STATISTICAL DECISION FUNCTIONS

By Abraham Wald[1]

Columbia University

Introduction and summary. The foundations of a general theory of statistical decision functions, including the classical non-sequential case as well as the sequential case, was discussed by the author in a previous publication [3]. Several assumptions made in [3] appear, however, to be unnecessarily restrictive (see conditions 1–7, pp. 297 in [3]). These assumptions, moreover, are not always fulfilled for statistical problems in their conventional form. In this paper the main results of [3], as well as several new results, are obtained from a considerably weaker set of conditions which are fulfilled for most of the statistical problems treated in the literature. It seemed necessary to abandon most of the methods of proofs used in [3] (particularly those in section 4 of [3]) and to develop the theory from the beginning. To make the present paper self-contained, the basic definitions already given in [3] are briefly restated in section 2.1.

In [3] it is postulated (see Condition 3, p. 207) that the space Ω of all admissible distribution functions F is compact. In problems where the distribution function F is known except for the values of a finite number of parameters, i.e., where Ω is a parametric class of distribution functions, the compactness condition will usually not be fulfilled if no restrictions are imposed on the possible values of the parameters. For example, if Ω is the class of all univariate normal distributions with unit variance, Ω is not compact. It is true that by restricting the parameter space to a bounded and closed subset of the unrestricted space, compactness of Ω will usually be attained. Since such a restriction of the parameter space can frequently be made in applied problems, the condition of compactness may not be too restrictive from the point of view of practical applications. Nevertheless, it seems highly desirable from the theoretical point of view to eliminate or to weaken the condition of compactness of Ω. This is done in the present paper. The compactness condition is completely omitted in the discrete case (Theorems 2.1–2.5), and replaced by the condition of separability of Ω in the continuous case (Theorems 3.1–3.4). The latter condition is fulfilled in most of the conventional statistical problems.

Another restriction postulated in [3] (Condition 4, p. 297) is the continuity of the weight function $W(F, d)$ in F. As explained in section 2.1 of the present paper, the value of $W(F, d)$ is interpreted as the loss suffered when F happens to be the true distribution of the chance variables under consideration and the decision d is made by the statistician. While the assumption of continuity of $W(F, d)$ in F may seem reasonable from the point of view of practical application, it is rather undesirable from the theoretical point of view for the following

[1] Work done under the sponsorship of the Office of Naval Research.

reasons. It is of considerable theoretical interest to consider simplified weight functions $W(F, d)$ which can take only the values 0 and 1 (the value 0 corresponds to a correct decision, and the value 1 to a wrong decision). Frequently, such weight functions are necessarily discontinuous. Consider, for example, the problem of testing the hypothesis H that the mean θ of a normally distributed chance variable X with unit variance is equal to zero. Let d_1 denote the decision to accept H, and d_2 the decision to reject H. Assigning the value zero to the weight W whenever a correct decision is made, and the value 1 whenever a wrong decision is made, we have:

$$W(\theta, d_1) = 0 \text{ for } \theta = 0, \text{ and } = 1 \text{ for } \theta \neq 0; W(\theta, d_2) = 0 \text{ for } \theta \neq 0, \text{ and } = 1 \text{ for } \theta = 0.$$

This weight function is obviously discontinuous. In the present paper the main results (Theorems 2.1–2.5 and Theorems 3.1–3.4) are obtained without making any continuity assumption regarding $W(F, d)$.

The restrictions imposed in the present paper on the cost function of experimentation are considerably weaker than those formulated in [3]. Condition 5 [3, p. 297] concerning the class Ω of admissible distribution functions, and condition 7 [3, p. 298] concerning the class of decision functions at the disposal of the statistician are omitted here altogether.

One of the new results obtained here is the establishment of the existence of so called minimax solutions under rather weak conditions (Theorems 2.3 and 3.2). This result is a simple consequence of two lemmas (Lemmas 2.4 and 3.3) which seem to be of interest in themselves.

The present paper consists of three sections. In the first section several theorems are given concerning zero sum two person games which go somewhat beyond previously published results. The results in section 1 are then applied to statistical decision functions in sections 2 and 3. Section 2 treats the case of discrete chance variables, while section 3 deals with the continuous case. The two cases have been treated separately, since the author was not able to find any simple and convenient way of combining them into a single more general theory.

