CHARACTERIZATION OF THE MINIMAL COMPLETE CLASS OF DECISION FUNCTIONS WHEN THE NUMBER OF DISTRIBUTIONS AND DECISIONS IS FINITE

A. WALD AND J. WOLFOWITZ
COLUMBIA UNIVERSITY

1. Introduction

The principal object of the present paper is to prove theorem 2 below. This theorem characterizes the minimal complete class in the problem under consideration, and improves on the result of theorem 1. Theorem 1 has been proved by one of us in much greater generality [1]. The proof given below is new and very expeditious. Another reason for giving the proof of theorem 1 here is that it is the first step in our proof of theorem 2. A different proof of theorem 1, based, like ours, on certain properties of convex bodies in finite Euclidean spaces, was communicated earlier to the authors by Dr. A. Dvoretzky. Theorem 3 gives another characterization of the minimal complete class.

Let x be the generic point of a Euclidean[1] space Z, and $f_1(x), \ldots, f_m(x)$ be any m (> 1) distinct cumulative probability distributions on Z. The statistician is presented with an observation on the chance variable X which is distributed in Z according to an unknown one of $f_1, \ldots, f_m$. On the basis of this observation he has to make one of l decisions, say $d_1, \ldots, d_l$. The loss incurred when x is the observed point, f_i is the actual (unknown) distribution, and the decision d_j is made, is $W_{ij}(x)$, where $W_{ij}(x)$ is a measurable function of x such that

$$\int_Z | W_{ij}(x) | \, df_i < \infty , \qquad i = 1, \ldots, m ; \quad j = 1, \ldots, l .$$

A randomized decision function $\eta(x)$, say, hereafter often called "test" for short, is defined as follows: $\eta(x) = [\eta_1(x), \eta_2(x), \ldots, \eta_l(x)]$ where

(a) $\eta(x)$ is defined for all x,

(b) $0 \leqq \eta_j(x), j = 1, \ldots, l,$

(c) $\sum_{j=1}^{l} \eta_j(x) = 1$ identically in x,

(d) $\eta_j(x)$ is measurable, $j = 1, \ldots, l$.

This research was done under a contract with the Office of Naval Research.

[1] The extension to general abstract spaces is trivial and we forego it. This entire paper could be given an abstract formulation without the least mathematical difficulty.

The statistical application of the function $\eta(x)$ is as follows: After the observation x has been obtained the statistician performs a chance experiment to decide which decision to make. The probability of making decision d_j according to this experiment is $\eta_j(x)$ $(j = 1, \ldots, l)$. The risk of the test $\eta(x)$, which will also be called its associated risk or risk point, is the complex $r(\eta) = (r_1, \ldots, r_m)$, where

$$r_i = \int_Z \left(\sum_{j=1}^{l} \eta_j(x) W_{ij}(x) \right) d f_i .$$

Let V be the totality of all risk points (in m-dimensional space) corresponding to all possible tests. It follows from results of Dvoretzky, Wald, and Wolfowitz[2] that the set V is closed and convex.

The test T with risk $r = (r_1, \ldots, r_m)$ is called uniformly better than the test T' with risk $r' = (r'_1, \ldots, r'_m)$ if $r_i \leqq r'_i$ for all i and the inequality sign holds for at least one i. A test T is called admissible if there exists no test uniformly better than T. A test which is not admissible may also be called inadmissible. A class C_0 of tests is called complete if, for any test T' not in C_0, there exists a test T in C_0 which is uniformly better than T'. A complete class is said to be minimal if no proper subclass of it is complete.

2. Proof of the complete class theorem

We first prove:

LEMMA. *The class C of all admissible tests is a minimal complete class.*[3]

If C is complete it is obviously minimal. Suppose C is not complete. Then there exists an inadmissible test T_1 such that no member of C is uniformly better than T_1. Since T_1 is inadmissible there exists an inadmissible test T_2 which is uniformly better than T_1. Consequently there exits a test T_3 which is uniformly better than T_2. Hence T_3 is uniformly better than T_1, and is therefore inadmissible. Proceeding in this manner we obtain a denumerable sequence $T_1, T_2, \ldots$ of tests, each test inadmissible and uniformly better than all its predecessors. Since the set V is closed it follows that there exists a test T_ω which is uniformly better than every member of the sequence $T_1, T_2, \ldots$. Hence T_ω is inadmissible. Repeating this procedure we obtain a nondenumerable well ordered set of inadmissible tests, each uniformly better than all its predecessors. Since each risk point has m components we can therefore obtain a nondenumerable well ordered set of real numbers, each smaller than any of its predecessors. Since this is impossible the lemma is proved.

