MOST ECONOMICAL MULTIPLE-DECISION RULES[1]

By Wm. Jackson Hall

University of North Carolina

0. Summary. This paper is concerned with non-sequential multiple-decision procedures for which the sample size is a minimum subject to either (1) lower bounds on the probabilities of making correct decisions or (2) upper bounds on the probabilities of making incorrect decisions. Such decision procedures are obtained by constructing artificial decision problems for which the minimax strategies provide solutions to problems (1) and (2). These are shown to be "likelihood ratio" and "unlikelihood ratio" decision rules, respectively. Thus, although problems (1) and (2) are formulated in the spirit of the classical Neyman-Pearson approach to two-decision problems, minimax theory is used as a tool for their solution.

Problems of both "simple" and "composite" discrimination are considered and some examples indicated. (Some multivariate examples are given in [4].) Various properties of the decision rules are derived, and relationships with works of Wald, Lindley, Rao and others are cited.

1. Simple discrimination.

A. *Formulation of the problem.* We are concerned with a sequence $X_1, X_2, \cdots$, of real- or vector-valued, independent, and identically distributed random variables, each having a density function f, belonging to some specified class Ω, w.r.t. a fixed measure μ.

The decision problem is to formulate a rule for choosing a non-negative integer n (completely non-random), and, after taking an observation

$$x = (x_1, \cdots, x_n)$$

on $X = (X_1, \cdots, X_n)$, for choosing one of m possible alternative decisions $A_1, \cdots, A_m$. A multiple decision rule (m-d.r.) for choosing among $A_1, \cdots, A_m$ on the basis of x is defined by an ordered set of non-negative, real-valued, measurable functions $\phi(x) = [\phi_1(x), \cdots, \phi_m(x)]$ on the space $\mathfrak{X}$ of x such that $\sum_i \phi_i = 1$ identically in x (for $n = 0$, the ϕ_i's are constants). A_i is then chosen with probability $\phi_i(x)$ when x is observed. For non-randomized d.r.'s (all ϕ_i's equal 0 or 1), the ϕ_i's are characteristic functions of mutually exclusive and exhaustive "acceptance" regions $R_1, \cdots, R_m$ in $\mathfrak{X}$, where A_i is accepted if $x \in R_i$.

Received January 24, 1956; revised November 15, 1957.

[1] This research was sponsored partly by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill and partly by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. 18(600)-458. Reproduction in whole or in part for any purpose of the United States Government is permitted.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

A subscript or superscript n denotes the corresponding sample size; $f^n(x)$ and μ^n are the joint density and product measure, respectively.

We suppose throughout Section 1 that Ω consists of a finite number, say l, of elements $f_1, \cdots, f_l$; we say that the corresponding decision problem is one of "simple discrimination" and a d.r. is a d.r. for "discriminating among $f_1, \cdots, f_l$." Here, if μ is non-atomic, only non-randomized d.r.'s need be considered [2].

A d.r. $D = D_n$ is characterized by the functions

$$p_{ij}(D) = \Pr(D \text{ chooses } A_j \mid f_i) = \int_{\mathfrak{X}} \phi_j(x) f_i^n(x)\, d\mu^n \quad (i = 1, \cdots, l; j = 1, \cdots, m).$$

We consider two different criteria for choosing a d.r. for simple discrimination. The first assumes that $l = m$ and that the decision A_i is to be preferred when f_i is true. Denote $p_{ii}(D) = p_i(D) = 1 - q_i(D)$, so that p_i is the probability of a "correct" decision and q_i the probability of an "incorrect" decision when f_i is true.

DEFINITION 1. Let $\alpha = (\alpha_1, \cdots, \alpha_m)$ 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 α for discriminating among $f_1, \cdots, f_m$ if it satisfies

$$p_i(D) \geqq \alpha_i \qquad (i = 1, \cdots, m) \tag{1}$$

and if N is the least integer n for which (1) 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.

We now no longer require that $l = m$, but 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 2. Let $\beta = (\beta_{ij})$ be a given $l \times m$ 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 $f_1, \cdots, f_l$ if it satisfies

$$p_{ij}(D) \leqq \beta_{ij} \quad (i = 1, \cdots, l; j = 1, \cdots, m) \tag{2}$$

and if N is the least integer n for which (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.

If $l = m$ and A_i is preferred when f_i is true, then an M.E. d.r. relative to β also controls the probabilities of correct decisions if $\sum_{j \neq i} \beta_{ij} < 1$ for all i.

If $l = m = 2$, both (1) and (2) reduce to upper bounds on the probabilities of the two kinds of error, and Definitions 1 and 2 define an M.E. 2-d.r. as one with minimum sample size subject to these bounds.

It is intuitively clear (and elementary to prove) that a necessary and sufficient condition for the existence of a M.E. m-d.r. relative to any α or β $(l = m)$ is that there exist uniformly consistent sequences of 2-d.r.'s for discriminating between every pair $\omega_i, \omega_j (i \neq j)$ [5].

We shall utilize elements of Wald's theory of decision functions as given in

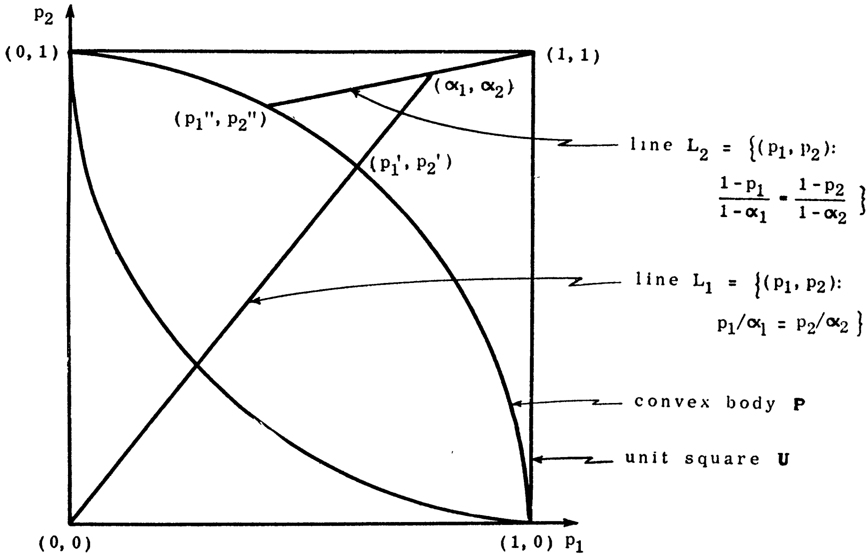

FIG. 1

[14], and shall use in particular some of the results of Sections 3.5 and 5.1.1, altering his notation slightly. The differences in the "data of the decision problem" assumed by Wald and here are only minor.

Let $\mathfrak{D}_n$ denote the class of all m-d.r.'s based on a sample of size n ($n = 0, 1, 2, \cdots$). Clearly, for $n \leqq N$, $\mathfrak{D}_n \subseteq \mathfrak{D}_N$; Lemma 1 follows almost immediately.

