends only on the sample size and is proportional to it. Suppose D^t has but two elements d^0 and d^1 corresponding to "F_0 is true" and "F_1 is true", respectively. Suppose the class of decision functions at the disposal of the experimenter is unrestricted. Let $w(F_i, d^j) = 1 - \delta_{ij}$ (Kroneker δ; $i, j = 0, 1$), and let $\beta(F_0) = \alpha$ and $\beta(F_1) = \beta$. Then, according to a theorem of Wald and Wolfowitz [23], the sequential probability ratio test of $H_0 : F = F_0$ against $H_1 : F = F_1$ is most economical where α and β are bounds on the two types of errors. In fact, it does more than minimize the maximum expected cost; it minimizes the expected cost at both F_0 and F_1.

Blyth [3] gives solutions to some estimation problems. His solutions may be used to find "M.E. estimators" for the mean of a variable which is (1) normal with known variance, or (2) rectangular with known range, and the cost is proportional to the sample size and the loss function is an arbitrary non-decreasing function of the absolute error of estimate.

Now let us consider non-sequential M.E. decision functions. We suppose $(X_1, \ldots, X_p)$, $(X_{p+1}, \ldots, X_{2p})$, $\ldots$, are independent and identically distributed random vectors. Let $\mathcal{D}_n$ be the class of decision functions for which the probability is one that the first np random variables in X are observed and no more, and suppose the class of decision functions at the disposal of the experimenter is $\mathcal{D} = \bigcup_n \mathcal{D}_n$ (that is, we only admit samples of a fixed size n, n = 0, 1, 2, ..., from the p-variate population). Suppose the cost function is a function of n only, say c(n). Denote decision functions in $\mathcal{D}_n$ by δ_n. Then a M.E. decision function δ_N is one which satisfies

$$(4.9) \qquad r_1(F, \delta) \leq \beta(F) \qquad \text{for all } F \in \underline{\Omega}$$

and $c(N) \leq c(n)$ for all n for which some $\delta \in \mathcal{D}_n$ satisfies (4.9). We then say that c(N) is the minimum cost and N is the M.E. sample size. If c(n) is an increasing function of n, then by defining a weight function and a β-function as in Sections 1.3.1, 1.3.2, 1.4, 2.3, or 2.4, this definition reduces to Definition 1.1, 1.2, 2.1, or 2.2, respectively.

To find such decision functions, Theorem 1 suggests the following procedure, analogously to the corresponding procedures given in Chapters I and II: Let δ_n^0 be a minimax solution w.r.t. W_1 (defined by (4.8)) relative to $\mathcal{D}_n$ (which surely exists if Assumption 4.1 holds). Suppose the minimum of c(n) over all n for which δ_n^0 satisfies (4.9) is attained for n = N. Then δ_N^0 is a M.E. decision function. Moreover, Theorem 4.1 also asserts that if δ_N' is any other M.E. decision function, then

$$\sup_{F \varepsilon \underline{\Omega}} \frac{r_1(F,\ \delta_N')}{\beta(F)} \geq \sup_{F \varepsilon \underline{\Omega}} \frac{r_1(F,\ \delta_N^0)}{\beta(F)}$$

or

$$(4.10) \qquad \inf_{F \varepsilon \underline{\Omega}} \frac{\beta(F) - r_1(F,\ \delta_N')}{\beta(F)} \leq \inf_{F \varepsilon \underline{\Omega}} \frac{\beta(F) - r_1(F,\ \delta_N^0)}{\beta(F)} .$$

No property analogous to (4.10) was proved for the special cases in Chapters I and II.

A possible extension of non-sequential M.E. decision functions as given above is to problems of k populations. We consider a cost function $c(\underline{n})$ where $\underline{n} = (n_1, \ldots, n_k)$ and n_i is the fixed sample size for the i^{th} population. We consider classes of decision functions $\mathcal{D}_c = \{\delta \ \varepsilon\ \mathcal{D} : c(\underline{n}) = c\}$, and let δ_c be a minimax solution w.r.t. W_1 relative to $\mathcal{D}_c$. Then, by Theorem 4.1, if we order the δ_c's according to increasing cost and choose the first δ_c in the sequence for which $r(F,\ \delta_c) \leq \beta(F)$ for all F, we will obtain a M.E. decision function.

4.4 Decision Functions with Bounds on the Maximum Expected Cost. We make the same definitions and assumptions as given in the first paragraph of Section 4.3, but instead of (4.8), we define the two weight functions

$$(4.11) \qquad W_1 = c(x;\ s)/\beta(F), \qquad W_2 = w(F,\ d^t).$$

Then a minimax solution w.r.t. W_2 relative to $\mathcal{D}^*$ will be a decision function which minimizes the maximum expected loss subject to the bounds on the expected cost: $r_2(F,\delta) \leq \beta(F)$ for all $F \varepsilon \underline{\Omega}$.

If $\beta(F)$ is independent of F, then, in terms of the weight functions (4.8), such decision functions are simply the minimax solutions w.r.t. W_1 relative to $\mathcal{D}_r$ (with r = 1) that were considered previously in obtaining M.E. decision functions. Thus, the approach to finding M.E. solutions by using Theorem 4.1 is similar to attacking this problem directly. Though Theorems 4.1 and 4.2 are applicable, their method of solving this problem does not appear to be very practical.