Let $\xi_0 = (\xi_{01}, \ldots, \xi_{0m})$ be an *a priori* probability distribution on the set consisting of $f_1, \ldots, f_m$. A test T_0 with the property that it minimizes

$$\sum_{i=1}^{m} \xi_{0i} r_i(T)$$

[2] A statement of some of the results is given in the *Proc. Nat. Acad. Sci. U.S.A.*, April, 1950. The fact that V is closed whatever be the f's follows from the complete results which, it is hoped, will be published shortly. The closure of V was also proved by one of the authors [2] under the assumption that the f's admit elementary probability laws.

[3] The fundamental idea of the proof of this lemma is already present in the proof of theorem 2.22 in the book by Wald [2]. Since the proof is so brief it is given here for completeness.

with respect to all tests T is called a Bayes solution with respect to ξ_0, or simply a Bayes solution when it is not necessary to specify ξ_0. Let $\xi_1, \ldots, \xi_h$ be a sequence of (*a priori*) probability distributions (each with m components). A Bayes solution with respect to the sequence $\xi_1, \ldots, \xi_h$ will be defined inductively as follows: When $h = 1$ it is a Bayes solution with respect to ξ_1. For $h > 1$ it is any test T_0 which minimizes

$$\sum_{i=1}^{m} \xi_{hi} r_i (T)$$

with respect to all tests T which are Bayes solutions with respect to the sequence $\xi_1, \ldots, \xi_{h-1}$. Since the set V of risk points is closed it follows that, for any sequence $\xi_1, \ldots, \xi_h$, a Bayes solution exists.

THEOREM 1. *Every admissible test is a Bayes solution with respect to some a priori distribution.* (Hence the class of Bayes solutions is complete.)

PROOF. Let $b = (b_1, \ldots, b_m)$ be a generic point in an m-dimensional Euclidean space. Let the set $B(b)$ be the set of all points $x = x_1, \ldots, x_m$ such that x is different from b and

$$x_i \leqq b_i, \qquad i = 1, \ldots, m .$$

Let the set $B'(b)$ be the set which consists of b and $B(b)$. Suppose T is an admissible test and $r = (r_1, \ldots, r_m)$ is its associated risk point. Since T is admissible r is a boundary[4] point of V and the set $VB(r)$ is empty.

Now V and $B'(r)$ are closed convex sets with only the boundary point r in common, and $B'(r)$ contains interior points. Hence there exists a plane π_1 through r, given by $\mu_1(b) = 0$, where

$$\mu_1 (b) = \sum_{i=1}^{m} t_{1i} (b_i - r_i) , \tag{1}$$

such that $\mu_1(b) \geqq 0$ in one of V and $B'(r)$, and $\mu_1(b) \leqq 0$ in the other. Reversing the signs of all the t_{1i}'s, if necessary, we can assume that some t_{1i}, say t_{1e}, is positive. Let $K(e)$ be the point each of whose coordinates is r_i except the e-th, which is K. When K is sufficiently small, $t_{1e}(K - r_e) < 0$. Hence for every point of $B(r)$ we have $\mu_1(b) \leqq 0$. From this it follows that every $t_{1i} \geqq 0$. For suppose that t_{1g}, say, were < 0. The point $K(g)$, with K sufficiently small, would be in $B(r)$ and yet $\mu_1[K(g)] > 0$. Thus every $t_{1i} \geqq 0$. We have

$$\mu_1(b) \geqq 0 \tag{2}$$

for every point in V. Hence the point r minimizes $\mu_1(b)$ for every point in V. Therefore T is a Bayes solution with respect to the *a priori* probability distribution ξ_1 whose i-th component ξ_{1i} $(i = 1, \ldots, m)$ is

$$\xi_{1i} = \frac{t_{1i}}{\sum_{i=1}^{m} t_{1i}} .$$

This proves the theorem.

