BAYES SOLUTIONS OF SEQUENTIAL DECISION PROBLEMS

By A. Wald and J. Wolfowitz
Columbia University

Summary. The study of sequential decision functions was initiated by one of the authors in [1]. Making use of the ideas of this theory the authors succeeded in [4] in proving the optimum character of the sequential probability ratio test. In the present paper the authors continue the study of sequential decision functions, as follows:

a) The proof of the optimum character of the sequential probability ratio test was based on a certain property of Bayes solutions for sequential decisions between two alternatives, the cost function being linear. This fundamental property, the convexity of certain important sets of a priori distributions, is proved in Theorem 3.9 in considerable generality. The number of possible decisions may be infinite.

b) Theorem 3.10 and section 4 discuss tangents and boundary points of these sets of a priori distributions.

(These results for finitely many alternatives were announced by one of us in an invited address at the Berkeley meeting of the Institute of Mathematical Statistics in June, 1948)[1]

c) Theorem 3.6 is an existence theorem for Bayes solutions. Theorem 3.7 gives a necessary and sufficient condition for a Bayes solution. These theorems generalize and follow the ideas of Lemma 1 of [4]

d) Theorems 3.8 and 3.8.1 are continuity theorems for the average risk function. They generalize Lemma 3 in [4]

e) Other theorems give recursion formulas and inequalities which govern Bayes solutions.

1. Introduction. In a previous publication of one of the authors [1] the decision problem was formulated as follows: Let $X = \{x_i\}$ ($i = 1, 2, \cdots$, ad inf.) be a sequence of chance variables. An observation on X is given by a sequence $x = \{x_i\}$ ($i = 1, 2, \cdots$, ad inf.) of real values, where x_i denotes the observed value of X_i. A sequence x is also called a sample or sample point, and the totality M of all possible sample points x is called the sample space. Let $G(x)$ denote the probability that $X_i < x_i$ for $i = 1, 2, \cdots$, ad inf.; i.e., G is the cumulative distribution function of X. In a statistical decision problem G is assumed to be unknown. It is merely known that G is an element of a given class Ω of distribution functions. There is given, furthermore, a space D^* whose elements d represent the possible decisions that can be made in the problem under consideration.

[1] A brief statement of some of the results of the present paper is to be found in the authors' paper of the same name in the *Proc. Nat. Acad. Sci. U. S. A.*, Vol. 35 (1949), pp. 99–102.

The problem is to construct a function $d = D(x)$, called **the decision** function, which associates with each sample point x an element d of D^* so that the decision $d = D(x)$ is made when x is observed.

Occasionally we shall use the symbol D to denote a decision function $D(x)$. This will be done especially when we want to emphasize that we mean the whole decision function and not merely a particular value of it corresponding to some particular x.

If $d = D(x)$ is the decision function adopted and if $x^0 = \{x_i^0\}$ $(i = 1, 2, \cdots)$ is the particular sample point observed, the number of components of x^0 we have to observe in order to reach a decision is equal to the smallest positive integer $n = n(x^0)$ with the property that $D(x) = D(x^0)$ for any x for which $x_1 = x_1^0, \cdots, x_n = x_n^0$. If no finite n exists with the above property, we put $n(x) = \infty$. If $d(x)$ is equal to a constant d, we put $n(x) = 0$. We shall call $n(x)$ the number of observations required by D when x is the observed sample. Of course, $n(x)$ depends also on the decision rule D adopted. To put this in evidence, we shall occasionally write $n(x, D)$ instead of $n(x)$. If D_0 is a decision function such that $n(x, D_0)$ has a constant value over the whole sample space M, we have the classical non-sequential case. If $n(x, D_0)$ is not constant, we shall say that D_0 is a sequential decision function.

In the remainder of this section we shall sketch briefly some of the fundamental notions of the theory without regard to regularity conditions. The latter will be discussed in the next section.

In [1] a weight function $W(G, d)$ was introduced which expresses the loss suffered by the statistician when G is the true distribution of X and the decision d is made. Let $c(n)$ denote the cost of making n observations; i.e., $c(n)$ is the cost of observing the values of $X_1, \cdots, X_n$. Then, if the decision function $d = D(x)$ is adopted and G is the true distribution of X, the expected value of the loss due to possible erroneous decisions plus the expected cost of experimentation is given by

$$r(G, D) = \int_M W[G, D(x)]\, dG(x) + \int_M c[n(x, D)]\, dG(x). \tag{1.1}$$

The above expression is called the risk when D is the decision function adopted and G is the true distribution.

Let ξ be an a priori probability distribution on Ω; i.e., ξ is a probability measure defined over a suitably chosen Borel field[2] of subsets of Ω. Then the expected value of $r(G, D)$ is given by

$$r(\xi, D) = \int_\Omega r(G, D)\, d\xi. \tag{1.2}$$

[2] A Borel field is an aggregate of sets such that a) the null set is a member of the field, b) the complement with respect to the entire space (here M) is a member of the field, c) the sum of denumerably many members of the field is itself in the field.

The above expression is called the risk when ξ is the a priori distribution on Ω and D is the decision function adopted.

We shall say that the decision function D_0 is a Bayes solution relative to the a priori distribution ξ if

$$r(\xi, D_0) \leqq r(\xi, D) \text{ for all } D. \tag{1.3}$$

If there existed an a priori distribution on Ω and if this distribution were known, we could put ξ equal to this a priori distribution and a Bayes solution relative to ξ would provide a very satisfactory solution of the decision problem. In most applications, however, not even the existence of an a priori distribution can be postulated. Nevertheless, the study of Bayes solutions corresponding to various a priori distributions is of great interest in view of some results given in [1]. It was shown in [1] that under rather general conditions the class C of the Bayes solutions corresponding to all possible a priori distributions ξ has the following property: If D_1 is a decision function that is not an element of C, there exists a decision function D_2 in C such that

$$r(G, D_2) \leqq r(G, D_1) \text{ for all } G \tag{1.4}$$

and

$$r(G, D_2) < r(G, D_1) \text{ for at least one } G. \tag{1.5}$$

It was furthermore shown in [1] that under general conditions a minimax solution D_0 of the decision problem is also a Bayes solution corresponding to some a priori distribution ξ. By a minimax solution we mean a decision function D_0 such that, for all D

$$\operatorname*{Sup}_G r(G, D_0) \leqq \operatorname*{Sup}_G r(G, D). \tag{1.6}$$

2. Regularity conditions and other assumptions. We shall make the following assumptions:

ASSUMPTION 1. *The chance variables are identically and independently distributed. The common distribution is either discrete or absolutely continuous.*

Let $p(a \mid F)$ denote the elementary probability law of X_i when F is the distribution of X_i; i.e., when F is discrete, $p(a \mid F)$ is the probability that $X_i = a$, and when F is absolutely continuous, $p(a \mid F)$ is the probability density of X_i at a.