It may be noted that Blyth's results are applicable to this problem also. He gives the solution to the two estimation problems referred to in Section 4.3.

APPENDIX

SOME "TWO-SIDED" TWO-DECISION RULES AND SYMMETRIC THREE-DECISION RULES FOR COMPOSITE DISCRIMINATION

A.1 Some "Two-Sided" Two-Decision Rules.

A.1.1 Introduction. Let A_1, A_2 denote two alternative decisions, the acceptance of A_i being preferred when an unknown parameter θ of the density function $f(x, \theta)$ of a random variable X, to be observed, is in a subset ω_i of the parameter space Ω (i = 1, 2). Given two numbers θ_1, θ_2 ($\theta_1 > \theta_2 \geq 0$), we define:

$$\omega_1 = \{\theta : |\theta| \geq \theta_1\} , \quad \omega_2 = \{\theta : |\theta| \leq \theta_2\} .$$

Given $\underline{\alpha} = (\alpha_1, \alpha_2)$, it is desired to construct a M.E. d.r. relative to $\underline{\alpha}$ for discriminating between ω_1, ω_2. We call such a decision problem a "two-sided" two-decision problem for obvious reasons.

We consider four different situations, using the above notation for each. M.E. d.r.'s for each situation are given below, and nomographs for obtaining such d.r.'s explicitly are given in Section A.2; some examples are given in Section A.4. The derivations of the d.r.'s are not given, however, since they are analogous to the derivations given in Section 2.5 and 2.6. Alternatively, they may be derived as tests which maximize the minimum power using a theorem of Hoeffding referred to in Section 2.1. (Many of these tests which maximize the minimum power have been derived in the literature, especially for the case of $\theta_2 = 0$; e.g., see Lehmann [15] and Hoeffding [7], [8], and [9].) Moreover, we feel that such two-sided problems are not usually very realistic except as approximations to the analogous symmetric three-decision problem where alternatives A_1', A_2', A_3' correspond to

$$\omega_1' = \{\theta : \theta \leq -\theta_1\}, \ \omega_2' = \omega_2, \ \omega_3' = \{\theta : \theta \geq \theta_1\},$$

respectively, and $\alpha_1' = \alpha_3' = \alpha_1$. We shall show in Section A.3 that this approximation is usually very good so that the solutions to the "two-sided" two-decision problem may be used as solutions to this symmetric three-decision problem.

The M.E. 2-d.r.'s referred to above are given below.

A.1.2 The Mean of a Normal Distribution, Variance Known. Suppose $f(x, \theta)$ is a normal density function with variance σ^2 (known) and mean μ; i.e., X is $N(\mu, \sigma^2)$. Letting μ be θ introduced above, a M.E. 2-d.r. relative to (α_1, α_2) is: Take a sample of size N and

$$\text{(A.1)} \qquad \text{choose} \quad \begin{matrix} A_1 \\ \\ A_2 \end{matrix} \quad \text{if} \quad |\bar{x}| \quad \begin{matrix} > \\ \\ \leq \end{matrix} \quad c = c^N$$

where (n, c^n) is a pair of solutions of

$$\text{(A.2)} \qquad \begin{aligned} &\Phi\left(\sqrt{n}\,\frac{\theta_1+c^n}{\sigma}\right) - \Phi\left(\sqrt{n}\,\frac{\theta_1-c^n}{\sigma}\right) \leq 1 - \alpha_1 \\ &\Phi\left(\sqrt{n}\,\frac{\theta_2+c^n}{\sigma}\right) - \Phi\left(\sqrt{n}\,\frac{\theta_2-c^n}{\sigma}\right) \geq \alpha_2 \end{aligned}$$

and N is the least integer n for which a pair of solutions exists. Solutions may be obtained by first solving the two equations

$$\text{(A.3)} \qquad \Phi\left(\sqrt{n}\,\frac{\theta_i+c^n}{\sigma}\right) - \Phi\left(\sqrt{n}\,\frac{\theta_i-c^n}{\sigma}\right) = \gamma_i \quad (i = 1, 2)$$

where $\gamma_1 = 1 - \alpha_1 \rho$, $\gamma_2 = \alpha_2 \rho$ with $\rho = 1$ for n and c and then, taking N to be the least integer $\geq$ n, re-solving (A.3) for c^N and ρ .

A M.E. 3-d.r. for the corresponding symmetric three-decision problem is given in Section 1.3.6, where, by symmetry, $c_1^n = -c_2^n = -c^n$; that is, take a sample of size N and

$$\text{choose} \begin{cases} A_1' & \text{if } \bar{x} < -c \\ A_2' & \text{if } |\bar{x}| \leq c \\ A_3' & \text{if } \bar{x} > c \end{cases}$$

where N and c are chosen as in Section 1.3.6.