LEMMA 1. *For every fixed sample size* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. and denote* $r_n = \max_i r(f_i, D_n^0)$, *where* $r(f_i, D_n)$ *is the risk w.r.t. some bounded loss function. Then the sequence* $\{r_n\}$, $n = 0, 1, 2, \cdots$, *is a non-increasing sequence.*

B. *Most economical decision rules relative to a vector* α. Later in this section, we shall apply Wald's theory to two specific loss functions and develop in each case a method of obtaining M.E. d.r.'s as defined by Definition 1. First, we motivate geometrically the selection of such loss functions so as to identify the minimax strategy with the desired one. This alternative approach may give some geometrical insight into the properties of the d.r.'s obtained.[2]

For fixed n, let $p(D) = (p_1(D), \cdots , p_m(D))$ denote a point in m-space, and $P_n = \{p(D): D \in \mathfrak{D}_n\}$. It can be shown ([2], [10]) that P_n is a convex body in the unit m-cube U containing all corners of U with coordinates summing to unity. The case $m = 2$ is illustrated. (Conditions under which P_n is a proper subset of $P_{n'}$ for $n < n'$ and for which P_n tends with increasing n to U are given elsewhere ([3], [5]).)

[2] The author is indebted to the referee for considerable improvement of this geometric presentation.

In the diagram, the point $\alpha \varepsilon P_n$; therefore n is smaller than the required sample size. The M.E. sample size is the smallest n for which $\alpha \varepsilon P_n$, in which case the points p', p'' and α coincide (approximately). To test whether or not $\alpha \varepsilon P_n$, we can examine the position of the points p' or p'' relative to the position of α.

The points on the "upper" surface of P_n (n fixed) include all points $p(D_n^*)$ corresponding to Bayes strategies D_n^* when the loss function is

$$W(f_i, A_j) = W_{ij} = -\delta_{ij}/\alpha_i \qquad (i, j = 1, \cdots, m), \tag{3}$$

where δ_{ij} denotes the Kronecker δ-function.[3] Then the risk w.r.t. W_{ij} is

$$r(f_i, D) = -\sum_{j=1}^{m} \delta_{ij}\, p_{ij}(D)/\alpha_i = -p_i(D)/\alpha_i \qquad (i = 1, \cdots, m). \tag{4}$$

If p' is not on the boundary of U, the least favorable distribution for the weight function W_{ij} will positively weigh each element in Ω. (This will occur if the region in $\mathfrak{X}$ of positive density is constant over Ω.) In this case, the minimax strategy D_n' will be such that $p(D_n')$ is on the line L_1 of constant risk; i.e., $p(D_n') = p'$. (To obtain the minimax geometry with the loss function W_{ij}, transform the diagram by dividing the ith coordinate by $-\alpha_i$; then the convex body P_n goes into the convex body of risk points $(r_1, \cdots, r_m)$, $r_i = r(f_i, D)$.)

Alternatively, the "upper" surface of P_n corresponds to Bayes strategies when the loss function is

$$W^*(f_i, A_j) = W_{ij}^* = (1 - \delta_{ij})/(1 - \alpha_j) \qquad (i, j = 1, 2, \cdots, m) \tag{5}$$

and the risk function is $r^*(f_i, D) = q_i(D)/\beta_i$, where $\beta_i = 1 - \alpha_i$. The least favorable distribution will likewise positively weight each element of Ω whenever p'' is not on the boundary of U. In this case, the minimax strategy D_n'' will be such that $p(D_n'')$ is on the line L_2 of constant risk; i.e., $p(D_n'') = p''$. (To obtain the minimax geometry with W_{ij}^*, transform the diagram by dividing the ith coordinate by $1 - \alpha_i$, again transforming P_n into the convex body of risk points.) This latter approach is similar to that used by Rao [11] for problems of classification in multivariate analysis.

When $l = m > 2$, there is an added complication for the latter loss function since the line (L_2) from $(1, 1, \cdots, 1)$ through α need not necessarily pierce P for $n < N$, the M.E. sample size. (Of course, if $\alpha \varepsilon P_n$, then the line certainly pierces P_n.) Thus the components of a least favorable distribution are not necessarily positive unless $n \geqq N$ and p'' is in the interior of U.

Thus, in one instance, minimax rules maximize the common ratio $p_1/\alpha_1 = \cdots = p_m/\alpha_m$ and, in the other, minimize the common ratio $q_1/\beta_1 = \cdots = q_m/\beta_m$. The M.E. sample size is the smallest one for which the common ratio is $\geqq 1$ or $\leqq 1$, respectively. We now formalize these results. (Wald's Theorem 5.3[4] asserts the existence of a minimax d.r. D^0 for any (fixed) sample size.)

[3] This loss function satisfies Wald's requirements although it is not necessarily zero when a correct decision is made nor necessarily positive otherwise, as intuitively suggested, but never required mathematically, by Wald.

[4] All references to Wald refer to [14] unless otherwise specified.

THEOREM 1. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (3) *for samples of fixed size n. Suppose for some n,*

$$\max_i r(f_i, D_n^0) \leqq -1 \tag{6}$$

and let N be the least such integer. Then D_N^0 *is an M.E. d.r. relative to the vector* α *for discriminating among* $f_1, \cdots, f_m$. *Conversely, if there exists an M.E. d.r. relative to* α *for discriminating among* $f_1, \cdots, f_m$, *and the M.E. sample size is N, then* D_N^0 *is an M.E. d.r.*

PROOF. From (4) and (6), it follows that D_N^0 satisfies (1). Now suppose for some $n < N$, there exists a d.r. D_n satisfying (1). Since D_n^0 is minimax, $\max_i r(f_i, D_n^0) \leqq \max_i r(f_i, D_n) = \max_i [-p_i(D_n)/\alpha_i]$. Since D_n satisfies (1), we have from above that $\max_i r(f_i, D_n^0) \leqq -1$, in contradiction to the fact that N is the least integer n for which this is true. Hence, D_N^0 is an M.E. d.r.

To prove the converse, suppose D_N is an M.E. d.r. Then

$$-1 \geqq \max_i [-p_i(D_N)/\alpha_i] = \max_i r(f_i, D_N) \geqq \max_i r(f_i, D_N^0)$$

since D_N^0 is a minimax d.r. Hence, (6) is satisfied for $n = N$, and since N is the M.E. sample size, D_N^0 is an M.E. d.r.

Lemma 1 assures us that any n for which (6) is violated is too small. Now let us consider the structure of minimax d.r.'s for a fixed sample size n.