1. Conditions for strict determinateness of a zero sum two person game. The normalized form of a zero sum two person game may be defined as follows (see [1, section 14.1]): there are two players and there is a bounded and real valued function $K(a, b)$ of two variables a and b given where a may be any point of a space A and b may be any point of a space B. Player 1 chooses a point a in A and player 2 chooses a point b in B, each choice being made in complete ignorance of the other. Player 1 then gets the amount $K(a, b)$ and player 2 the amount $-K(a, b)$. Clearly, player 1 wishes to maximize $K(a, b)$ and player 2 wishes to minimize $K(a, b)$.

Any element a of A will be called a pure strategy of player 1, and any element

b of B a pure strategy of player 2. A mixed strategy of player 1 is defined as follows: instead of choosing a particular element a of A, player 1 chooses a probability measure ξ defined over an additive class $\mathfrak{A}$ of subsets of A and the point a is then selected by a chance mechanism constructed so that for any element α of $\mathfrak{A}$ the probability that the selected element a will be contained in α is equal to $\xi(\alpha)$. Similarly, a mixed strategy of player 2 is given by a probability measure η defined over an additive class $\mathfrak{B}$ of subsets of B and the element b is selected by a chance mechanism so that for any element β of $\mathfrak{B}$ the probability that the selected element b will be contained in β is equal to $\eta(\beta)$. The expected value of the outcome $K(a, b)$ is then given by

$$K^*(\xi, \eta) = \int_B \int_A K(a, b)\, d\xi\, d\eta. \tag{1.1}$$

We can now reinterpret the value of $K(a, b)$ as the value of $K^*(\xi_a, \eta_b)$ where ξ_a and η_b are probability measures which assign probability 1 to a and b, respectively. In what follows, we shall write $K(\xi, \eta)$ for $K^*(\xi, \eta)$, $K(a, b)$ will be used synonymously with $K(\xi_a, \xi_b)$, $K(a, \eta)$ synonymously with $K(\xi_a, \eta)$ and $K(\xi, b)$ synonymously with $K(\xi, \eta_b)$. This can be done without any danger of confusion.

A game is said to be strictly determined if

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta). \tag{1.2}$$

The basic theorem proved by von Neumann [1] states that if A and B are finite the game is always strictly determined, i.e., (1.2) holds. In some previous publications (see [2] and [3]) the author has shown that (1.2) always holds if one of the spaces A and B is finite or compact in the sense of some intrinsic metric, but does not necessarily hold otherwise. A necessary and sufficient condition for the validity of (1.2) was given in [2] for spaces A and B with countably many elements. In this section we shall give sufficient conditions as well as necessary and sufficient conditions for the validity of (1.2) for arbitrary spaces A and B. These results will then be used in later sections.

In what follows, for any subset α of A the symbol ξ_α will denote a probability measure ξ in A for which $\xi(\alpha) = 1$. Similarly, for any subset β of B, ξ_β will stand for a probability measure η in B for which $\eta(\beta) = 1$. We shall now prove the following lemma.

LEMMA 1.1. *Let* $\{\alpha_i\}$ $(i = 1, 2, \cdots,$ *ad inf.*) *be a sequence of subsets of* A *such that* $\alpha_i \subset \alpha_{i+1}$ *and let* $\alpha = \sum_{i=1}^{\infty} \alpha_i$. *Then*

$$\lim_{i\to\infty} \operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta). \tag{1.3}$$

PROOF: Clearly, the limit of $\operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta)$ exists as $i \to \infty$ and cannot exceed the value of the right hand member in (1.3). Put

$$\lim_{i\to\infty} \operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta) = \rho \tag{1.4}$$

and

$$\text{(1.5)} \qquad \operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) = \rho + \delta \qquad (\delta \geq 0).$$

Suppose that $\delta > 0$. Then there exists a probability measure ξ_α^0 such that

$$\text{(1.6)} \qquad K(\xi_\alpha^0, \eta) \geqq \rho + \frac{\delta}{2} \qquad \text{for all } \eta.$$