[4] Here the notions of inner point and boundary point are relative to the surrounding m-dimensional space.

3. First characterization of admissible tests

We now prove the main result:

THEOREM 2. *In order that a test T be admissible it is necessary and sufficient that it be a Bayes solution with respect to a sequence of h ($\leqq m$) a priori probability distribution functions ($\xi_1, \ldots, \xi_h$), such that the matrix $\{\xi_{ij}\}$, $i = 1, \ldots, h$; $j = 1, \ldots, m$, has the following properties:* (a) *for any j there exists an i such that $\xi_{ij} > 0$,* (b) *the matrix $\{\xi_{ij}\}$, $i = 1, \ldots, (h - 1)$; $j = 1, \ldots, m$, does not possess property* (a).

PROOF. The sufficiency of the above condition is easy to see. We proceed at once to the proof of necessity.

Let therefore r be the risk point of an admissible test T. By theorem 1 T is a Bayes solution with respect to ξ_1. We shall carry over the notation of theorem 1, except that, for typographical simplicity, we shall put r at the origin. (We may do this without loss of generality.) The origin will be written for short as the point 0. Let V_1 be the intersection of V with the plane π_1 defined by $\mu_1(b) = 0$. V_1 is convex and closed. Suppose it is of dimensionality $m - c_1$, $2 \leqq c_1 \leqq m$. Let the vector α denote the generic point in V_1. Let the vector β be any point in the plane π_1 and not in $B(0)$ such that the convex hull V_1' of V_1 and β is of dimensionality $m - c_1 + 1$. Let V_1'' be the convex hull of V_1 and $(-\beta)$. We now assert that either V_1' or V_1'' has no points in common with $B(0)$. For suppose to the contrary that

$$q_1\alpha_1 + (1 - q_1)\beta$$

and

$$q_2\alpha_2 - (1 - q_2)\beta$$

with

$$\alpha_1 \in V_1, \quad \alpha_2 \in V_1, \quad 0 \leqq q_1 \leqq 1, \quad 0 \leqq q_2 \leqq 1,$$

are both in $B(0)$. Moreover, $q_1 \neq 0$ since $\beta \notin B(0)$, and $q_1 \neq 1$, $q_2 \neq 1$ since 0 is admissible. Hence

$$q_1 (1 - q_2)\, \alpha_1 + (1 - q_1)(1 - q_2)\, \beta$$

and

$$q_2 (1 - q_1)\, \alpha_2 - (1 - q_1)(1 - q_2)\, \beta$$

are each in $B(0)$. Hence

$$q_1 (1 - q_2)\, \alpha_1 + q_2 (1 - q_1)\, \alpha_2$$

is also in $B(0)$, and consequently

$$\alpha_0 = \frac{q_1 (1 - q_2)\, \alpha_1 + q_2 (1 - q_1)\, \alpha_2}{q_1 (1 - q_2) + q_2 (1 - q_1)}$$

is in $B(0)$. Now α_0 lies in the line segment from α_1 to α_2, and hence is in V_1. This contradicts the fact that 0 is admissible and proves our assertion that either V_1' or V_1'' has no points in common with $B(0)$.

We repeat the above procedure $(c_1 - 1)$ times and conclude: There exists a closed convex set V_1^* which contains V_1, lies entirely in π_1, is of dimensionality $m - 1$, and has no points in common with $B(0)$.

Suppose γ_1 of the t_{1i} are positive. If $\gamma_1 = m$ the theorem is proved. Assume therefore that $\gamma_1 < m$. Without loss of generality we assume