In the space M of sequences x let B be the smallest Borel field which contains all sets of points x which are defined by the relations

$$x_i < a_i \qquad i = 1, 2, \cdots \text{ ad inf.},$$

where the a_i are real numbers or $+\infty$. Each admissible[3] F induces a probability measure $F^*(B)$ on M; the totality of these probability measures is Ω. Let H^*

[3] An F or F* is admissible if F* is in Ω.

be a given Borel field of subsets of Ω. The only subsets of Ω which we shall discuss in this paper will be members of H^*, and all probability measures on Ω which we shall discuss will be measurable (H^*). This will henceforth be assumed without further repetition.

Let A^* be any set in H^*, and A the set of F which corresponds to the F^* in A^*. The sets A form a Borel field, say H. By definition, the probability measure of a set A according to a probability measure $\xi(H^*)$ on Ω is to be the same as the probability measure of A^* according to ξ.

Let $M \times \Omega$ be the Cartesian product of M and Ω ([5], page 82), and K be the smallest Borel field of subsets of $M \times \Omega$ which contains the Cartesian product of any member of B by any member of H^*.

For a given decision function $d = D(x)$, $W(F, D(x))$ is a function of F and x. Hereafter in this paper we shall limit ourselves to functions $D(x)$ such that $W(F, D(x))$ is measurable (K), and $n(x, D)$ is measurable (B).

It is true that in Section 1, W was given as a function of G, the distribution of X. Because of Assumption 1, $G = F^*$, and there is a one-to-one correspondence between F and F^*. Thus we may, in appropriate places, interchange them freely.

ASSUMPTION 2. *For every real a, except possibly on a Borel set*[4] *whose probability is zero according to every admissible F, $p(a \mid F)$ exists and is a function of a and F which is measurable (K). If the admissible distributions F are discrete, there exists a fixed sequence $\{b_i\}$ ($i = 1, 2, \cdots$, ad inf.) of real values such that $\sum_{i=1}^{\infty} p(b_i \mid F) = 1$ for all admissible F.*

ASSUMPTION 3. *$W(F, d)$ is bounded. For every d in D^*, $W(F, d)$ is a function of F which is measurable (H).*

In what follows ξ will always denote a probability measure (H^*) on Ω. Thus

$$W(\xi, d) = \int_{\Omega} W(F, d)\, d\xi$$

exists.

ASSUMPTION 4. *The function $c(n) = cn$. Without loss of generality we may take $c = 1$, so that $c(n) = n$.*

We shall introduce the following convergence definition in the space D^*: the sequence $\{d_i\}$ converges to d_0 if

$$\lim_{i \to \infty} W(F, d_i) = W(F, d_0)$$

uniformly in the admissible F's.

ASSUMPTION 5. *The space D^* is compact in the sense of the above convergence definition.*

One can easily verify that, if $\lim_{i \to \infty} d_i = d_0$, then

$$\lim_{i \to \infty} W(\xi, d_i) = W(\xi, d_0);$$

---

[4] A Borel set is a member of the smallest Borel field which contains all the open sets of the real line.

i.e., $W(\xi, d)$ is a continuous function of d. Thus, because of Assumption 5, the minimum of $W(\xi, d)$ with respect to d exists.

We shall now show that, under the above conditions

$$\int_M W[F^*, D(x)]\, dF^*(x) \tag{2.1}$$

exists and is a function of F^* measurable (H^*). For any j let R_j be the set in B such that $n(x, D) = j$. Then it is enough to show that, for any j,

$$\int_{R_j} W[F^*, D(x)]\, dF^*(x) \tag{2.2}$$

exists and is a function of F^* measurable (H^*).

In the discrete case, the integral (2.2) is equal to the sum[5]

$$\sum_{(x_1, \dots, x_j) \epsilon R_j} W[F^*, D(x)] p(x_1 \mid F) \cdots p(x_j \mid F). \tag{2.3}$$

For fixed values of $x_1, \cdots, x_j$, the expression under the summation sign is obviously a function of F^* measurable (H^*). Since, because of Assumption 2, there are only countably many points $(x_1, \cdots, x_j)$ in R_j, the sum (2.3) must be a function of F^* measurable (H^*).

In the absolutely continuous case, the integral (2.2) is equal to (2.4)

$$\int_{R_j} W[F^*, D(x)] \prod_{i=1}^{j} p(x_i \mid F)\, d\nu(j) \tag{2.4}$$

where $\nu(j)$ is Borel measure in the j-dimensional Euclidean space. The integrand is measurable (K). Hence, the integral (2.4) exists and is a function of F^* measurable (H^*) (see [5], Chapter III, Theorems 9.3 and 9.8).

3. Some results concerning Bayes solutions. If ξ is the a priori probability measure on Ω, the a posteriori probability of a subset ω of Ω for given values $x_1, \cdots, x_m$ of the first m chance variables is given by

$$\xi(\omega \mid \xi, x_1, \cdots, x_m) = \frac{\int_\omega p(x_1 \mid F) \cdots p(x_m \mid F)\, d\xi}{\int_\Omega p(x_1 \mid F) \cdots p(x_m \mid F)\, d\xi}. \tag{3.1}$$

Let

$$\rho_0(\xi) = \operatorname*{Min}_d W(\xi, d). \tag{3.2}$$

For any positive integral value m, let $\rho_m(\xi)$ denote the infimum of $r(\xi, D)$ with respect to D where D is restricted to decision functions for which $n(x, D) \leqq m$ for all x. For any positive integer m, let $d = D^m(x)$ denote a decision function

[5] Because of the definition of R_j we may, in the expressions (2.3) and (2.4), proceed as if R_j were a Borel set in j-dimensional Euclidean space.