DEFINITION 3. A d.r. D defined by $\phi(x)$ is said to be a likelihood ratio d.r. if there exist positive constants $a_1, \cdots, a_m$ such that for any j and any x for which $\phi_j(x) > 0$, $a_j f_j^n(x) \geqq a_i f_i^n(x)$ for all $i \neq j$.
(Note that $a_1, \cdots, a_m$ determine ϕ completely except in sets of x for which $a_i f_i^n(x) = \max_j a_j f_j^n(x)$ for more than one value of i.) Setting $a_j = \xi_j/\alpha_j$, where $\xi = (\xi_1, \cdots, \xi_m)$ is an a priori distribution over $\Omega = (f_1, \cdots, f_m)$, it follows from Wald's Theorem 5.1 (with (5.6) replaced by (5.7)) that a Bayes d.r. relative to any ξ for which all $\xi_i > 0$ is a likelihood ratio d.r., and conversely.

Wald's Theorem 5.3 asserts the existence of a minimax d.r. and a least favorable distribution, and that any minimax d.r. is a Bayes d.r. relative to any least favorable distribution. Moreover, it follows from (4) and Wald's Theorem 5.3 (iii) that if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$p_1(D^0)/\alpha_1 = \cdots = p_m(D^0)/\alpha_m. \tag{7}$$

We shall give sufficient conditions for this to be true.

ASSUMPTION 1. If R is a subset of $\mathfrak{X}$ for which $\int_R f_i^n(x)\, d\mu^n = 0$ for some i, then $\int_R f_i^n(x)\, d\mu^n = 0$ for all values of i. (Whenever this assumption is made, we shall tacitly assume that $\mathfrak{X}$ is redefined so that $f_i^n(x) > 0$ for all i and $x \in \mathfrak{X}$.)
We state a theorem analogous to Wald's Theorem 5.4;[5] the proof (not given) is also analogous.

[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 $\delta = 1/u$ $(i = 1, \cdots, u)$ identically in x, and then $r(F_j, \delta) = (u-1)/u < 1$ for $j = 1, \cdots, k$.

THEOREM 2. *If Assumption 1 holds, all components of a least favorable distribution ξ^0 w.r.t. the weight function w_{ij} are positive.*

Hence, under Assumption 1, an M.E. d.r. may be obtained by the following method: for each sample size n, find a likelihood ratio d.r. D_n^0 for the constants $a_1, \cdots, a_m$ determined by Eqs. (7), and then choose N as the minimum n for which $p_1(D_n^0) \geqq \alpha_1$.

As an alternative approach, we can consider the weight function W_{ij}^* and proceed analogously to the first approach, giving a theorem identical to Theorem 1 with (6) replaced by $\max_i r(f_i, D_n^0) \leqq 1$; and, replacing $a_j = \xi_j/\alpha_j$ by ξ_j/β_j, it follows analogously that a Bayes d.r. relative to any ξ for which all $\xi_i > 0$ is a likelihood ratio d.r., and conversely. Moreover, if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$q_1(D^0)/\beta_1 = \cdots = q_m(D^0)/\beta_m. \tag{8}$$

We shall give sufficient conditions for this to be true. Analogously to Wald's Theorem 5.4, we have:

LEMMA 2. *If Assumption 1 holds, and if there exists some d.r. D for which*

$$r(f_i, D) < 1/\max_{1 \leqq j \leqq m} \beta_j \qquad (i = 1, \cdots, m),$$

then all components of a least favorable distribution are positive.

The following lemma may be useful in this regard:

LEMMA 3. *If $\beta_i < [1/(m-1)]\sum_{j=1}^m \beta_j$ (i.e., $\alpha_i > [\sum \alpha_j - 1]/[m-1]$) for all i, then there exists a d.r. D for which $r(f_i, D) < 1/\max_j \beta_j$ for all i.*

The proof follows by considering a d.r. defined by $\phi_i(x) = 1 - (m-1)\beta_i/\sum \beta_j > 0$ identically in x $(i = 1, \cdots, m)$.

THEOREM 3. *Suppose Assumption 1 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 $n \geqq N$, the M.E. sample size, and that D_n^0 is a minimax d.r. for samples of size n; then, using Lemma 1 and the theorem analogous to Theorem 1, D_n^0 satisfies (1). Use of Lemma 2 completes the proof.

Hence, under Assumption 1, D_N^0 is a likelihood ratio d.r., and an M.E. d.r. may be obtained by considering likelihood ratio d.r.'s D_n^0 for each n for constants $a_1, \cdots, a_m$ determined by (8), and then choosing N as the minimum n for which $q_1(D_n^0) \leqq \beta_1$. If for some n one of the components of a least favorable distribution is zero, we know that n is less than the M.E. sample size (Lemma 1).

A Bayes d.r. relative to any ξ of which all components are positive is admissible [15]. Hence, any likelihood ratio d.r. is admissible, and under Assumption 1 M.E. d.r.'s obtained by either of the above approaches are admissible. Thus, denoting an M.E. d.r. by D_N^0, there does not exist a d.r. D_N' for which $p_i(D_N') \geqq p_i(D_N^0)$ $(i = 1, \cdots, m)$ with strict inequality for at least one i (under Assumption 1).

Suppose now that a real-valued statistic $t = t(x_1, \cdots, x_n)$ exists which is sufficient for the class $\{f_i^n\}$ $(i = 1, \cdots, m)$, and suppose that t has a monotone likelihood ratio for some ordering of the elements of Ω; i.e., if $g_i(t)$ is the density of t corresponding to $f_i(x)$, then, for some ordering of the subscripts,

$$g_i(t_1)g_j(t_2) \geqq g_i(t_2)g_j(t_1)$$

for $i > j$ and $t_1 > t_2$ [8]. It follows almost immediately that for any $\phi(x)$ which defines a likelihood ratio d.r. there exist constants $\{c_i\}$, $-\infty = c_0 \leqq c_1 \leqq \cdots \leqq c_{m-1} \leqq c_m = \infty$, such that $\phi_i(x) > 0$ implies $c_{i-1} \leqq t(x) \leqq c_i$. Moreover, $\phi_i(x) = 1$ if the latter inequalities are strict, so that randomization may be required only at the points $t = c_i$ and only then if such points have positive probability. Such d.r.'s have been called monotone [1], [8]. If, for example, f_i is of the exponential type $f_i = \beta(\theta_i)e^{\theta_i x}r(x)$, $r \geqq 0$ and θ_i real, for all i, the above conditions are satisfied [1].

Example 1. Suppose f_i is a normal density function with mean θ_i $(-\infty < \theta_1 < \cdots < \theta_m < \infty)$ and known variance σ^2. Then $t = \bar{x}$ is sufficient and the c_i's and N may be obtained by first solving the following equations (iteratively) for the c_i^n's and n with $\rho_n = 1$:

$$p_i(D_n) = \Phi[\sqrt{n}(c_i^n - \theta_i)/\sigma] - \Phi[\sqrt{n}(c_{i-1}^n - \theta_i)/\sigma] = \alpha_i\rho_n \quad (i = 1, \cdots, n), \tag{9}$$

where Φ denotes the standard normal distribution function, and then, choosing N to be the least integer $\geqq n$, re-solving for the c_i's and ρ_N. Such a monotone rule will be minimax w.r.t. W_{ij} for the M.E. sample size. Alternatively, (9) may be replaced by equations of the form $1 - p_i(D_n) = (1 - \alpha_i)\rho_n'$, and a solution obtained which will be minimax w.r.t. W_{ij}^*.