A.1.3 The Mean of a Normal Distribution, Variance Unknown. Again, we suppose X is $N(\mu, \sigma^2)$, but now both μ and σ are unknown. We let $\mu/\sigma = \theta$ and denote $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n - 1)$. A M.E. 2-d.r. relative to (α_1, α_2) is: Take a sample of size N and choose A_1 or A_2 according to (A.1) with $\bar{x}$ replaced by $\bar{x}/s$, and where N and c are chosen as in Section A.1.2 with (A.2) replaced by

$$\text{(A.4)} \quad \begin{aligned} T_{n-1,\sqrt{n}\theta_1}(\sqrt{n}c) - T_{n-1,\sqrt{n}\theta_1}(-\sqrt{n}c) &\leq 1 - \alpha_1 \\ T_{n-1,\sqrt{n}\theta_2}(\sqrt{n}c) - T_{n-1,\sqrt{n}\theta_2}(-\sqrt{n}c) &\geq \alpha_2 \end{aligned}$$

where $T_{f,\delta}(t)$ denotes the non-central t distribution function with f degrees of freedom and non-centrality factor δ (i.e., $t = \sqrt{f}(z + \delta)/\chi_f$ where z is N(0, 1) and χ_f^2 has a chi-squared distribution with f degrees of freedom). (A.4) may be solved as in Section A.1.2 with (A.3) replaced by

(A.5) $T_{n-1,\sqrt{n}\,\theta_i}(\sqrt{n}\ c) - T_{n-1,\sqrt{n}\,\theta_i}(-\sqrt{n}\ c) = \gamma_i \quad (i = 1,2).$

A.1.4 The Parameter of a Binomial Distribution. Suppose $f(x, \theta)$ is a point binomial distribution with parameter p. Let $\theta = p - 1/2 \quad (1/2 > \theta_1 > \theta_2 \geq 0)$, and denote by t_n the number of observations in a sample of size n that are ≤ 0. We restrict ourselves to non-randomized d.r.'s; a M.E. randomized d.r. may be obtained by making obvious modifications.

A M.E. non-randomized 2-d.r. relative to (α_1, α_2) is: Take a sample of size N and choose A_1 or A_2 according to (A.1) with $\bar{x}$ replaced by $(t_n/n - 1/2)$, and where N and c are chosen as before with (A.2) replaced by

$$\text{(A.6)} \quad \begin{aligned} B_{n,1/2+\theta_1}(n/2 + nc) - B_{n,1/2+\theta_1}(n/2 - nc - 1) &\leq 1 - \alpha_1 \\ B_{n,1/2+\theta_2}(n/2 + nc) - B_{n,1/2+\theta_2}(n/2 - nc - 1) &\geq \alpha_2 \end{aligned}$$

and (A.3) replaced by

$$\text{(A.7)} \quad B_{n,1/2+\theta_i}(n/2+nc) - B_{n,1/2+\theta_i}(n/2-nc-1) = \gamma_i \quad (i = 1,2).$$

$B_{n,p}(t)$ denotes the binomial distribution function as in Chapter I.

A.1.5 The Median of an Unspecified Distribution. Suppose X has an unknown density function $f(x)$ w.r.t. a measure μ on the real line such that $\mu\{x \leq 0\} > 0$ and $\mu\{x > 0\} > 0$, and denote $\theta = \theta(f) = \int_{-\infty}^{0} f(x)\, d\mu - 1/2$. Denote by t_n the number of observations in a sample of size n which are ≤ 0. Restricting ourselves to non-randomized d.r.'s, a M.E. 2-d.r. relative to (α_1, α_2) is: Take a sample of size N and choose A_1 or A_2 according to (A.1) with $\bar{x}$ replaced by $(t_n/n - 1/2)$ and where N and c are chosen as before with (A.2) replaced by (A.6) and (A.3) replaced by (A.7); that is, a M.E. d.r. for the parameter of a binomial distribution is also a M.E. d.r. for the median of an unspecified distribution, in the above sense.

A.2 Nomographic Solutions. In this section we develop a nomographic method for solving the equations in (A.3) for n and c^n. Then, by using normal approximations to the non-central t and binomial distributions, we use the same method for solving the equations in (A.5), (A.7), and (A.9), thus making it possible to obtain explicitly the M.E. d.r.'s of Section A.1.

A.2.1 The Nomographs. Consider the function

$$\gamma(x, y) = \Phi(y + x) - \Phi(y - x) \tag{A.8}$$

and the inverse function $y = y(x, \gamma)$. Figure A.1 below is a graph of y as a function of x for various values of γ. Figure A.2 below is a graph of the function

$$z(x; \gamma_1, \gamma_2) = y(x, \gamma_2)/y(x, \gamma_1)$$

as a function of x for various pairs of values (γ_1, γ_2).

Now suppose x is a function of n and c, say

$$x = f(n, c). \tag{A.9}$$

Suppose further that y may be factored as:

$$y = y(f(n,c), \gamma) = g(n,c)\cdot h(\gamma). \tag{A.10}$$

Then we have $z = z(f(n,c); \gamma_1, \gamma_2) = h(\gamma_2)/h(\gamma_1)$, independent of n and c.