$$t_{1i} > 0, \qquad i \leqq \gamma_1$$
$$t_{1i} = 0, \qquad i > \gamma_1 .$$

Let $B(0; 1)$ be the set of all points b different from 0 such that

$$b_i = 0 , \quad i \leqq \gamma_1$$
$$b_i \leqq 0 , \quad i > \gamma_1 .$$

Let $B'(0; 1)$ be the set of points consisting of 0 and $B(0; 1)$. The closed convex sets V_1^* and $B'(0, 1)$ both lie in π_1 and have only the origin 0 in common. 0 is obviously a boundary[5] point of $B'(0; 1)$. It is also a boundary point of V_1^* because $B'(0; 1)$ is of dimensionality $\geqq 1$ $(\gamma_1 < m)$, and $V_1^* B(0)$ is empty. The set V_1^* has inner[5] points. Hence there exists an $(m - 2)$-dimensional linear subspace π_2 of π_1, defined by $\mu_1(b) = 0$ and $\mu_2(b) = 0$, where

$$\mu_2 (b) = \sum_{i=1}^{m} t_{2i} b_i \tag{3}$$

such that $\mu_2(b) \geqq 0$ in one of V_1^* and $B'(0, 1)$, and $\mu_2(b) \leqq 0$ in the other.

We now consider two cases:

(a) $t_{2i} \neq 0$ for some $i > \gamma_1$. Without loss of generality we assume $t_{2(\gamma_1+1)} \neq 0$. Hence there exists a real number $\lambda_2 \neq 0$ such that

$$t_{1i} + \lambda_2 t_{2i} > 0 , \quad i \leqq \gamma_1 + 1 . \tag{4}$$

The space π_2 can also be defined by $\mu_1(b) = 0$ and $\mu_1(b) + \lambda_2\mu_2(b) = 0$. We will now show that

$$t_{1i} + \lambda_2 t_{2i} \geqq 0 , \quad i > \gamma_1 + 1 . \tag{5}$$

For suppose that, say,

$$t_{1e} + \lambda_2 t_{2e} < 0 , \quad e > \gamma_1 + 1 . \tag{6}$$

By the definition of π_2 the sign of $\mu_1(b) + \lambda_2\mu_2(b)$ does not change in $B(0; 1)$. Using the point $K(\gamma_1 + 1)$ with K negative we see that $\mu_1(b) + \lambda_2\mu_2(b) \leqq 0$ for b in $B(0; 1)$. If (6) held we would have

$$\mu_1 [K (e)] + \lambda_2 \mu_2 [K (e)] > 0$$

for K negative, in violation of what we have just proved. Hence (5) must hold.

(b) $t_{2i} = 0$ for $i > \gamma_1$.

Consider the expressions

$$M_1 (b) = \mu_1 (b) + \lambda \mu_2 (b) \tag{7}$$

and

$$M_2 (b) = \mu_1 (b) - \lambda \mu_2 (b) . \tag{8}$$

For sufficiently small positive λ all their coefficients are nonnegative. Both expressions do not change sign in V_1^*, because of the definition of π_2. We assert that either $M_1(b) \geqq 0$ in V_1^* or $M_2(b) \geqq 0$ in V_1^*. For $M_1(b)$ and $M_2(b)$ cannot be identically zero on V_1^*, because V_1^* lies in π_1 and is of dimensionality $(m - 1)$. Let b_0 be some point in V_1^* where $M_1(b_0) \neq 0$. Since $M_1(b_0) + M_2(b_0) = 2\mu_1(b_0) = 0$, it follows that $M_2(b_0) \neq 0$, and either $M_1(b_0)$ or $M_2(b_0)$ is positive.

[5] Here the notions of inner point and boundary point are relative to the surrounding space π_1.

But then either $M_1(b)$ or $M_2(b)$ is nonnegative for every b in V_1^*, which is the assertion to be proved. Let $M(b)$ denote that one of $M_1(b)$ and $M_2(b)$ for which $M(b) \geqq 0$ for every point b in V_1^*, and let λ_2 denote that one of λ, $-\lambda$ which is associated with $M(b)$.

In both case (a) and case (b) we have that the test T with associated risk point 0 is a Bayes solution with respect to ξ_1, ξ_2, where

$$\xi_{2i} = \frac{t_{1i} + \lambda_2 t_{2i}}{\sum_{i=1}^{m} (t_{1i} + \lambda_2 t_{2i})} .$$

We redefine $\mu_2(b)$ so that $t_{2i} = \xi_{2i}$. This will help to simplify the notation.

If ξ_1 and ξ_2 do not fulfill the conditions of the theorem for $h = 2$ the above procedure can be repeated. We shall sketch the procedure which yields π_3, π_1 and π_2 having been previously obtained.