D for which $n(x, D) \leqq m$ for all x. Thus, we can write

$$\rho_m(\xi) = \operatorname*{Inf}_{D^m} r(\xi, D^m) \quad (m = 1, 2, \cdots, \text{ad inf.}). \tag{3.3}$$

Let

$$\rho(\xi) = \operatorname*{Inf}_{D} r(\xi, D). \tag{3.4}$$

We shall first prove several theorems concerning the functions $\rho_0(\xi)$, $\rho_m(\xi)$, and $\rho(\xi)$.

THEOREM 3.1. *The following recursion formula holds:*[6]

$$\rho_{m+1}(\xi) = \text{Min}\left[\rho_0(\xi), 1 + \int_{-\infty}^{\infty} \rho_m(\xi_a)\, p(a \mid \xi)\, da\right] \tag{3.5}$$

$$(m = 0, 1, 2, \cdots, \text{ad inf.})$$

where

$$\xi_a(\omega) = \xi(\omega \mid \xi, a) \text{ and } p(a \mid \xi) = \int_\Omega p(a \mid F)\, d\xi. \tag{3.6}$$

PROOF: Let $\rho_m^*(\xi)$ ($m = 1, 2, \cdots$, ad inf.) denote the infimum of $r(\xi, D)$ with respect to D where D is subject to the restriction that $n(x, D) \geqq 1$ and $\leqq m$ for all x. Clearly,

$$\rho_{m+1}(\xi) = \text{Min}[\rho_0(\xi), \rho_{m+1}^*(\xi)]. \tag{3.7}$$

Let $\rho_m^*(\xi \mid a)$ denote the infimum with respect to D of the conditional risk (conditional expected value of $W[F, D(x)] + n(x, D)$) when the first observation x_1 on X_1 is a and D is restricted to decision functions for which $n(x, D) \geqq 1$ and $\leqq m$ for all x. Let $\bar{D}(m)$ be the temporary generic designation of such a decision function. Let $\bar{D}(m \mid a)$ be the decision function which is obtained from $\bar{D}(m)$ when the first observation is a. Finally let $r(\xi, D \mid a)$ be the conditional risk when the a priori distribution function is ξ, D is the decision function and requires at least one observation, and the first observation is a. We then have that

$$r(\xi, \bar{D}(m + 1) \mid a) = r(\xi_a, \bar{D}(m + 1 \mid a)) + 1.$$

Hence

$$\rho_{m+1}^*(\xi \mid a) = \rho_m(\xi_a) + 1. \tag{3.8}$$

The unconditional quantity $\rho_{m+1}^*(\xi)$ must clearly be equal to the average value of the infimum of the conditional risk. Thus we have

$$\rho_{m+1}^*(\xi) = \int_{-\infty}^{\infty} \rho_{m+1}^*(\xi \mid a) p(a \mid \xi) da. \tag{3.9}$$

[6] If the distribution of X is discrete, the integration with respect to a is to be replaced by summation with respect to a. This remark refers also to subsequent formulas.

Equation (3.5) follows from (3.7), (3.8) and (3.9).

THEOREM 3.2. *The function $\rho(\xi)$ satisfies the following equation:*

$$\rho(\xi) = \text{Min}\left[\rho_0(\xi), \int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da + 1\right]. \tag{3.10}$$

The proof of this theorem is omitted, since it is essentially the same as that of Theorem 3.1.

THEOREM 3.3.[7] *The following inequalities hold:*

$$0 \leqq \rho_m(\xi) - \rho(\xi) \leqq \frac{W_0^2}{m} \qquad (m = 1, 2, \cdots, \text{ad inf.}) \tag{3.11}$$

where W_0 is the least upper bound of $W(F, d)$.

PROOF: Let $\{D_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of decision functions such that

$$\lim_{i=\infty} r(\xi, D_i) = \rho(\xi). \tag{3.12}$$

Let, furthermore, $P_i(\xi)$ denote the probability that at least m observations will be made when D_i is the decision function adopted and ξ is the a priori probability measure on Ω. Since $\rho(\xi) \leqq W_0$ and since

$$r(\xi, D_i) \geqq mP_i(\xi), \tag{3.13}$$

it follows from (3.12) that

$$\limsup_{i=\infty} P_i(\xi) \leqq \frac{W_0}{m}. \tag{3.14}$$

Let D_i^m be the decision function obtained from D_i as follows: $D_i^m(x) = D_i(x)$ for all x for which $n(x, D_i) \leqq m$. $D_i^m(x)$ is equal to a fixed element d_0 for all x for which $n(x, D_i) > m$.[8]

Clearly,

$$r(\xi, D_i^m) \leqq r(\xi, D_i) + P_i(\xi)W_0. \tag{3.15}$$

From (3.12), (3.14) and (3.15) it follows that

$$\limsup_{i=\infty} r(\xi, D_i^m) \leqq \rho(\xi) + \frac{W_0^2}{m}. \tag{3.16}$$

Since $\rho_m(\xi)$ cannot exceed the left hand member of (3.16), the second half of (3.11) follows from (3.16). The first half of (3.11) is obvious.

---

[7] This theorem is essentially the same as Lemma 2.1 in [6].

[8] We verify that $W(F, D_i^m)$ is measurable (K), as follows: Consider the set V of couples (F, x) such that $W(F, D_i^m(x)) < c$, where c is some real constant. We want to show that $V \epsilon K$. For this purpose let V_0 be the set of couples (F, x) such that $W(F, D_i(x)) < c$. Then $V_0 \epsilon K$. Let V_1 be the set of x's such that $n(x, D_i) \leqq m$. Then $V_1 \epsilon B$, $(\Omega \times V_1) = V_2 \epsilon K$, $V_0V_2 \epsilon K$. Let $V_3 = M - V_1$. For every $x \epsilon V_3$ we have $W(F, D_i^m(x)) = W(F, d_0)$. Let V_4 be the set of F's such that $W(F, d_0) < c$. Then $V_4 \epsilon H$ by Assumption 3. Finally we have $V = V_0V_2 + V_4 \times V_3$, so that $V \epsilon K$.

The immediate consequence of Theorem 3.3 is the relation[9]

(3.17) $$\lim_{m \to \infty} \rho_m(\xi) = \rho(\xi).$$

THEOREM 3.4. *If ξ_1 and ξ_2 are two probability measures on Ω such that*[10]

(3.18) $$\frac{\xi_1(\omega)}{\xi_2(\omega)} \leqq 1 + \epsilon \text{ for all } \omega,$$

then

(3.19) $$\rho(\xi_1) \leqq (1 + \epsilon)\rho(\xi_2).$$

PROOF: It follows from (3.18) that

(3.20) $$r(\xi_1, D) \leqq (1 + \epsilon)r(\xi_2, D) \text{ for all } D.$$

Hence, (3.19) must hold.