Suppose we are given two values of γ, (γ_1, γ_2), and we wish to solve the two corresponding simultaneous equations expressed in (A.8) for (n, c) where x and y are given by (A.9) and (A.10). We proceed as follows: Enter Figure A.2 at $z = h(\gamma_2)/h(\gamma_1)$ and read off $x = x_0$ as the abscissa of the (γ_1,γ_2)-curve. Enter Figure A.1 at x_0 and read off $y(x_0, \gamma_1) = y_0$ as the

FIGURE A.1[1]

$y = y(x,\gamma)$ Defined by $\gamma(x,y) = \Phi(y+x) - \Phi(y-x)$

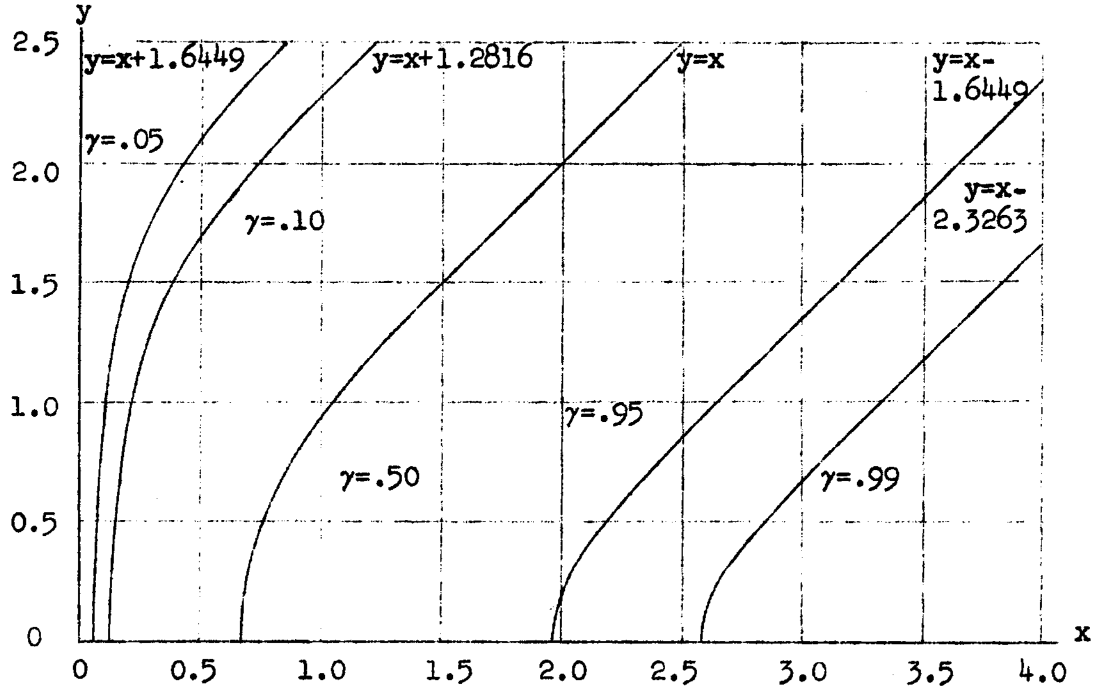
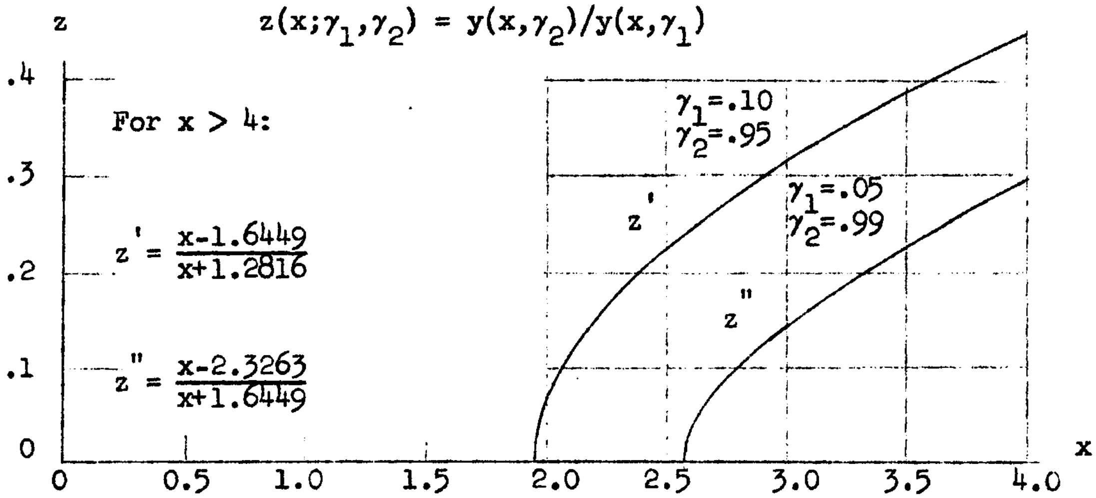


FIGURE A.2

$z(x;\gamma_1,\gamma_2) = y(x,\gamma_2)/y(x,\gamma_1)$



1. This figure was constructed with the aid of Tables of Normal Probability Functions [26].

ordinate of the γ_1-curve. Then solve $x_0 = f(n, c)$ and $y_0 = g(n, c) \cdot h(\gamma_1)$ for n and c.