Let V_2 be the intersection of π_2 and V_1^*. If V_2 is of dimensionality less than $(m - 2)$ proceed as before to obtain V_2^* which is closed, convex, contains V_2, has no point in common with $B(0)$, and is of dimensionality $(m - 2)$. Let U be the set of integers $i \leqq m$ such that

$$\xi_{1i} = \xi_{2i} = 0$$

and let $\bar{U}$ be the complementary set. The set U is not empty, for else the theorem would be already proved. Let $B(0; 2)$ be the set of all points b different from 0 such that

$$b_i = 0 , \qquad i \in \bar{U}$$

$$b_i \leqq 0 , \qquad i \in U .$$

Let $B'(0; 2)$ be the set of points consisting of 0 and $B(0; 2)$. The closed convex sets $B'(0; 2)$ and V_2^* are separated by an $(m - 3)$-dimensional linear subspace π_3 of π_2 which passes through 0 and may be defined by $\mu_1(b) = 0$, $\mu_2(b) = 0$, and $\mu_3(b) = 0$, where,

$$\mu_3(b) = \sum_{i=1}^{m} t_{3i} b_i .$$

As before, we distinguish two cases. Case (a) occurs when $t_{3i} \neq 0$ for some index $i \in U$. As before, we prove that for suitable $\lambda \neq 0$ the expression

$$\mu_1(b) + \mu_2(b) + \lambda \mu_3(b) \tag{9}$$

has all coefficients nonnegative. Case (b) occurs when $t_{3i} = 0$ for every $i \in U$. For $|\lambda|$ sufficiently small we have then that (9) has all coefficients nonnegative, and either (9) or

$$\mu_1(b) + \mu_2(b) - \lambda \mu_3(b) \tag{10}$$

can be shown as before to be nonnegative in V_2^*. We obtain ξ_3 in this manner.

The above procedure can be repeated as long as the corresponding set U is not empty. However, when the set U is empty, the theorem is proved. The set U will be empty in at most m steps of the procedure.

Suppose for a moment that the f_i all possess density functions f_i^*. A Bayes solution $\eta(x)$ with respect to ξ_1 may be found as follows: $\eta_j(x) = 0$ for all j for which

$$\nu_{1j}(x) = \sum_{i=1}^{m} \xi_{1i} f_i^*(x) W_{ij}(x)$$

is not a minimum with respect to j ($j = 1, \ldots, l$); $\eta_j(x)$ is defined arbitrarily between zero and one, inclusive, for all other j, provided only that every component of the resulting $\eta(x)$ is measurable and the sum is always one. If a Bayes solution with respect to ξ_1, ξ_2 is desired one can proceed as follows: First, define $\eta_j(x) = 0$ for all j for which $\nu_{1j}(x)$ is not a minimum. Among the remaining j define $\eta_j(x) = 0$ for those j for which

$$\nu_{2j}(x) = \sum_{i=1}^{m} \xi_{2i} f_i^* W_{ij}(x)$$

is not a minimum (for these j). Define $\eta_j(x)$ arbitrarily between zero and one, inclusive, for all other j, subject to the requirements of measurability and the fact that the components must add to one. A Bayes solution with respect to $\xi_1, \ldots, \xi_h$ can be obtained similarly.

If the f_i are not absolutely continuous we can proceed as follows: Let τ be the finite measure defined for any Borel set Σ by

$$\tau(\Sigma) = \sum_{i=1}^{m} P\{\Sigma \mid f_i\} .$$

Then every f_i is absolutely continuous with respect to τ and hence, by the Radon-Nikodym theorem, possesses a density function with respect to τ. We can then proceed as before.