The above theorem permits the computation of a simple and in many cases useful lower bound of $\int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da$ as follows:

For any real value a, let ϵ_a be a non-negative value (not necessarily finite) determined such that

(3.21) $$\frac{\xi(\omega)}{\xi_a(\omega)} \leqq 1 + \epsilon_a \text{ for all } \omega.$$

Then

(3.22) $$\int_{-\infty}^{\infty} \rho(\xi_a)\, p(a \mid \xi)\, da \geqq \int_{-\infty}^{\infty} \frac{\rho(\xi)}{1 + \epsilon_a}\, p(a \mid \xi)\, da = \rho(\xi) \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da.$$

Since $\epsilon_a \geqq 0$ and since $\rho_0(\xi) \geqq \rho(\xi)$, we obviously have

(3.23) $$\rho(\xi) \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da \geqq \rho(\xi) - \left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right]\rho_0(\xi).$$

Hence, we obtain the inequality

(3.24) $$\int_{-\infty}^{\infty} \rho(\xi_a)\, p(a \mid \xi)\, da \geqq \rho(\xi) - \rho_0(\xi)\left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right].$$

An upper bound of the left hand member in (3.24) is obtained by replacing ρ by ρ_0; i.e.,

(3.25) $$\int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da \leqq \int_{-\infty}^{\infty} \rho_0(\xi_a)p(a \mid \xi)\, da.$$

---

[9] A proof of (3.17) is contained implicitly in the work of Arrow, Blackwell and Girshick ([2], Section 1.3).

[10] The left member of (3.18) is defined to be equal to 1 when $\xi_1(\omega) = \xi_2(\omega) = 0$.

The bounds given in (3.24) and (3.25) may be useful in constructing Bayes solutions, since the following theorem holds:

THEOREM 3.5. *If*

$$\rho_0(\xi) > \int_{-\infty}^{\infty} \rho_0(\xi_a)p(a \mid \xi)\, da + 1, \tag{3.26}$$

then $\rho(\xi) < \rho_0(\xi)$. *If*

$$\rho_0(\xi)\left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right] < 1, \tag{3.27}$$

then $\rho(\xi) = \rho_0(\xi)$.

The above theorem is an immediate consequence of (3.10), (3.24) and (3.25).

A decision procedure relative to a given a priori probability measure ξ_0 will be given with the help of the function $\rho(\xi)$ as follows: If $\rho(\xi_0) = \rho_0(\xi_0)$, take a final decision d for which $W(\xi_0, d)$ is minimized. If $\rho(\xi_0) < \rho_0(\xi_0)$, take an observation on X_1 and compute the a posteriori probability measure ξ_1. If $\rho(\xi_1) = \rho_0(\xi_1)$, stop experimentation with a final decision d for which $W(\xi_1, d)$ is minimized. If $\rho(\xi_1) < \rho_0(\xi_1)$, take an observation on X_2 and compute the a posteriori probability measure ξ_2 corresponding to the observed values of X_1 and X_2, and so on. The above decision procedure will be shown later to be a Bayes solution. Theorem 3.5 permits one to decide whether $\rho(\xi) < \rho_0(\xi)$ or $= \rho_0(\xi)$ whenever ξ satisfies (3.26) or (3.27). Theorem 3.5 will be useful when the class of all ξ's for which neither (3.26) nor (3.27) holds is small.

For the purposes of the next theorem let $\hat{D}$ designate the decision procedure described in the preceding paragraph. (We shall shortly show that $\hat{D}$ is a decision function in the sense of our definition.)

Let $\hat{D}^0$ be the decision procedure where the first observation is taken and then one proceeds according to $\hat{D}$.

We shall now prove that $\hat{D}$ and $\hat{D}^0$ are Bayes solutions. More precisely, we shall prove the following theorem:[11]

THEOREM 3.6. *For any* ξ, $\hat{D}$ *and* $\hat{D}^0$ *as defined above are decision functions. Let* D *be any decision function for which* $n(x, D) \geqq 1$ *and let*

$$\rho^*(\xi) = \operatorname*{Inf}_D r(\xi, D).$$

Then

$$r(\xi, \hat{D}) = \rho(\xi)$$

and

$$r(\xi, \hat{D}^0) = \rho^*(\xi).$$

[11] This theorem follows also from some earlier more general existence theorems ([6], Theorems 2.4 and 3.3). (See also [4], Lemma 1.) The validity of Theorem 3.6 was proved also by Arrow, Blackwell and Girshick [2].

In view of this theorem, the operation "infimum with respect to D" in the definitions of $\rho(\xi)$, and $\rho^*(\xi)$ can be replaced by "minimum with respect to D."

First we shall establish the measurability properties of $\hat{D}$ and $\hat{D}^0$. Since the proofs are similar, we restrict ourselves to consideration of $\hat{D}$. Let $\xi_{x_1,\dots,x_m}$ be the a posteriori distribution (3.1). From the (B) measurability of $\rho_0(\xi_{x_1,\dots,x_m})$ and $\rho(\xi_{x_1,\dots,x_m})$ it follows easily that $n(x, \hat{D})$ is measurable (B). It remains to prove that $W(F, \hat{D}(x))$ is measurable (K). For this purpose, let $L^i = (d_1^i, \cdots, d_{k_i}^i)$ be a sequence $\frac{1}{i}$ dense in D^*, i.e., for any $d \in D^*$ there exists a $g \in D^*$ such that $g \in L^i$ and $| W(F, d) - W(F, g) | < \frac{1}{i}$ uniformly in F. (The existence of such a sequence follows from Assumption 5.) Let now $D_i(x)$ be a decision function defined as follows:

$$n(x, D_i) = n(x, \hat{D}).$$

Suppose $n(x, \hat{D}) = m$ when the observations are $x_1, \cdots, x_m$. We define $D_i(x)$ to be such that $D_i(x)$ is an element of L^i and

$$W(\xi_{x_1,\dots,x_m}, D_i(x)) = \min_{d \in L^i} W(\xi_{x_1,\dots,x_m}, d), \tag{3.28}$$

i.e., $D_i(x)$ takes the minimizing value of d. For any fixed d, the set of x's satisfying the equation $D_i(x) = d$ is without difficulty shown to be (B) measurable. Since $D_i(x)$ assumes only a finite number of values in D^*, it follows from Assumption 3 that $W(F, D_i(x))$ is measurable (K). Now

$$\lim_{i \to \infty} W(F, D_i(x)) = W(F, \hat{D}(x)),$$

so that $W(F, \hat{D}(x))$ is measurable (K).