Let $\xi_{\alpha_i}^0$ be the probability measure defined as follows: for any subset α^* of α_i we have

$$\text{(1.7)} \qquad \xi_{\alpha_i}^0(\alpha^*) = \frac{\xi_\alpha^0(\alpha^*)}{\xi_\alpha^0(\alpha_i)}.$$

Then, since $\lim_{i=\infty} \xi_\alpha^0 (\alpha - \alpha_i) = 0$, we have

$$\text{(1.8)} \qquad \lim_{i=\infty} K(\xi_{\alpha_i}^0, \eta) = K(\xi_\alpha^0, \eta)$$

uniformly in η. Hence, for sufficiently large i, we have

$$\text{(1.9)} \qquad \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}^0, \eta) \geq \rho + \frac{\delta}{3},$$

which is a contradiction to (1.4). Thus, $\delta = 0$ and Lemma 1.1 is proved. Interchanging the role of the two players, we obtain the following lemma.

LEMMA 1.2. *Let $\{\beta_i\}$ be a sequence of subsets of B such that $\beta_i \subset \beta_{i+1}$ and let $\sum_{i=1}^{\infty} \beta_i = \beta$. Then*

$$\text{(1.10)} \qquad \lim_{i=\infty} \operatorname*{Inf}_{\eta_{\beta_i}} \operatorname*{Sup}_{\xi} K(\xi, \eta_{\beta_i}) = \operatorname*{Inf}_{\eta_\beta} \operatorname*{Sup}_{\xi} K(\xi, \eta_\beta).$$

We shall now prove the following lemma.

LEMMA 1.3. *The inequality*[2]

$$\text{(1.11)} \qquad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) \leq \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta)$$

always holds.

PROOF: for any given $\epsilon > 0$, it is possible to find probability measures ξ^0 and η^0 such that

$$\text{(1.12)} \qquad \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta) \geqq \operatorname*{Sup}_{\xi} K(\xi, \eta^0) - \epsilon$$

and

$$\text{(1.13)} \qquad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) \leqq \operatorname*{Inf}_{\eta} K(\xi^0, \eta) + \epsilon.$$

[2] This inequality was given by v. Neumann [1] for finite spaces A and B.

Then we have,

$$(1.14)\quad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) \leq \operatorname*{Inf}_{\eta} K(\xi^0, \eta) + \epsilon \leq K(\xi^0, \eta^0) + \epsilon$$
$$\leq \operatorname*{Sup}_{\xi} K(\xi, \eta^0) + \epsilon \leqq \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta \quad) + 2\,\epsilon.$$

Since ϵ can be chosen arbitrarily small, Lemma 1.3 is proved.

THEOREM 1.1. *If α is a subset of A such that*

$$\operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_\alpha} K(\xi_\alpha, \eta)$$

and

$$\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_\alpha} K(\xi_\alpha, \eta) = \operatorname*{Inf}_{\eta} \operatorname{Sup} K(\xi, \eta),$$

then

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta).$$

PROOF: Clearly,

$$(1.15)\quad \operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) \leq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta)$$

and

$$(1.16)\quad \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_\alpha} K(\xi_\alpha, \eta) \leq \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta).$$

If the left hand members of (1.15) and (1.16) are equal to each other and equal to the right member of (1.16), then

$$(1.17)\quad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) \geq \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta).$$

Because of Lemma 1.3 the equality sign must hold and Theorem 1.1 is proved.

Interchanging the two players, we obtain from Theorem 1.1:

THEOREM 1.2. *If β is a subset of B such that* $\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta_\beta} K(\xi, \eta_\beta) = \operatorname*{Inf}_{\eta_\beta} \operatorname*{Sup}_{\xi} K(\xi, \eta_\beta)$ *and* $\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta_\beta} K(\xi, \eta_\beta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta)$,

then

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta).$$

We shall now prove the following theorem.