Note that if $x_0 + y_0$ is large, say > 3, then $\Phi(y_0+x_0) \doteq 1$ so that $\gamma_1 = \Phi(y_0 + x_0) - \Phi(y_0 - x_0) \doteq 1 - \Phi(y_0 - x_0)$ or $y_0 \doteq x_0 + \Phi^{-1}(\alpha_1)$. (See Section A.3.) This gives an alternative method for obtaining y_0 without using Figure A.1. (By using Figure A.1, additional curves may be plotted on Figure A.2 if needed.)

We now apply this method to solving the equations of Section A.1.

A.2.2 <u>The Mean of a Normal Distribution, Variance Known.</u> We shall use the nomographs of the previous section for solving the equations in (A.3) for n and c^n for specified values of α_1 and α_2 with $\rho = 1$.

Note that the equations in (A.3) are of the form (A.8) with $x = f(n, c) = \sqrt{n}\, c/\sigma$ where $h(\gamma_i) = \theta_i$ $(i = 1, 2)$. Therefore, by the procedure of Section A.2.1, we may obtain $x_0 = \sqrt{n}\, c/\sigma$ and $y_0 = y(x_0, \gamma_1) = \sqrt{n}\, \theta_1/\sigma$ and then solve for n and c, obtaining

$$c = \theta_1 x_0/y_0 \qquad\qquad n = (y_0\sigma/\theta_1)^2 .$$

The least integer $\geq$ n is the M.E. sample size.

A.2.3 The Mean of a Normal Distribution, Variance Unknown. Using a normal approximation to the non-central t distribution given by Johnson and Welch [10]:

$$T_{f,\delta}(t) \doteq \Phi\left(\frac{t - \delta}{\sqrt{1 + \frac{t^2}{2f}}}\right),$$

the equations in (A.5) may be approximated by

$$\Phi\left(\sqrt{n}\,\frac{c - \theta_i}{\sqrt{1 + \frac{nc^2}{2n-2}}}\right) - \Phi\left(\sqrt{n}\,\frac{-c - \theta_i}{\sqrt{1 + \frac{nc^2}{2n-2}}}\right)$$

$$= \Phi\left(\sqrt{n}\,\frac{\theta_i + c}{\sqrt{1 + \frac{nc^2}{2n-2}}}\right) - \Phi\left(\sqrt{n}\,\frac{\theta_i - c}{\sqrt{1 + \frac{nc^2}{2n-2}}}\right)$$

$$= \gamma_i \qquad (i = 1,\ 2)$$

and are thus of the form (A.8) with $x = f(n,c) = \sqrt{n}c/\sqrt{1+nc^2/(2n-2)}$ where $h(\gamma_i) = \theta_i$ $(i = 1,\ 2)$. We may obtain x_0 and $y_0 = y(x_0,\gamma_1)$ by the procedure of Section A.2.1 and then solve for n and c, obtaining

$$c = \theta_1 x_0 / y_0$$

$$n = \frac{y_0^2}{\theta_1^2}\left[1 + nc^2/(2n-2)\right] > \frac{y_0^2}{\theta_1^2}(1 + c^2/2) = \frac{y_0^2}{\theta_1^2} + \frac{x_0^2}{2} ,$$

the latter giving a satisfactory approximation except for very small n.

However, these solutions are only approximations since the non-central t distribution has been approximated by a normal distribution. They may be improved iteratively, or simply checked by seeing if a solution to (A.4) exists for $n = N - 1$, computing the non-central t distribution function more accurately from its Edgeworth Type A Series, as discussed in [10] (p. 388).

A.2.4 <u>The Parameter of a Binomial Distribution and the Median of an Unspecified Distribution</u>. The equations to be solved to obtain M.E. d.r.'s in Sections A.1.4 and A.1.5 are identical so the two problems are treated simultaneously here. Using the normal approximation to the binomial distribution,

$$B_{n,p}(t) \doteq \Phi\left(\frac{k + 1/2 - np}{\sqrt{np(1 - p)}} \right) ,$$

the equations in (A.7) may be approximated by

$$\Phi\left(\frac{nc + 1/2 - n\theta_i}{\sqrt{n(1/4 - \theta_i^2)}}\right) - \Phi\left(\frac{-nc - 1 + 1/2 - n\theta_i}{\sqrt{n(1/4 - \theta_i^2)}}\right)$$

$$= \Phi\left(\sqrt{n}\,\frac{\theta_i + c + 1/2n}{\sqrt{1/4 - \theta_i^2}}\right) - \Phi\left(\sqrt{n}\,\frac{\theta_i - c - 1/2n}{\sqrt{1/4 - \theta_i^2}}\right)$$