We shall now prove that $\hat{D}$ is a Bayes solution, i.e., that

$$\rho(\xi) = r(\xi, \hat{D}). \tag{3.29}$$

In a similar way it can be proved that

$$\rho^*(\xi) = r(\xi, \hat{D}^0). \tag{3.30}$$

If $\rho_0(\xi) = \rho(\xi)$, there can be no better decision function (from the point of view of reducing the risk) than $\hat{D}$, i.e., $\hat{D}$ is a Bayes solution. Suppose then that

$$\rho_0(\xi) > \rho(\xi). \tag{3.31}$$

If (3.31) holds and $\hat{D}$ is not a Bayes solution, there exists a decision function $\bar{D}_1$ such that

$$r(\xi, \bar{D}_1) < r(\xi, \hat{D}) \tag{3.32}$$

and

$$r(\xi, \bar{D}_1) < \frac{\rho_0(\xi) + \rho(\xi)}{2}. \tag{3.33}$$

Now $\bar{D}_1$ must require that at least one observation be taken, else (3.33) could not hold. Thus $\hat{D}$ and $\bar{D}_1$ both require at least one observation.

Suppose one observation is taken. Let $r(\xi, D \mid a)$ denote the conditional risk of proceeding according to D when ξ is the a priori distribution and a is the first observation. For a given D we have that $r(\xi, D \mid a)$ is a function only of ξ_a. In particular $r(\xi, \hat{D} \mid a)$ and $r(\xi, \bar{D}_1 \mid a)$ are functions only of ξ_a.

We can now apply to $r(\xi, \hat{D} \mid a)$ and $r(\xi, \bar{D}_1 \mid a)$ the same argument that was applied above to $r(\xi, \hat{D})$ and $r(\xi, \bar{D}_1)$, and conclude again as follows: whenever $\rho_0(\xi_a) = \rho(\xi_a)$ (when one takes no more observations according to $\hat{D}$), taking additional observations cannot diminish the conditional risk below $r(\xi, \hat{D} \mid a)$ ($\bar{D}_1$ may require an additional observation without having

$$r(\xi, \bar{D}_1 \mid a) > r(\xi, \hat{D} \mid a).$$

This can happen when $\rho_0(\xi_a) = \rho^*(\xi_a)$). Whenever $\rho_0(\xi_a) > \rho(\xi_a)$ (when $\hat{D}$ requires us to take another observation) two cases may occur: either a) $\bar{D}_1$ requires us to take another observation, in which case its decision is the same as that of $\hat{D}$, or b) $\bar{D}_1$ requires us to stop taking observations. There exists then another decision function whose conditional risk is less than

$$\frac{\rho_0(\xi_a) + \rho(\xi_a)}{2} + 1.$$

Both this decision function and $\hat{D}$ require that another observation be taken. We conclude that up to and including the first observation, $\hat{D}$ coincides either with $\bar{D}_1$ or with another decision function $\bar{D}_2$ whose risk is not greater than that of $\bar{D}_1$.

We continue in this manner for 2, 3, $\cdots$ observations. The above argument is always valid because of Assumption 4 and because the past history of the process (the sequence of observations) enters only through the a posteriori probability. Thus we conclude that for any positive integer k there exists a decision function $\bar{D}_k$ such that up to and including the k-th observation $\hat{D}$ gives the same decision as $\bar{D}_k$ and the risk corresponding to $\bar{D}_k$ does not exceed the risk corresponding to $\bar{D}_1$. Since $\lim_{k\to\infty} r(\xi, \bar{D}_k) \geqq r(\xi, \hat{D})$, (3.32) cannot hold. Hence (3.29) holds and $\hat{D}$ is a Bayes solution.

For any probability measure ξ on Ω one of the following three conditions must hold:

(1) $\text{Min}_d\, W(\xi, d) < r(\xi, D)$ for any D for which $n(x, D) \geqq 1$.

(2) $\text{Min}_d\, W(\xi, d) \leqq r(\xi, D)$ for all D for which $n(x, D) \geqq 1$, and the equality sign holds for at least one D with $n(x, D) \geqq 1$.

(3) There exists a D with $n(x, D) \geqq 1$ such that $\text{Min}_d\, W(\xi, d) > r(\xi, D)$.

In view of Theorem 3.6, the conditions (1), (2) and (3) can be expressed by: (1) $\rho_0(\xi) < \rho^*(\xi)$, (2) $\rho_0(\xi) = \rho^*(\xi)$ and (3) $\rho_0(\xi) > \rho^*(\xi)$, respectively.

We shall say that a probability measure ξ on Ω is of the first type if it satisfies (1), of the second type if it satisfies (2), and of the third type if it satisfies (3). Since the a posteriori probability defined in (3.1) is also a probability measure on Ω, any a posteriori probability measure will be one of the three types mentioned above.

We shall now prove the following characterization theorem:

THEOREM 3.7.[12] *A necessary and sufficient condition for a decision function* $d = D_0(x)$ *to be a Bayes solution relative to a given a priori distribution* ξ_0 *is that the following three relations be fulfilled for any sample point* x, *except perhaps on a set whose probability measure is zero when* ξ_0 *is the a priori distribution in* Ω:

(a) *For any* $m < n(x, D_0)$, *the a posteriori distribution* $\xi(\omega \mid \xi_0, x_1, \cdots, x_m)$ is *either of the second or of the third type,*

(b) *For* $m = n(x, D_0)$, *the a posteriori distribution* $\xi(\omega \mid \xi_0, x_1, \cdots, x_m)$ *is either of the first or the second type,*

(c) *For* $m = n(x, D_0)$, *we have*

$$\operatorname*{Min}_{d} W(\xi_{x_1,\dots,x_m}, d) = W(\xi_{x_1,\dots,x_m}, D_0(x)),$$

where $\xi_{x_1,\dots,x_m}$ *stands for an a priori distribution that is equal to the a posteriori distribution corresponding to* $\xi_0, x_1, \cdots, x_m$.