Other examples may be treated analogously, allowing for randomization in the discrete cases if desired.

C. *Most economical decision rules relative to a matrix* β. To obtain M.E. d.r.'s as defined by Definition 2, we shall construct an artificial decision problem whose minimax solution will have the properties desired. For convenience, we replace each β_{ij} which is equal to unity by $+\infty$.

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

$$W_{ijk} = 1/\beta_{ij} \quad \text{if } j = k \ (i = 1, \cdots, l; j, k = 1, \cdots, m) \text{ and } 0 \text{ otherwise.} \tag{10}$$

We consider the artificial decision problem of choosing among $A_1, \cdots, A_m$ when one of the $l' = lm$ density functions g_{ij} is "true", and where the "loss" incurred by choosing A_k when g_{ij} is "true", is $W(g_{ij}, A_k)$. The risk function is $r(g_{ij}, D) = \sum_k W_{ijk}p'_{ijk}(D)$, where $p'_{ijk}(D) = \Pr(D \text{ chooses } A_k \mid g_{ij}) = p_{ik}(D)$; thus $r(g_{ij}, D) = p_{ij}(D)/\beta_{ij}$ $(i = 1, \cdots, l;\ j = 1, \cdots, m)$. Wald's Theorem 5.3 asserts the existence of minimax d.r.'s.

THEOREM 4. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (10) *for discriminating among* $g_{11}, g_{12}, \cdots, g_{lm}$ *for samples of fixed size* n. *Suppose for some* n, $\max_{i,j} r(g_{ij}, D_n^0) \leqq 1$, *and let* N *be the least such integer.*

Then D_N^0 *is an M.E. d.r. relative to the matrix* β *for discriminating among* $f_1, \cdots, f_l$. *Conversely, if there exists an M.E. d.r. relative to* β *and N is the M.E. sample size, then* D_N^0 *is an M.E. d.r.*

The theorem may be proved in a similar manner to Theorem 1. Now let us consider the structure of these minimax solutions w.r.t. W_{ijk}.

DEFINITION 4. A d.r. D defined by $\phi(x)$ is said to be an unlikelihood ratio d.r. if there exist non-negative constants a_{ij} $(i \neq j;\ i = 1, \cdots, l;\ j = 1, \cdots, m)$, where for each i at least one $a_{ij} > 0$, such that for any k and any x for which $\phi_k(x) > 0$, $\sum_{i \neq k} a_{ik} f_i^n(x) \leqq \sum_{i \neq j} a_{ij} f_i^n(x)$ for all $j \neq k$.

Setting $a_{ij} = \xi_{ij} / \beta_{ij}$, where $\xi = (\xi_{11}, \xi_{12}, \cdots, \xi_{lm})$ denotes an a priori distribution over Ω', we have from Wald's Theorem 5.1 that any Bayes d.r. relative to ξ is an unlikelihood ratio d.r. and conversely. Lindley [10] introduced such d.r.'s, obtained by his "method of minimum unlikelihood." Hereafter, we shall suppose $\xi_{ij} = 0$ for every i, j for which $\beta_{ij} = \infty$ without loss of generality.

Wald's Theorem 5.3 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,

$$p_{ij}(D^0)/\beta_{ij} = \max_{i,j} [p_{ij}(D^0)/\beta_{ij}] \qquad \text{for any } i, j \text{ for which } \xi_{ij}^0 > 0. \tag{11}$$

Apparently, however, there are no general conditions under which all $\xi_{ij}^0 > 0$, and consequently we have no proof of the admissibility of a minimax d.r. In fact, supposing $l = m$ and the β_{ij}'s satisfy $\sum_j^m \beta_{ij} = 1$ for every i, then $\xi_{ij}^0 > 0$ for all i, j would imply $p_{ij} = \beta_{ij}$, regardless of the sample size! Geometrically, the convex body in the $l \cdot m$-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. However, we do have the following theorem in this regard, assuming $l = m$ and A_i is "correct" when f_i is true $(i = 1, \cdots, m)$.

THEOREM 5. *Suppose Assumption* 1 *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* $\sum_{i=1}^m \xi_{ij}^0 > 0$ *for every* j.

The theorem may be proved by a contradiction, using Assumption 1, Definition 4, Lemma 1, and constructing a Bayes d.r. relative to ξ^0. From Theorem 5 and (11), 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.

Example 2. We shall consider unlikelihood ratio d.r.'s for samples of size n for Example 1 above. For simplicity, suppose $\sigma = 1$, $l = m = 3$, and $\theta_2 = 0$, the alternatives A_1, A_2, A_3 corresponding respectively to the densities f_1, f_2, f_3.

A d.r. with acceptance regions

$$R_1^n = \{x : h_1^n(x) \leqq h_2^n(x), h_1^n(x), \leqq h_3^n(x)\},$$

$$R_2^n = \{x : h_2^n(x) < h_1^n(x), h_2^n(x) \leqq h_3^n(x)\},$$

$$R_3^n = \{x : h_3^n(x) < h_1^n(x), h_3^n(x) < h_2^n(x)\},$$

where $h_1^n(x) = a_{ji}f_j^n + a_{ki}f_k^n$ and (i, j, k) is a permutation of (1, 2, 3), is an unlikelihood ratio d.r. for the weights (a_{ij}). Denoting the sample mean by $\bar{x}$, we may replace h_i^n by $g_i^n = [a_{ji} \exp (n\theta_j\bar{x} - n\theta_j^2/2) + a_{ki} \exp (n\theta_k\bar{x} - n\theta_k^2/2)]$. Now g_1^n is an increasing function of $\bar{x}$ and g_3^n a decreasing function; g_2^n has a single stationary point, a minimum. 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 $R_1 = \{x:\bar{x} \leqq c_1 \text{ or } c_3 \leqq \bar{x} \leqq c_4\}$, $R_2 = \{x:c_2 \leqq \bar{x} \leqq c_3\}$, $R_3 = \{x:c_1 \leqq \bar{x} \leqq c_2 \text{ or } \bar{x} \geqq c_4\}$, where either $c_1 = c_2$, or $c_3 = c_4$, or both. (Equality signs have been assigned everywhere in the R_i's for simplicity.) Let c (=2 or 3) denote the number of c_i's to be determined. The c_i's may be obtained by solving $c + 1$ of the six equations $p_{ij} = \rho\beta_{ij}$ for the c_i's and ρ, the choice of the equations to be solved being such that $p_{ij} \leqq \rho\beta_{ij}$ for all six pairs of subscripts. Theorem 5 may be helpful in this choice of equations. To obtain an M.E. d.r., the sample size n is to be minimized subject to $\rho = \rho_n \leqq 1$. Similar methods may be applied to simple discrimination problems concerning any distribution of the exponential type.