THEOREM 1.3. *If $\{\alpha_i\}$ is a sequence of subsets of A such that $\alpha_i \subset \alpha_{i+1}$ and $\sum_{i=1}^{\infty} \alpha_i = A$, and if*

$$(1.18)\quad \operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta)$$

for each i, then a necessary and sufficient condition for the validity of

$$\text{(1.19)} \qquad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta)$$

is that

$$\text{(1.20)} \qquad \lim_{i=\infty} \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta).$$

PROOF: Because of (1.18) and Lemma 1.1 we have

$$\text{(1.21)} \qquad \lim_{i=\infty} \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta).$$

Hence, (1.20) implies (1.19) and (1.19) implies (1.20). This proves Theorem 1.3.

Interchanging the role of the two players, we obtain from Theorem 1.3 the following theorem.

THEOREM 1.4. *If* $\{\beta_i\}$ *is a sequence of subsets of B such that* $\beta_i \subset \beta_{i+1}$ *and* $\sum_{i=1}^{\infty} \beta_i = \beta$, *and if*

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta_{\beta_i}} K(\xi, \eta_{\beta_i}) = \operatorname*{Inf}_{\eta_{\beta_i}} \operatorname*{Sup}_{\xi} K(\xi, \eta_{\beta_i}),$$

then a necessary and sufficient condition for the validity of (1.19) *is that*

$$\text{(1.22)} \qquad \lim_{i=\infty} \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta_{\beta_i}} K(\xi, \eta_{\beta_i}) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta).$$

In [3] an intrinsic metric was introduced in the spaces A and B. The distance of two elements a_1 and a_2 of A is defined by

$$\text{(1.23)} \qquad \delta(a_1, a_2) = \operatorname*{Sup}_{b} | K(a_1, b) - K(a_2, b) | .$$

Similarly, the distance between two points b_1 and b_2 of B is defined by

$$\text{(1.24)} \qquad \delta(b_1, b_2) = \operatorname*{Sup}_{a} | K(a, b_1) - K(a, b_2) | .$$

Suppose that there exists a sequence $\{\alpha_i\}$ of subsets of A such that α_i is conditionally compact, $\alpha_i \subset \alpha_{i+1}$ and $\sum_{i=1}^{\infty} \alpha_i = A$.[3] It was shown in [3] that for any conditionally compact subset α_i the relation (1.18) holds. Hence, according to Theorem 1.3, a necessary and sufficient condition for the validity of (1.19) is that (1.20) holds for a sequence $\{\alpha_i\}$ where α_i is conditionally compact, $\alpha_i \subset \alpha_{i+1}$ and $\sum_{i=1}^{\infty} \alpha_i = A$. Similar remarks can be made concerning the space B. The distance definitions given in (1.23) and (1.24) can be extended to the spaces of the probability measures ξ and η, respectively. That is,

$$\text{(1.25)} \qquad \delta(\xi_1, \xi_2) = \operatorname*{Sup}_{\eta} | K(\xi_1, \eta) - K(\xi_2, \eta) |$$

[3] For a definition of compact and conditionally compact sets, see F. Hausdorff, *Mengenlehre* (3rd edition), p. 107, or [3, p. 296].

and

(1.26) $$\delta(\eta_1, \eta_2) = \operatorname*{Sup}_{\xi} | K(\xi, \eta_1) - K(\xi, \eta_2) | .$$

We shall say that a probability measure ξ is discrete if there exists a denumerable subset α of A such that $\xi(\alpha) = 1$. Similarly, a probability measure η will be said to be discrete if $\eta(\beta) = 1$ for some denumerable subset β of B. We shall now prove the following theorem.

THEOREM 1.5. *If the choice of player* 1 *is restricted to elements of a class C of probability measures ξ in which the class of all discrete probability measures ξ is dense, then a necessary and sufficient condition for the game to be strictly determined is that there exists a sequence $\{a_i\}$ of elements of A such that*

(1.27) $$\lim_{i=\infty} \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} K(\xi, \eta)$$

where

$$\alpha_i = \{a_1, a_2, \cdots, a_i\}.$$

PROOF: Since the class of all discrete probability measures ξ lies dense in the class C, there exists a sequence $\alpha = \{a_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ such that