$$= \gamma_i \qquad (i = 1.\ 2),$$

and are of the form (A.8) with $x_i = \sqrt{n}\ (c + 1/2n)/\sqrt{1/4 - \theta_i^2}$ and $y_i = \sqrt{n}\ \theta_i/\sqrt{1/4 - \theta_i^2} = g(n,\ c)\cdot h(\gamma_i)$ where $h(\gamma_i)$ $= \theta_i/\sqrt{1/4 - \theta_i^2}$, $g(n,c) = \sqrt{n}$ $(i = 1,\ 2)$. But x is not a function of (n, c) only, so that the above method cannot be applied directly. Instead, let $x_i = f'(n,\ c,\ \gamma_i)$, $y_i = y(x_i,\ \gamma_i)$ $= g(n,\ c)\cdot h(\gamma_i)$, and $z = \dfrac{y(x_2,\ \gamma_2)}{y(x_1,\ \gamma_1)} = \dfrac{h(\gamma_2)}{h(\gamma_1)} = \dfrac{\theta_2\sqrt{1/4 - \theta_1^2}}{\theta_1\sqrt{1/4 - \theta_2^2}}$. Consider γ_2 fixed (implying θ_2 fixed), and let $x' = f(n,\ c)$ $= \sqrt{n}\dfrac{c + 1/2n}{\sqrt{1/4 - \theta_1^2}}$. Construct a graph of z as a function of x' from Figure A.1 by taking the ratio of the ordinates of the

γ_2-curve at the abscissi $x = \sqrt{(1/4 - \theta_1^2)/(1/4 - \theta_2^2)}\, x'$ and the ordinates of the γ_1-curve at the abscissi $x = x'$. Entering this graph at $z = \dfrac{\theta_2\sqrt{1/4 - \theta_1^2}}{\theta_1\sqrt{1/4 - \theta_2^2}}$, we obtain x', and $y' = y(x', \gamma_1)$ may be obtained from Figure A.1; then x' and y' as functions of n and c may be solved, obtaining

$$n = \frac{1/4 - \theta_1^2}{\theta_1^2}\, y'^2$$

$$c = \frac{x'}{\sqrt{n}}\sqrt{1/4 - \theta_1^2} - \frac{1}{2n} = \theta_1 x'/y' - 1/2n \, .$$

To check or improve the accuracy of the solutions, tables of the binomial distribution function may be used.

A.3 Approximate Equivalence with Symmetric Three-Decision Rules.

We first consider the case of decision rules concerning the mean of a normal distribution, variance known. The equations to be solved for n and c to obtain a M.E. symmetric 3-d.r. are, from Sections 1.3.6 and A.1.2, $\Phi[\sqrt{n}\,(\theta_1 - c^n)/\sigma] = \alpha_1$ and

$\Phi[\sqrt{n}\,(\theta_2 + c^n)/\sigma] - \Phi[\sqrt{n}\,(\theta_2 - c^n)/\sigma] = \alpha_2$; and to obtain a M.E. two-sided 2-d.r., (A.3) must be solved with $\rho = 1$. Clearly, if $\Phi[\sqrt{n}\,(\theta_1 + c^n)/\sigma] \doteq 1$, then the solutions (n, c) of the two pairs of equations will be equal. We shall show for various lower bounds on α_1 and α_2 (see Table A.1) that $1 - \Phi[\sqrt{n}\,(\theta_1 + c^n)/\sigma] < .0005$ independently of θ_1, θ_2, and σ.

Consider the equations

$$\text{(A.11)} \qquad \Phi(y_i + x) - \Phi(y_i - x) = \gamma_i \qquad (i = 1, 2)$$

$(y_1 > y_2 \geq 0,\ x > 0)$ where $\gamma_1 = 1 - \alpha_1$ and $\gamma_2 = \alpha_2$. It is sufficient to show that $y_1 + x$ is "large" for $\alpha_i \geq \alpha_i^0$ $(i = 1,2)$, say, and for all θ_1, θ_2, σ $(\theta_1 > \theta_2 \geq 0)$. Denoting the standard normal density function by ϕ and considering (A.11) as $\gamma_i = \gamma_i(x, y_i)$, we have

$$\frac{\partial\gamma_i}{\partial x} = \phi(y_i + x) + \phi(y_i - x) > 0$$

$$\frac{\partial\gamma_i}{\partial y_i} = \phi(y_i + x) - \phi(y_i - x) < 0 \text{ since } |y+x| > |y-x|$$

$$\frac{\partial x}{\partial y_i} = -\frac{\partial\gamma_i}{\partial y_i} \Big/ \frac{\partial\gamma_i}{\partial x} > 0 \qquad (i = 1, 2)$$

so that γ_i is an increasing function of x for fixed y_i and a decreasing function of y_i for fixed x, and x increases with y_i for fixed γ_i.

Writing x as a function of y_i and γ_i, the two equations in (A.11) may be written

$$(A.12) \qquad x(y_1, \gamma_1) = x(y_2, \gamma_2).$$

From the above monotonicity relations, we thus have

$$(A.13) \qquad x = x(y_2, \gamma_2) \geq x(\min y_2, \min \gamma_2) = x(0, \alpha_2^0) = x^0, \text{ say.}$$

Writing y_1 as a function of x and γ_1, we therefore have

$$(A.14) \qquad y_1(x, \gamma_1) \geq y_1(x^0, \max \gamma_1) = y_1(x^0, 1 - \alpha_1^0) = y_1^0 \text{ , say.}$$

Thus, x^0 is the solution of $\Phi(x^0) - \Phi(-x^0) = \alpha_2^0$, and y_1^0 is the solution of $\Phi(y_1^0 + x^0) - \Phi(y_1^0 - x^0) = 1 - \alpha_1^0$, and for all $\alpha_i \geq \alpha_i^0$ $(i = 1, 2)$, $y_1 + x \geq y_1^0 + x^0$ or $\Phi(y_1 + x) \geq \Phi(y_1^0 + x^0)$.