2. Composite discrimination.

A. *The problem.* In this section we allow a continuum of possible density functions. For specificity, assume Ω to be the space of a real- or vector-valued parameter θ indexing the class of density functions w.r.t. μ with elements $f(x, \theta)$.

We further suppose that disjoint subsets $\omega_1, \cdots, \omega_l$ of Ω are specified such that for every pair i, j $(i = 1, \cdots, m; j = 1, \cdots, l)$ there is a definite preference for or against the decision A_j if the true $\theta \varepsilon \omega_i$. We suppose that none of the decisions is definitely preferred if θ is not in some ω_i; this "indifference region" is excluded from Ω for convenience. Under these assumptions, we say that the corresponding decision problem is one of "composite discrimination" and a d.r. is a d.r. for "discriminating among $\omega_1, \cdots, \omega_l$." A d.r. $D = D_n$ is characterized by the functions

$$p_j(\theta, D) = \Pr (D \text{ chooses } A_j|\theta) = \int_{\mathfrak{X}} \phi_j(x)f^n(x, \theta)d\mu^n \quad (j = 1, \cdots, m),$$

defined for all $\theta \varepsilon \Omega$.

We again consider two criteria for choosing a d.r. for composite discrimination. The first requires $l = m$ and A_i to be a "correct" decision if $\theta \varepsilon \omega_i$ and "incorrect" if $\theta \varepsilon \omega_j (j \neq i)$. For the second criterion, we suppose that corresponding to each ω_i one or more alternatives A_j is preferable when $\theta \varepsilon \omega_i$.

The definitions and comments of Section 1.A may be restated, substituting only ω_i for f_i, $\inf_{\theta\varepsilon\omega_i} p_i(\theta, D)$ for $p_i(D)$, and $\sup_{\theta\varepsilon\omega_i} q_i(\theta, D)$ for $q_i(D)$. When $l = m = 2$, an M.E. 2-d.r. may be considered as a test of the hypothesis that $\theta \varepsilon \omega_1$ against the class of alternatives $\theta \varepsilon \omega_2$, satisfying bounds on the two kinds of error; such a d.r. may be obtained by considering, for each n, tests of size $1 - \alpha_1$ 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 α_2 [7].

Before extending this result to m-d.r.'s for composite discrimination, we require some results in minimax decision theory.

B. *Minimax decision rules for fixed sample sizes.* We prove three theorems which may be useful in finding minimax d.r.'s. Also, if a sufficient statistic with a monotone likelihood ratio exists, Karlin and Rubin's complete class theorem may be applicable [1], [8]. Sverdrup's results [13] should also be noted.

We shall use a number of Wald's results in [14], Section 3.5 and 5.1.4, with some alteration in his assumptions and notation. We denote a weight function by $W(\theta, A_j) = W_j(\theta)$ $(j = 1, \cdots, m)$ and the corresponding risk function when using a d.r. D by $r(\theta, D)$. An a priori distribution over the Borel subsets $\{\omega\}$ of Ω is denoted by $\Xi = (\xi, \lambda)$, where $\Xi(\omega) = \Pr(\theta \varepsilon \omega) = \sum_{i=1}^{l} \xi_i \lambda_i(\omega)$ and $\xi_i = \Xi(\omega_i)$, $\lambda_i(\omega) = \Pr(\theta \varepsilon \omega | \theta \varepsilon \omega_i)$ $(i = 1, \cdots, l)$. The average risk relative to Ξ is denoted by $r(\Xi, D)$. Other terminology and notation will be self-evident. Wald's Assumptions 5.1 and 5.6, his remarks on page 148 characterizing a Bayes solution, and his theorems 5.11, 5.12, 3.8, 3.9, and 3.10 characterizing minimax solutions are especially pertinent to what follows. Lehmann's existence theorem for least favorable distributions [9] might also be noted.

Assumption 2. *For each i, j pair $(i = 1, \cdots, l; j = 1, \cdots, m)$, $W_j(\theta)$ equals a constant, say W_{ij}, for all $\theta \varepsilon \omega_i$.*
(That is, for each alternative, the loss varies only from subset to subset among $\omega_1, \cdots, \omega_l$ and not within any subset.) This assumption is sufficient to imply the validity of Wald's Assumptions 3.1 to 3.6 (see his remarks on page 148). For a given set of conditional distributions $\lambda = (\lambda_1, \cdots, \lambda_l)$, we denote

$$f_i^\lambda(x) = \int_{\omega_i} f^n(x, \theta)\, d\lambda_i \qquad (i = 1, \cdots, l); \tag{12}$$

n is fixed and need not be evident in the notation.

Theorem 6. *If Assumption 2 holds, a necessary and sufficient condition for a d.r. D^* to be a Bayes d.r. relative to $\Xi = (\xi, \lambda)$ for discriminating among $\omega_1, \cdots, \omega_l$ is that D^* be a Bayes d.r. relative to ξ for discriminating among $f_1^\lambda, \cdots, f_l^\lambda$ w.r.t. the weight function W_{ij}. The average risk in the two cases are equal.*

Proof. Using Assumption 2 and (12), we have

$$\int_\Omega W_j(\theta) f^n(x, \theta)\, d\Xi = \sum_{i=1}^{l} \xi_i W_{ij} f_i^\lambda(x).$$

The first part of the theorem follows immediately, using Wald's Theorem 5.1 and second paragraph on page 148. By expressing $r(\theta, D)$ as in Wald's (5.81), interchanging the order of integration, using (12) and Wald's (5.2), we have for any d.r. D,

$$\int_{\omega_i} r(\theta, D)\, d\lambda_i = r(f_i^\lambda, D). \tag{13}$$

Denoting by $r_\lambda(\xi, D)$ the average risk relative to ξ when discriminating among $f_1^\lambda, \cdots, f_l^\lambda$, we have

$$r(\Xi, D) = r_\lambda(\xi, D), \tag{14}$$

completing the proof.

THEOREM 7. *Suppose Assumption 2 holds. Necessary and sufficient conditions that* $\Xi^0 = (\xi^0, \lambda^0)$ *be a least favorable distribution and* D^0 *a minimax d.r. for discriminating among* $\omega_1, \cdots, \omega_l$ *are that*

(i) ξ^0 *is a least favorable distribution and* D^0 *is a minimax d.r. w.r.t.* W_{ij} *for discriminating among* $f_1^{\lambda^0}, \cdots, f_l^{\lambda^0}$; *and*

(ii) *for any i for which* $\xi_i^0 > 0$, $\int_{\omega_i} r(\theta, D^0)\, d\lambda_i^0 = \sup_{\theta \varepsilon \omega_i} r(\theta, D^0)$. *Moreover, the maximum risks in the two cases are equal; i.e.,*

$$\sup_{\Omega} r(\theta, D^0) = \max_{1 \leq i \leq l} r(f_i^{\lambda^0}, D^0). \tag{15}$$

PROOF. Necessity: Since Ξ^0 is least favorable, $\inf_D r(\Xi^0, D) \geqq \inf_D r((\xi, \lambda^0), D)$ for any ξ, so that, using (14), $\inf_D r_{\lambda^0}(\xi^0, D) \geqq \inf_D r_{\lambda^0}(\xi, D)$; that is, ξ^0 is least favorable. Using Wald's Theorem 3.9 and then Theorem 6, D^0 is a Bayes d.r. relative to Ξ^0 and a minimax d.r. for discriminating among $f_1^{\lambda 0}, \cdots, f_l^{\lambda 0}$.