(1.28) $$\operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta).$$

Since $\alpha_i = \{a_1, \cdots, a_i\}$ is finite, we have

(1.29) $$\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi_{\alpha_i}} \operatorname*{Inf}_{\eta} K(\xi_{\alpha_i}, \eta).$$

It then follows from Lemma 1.1 that

(1.30) $$\lim_{i=\infty} \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi_{\alpha_i}} K(\xi_{\alpha_i}, \eta) = \operatorname*{Sup}_{\xi_\alpha} \operatorname*{Inf}_{\eta} K(\xi_\alpha, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta).$$

Clearly, (1.30) and strict determinateness of the game implies (1.27). On the other hand, any $\alpha = \{a_i\}$ that satisfies (1.27), will satisfy also (1.28) and (1.30). But (1.27) and (1.30) imply that the game is strictly determined. Thus, Theorem 1.5 is proved.

THEOREM 1.6. *If the choice of player* 2 *is restricted to elements of a class C of probability measure η in which the class of all discrete probability measures η lies dense, then a necessary and sufficient condition for the strict determinateness of the game is that there exists a sequence $\beta = \{b_i\}$ of elements of B such that*

(1.31) $$\lim_{i=\infty} \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta_{\beta_i}} K(\xi, \eta_{\beta_i}) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} K(\xi, \eta)$$

where

$$\beta_i = \{b_1, \cdots, b_i\}.$$

This theorem is obtained from Theorem 1.5 by interchanging the players 1 and 2.

2. Statistical decision functions: the case of discrete chance variable.

2.1. *The problem of statistical decisions and its interpretation as a zero sum two person game.* In some previous publications (see, for example, [3]) the author has formulated the problem of statistical decisions as follows: Let $X = \{X^i\}$ ($i = 1, 2, \cdots$, ad inf.) be an infinite sequence of chance variables. Any particular observation x on X is given by a sequence $x = \{x^i\}$ of real values where x^i denotes the observed value of X^i. Suppose that the probability distribution $F(x)$ of X is not known. It is, however, known that F is an element of a given class Ω of distribution functions. There is, furthermore, a space D given whose elements d represent the possible decisions that can be made in the problem under consideration. Usually each element d of D will be associated with a certain subset ω of Ω and making the decision d can be interpreted as accepting the hypothesis that the true distribution is included in the subset ω. The fundamental problem in statistics is to give a rule for making a decision, that is, a rule for selecting a particular element d of D on the basis of the observed sample point x. In other words, the problem is to construct a function $d(x)$, called decision function, which associates with each sample point x an element $d(x)$ of D so that the decision $d(x)$ is made when the sample point x is observed.

This formulation of the problem includes the sequential as well as the classical non-sequential case. For any sample point x, let $n(x)$ be the number of components of x that must be known to be able to determine the value of $d(x)$. In other words, $n(x)$ is the smallest positive integer such that $d(y) = d(x)$ for any y whose first n coordinates are equal to the first n coordinates of x. If no finite n exists with the above property, we put $n = \infty$. Clearly, $n(x)$ is the number of observations needed to reach a decision. To put in evidence the dependence of $n(x)$ on the decision rule used, we shall occasionally write $n(x; \mathfrak{D})$ instead of $n(x)$ where $\mathfrak{D}$ denotes the decision function $d(x)$ used. If $n(x)$ is constant over the whole sample space, we have the classical case, that is the case where a decision is to be made on the basis of a predetermined number of observations. If $n(x)$ is not constant over the sample space, we have the sequential case. A basic question in statistics is this: What decision function should be chosen by the statistician in any given problem? To set up principles for a proper choice of a decision function, it is necessary to express in some way the degree of importance of the various wrong decisions that can be made in the problem under consideration. This may be expressed by a non-negative function $W(F, d)$, called weight functions, which is defined for all elements F of Ω and all elements d of D. For any pair (F, d), the value $W(F, d)$ expresses the loss caused by making the decision d when F is the true distribution of X. For any positive integer n, let $c(n)$ denote the cost of making n observations. If the decision function $\mathfrak{D} = d(x)$ is used the expected loss plus the expected cost of experimentation is given by

$$(2.1) \qquad r[F, \mathfrak{D}] = \int_M W[F, d(x)]\, dF(x) + \int_M c(n(x))\, dF(x)$$

where M denotes the sample space, i.e. the totality of all sample points x. We shall use the symbol $\mathfrak{D}$ for $d(x)$ when we want to indicate that we mean the whole decision function and not merely a value of $d(x)$ coresponding to some x.