4. Second characterization of admissible tests

We return to the problem of characterizing admissible solutions and shall describe another procedure of doing so. Let $\xi_1, \ldots, \xi_u$ be any sequence of *a priori* distributions with the property that for each j ($j = 1, \ldots, m$), there exists exactly one i ($i = 1, \ldots, u$) such that $\xi_{ij} > 0$. We shall call this property the property U. Let $v[i, 1], \ldots, v[i, h(i)]$ be the set of integers j ($j = 1, \ldots, m$), for which $\xi_{ij} > 0$ ($i = 1, \ldots, u$). Let

$$r(1) = [r(1, 1), \ldots, r(1, m)]$$

be the risk point of any Bayes solution with respect to ξ_1. Let $\bar{V}_1$ be the intersection of V with the planes

$$b_{v[1,1]} = r(1, v[1, 1])$$
$$\cdots\cdots\cdots$$
$$b_{v[1,h(1)]} = r(1, v[1, h(1)]) .$$

Let

$$r(2) = (r(2, 1), \ldots, r(2, m))$$

be any Bayes solution with respect to ξ_2 among the elements of $\bar{V}_1$. (Since $\bar{V}_1$ is closed and bounded at least one such solution exists.) Let $\bar{V}_2$ be the intersection of $\bar{V}_1$ with the planes

$$b_{v[2,1]} = r(2, v[2, 1])$$

$$\cdots \quad \cdots\cdots$$

$$b_{v[2,h(2)]} = r(2, v[2, h(2)]) .$$

Let $r(3)$ be any Bayes solution with respect to ξ_3 among the elements of $\bar{V}_2$, etc., etc. The end product of this procedure is the set $\bar{V}_u$.

THEOREM 3. *The class C of all admissible tests coincides with the class of tests with risk points in* $\bar{V}_u$ *for all sequences* $(\xi_1, \ldots, \xi_u)$ *with the property U.*

PROOF. First we prove that any risk point z in $\bar{V}_u$ is admissible. Suppose that it is not admissible, that z' is uniformly better than z, and that i $(i = 1, 2, \ldots, u)$ is the smallest integer such that for an index j which is a member of

$$v[i, 1], \ldots, v[i, h(i)]$$

the j-th coordinate of z' is less than that of z. We see that z' must lie in

$$\bar{V}_1, \bar{V}_2, \ldots, \bar{V}_{i-1} .$$

Consequently z cannot lie in $\bar{V}_i$, which contradicts the hypothesis that z is in $\bar{V}_u \subset \bar{V}_i$.

Let now z be any point in C. It must be a Bayes solution with respect to, say, ξ_1. If z is the unique Bayes solution with respect to ξ_1, or if ξ_{1j} is positive for every j from 1 to m, there is nothing left to prove. Assume therefore that neither of these is true. We define ξ_1 as the first member of the sequence $\xi_1, \ldots, \xi_u$ which we want to construct. Define $\bar{V}_1$ as before. A reëxamination of the proof of theorem 1 shows that the only property of V that was used in the proof is that V is a convex body (that is, a closed, convex, bounded set). But $\bar{V}_1$ has this property. Hence the argument of theorem 1 can be applied to $\bar{V}_1$ to obtain an *a priori* distribution ξ_2 such that (a) z is a Bayes solution with respect to ξ_2, if one limits one's self to points of $\bar{V}_1$, (b) $\xi_{2j} = 0$ for any j such that $\xi_{1j} > 0$. Repeating the above procedure we obtain the desired result.

5. Concluding remarks

I) The only property of V used in theorems 1, 2, and 3 is that V is a convex body. Suppose that, for some reason, the statistician is limited to choosing one of a given proper subclass of the class of all tests. If the set of risk points of this subclass is a convex body, theorems 1, 2, and 3 will hold for this subclass.

II) The only use that was made in the preceding arguments of the fact that the number l of possible decisions is finite, was in invoking the result of Dvoretzky, Wald, and Wolfowitz that V is a convex body. Suppose now that the number of possible decisions is no longer finite. For each x, $\eta(x)$ is then a probability measure on a Borel field of subsets of the space D of decisions (see [2], for example). The risk point of a test is defined appropriately. If V is a convex body then theorems 1, 2, and 3 remain valid. If V is a convex body for a subclass of the class of all tests then theorems 1, 2, and 3 are valid for this subclass. If the class of available tests is the class of all possible tests it is obvious that V is convex. Whether V is closed will in general depend upon W and the space D.

REFERENCES

[1] A. WALD, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), pp. 279–313.
[2] ———, *Statistical Decision Functions*, Wiley, New York, 1950.