We shall now verify (15). Using Wald's Theorem 5.3 (iii), $\max_i r(f_i^{\lambda 0}, D^0) = \sum_i \xi_i^0 r(f_i^{\lambda 0}, D^0) = r_{\lambda^0}(\xi^0, D^0)$, so that together with (14) and Wald's Theorem 3.10, we have $\max_i r(f_i^{\lambda 0}, D^0) = r(\Xi^0, D^0) = \sup_\Omega r(\theta, D^0)$. Continuing with the necessity, for any i for which $\xi_i^0 > 0$, we have $r(f_i^{\lambda^0}, D^0) = \max_i r(f_i^{\lambda^0}, D^0)$ and $\sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$ by Wald's Theorem 3.10, which, together with (15) and (13), prove (ii).

Sufficiency: By Wald's Theorem 3.9 and Theorem 6, D^0 is a Bayes d.r. relative to $\Xi^0 = (\xi^0, \lambda^0)$; i.e., $r(\Xi^0, D^0) = \inf_D r(\Xi^0, D)$. Hence, we need only prove that Ξ^0 is a least favorable distribution. Suppose it is not; then there exists a $\Xi = (\xi, \lambda)$ such that $\inf_D r(\Xi^0, D) < \inf_D r(\Xi, D)$. But $\inf_D r(\Xi, D) \leqq r(\Xi, D^0) = \sum_i \xi_i \int_{\omega_i} r(\theta, D^0)\, d\lambda_i \leqq \sum_i \xi_i \sup_{\omega_i} r(\theta, D^0) \leqq \sup_\Omega r(\theta, D^0)$. Combining these last three results, $r(\Xi^0, D^0) < \sup_\Omega r(\theta, D^0)$.

By Wald's Theorem 5.3 (iii), 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 (13) 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),

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

a contradiction. Q.E.D.

THEOREM 8. *Suppose Assumption 2 holds, and suppose* $\{\lambda^\nu\}$ *is a sequence of sets of conditional a priori distributions and* D^0 *a d.r. such that*

$$\lim_{\nu=\infty} \int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu = \sup_{\omega_i} r(\theta, D^0) \qquad (i = 1, \cdots, l), \tag{16}$$

where for each $\nu = 1, 2, \cdots$, D^ν *is a minimax d.r. for discriminating among* $f_1^{\lambda^\nu}, \cdots, f_l^{\lambda^\nu}$. *Then* D^0 *is a minimax d.r. for discriminating among* $\omega_1, \cdots, \omega_l$.

PROOF. By Wald's Theorem 5.3, for each ν there exists a least favorable distribution ξ^ν, and D^ν is a Bayes d.r. relative to ξ^ν for discriminating among $f_1^{\lambda^\nu}, \cdots, f_l^{\lambda^\nu}$; i.e., for any d.r. D, $r_{\lambda^\nu}(\xi^\nu, D^\nu) \leqq r_{\lambda^\nu}(\xi^\nu, D)$, and hence, using (14),

$$\sum_i \xi_i^\nu \int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu \leqq \sum_i \xi_i^\nu \int_{\omega_i} r(\theta, D)\, d\lambda_i^\nu \leqq \sup_\Omega r(\theta, D) \tag{17}$$

Now each sequence $\{\xi_i^\nu\}$ has at least one limit point; let $\{\Xi^{\nu_j}\}$, $j = 1, 2, \cdots$, be a sub-sequence of $\{\Xi^\nu = (\xi^\nu, \lambda^\nu)\}$ for which each $\xi_i^{\nu_j}$ converges to a limit, say ξ_i^0; then $\sum_i \xi_i^0 = 1$. By Wald's Theorem 5.3 (iii) and (13), 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 (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 (16), $\lim_{j=\infty} \sum_i \xi_i^{\nu_j} \int_{\omega_i} r(\theta, D^{\nu_j})\, d\lambda_i^{\nu_j} = \sum_i \xi_i^0 \sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$, which, together with (22), asserts $\sup_\Omega r(\theta, D^0) \leqq \sup_\Omega r(\theta, D)$ for any D. Q.E.D.

If a least favorable distribution exists, the problem reduces to one ot simple discrimination, so that if μ is non-atomic only non-randomized d.r.'s need be considered. A lemma for the case of composite discrimination analogous to Lemma 1 may be derived.

C. *Most economical decision rules relative to a vector* α. As in Section 1.B, we shall apply the above theory to two specific weight functions $W_j(\theta)$ and develop in each case a method of obtaining M.E. d.r.'s relative to α. We assume $l = m$. First, let

$$W(\theta, A_j) = W_j(\theta) = -1/\alpha_j \qquad \text{if } \theta \in \omega_j \text{ and 0 otherwise.}[6] \tag{18}$$

The risk w.r.t. $W_j(\theta)$ is $r(\theta, D) = -p_i(\theta, D)/\alpha_i$ if $\theta \in \omega_i$ $(i = 1, \cdots, m)$, and $\sup_{\omega_i} r(\theta, D) = -\inf_{\omega_i} p_i(\theta, D)/\alpha_i$ $(i = 1, \cdots, m)$. By Wald's Theorem 5.12 (i), there exists a minimax d.r. D^0 for any (fixed) sample size.

THEOREM 9. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (18) *for samples of fixed size n. Suppose for some n,* $\sup_\Omega r(\theta, D_n^0) \leqq -1$ *and let N be the least such integer. Then* D_n^0 *is an M.E. d.r. relative to* α *for discriminating among* $\omega_1, \cdots, \omega_m$. *Conversely, if there exists an M.E. d.r. relative to* α *for discriminating among* $\omega_1 . \cdots, \omega_m$, *and the M.E. sample size is N, then* D_N^0 *is an M.E. d.r.*

The proof is like that of Theorem 1, replacing $p_i(D_n)$ by $\inf_{\omega_i} p_i(\theta, D_n)$.