The above expression (2.1) is called the risk. Thus, the risk is a real valued non-negative function of two variables F and $\mathfrak{D}$ where F may be any element of Ω and $\mathfrak{D}$ any decision rule that may be adopted by the statistician.

Of course, the statistician would like to make the risk r as small as possible. The difficulty he faces in this connection is that r depends on two arguments F and $\mathfrak{D}$, and he can merely choose $\mathfrak{D}$ but not F. The true distribution F is chosen, we may say, by Nature and Nature's choice is usually entirely unknown to the statistician. Thus, the situation that arises here is very similar to that of a zero sum two person game. As a matter of fact, the statistical problem may be interpreted as a zero sum two person game by setting up the following correspondence:

Two Person Game	*Statistical Decision Problem*
Player 1	Nature
Player 2	Statistician
Pure strategy a of player 1	Choice of true distribution F by Nature
Pure strategy b of player 2	Choice of decision rule $\mathfrak{D} = d(x)$
Space A	Space Ω
Space B	Space Q of decision rules $\mathfrak{D}$ that can be used by the statistician.
Outcome $K(a, b)$	Risk $r(F, \mathfrak{D})$
Mixed strategy ξ of player 1	Probability measure ξ defined over an additive class of subsets of Ω (a priori probability distribution in the space Ω)
Mixed strategy η of player 2	Probability measure η defined over an additive class of subsets of the space Q. We shall refer to η as randomized decision function.
Outcome $K(\xi, \eta)$ when mixed strategies are used.	Risk $r(\xi, \eta) = \int_Q \int_\Omega r(F, \mathfrak{D})\, d\xi\, d\eta$.

2.2. *Formulation of some conditions concerning the spaces Ω, D, the weight function $W(F, d)$ and the cost function of experimentation.* A general theory of statistical decision functions was developed in [3] assuming the fulfillment of seven conditions listed on pp. 297–8.[4] The conditions listed there are unnecessarily restrictive and we shall replace them here by a considerably weaker set of conditions.

In this chapter we shall restrict ourselves to the study of the case where each of the chance variables $X^1, X^2, \cdots$, ad inf. is discrete. We shall say that a chance

[4] In [3] only the continuous case is treated (existence of a density function is assumed), but all the results obtained there can be extended without difficulty to the discrete case.

variable is discrete if it can take only countably many different values. Let $a_{i1}, a_{i2}, \cdots$, ad inf. denote the possible values of the chance variable X^i. Since it is immaterial how the values a_{ij} are labeled, there is no loss of generality in putting $a_{ij} = j (j = 1, 2, 3, \cdots$, ad inf.). Thus, we formulate the following condition.

CONDITION 2.1. *The chance variable X^i ($i = 1, 2, \cdots$, ad inf.) can take only positive integral values.*

As in [3], also here we postulate the boundedness of the weight functior, i.e., we formulate the following condition.

CONDITION 2.2. *The weight function $W(F, d)$ is a bounded function of F and d.*

To formulate condition 2.3, we shall introduce some definitions. Let ω be a given subset of Ω. The distance between two elements d_1 and d_2 of D relative to ω is defined by

$$\delta(d_1, d_2; \omega) = \operatorname*{Sup}_{F \epsilon \omega} | W(F, d_1) - W(F, d_2) |. \tag{2.2}$$

We shall refer to $\delta(d_1, d_2; \Omega)$ as the absolute distance, or more briefly, the distance between d_1 and d_2. We shall say that a subset D^* of D is compact (conditionally compact) relative to ω, if it is compact (conditionally compact) in the sense of the metric $\delta(d_1, d_2; \omega)$. If D^* is compact relative to Ω, we shall say briefly that D^* is compact.