PROOF: We shall omit the proof of the sufficiency of the conditions (a), (b) and (c), since it is essentially the same as that of Theorem 3.6. To prove the necessity of these conditions, let $d = D_0(x)$ be a decision function and let M^* denote the set of all sample points x for which at least one of the relations (a), (b) and (c) is violated. First, we shall show tht M^* is a set measurable (B). Let M_1^* be the set of all x's for which (a) is violated, M_2^* the set of all x's for which (b) is violated, and M_3^* the set of all x's for which (c) is violated. Clearly, M^* is shown to be measurable (B) if we can show that $M_i^* (i = 1, 2, 3)$ is measurable (B). Let $M_{ir}^* (r = 1, 2, \cdots$, ad inf) denote the subset of M_i^* for which the first violation of the corresponding condition occurs for the sample $x_1, \cdots, x_r$. We merely have to show that M_{ir}^* is measurable (B) for all i and r. The measurability of M_{3r}^* follows from the fact that $\operatorname{Min}_d W(\xi_{x_1,\dots,x_r}, d)$ and

$$W[\xi_{x_1,\dots,x_r}, D_0(x)]$$

are functions of x measurable (B). To show the measurability of M_{1r}^* and M_{2r}^*, it is sufficient to show that the set of all samples $x_1, \cdots, x_r$ for which $\xi_{x_1,\dots,x_r}$ is of type $i (i = 1, 2, 3)$ is measurable (B). But this follows from the fact that $\rho_0(\xi_{x_1,\dots,x_r})$ and $\rho^*(\xi_{x_1,\dots,x_r})$ are functions of $(x_1, \cdots, x_r)$ measurable (B). Hence, M^* is proved to be measurable (B).

For any x in M^* let $m(x)$ be the smallest positive integer such that at least one of the relations (a), (b) and (c) is violated for the finite sample

$$x_1, x_2, \cdots, x_{m(x)}.$$

Clearly, if x is a point in M^*, then also any sample point y is in M^* for which $y_1 = x_1, \cdots, y_{m(x)} = x_{m(x)}$. Let x^0 be any particular sample point in M^* and let $\bar{r}(\xi_0, D_0, x_1^0, \cdots, x_{m(x^0)}^0)$ denote the conditional risk when ξ_0 is the a priori

[12] See also the proof of Lemma 1 in [4].

distribution in Ω, D_0 is the decision function adopted and the first $m(x^0)$ observations are equal to $x_1^0, \cdots, x_{m(x^0)}^0$, respectively; i.e., $r(\xi_0, D_0, x_1^0, \cdots, x_{m(x^0)}^0)$ is the conditional expected value of $W(F, D_0(x)) + n(x, D_0)$, when ξ_0 is the a priori distribution in Ω, D_0 is the decision function adopted and $x_1^0, \cdots, x_{m(x^0)}^0$ are the first $m(x^0)$ observations.

Let $D_1(x)$ be the decision function determined as follows: for any x not in M^* we put $D_1(x) = D_0(x)$. For any x in M^*, let $n(x_1, D_1)$ be equal to the smallest integer $n(x) \geqq m(x)$ for which

$$\rho_0(\xi_{x_1, \ldots, x_{n(x)}}) = \rho(\xi_{x_1, \ldots, x_{n(x)}})$$

and the value of $D_1(x)$ is determined so that condition (c) of our theorem is fulfilled. Since, for any positive integer m, the subset of M^* where $m(x) = m$ is (B) measurable, $D_1(x)$ has the proper measurability properties. Applying Theorem 3.6, we see that

$$r(\xi_0, D_1, x_1, \cdots, x_{m(x)}) = \rho(\xi_{x_1, \ldots, x_{m(x)}}) \tag{3.34}$$

for any x in M^*. On the other hand, since D_0 violates at least one of the conditions (a), (b), and (c) at every point x in M^*, we have

$$r(\xi_0, D_0, x_1, \cdots, x_{m(x)}) > \rho(\xi_{x_1, \ldots, x_{m(x)}}) \tag{3.35}$$

for every x in M^*. If the probability measure of M^* is positive when ξ_0 is the a priori probability measure, the above two relations imply that

$$r(\xi_0, D_0) > r(\xi_0, D_1).$$

Thus, D_0 is not a Bayes solution and the proof of Theorem 3.7 is complete.

We shall now prove the following continuity theorem.[13]

THEOREM 3.8. *Let* $\{\xi_i\}$ $(i = 0, 1, 2, \cdots, \textit{ad inf.})$ *be a sequence of probability measures on* Ω *such that*

$$\lim_{i=\infty} \frac{\xi_i(\omega)}{\xi_0(\omega)} = 1 \textit{ uniformly in } \omega. \tag{3.36}$$

Then

$$\lim_{i=\infty} \rho(\xi_i) = \rho(\xi_0). \tag{3.37}$$

PROOF: It follows from (3.36) that for any $\epsilon > 0$, we have for almost all values i

$$\frac{\xi_i(\omega)}{\xi_0(\omega)} < 1 + \epsilon \text{ and } \frac{\xi_0(\omega)}{\xi_i(\omega)} < 1 + \epsilon \text{ for all } \omega. \tag{3.38}$$

Our theorem is an immediate consequence of (3.38) and Theorem 3.4.

---

[13] A proof of this theorem for finite Ω was given by G. W. Brown and is included in [2]. See also Lemma 3 in [4].

A stronger continuity theorem is the following:

THEOREM 3.8.1. *Let* $\{\xi_i\}$, $(i = 0, 1, 2, \cdots$, *ad inf.) be a sequence of probabitity measures on* Ω *such that*

$$\lim_{i \to \infty} \xi_i(\omega) = \xi_0(\omega)$$

uniformly in ω. *Then* (3.37) *holds.*

PROOF: It follows from (3.11) that

$$\lim_{m \to \infty} \rho_m(\xi) = \rho(\xi)$$

uniformly in ξ. Hence it is sufficient to prove that, under the conditions of the theorem,

$$\lim_{i \to \infty} \rho_m(\xi_i) = \rho_m(\xi_0)$$

for any m. Let $D^m(x)$ denote a decision function for which $n(x, D^m) \leqq m$ for all x. It follows that, for a fixed m, $r(F, D^m)$ is bounded, uniformly in F and D^m (Assumptions 3 and 4). From the hypothesis on $\{\xi_i\}$ it then follows that

$$\lim_{i \to \infty} r(\xi_i, D^m) = r(\xi_0, D^m)$$

uniformly in D^m. From this the desired result follows readily.