Note that (18) satisfies Assumption 2 with W_{ij} given by (3). Hence, if a least favorable distribution $\Xi^0 = (\xi^0, \lambda^0)$ exists, Theorems 6 and 7 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 Section 1 will be applicable. If a least favorable distribution does not exist, Theorem 8 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 $\Xi^0 = (\xi^0, \lambda^0)$ exists. By Theorem 7(ii),

$$\int_{\omega_i} p_i(\theta, D^0)\, d\lambda_i^0 = \inf_{\theta \in \omega_i} p_i(\theta, D^0) \qquad \text{for any } i \text{ for which } \xi_i^0 > 0.$$

ASSUMPTION 3. If R is a subset of $\mathfrak{X}$ for which $\int_R f^n(x, \theta)\, d\mu^n = 0$ for some $\theta \in \Omega$, then $\int_R f^n(x, \theta)\, d\mu^n = 0$ for all $\theta \in \Omega$.
This assumption implies Assumption 1 for the density functions $f_1^\lambda, \cdots, f_m^\lambda$,

[6] See footnote 3.

defined by (12), for any set of conditional distributions λ. If Assumption 3 holds, and if a least favorable distribution exists, it follows from Theorem 2, Wald's Theorem 5.3(iii) and (18) that

$$\frac{1}{\alpha_1} \inf_{\theta \varepsilon \omega_1} p_1(\theta, D^0) = \cdots = \frac{1}{\alpha_m} \inf_{\theta \varepsilon \omega_m} p_m(\theta, D^0), \tag{19}$$

where D^0 is a minimax d.r.

As a second approach, consider the weight function:

$$W(\theta, A_j) = W_j(\theta) = 1/\beta_i \qquad \text{if } \theta \varepsilon \omega_i,\ i \neq j, \text{ and 0 otherwise,} \tag{20}$$

where $\beta_i = 1 - \alpha_i$ as before. Then $r(\theta, D) = q_i(\theta, D)/\beta_i$ if $\theta \varepsilon \omega_i$ $(i = 1, \cdots, m)$. We may proceed analogously to the first approach, making changes corresponding to those made analogously in Section 1. We thus obtain a theorem analogous to Theorem 9 and also

THEOREM 10. *Suppose Assumption 3 holds and that a least favorable distribution exists. For any sample size greater than or equal to the M.E. sample size,*

$$\frac{1}{\beta_1} \sup_{\theta \varepsilon \omega_1} q_1(\theta, D^0) = \cdots = \frac{1}{\beta_m} \sup_{\theta \varepsilon \omega_m} q_m(\theta, D^0) \tag{21}$$

where D^0 is a minimax d.r.

No proof of admissibility of the M.E. d.r.'s derived in this section has been obtained. However, if Assumption 3 holds and there exists a least favorable distribution, it can easily be verified that there does not exist a d.r. D_N' for which $\inf_{\omega_i} p_i(\theta, D_N') \geqq \inf_{\omega_i} p_i(\theta, D_N^0)$ $(i = 1, \cdots, m)$ with strict inequality for at least one i, where D_N^0 is an M.E. d.r. obtained by either of the minimax methods.

D. *Most economical decision rules relative to a matrix β.* Just as the approach of Section 1.B was extended in Section 1.C, we shall extend the approach of Section 2.C in this section to the consideration of M.E. d.r.'s for composite discrimination relative to $\beta = (\beta_{ij})$.

Suppose n is fixed, and consider parameter spaces $\Omega_1, \cdots, \Omega_m$, each Ω_j being identical to Ω, and denote $\Omega' = U_j\Omega_j$. For each j, denote the corresponding subsets by $\omega_{1j}, \cdots, \omega_{lj}$. Define a weight function $W(\theta, A_k) = W_k(\theta)$ for $k = 1, \cdots, m$, by

$$W_k(\theta) = 1/\beta_{ij} \qquad \text{if } \theta \varepsilon \omega_{ij} \text{ and } j = k \ (i = 1, \cdots, l;\ j = 1, \cdots, m), \text{ and 0 otherwise.} \tag{22}$$

Then $r(\theta, D) = p_j(\theta, D)/\beta_{ij}$ if $\theta \varepsilon \omega_{ij}$. Let Ξ be an a priori distribution over Ω' with components $\xi_{ij} = \Xi(\omega_{ij})$ and $\lambda_{ij}(\omega) = \Pr(\theta \varepsilon \omega \mid \theta \varepsilon \omega_{ij})$. For a given set of λ's, denote

$$g_{ij}^{\lambda}(x) = \int_{\omega_{ij}} f^n(x, \theta)\, d\lambda_{ij}. \tag{23}$$

Theorem 9 may be restated and proved, substituting only (22) for (18), $+1$ for -1, and β for α. The theorems of Section 2.B may be applied to obtain minimax d.r.'s for composite discrimination w.r.t. the weight function (22) by replacing l in the theorems by $l' = l \cdot m$ 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 (23) w.r.t. a set of "least favorable conditional distributions" λ, and Theorem 5 and the remarks of Section 1.C are applicable. Thus, this method of solution gives unlikelihood ratio 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 8 may be applicable.

Example 3. Suppose $f(x, \theta)$ is a normal density function with variance σ^2 (known) and mean θ, and

$$\omega_1 = \{\theta : \theta \leqq \theta_1\}, \qquad \omega_2 = \{\theta : \theta_2' \leqq \theta \leqq \theta_2''\}, \qquad \omega_3 = \{\theta : \theta \geqq \theta_3\},$$

for some specified $\theta_1 < \theta_2' \leqq \theta_2'' < \theta_3$. It may be shown that the least favorable conditional distributions over ω_1, ω_2, ω_3 (Theorem 7) assign probability one to θ_1, θ_2, θ_3, where $\theta_2 = \theta_2'$ or θ_2'' determined below. Thus, this example reduces to Example 1. (Karlin and Rubin's results [8] also imply that a minimax rule will be monotone in $\bar{x}$; determining the explicit form of the monotone rule is equivalent to showing that the above distribution is least favorable.)

θ_2 is determined as follows:

$$\text{(24)} \qquad \begin{aligned} \theta_2 &= \theta_2' \quad \text{if} \quad p_2(\theta_2', D') \leqq p_2(\theta_2'', D'), \\ \theta_2 &= \theta_2'' \quad \text{if} \quad p_2(\theta_2'', D'') < p_2(\theta_2', D''), \end{aligned}$$

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 that such a determination of θ_2 is complete and consistent 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 (9), with either a prime or double-prime on ρ, D, c_1, and c_2, we have $c_1 = \theta_1 + \sigma\Phi^{-1}(\alpha_1\rho)/\sqrt{n}$ and

$$c_2 = \theta_3 + \sigma\Phi^{-1}(1 - \alpha_3\rho)/\sqrt{n},$$

where $\Phi^{-1}(x) = t$ is defined by $\Phi(t) = x$. Substituting in $p_2(\theta, D)$ it becomes clear that it is a decreasing function of ρ for fixed θ. Now $\alpha_2\rho' = p_2(\theta_2', D')$ and $\alpha_2\rho'' = p_2(\theta_2'', D'')$ so that

$$\text{(25)} \qquad \alpha_2(\rho'' - \rho') = p_2(\theta_2'', D'') - p_2(\theta_2', D').$$

Suppose $p_2(\theta_2'', D'') > p_2(\theta_2', D'')$; substituting in (25), it follows that $\rho'' > \rho'$ since p_2 is a decreasing function of ρ. For the same reasons,

$$0 < \alpha_2(\rho'' - \rho') < p_2(\theta_2'', D') - p_2(\theta_2', D').$$

Conversely, in the same manner, if $p_2(\theta_2'', D') > p_2(\theta_2', D')$, then

$$\alpha_2(\rho'' - \rho') > p_2(\theta_2'', D'') - p_2(\theta_2'', D'),$$

and ρ'' must be greater than ρ'; hence,

$$0 < \alpha_2(\rho'' - \rho') < p_2(\theta_2'', D'') - p_2(\theta_2', D'').$$

Other examples with exponential density functions may be treated analogously, and also similar examples for Section 2.D.