Table A.1 below gives pairs (α_1^0, α_2^0) of lower bounds on (α_1, α_2) for which $\Phi(y_1^0 + x^0) > .9995$ so that for $\alpha_i \geq \alpha_i^0$ $(i = 1, 2)$, $1 - \Phi(y_1 + x) < .0005$ for all θ_1, θ_2, σ. For all such values (α_1, α_2), the symmetric three-decision problem is virtually equivalent to the two-sided two-decision problem.

TABLE A.1

Lower Bounds on (α_1, α_2) for which $1 - \Phi[\sqrt{n}(\theta_1 + c^n)/\sigma]$

$< .0005$

.001, .999	.501, .900	.950, .600
.035, .990	.800, .800	.990, .300
.265, .950	.900, .700	.999, .100

If the variance is unknown, and we assume the normal approximation to the non-central t-distribution given in Section A.2.3 to be sufficiently accurate, then the above argument holds for this case as well so that the symmetric three-decision problem and the two-sided two-decision problem are virtually equivalent if α_1, α_2 satisfy bounds such as those given in Table A.1.

The cases of decision rules concerning the parameter of a binomial distribution or the median of an unspecified distribution can be treated in an analogous manner with one minor modification, assuming the normal approximation to the binomial distribution is sufficiently accurate. As in Section A.2.4, we see that a subscript i is required on x in (A.11) and the argument following it. Equation (A.12) must be replaced by

$$x_1(y_1, \gamma_1) = \frac{\sqrt{1 - \theta_2^2}}{\sqrt{1 - \theta_1^2}} x_2(y_2, \gamma_2) ,$$

and hence $x_1 > x_2$. Hence, (A.13) may be replaced by $x_1 > x_2 \geq x_2(0, \alpha_2^0) = x^0$ and thus (A.14) still holds. Hence, the results of the argument are the same and the two decision problems are virtually equivalent for α_1, α_2 satisfying bounds such as those in Table A.1.

A.4 Tables of Most Economical Sample Sizes. In this section, we give some tables of M.E. sample sizes, exemplifying the decision problems considered above. Table A.2 gives M.E. sample sizes relative to the vectors $\underline{\alpha} = (.90, .95)$ and $\underline{\alpha} = (.95, .99)$ for various values of $\tau_1 = \theta_1/\sigma$ and $\tau_2 = \theta_2/\sigma$ for the "two-sided" decision problem concerning the mean of a normal distribution, variance known, as introduced in Section A.1.2. The sample sizes have been computed from the nomographs as indicated in Section A.2.2. Table A.2 also gives the M.E. sample sizes as computed from the nomographs for the analogous decision problem when the variance is known, as set out in Section A.1.3, where now $\tau_i = \theta_i$ $(i = 1, 2)$. However, these latter sample sizes are not necessarily exact since their computation utilized the normal approximation of the non-central t distribution given in Section A.2.3; but, by using an Edgeworth expansion of the non-central t distribution, several of the values were checked, all of which were found to be correct. In any case, the relative sizes indicated may be of more interest.

Table A.3 gives the M.E. sample sizes for "two-sided" non-randomized d.r.'s concerning the parameter of a binomial distribution or the median of an unspecified distribution as defined in Section A.1.4 and A.1.5. All sample sizes between 0 and 50 listed were taken from Tables of the Binomial Probability Distribution [25] and are accurate. All other values were computed from the nomographs as outlined in Section A.2.4, and the last digit is very approximate due to the use of a normal approximation to the binomial distribution. A better approximation might be expected for randomized d.r.'s since the introduction of randomization essentially has the effect of making a discrete distribution continuous, just as the normal approximation does. In any case, the relative values may be useful.

Of particular interest in both tables is the increase in the M.E. sample size when θ_2 (or τ_2) is increased from zero to a slightly larger number.

TABLE A.2

Most Economical Sample Sizes:

Normal Mean, Variance Known and Unknown

$\tau_i = \begin{cases} \theta_i/\sigma \text{ if } \sigma \text{ known} \\ \theta_i \text{ if } \sigma \text{ unknown} \end{cases}$		Most Economical Sample Size			
		$\alpha_1 = .90,\ \alpha_2 = .95$		$\alpha_1 = .95,\ \alpha_2 = .99$	
τ_1	τ_2	σ known	σ unknown	σ known	σ unknown
2.0	0	3	5	5	8
2.0	.10	3	5	5	9
2.0	.50	4	8	8	14
1.0	0	11	13	18	22
1.0	.10	12	14	20	24
1.0	.50	35	45	64	83
0.5	0	43	44	72	75
0.5	.10	54	57	99	104
0.5	.20	96	102	176	188
0.2	0	263	265	446	449
0.2	.05	382	385	701	708
0.2	.10	857	868	1578	1598
0.1	0	1051	1053	1782	1785
0.1	.02	1347	1350	2466	2471
0.1	.05	3428	3439	6310	6329