A class C of probability measures ξ on Ω will be said to be convex if for any two elements ξ_1 and ξ_2 of C and for any positive value $\lambda < 1$, the probability measure $\xi = \lambda\xi_1 + (1 - \lambda)\xi_2$ is an element of C.

For any element d_0 of D, let C_{i,d_0} denote the class of all probability measures ξ of type i $(i = 1, 2, 3)$ for which $W(\xi, d_0) = \operatorname{Min}_d W(\xi, d)$. Let C_d denote the set-theoretical sum of $C_{1,d}$ and $C_{2,d}$. We shall now prove the following theorem.

THEOREM 3.9. *For any element* d, *the classes* $C_{1,d}$ *and* C_d *are convex.*[14]

Let ξ_1 and ξ_2 be two elements of $C_{1,d}$. Then for any decision function $D(x)$ which requires at least one observation we have

$$W(\xi_1, d) < r(\xi_1, D) \text{ and } W(\xi_2, d) < r(\xi_2, d). \tag{3.39}$$

Let $\xi = \lambda\xi_1 + (1 - \lambda)\xi_2$ where λ is a positive number <1. Clearly,

$$W(\xi, d) = \lambda W(\xi_1, d) + (1 - \lambda) W(\xi_2, d) \tag{3.40}$$

and

$$r(\xi, D) = \lambda r(\xi_1, D) + (1 - \lambda) r(\xi_2, D). \tag{3.41}$$

From (3.39), (3.40) and (3.41) we obtain

$$W(\xi, d) < r(\xi, D) \quad \text{and} \quad W(\xi, d) = \operatorname{Min}_{d^*} W(\xi, d^*). \tag{3.42}$$

Hence ξ is an element of $C_{1,d}$ and the convexity of $C_{1,d}$ is proved. The convexity of C_d can be proved in the same way by replacing $<$ by $\leqq$ in (3.39) and (3.42).

[14] See also Lemma 2 in [4].

We shall say that a set L of probability measures ξ is a linear manifold if for any two elements ξ_1 and ξ_2 of L, $\xi = \alpha\xi_1 + (1 - \alpha)\xi_2$ is also an element of L for any real value α for which $\alpha\xi_1 + (1 - \alpha)\xi_2$ is a probability measure. A linear manifold L will be said to be tangent to C_d if the intersection of L and $C_{2,d}$ is not empty, but the intersection of L and $C_{1,d}$ is empty.

For any decision function $D(x)$ and for any element d of D^*, let $L(D, d)$ denote the linear manifold consisting of all ξ which satisfy the equation

$$W(\xi, d) = r(\xi, D). \tag{3.43}$$

THEOREM 3.10. *Let ξ_0 be an element of $C_{2,d}$ and let $D_0(x)$ be a decision function that requires at least one observation and is such that $W(\xi_0, d) = r(\xi_0, D_0)$. Then the linear manifold $L(D_0, d)$ is tangent to C_d.*

PROOF: ξ_0 is obviously an element of $L(D_0, d)$. Thus the intersection of $L(D_0, d)$ and $C_{2,d}$ is not empty. For any element ξ_1 of $C_{1,d}$ we have $W(\xi_1, d) < r(\xi_1, D)$ for any D that requires at least one observation. Hence, $W(\xi_1, d) < r(\xi_1, D_0)$ and, therefore, ξ_1 cannot be an element of $L(D_0, d)$. This proves our theorem.

4. Applications to the case where Ω and D^* are finite. In this section we shall apply the general results of the preceding section to the following special case: the space Ω consists of a finite number of elements, $F_1, \cdots, F_k$ (say), and the space D^* consists of the elements $d_1, \cdots, d_k$ where d_i denotes the decision to accept the hypothesis H_i that F_i is the true distribution. Let

$$W(F_i, d_j) = W_{ij} = 0 \text{ for } i = j \text{ and } >0 \text{ for } i \neq j. \tag{4.1}$$

It will be sufficient to discuss the cases $k = 2$ and $k = 3$, since the extension to $k > 3$ will be obvious. We shall first consider the case $k = 2$. In this case any a priori distribution ξ is represented by two numbers g_1 and g_2 where g_i is the a priori probability that F_i is true ($i = 1, 2$). Thus, $g_i \geqq 0$ and $g_1 + g_2 = 1$. Let ξ_i denote the a priori distribution corresponding to $g_i = 1$ ($i = 1, 2$). Clearly C_{d_1} contains ξ_1 but not ξ_2, and C_{d_2} contains ξ_2 but not ξ_1. Because of Theorems 3.9 and 3.7, C_{d_1} and C_{d_2} are closed and convex. Furthermore, we obviously have

$$g_2W_{21} \leqq g_1W_{12} \text{ for all } \xi \text{ in } C_{d_1} \tag{4.2}$$

and

$$g_2W_{21} \geqq g_1W_{12} \text{ for all } \xi \text{ in } C_{d_2}. \tag{4.3}$$

Let $\xi_0 = (g_1^0, g_2^0)$ be the a priori distribution for which

$$g_2^0W_{21} = g_1^0W_{12}. \tag{4.4}$$

It follows from (4.2) and (4.3) that there exist two positive numbers c' and c'' such that

$$0 < c' \leqq g_2^0 \leqq c'' < 1 \tag{4.5}$$

and such that the class C_{d_1} consists of all ξ for which $g_2 \leqq c'$, and the class C_{d_2} consists of all ξ for which $g_2 \geqq c''$.

Thus, the following decision procedure will be a Bayes solution relative to the a priori distribution $\xi = (g_1, g_2)$: *If $g_2 \leqq c'$ or $\geqq c''$, do not take any observations and make the corresponding final decision. If $c' < g_2 < c''$, continue taking observations until the a posteriori probability of H_2 is either $\geqq c''$ or $\leqq c'$. If this a posteriori probability is $\geqq c''$, accept H_2, and if it is $\leqq c'$, accept H_1.*

The a posteriori probability of H_2 after the first m observations have been made is given by

$$(4.6) \qquad g_{2m} = \frac{g_2 p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}{g_1 p(x_1 \mid F_1) \cdots p(x_m \mid F_1) + g_2 p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}.$$

If $c' < g_2 < c''$ and if the probability (under F_1 as well as under F_2) is zero that $g_{2m} = c'$ or $= c''$ for some m, then it follows from Theorem 3.8 that the above described Bayes solution is essentially unique; i.e., any other Bayes solution can differ from the one given above only on a set whose probability measure is zero under both F_1 and F_2.