Example 4. Now suppose σ is also unknown; denote the mean by μ and replace θ in the ω_i's defined in Example 3 by μ/σ.

Denoting Student's ratio by t and the sample sum of squares by s^2, (t, s) is sufficient for $\theta = (\mu, \sigma)$. If we invoke invariance (under changes in scale), it follows from Blackwell and Girshick's work [1] that a minimax invariant rule must be monotone in t. Theorem 8.8.1 in [1] proves, for the m-decision case as well as the 2-decision case, that invariance is no restriction when discriminating among $\theta_1, \cdots, \theta_m$, where $\theta = \mu/\sigma$. Thus a minimax d.r. for discriminating among $\theta_1, \theta_2, \theta_3$ is monotone in t. By showing that the risk for a monotone rule is a maximum in ω_i at $\mu/\sigma = \theta_i$ (with θ_2 determined as in Example 3), it will follow that a monotone rule in t, with c_1, c_2 and ρ determined by equations of the form (9) with the Φ's replaced by non-central t distribution functions, will be minimax for discriminating among $\omega_1, \omega_2, \omega_3$.

Alternatively, this same result may be obtained by an application of our Theorem 8, letting λ_i^ν assign probability one to sets of (μ, σ) in which

$$\mu/\sigma = \theta_i$$

and letting σ^{-2} be distributed as χ^2 with degrees of freedom tending to 0 as $\nu \to \infty$. The details appear in [3], adapted from a 2-d.r. argument by Hoeffding.

Example 5. We shall derive a three-decision extension of the sign test for the median of an arbitrary distribution function by adapting an example of Hoeffding [6]. (See also [12].) Analogously, an M.E. d.r. concerning any quartile of an arbitrary distribution may be derived.

Let Ω be the class of all density functions f w.r.t. a fixed measure μ on the real line such that $\mu\{x \leqq 0\} > 0$, $\mu\{x > 0\} > 0$. Denote $\theta(f) = \int_{-\infty}^{0} f(x)\, d\mu$. Given $\theta_1, \theta_2', \theta_2'', \theta_3$ $(0 < \theta_1 < \theta_2' \leqq \frac{1}{2} \leqq \theta_2'' < \theta_3 < 1)$, let $\omega_1 = \{f:\theta(f) \leqq \theta_1\}$, $\omega_2 = \{f:\theta_2' \leqq \theta(f) \leqq \theta_2''\}$, $\omega_3 = \{f:\theta(f) \geqq \theta_3\}$. The alternatives A_1, A_2, A_3, corresponding to $\omega_1, \omega_2, \omega_3$, might be that the median of the unknown distribution is "appreciably" less than zero, "close" to zero, "appreciably" greater than zero, respectively.

Let $f(x, \theta) = \theta^{b(x)}(1 - \theta)^{1-b(x)}/c$ if $x \leqq c$ and 0 otherwise where c is an arbitrary positive constant and $b(x) = 1$ if $x \leqq 0$ and 0 otherwise, and let $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ be a set of conditional distributions over $\omega_1, \omega_2, \omega_3$, respectively, where λ_i assigns probability 1 to $f(x, \theta_i)$ and where θ_2 is to be determined as in Example 3. It is easily verified that a minimax d.r. D_n (n fixed) for discriminating among $f_1^\lambda, f_2^\lambda, f_3^\lambda$ is monotone in $t(x) = \sum_k b(x_k)$, the number of non-positive observations, with c_1, c_2 and values of ϕ_i when $t = c_1$ or c_2 determined so that $p_i(\theta_i, D_n) = \alpha_i\rho$ $(i = 1, 2, 3)$ for some ρ; and

$$p_1(\theta, D_n) = B(c_1 - 1) + a_1 b(c_1), \qquad p_3(\theta, D_n) = 1 - B(c_2) + (1 - a_2)b(c_2),$$

$$p_2(\theta, D_n) = B(c_2 - 1) + a_2 b(c_2) - B(c_1) + (1 - a_1)b(c_1),$$

where $B = B_{n,\theta}$ and $b = b_{n,\theta}$ denote the binomial distribution function and probability function, respectively, and $a_i = \phi_i(c_i) = 1 - \phi_{i+1}(c_i)$. (It may be shown that D_n defined above is also minimax for discriminating among

$$b_{n,\theta_1}, \quad b_{n,\theta_2}, \quad b_{n,\theta_3}.)$$

This λ may be shown to be least favorable, and an M.E. d.r. may be obtained according to Theorem 9 (see Example 1).

3. Acknowledgments. 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.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[2] A. Dvoretzky, A. Wald and J. Wolfowitz, "Elimination of randomization in certain statistical decision procedures and zero-sum two-person games," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 1–21.

[3] W. J. Hall, "Most economical multiple-decision rules," Institute of Statistics Mimeograph Series No. 115, University of North Carolina, 1954.

[4] W. J. Hall, "An optimum property of Bechofer's single-sample multiple-decision procedure for ranking means and some extensions," Institute of Statistics Mimeograph Series No. 118, University of North Carolina, 1954.

[5] W. J. Hall, "On uniformly consistent sequences of multiple-decision rules," abstract, *Ann. Math. Stat.*, Vol. 27 (1956), p. 221.

[6] Wassily Hoeffding, " 'Optimum' nonparametric tests," *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 83–92.

[7] Wassily Hoeffding, "Lectures on estimation and testing of hypotheses," University of North Carolina, fall quarter 1952 (unpublished).

[8] S. Karlin and H. Rubin, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

[9] E. L. Lehmann, "On the existence of least favorable distributions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 408–416.

[10] D. V. Lindley, "Statistical inference," *J. Royal Stat. Soc.*, Series B, Vol. 15 (1953), pp. 30–76.

[11] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952, p. 311.

[12] Erik Ruist, "Comparison of tests for non-parametric hypotheses," *Arkiv for Matematik*, Vol. 3 (1954), pp. 133–163.

[13] Erling Sverdrup, "Weight functions and minimax procedures in the theory of statistical inference," *Archiv for Mathematik og Naturvidenskab*, Vol. 51 (1952), pp. 1–73.

[14] Abraham Wald, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

[15] A. Wald and J. Wolfowitz, "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, pp. 149–157.