TABLE A.3

Most Economical Sample Sizes:

Binomial Parameter or Median of Unspecified Distribution

θ_1	θ_2	M.E. Sample Size $\alpha_1 = .90$ $\alpha_2 = .95$	M.E. Sample Size $\alpha_1 = .95$ $\alpha_2 = .99$	θ_1	θ_2	M.E. Sample Size $\alpha_1 = .90$ $\alpha_2 = .95$	M.E. Sample Size $\alpha_1 = .95$ $\alpha_2 = .99$
.40	0	11	19	.20	.05	89	163
.40	.05	15	25	.20	.10	195	359
.40	.10	17	32	.10	0	259	439
.40	.20	37	61	.10	.01	277	485
.30	0	24	40	.10	.02	331	606
.30	.05	32	53	.10	.05	837	1542
.30	.10	45	81	.05	0	1484	1775
.30	.20	161	298	.05	.01	1333	2454
.20	0	62	105	.05	.02	2367	4941
.20	.02	66	115	.02	0	6564	11128

BIBLIOGRAPHY

[1] Berger, Agnes, "On Uniformly Consistent Tests", Annals of Mathematical Statistics, 22 (1951), 289-293.

[2] Berger, A., and Wald, A., "On Distinct Hypotheses", Annals of Mathematical Statistics, 20 (1949), 104-109.

[3] Blyth, Colin R., "On Minimax Statistical Decision Procedures and Their Admissibility", Annals of Mathematical Statistics, 22 (1951), 22-42.

[4] Dvoretzky, A., Wald, A., and Wolfowitz, J., "Elimination of Randomization in Certain Statistical Decision Procedures and Zero-Sum Two-Person Games", Annals of Mathematical Statistics, 22 (1951), 1-21.

[5] Ferris, C. D., Grubbs, F. E., and Weaver, C. L., "Operating Characteristics for the Common Statistical Tests of Significance", Annals of Mathematical Statistics, 17 (1946), 178-197.

[6] Ghosh, M. N., "An Extension of Wald's Decision Theory to Unbounded Weight Functions", Sankhya, 12 (1952), 8-26.

[7] Hoeffding, Wassily, "'Optimum' Nonparametric Tests", Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1951, 83-92.

[8] Hoeffding, Wassily, "Lectures on Estimation and Testing of Hypotheses", University of North Carolina, Fall Quarter 1952 (unpublished).

[9] Hoeffding, Wassily, "Lectures on Non-Parametric Inference", University of North Carolina, Winter Quarter 1953 (unpublished).

[10] Johnson, N. L., and Welch, B. L., "Applications of the Non-Central t-Distribution", Biometrika, 31 (1939), 362-389.

[11] Konijn, H. S., "On Certain Classes of Statistical Decision Procedures", Annals of Mathematical Statistics, 24 (1953), 440-448.

[12] Kruskal, William, "The Monotonicity of the Ratio of Two Noncentral t Density Functions", Annals of Mathematical Statistics, 25 (1954), 162-165.

[13] Lehmann, E. L., Theory of Testing Hypotheses, Associated Student Stores, University of California.

[14] Lehmann, E. L., "On the Existence of Least Favorable Distributions", Annals of Mathematical Statistics, 23 (1952), 408-416.

[15] Lehmann, E. L., "Some Principles of the Theory of Testing Hypotheses", Annals of Mathematical Statistics, 21 (1950), 1-26.

[16] Lindley, D. V., "Statistical Inference", Journal of the Royal Statistical Society, Series B, 15 (1953), 30-76.

[17] Rao, C. R., Advanced Statistical Methods in Biometric Research, New York, John Wiley and Sons, 1952.

[18] Ruist, Erik, "Comparison of Tests for Non-Parametric Hypotheses", Arkiv för Matematik, 3 (1954), 133-163.

[19] Savage, L. J., "Theory of Statistical Decision", Journal of the American Statistical Association, 46 (1951), 61.

[20] Sverdrup, Erling, "Weight Functions and Minimax Procedures in the Theory of Statistical Inference", Archiv for Mathematik og Naturvidenskab, 51 (1952), 1-73.

[21] Wald, Abraham, Sequential Analysis, New York, John Wiley and Sons, 1947.

[22] Wald, A., Statistical Decision Functions, New York, John Wiley and Sons, 1950.

[23] Wald, A., and Wolfowitz, J., "Optimum Character of the Sequential Probability Ratio Test", Annals of Mathematical Statistics, 19 (1948), 326-339.

[24] Wald, A., and Wolfowitz, J., "Characterization of the Minimal Complete Class of Decision Functions when the Number of Distributions and Decisions is Finite", Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1951, 149-157.

[25] Tables of the Binomial Probability Distribution, National Bureau of Standards Applied Mathematics Series, 6, 1950.

[26] Tables of Normal Probability Functions, National Bureau of Standards Applied Mathematics Series, 23, 1953.