Provided that at least one observation is made, one can easily verify that the above described Bayes solution is identical with a sequential probability ratio test for testing H_2 against H_1. The sequential probability ratio test is defined as follows (see [3]): Two positive constants A and B ($B < A$) are chosen. Experimentation is continued as long as the probability ratio

$$(4.7) \qquad \frac{p_{2m}}{p_{1m}} = \frac{p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}{p(x_1 \mid F_1) \cdots p(x_m \mid F_1)}$$

satisfies the inequality $B < \frac{p_{2m}}{p_{1m}} < A$. If $\frac{p_{2m}}{p_{1m}} \geqq A$, accept H_2. If $\frac{p_{2m}}{p_{1m}} \leqq B$, accept H_1. The Bayes solution described above coincides with this probability ratio test for properly chosen values of the constants A and B.

The results described above for $k = 2$ are essentially the same as those contained in Lemmas 1 and 2 of an earlier publication [4] of the authors.

We shall now discuss the case $k = 3$. Any a priori distribution ξ can be represented by a point with the barycentric coordinates g_1, g_2 and g_3, where g_i is the a priori probability of $H_i (i = 1, 2, 3)$. The totality of all possible a priori distributions ξ will fill out the triangle T with the vertices 0_1, 0_2 and 0_3 where 0_i represents the a priori distribution corresponding to $g_i = 1$ (see Figure 1).

Clearly, the vertex 0_i is contained in C_{d_i}. Thus, because of Theorem 3.9, $C_{d_i} (i = 1, 2, 3)$ is a convex subset of T containing the vertex 0_i, as indicated in Figure 1.

If one of the components of $\xi = (g_1, g_2, g_3)$ is zero, say $g_i = 0$, then H_i can be disregarded and the problem of constructing Bayes solutions reduces to the previously considered case where $k = 2$. Thus, in particular, the determination of the boundary points $P_1, P_2, \cdots, P_6$ of C_{d_1}, C_{d_2} and C_{d_3} which are on the boundary of the triangle T, reduces to the previously considered case, $k = 2$.

It follows from Theorems 3.8 and 3.9 that the intersection of C_{d_i} with any straight line T_i through 0_i is a closed segment. One endpoint of this segment is, of course, 0_i. Let B_i denote the other endpoint. It follows from Theorem 3.7 that B_i must be a point of C_{2,d_i}. Any interior point of 0_iB_i can be shown to be an element of C_{1,d_i}. The proof of this is very similar to that of Theorem 3.9.

We shall now show how tangents to the sets C_{d_1}, C_{d_2} and C_{d_3} can be constructed at the boundary points $P_1, P_2, \cdots, P_6$. Consider, for example, the boundary point P_1 of C_{d_1} that lies on the line $0_1 0_2$. Let ξ_1 be the a priori distribution represented by the point P_1. Since the a priori probability of H_3 is zero according to ξ_1, we can disregard H_3 in constructing Bayes solutions relative to ξ_1. Let $D_1(x)$ be a sequential probability ratio test for testing H_1 against H_2

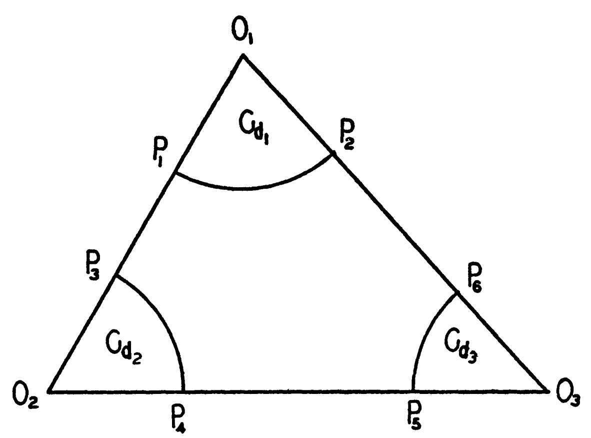

FIG. 1

which requires at least one observation and which is a Bayes solution relative to ξ_1. Since ξ_1 is a boundary point, such a decision function D_1 exists. Thus, we have

$$W(\xi_1, d_1) = r(\xi_1, D_1) = \operatorname*{Inf}_D r(\xi_1, D). \tag{4.8}$$

Let α_{ij} denote the probability of accepting H_j when H_i is true and D_1 is the decision function adopted. Let, furthermore, n_i denote the expected number of observations required by the decision procedure when F_i is true and D_1 is adopted. Then, for any a priori distribution $\xi = (g_1, g_2, g_3)$ we have

$$r(\xi, D_1) = \sum_{i,j} g_i W_{ij} \alpha_{ij} + \sum_i g_i n_i \tag{4.9}$$

and

$$W(\xi, d_1) = \sum_i g_i W_{i1}. \tag{4.10}$$

Thus, the linear manifold $L(D_1, d_1)$ is simply the straight line given by the equation

$$\sum_i g_i W_{i1} = \sum_{i,j} g_i W_{ij}\alpha_{ij} + \sum_i g_i n_i . \tag{4.11}$$

This straight line goes through P_1 and, because of Theorem 3.10, it is tangent to C_{d_1}. Tangents at the same points $P_2, \cdots, P_6$ can be constructed in a similar way.

The convexity properties of the sets $C_{d_i}(i = 1, 2, \cdots, k)$ were established by the authors prior to the more general results described in Sections 2 and 3 and were stated by one of the authors in an address given at the Berkeley meeting of the Institute of Mathematical Statistics, June, 1948. More general results when Ω and D^* are finite, admitting also non-linear cost functions, were obtained later by Arrow, Blackwell and Girshick [2].

REFERENCES

[1] A. WALD, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), pp. 279–313.

[2] K. J. ARROW, D. BLACKWELL, M. A. GIRSHICK, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213–244.

[3] A. WALD, "*Sequential analysis*," John Wiley & Sons, New York, 1947.

[4] A. WALD AND J. WOLFOWITZ, "Optimum character of the sequential probability ratio test," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 326–339.

[5] S. SAKS, "*Theory of the integral*," Hafner Publishing Company, New York.

[6] A. WALD, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 165–205.