An element d of D is said to be uniformly better than the element d' of D relative to a subset ω of Ω if

$$W(F, d) \leqq W(F, d') \text{ for all } F \text{ in } \omega$$

and if

$$W(F, d) < W(F, d') \text{ for at least one } F \text{ in } \omega.$$

A subset D^* of D is said to be complete relative to a subset ω of Ω if for any d outside D^* there exists an element d^* in D^* such that d^* is uniformly better than d relative to ω.

CONDITION 2.3. *For any positive integer i and for any positive ϵ there exists a subset $D^*_{i,\epsilon}$ of D which is compact relative to Ω and complete relative to $\omega_{i,\epsilon}$ where $\omega_{i,\epsilon}$ is the class of all elements F of Ω for which prob $\{X^1 \leqq i\} \geqq \epsilon$.*

If D is compact, then it is compact with respect to any subset ω of Ω and Condition 2.3 is fulfilled. For any finite space D, Condition 2.3 is obviously fulfilled. Thus, Condition 2.3 is fulfilled, for example, for any problem of testing a statistical hypothesis H, since in that case the space D contains only two elements d_1 and d_2 where d_1 denotes the decision to reject H and d_2 the decision to accept H.

In [3] it was assumed that the cost of experimentation depends only on the number of observations made. This assumption is unnecessarily restrictive. The cost may depend also on the decision rule $\mathfrak{D}$ used. For example, let $\mathfrak{D}_1$ and $\mathfrak{D}_2$ be two decision rules such that $n(x; \mathfrak{D}_1)$ is equal to a constant n_0, while

$\mathfrak{D}_2$ is such that at any stage of the experimentation where $\mathfrak{D}_2$ requires taking at least one additional observation the probability is positive that experimentation will be terminated by taking only one more observation. Let x^0 be a particular sample point for which $n(x^0; \mathfrak{D}_2) = n(x^0, \mathfrak{D}_1) = n_0$. There are undoubtedly cases where the cost of experimentation is appreciably increased by the necessity of having to look at the observations at each stage of the experiment before we can decide whether or not to continue taking additional observations. Thus in many cases the cost of experimentation when x^0 is observed may be greater for $\mathfrak{D}_2$ than for $\mathfrak{D}_1$. The cost may also depend on the actual values of the observations made. Thus, we shall assume that the cost c is a single valued function of the observations $x^1, \cdots, x^m$ and the decision rule $\mathfrak{D}$ used, i.e., $c = c(x^1, \cdots, x^m, \mathfrak{D})$.

CONDITION 2.4. *The cost $c(x^1, \cdots, x^m, \mathfrak{D})$ is non-negative and* lim $c(x^1, \cdots, x^m, \mathfrak{D}) = \infty$ *uniformly in $x^1, \cdots, x^m, \mathfrak{D}$ as $m \to \infty$. For each positive integral value m, there exists a finite value c_m, depending only on m, such that $c(x^1, \cdots, x^m, \mathfrak{D}) \leqq c_m$ identically in $x^1, \cdots, x^m, \mathfrak{D}$. Furthermore, $c(x^1, \cdots, x^m, \mathfrak{D}_1) = c(x^1, \cdots, x^m, \mathfrak{D}_2)$ if $n(x; \mathfrak{D}_1) = n(x; \mathfrak{D}_2)$ for all x. Finally, for any sample point x we have $c(x^1, \cdots, x^{n(x,\mathfrak{D}_1)}, \mathfrak{D}_1) \leqq c(x^1, \cdots, x^{n(x,\mathfrak{D}_2)}, \mathfrak{D}_2)$ if there exists a positive integer m such that $n(x, \mathfrak{D}_1) = n(x, \mathfrak{D}_2)$ when $n(x, \mathfrak{D}_2) < m$ and $n(x, \mathfrak{D}_1) = m$ when $n(x, \mathfrak{D}_2) \geqq m$.*

2.3 *Alternative definition of a randomized decision function, and a further condition on the cost function.* In Section 2.1 we defined a randomized decision function as a probability measure η defined over some additive class of subsets of the space Q of all decision functions $d(x)$. Before formulating an alternative definition of a randomized decision function, we have to make precise the meaning of η by stating the additive class C_Q of subsets of Q over which η is defined. Let C_D be the smallest additive class of subsets of D which contains all subsets of D which are open in the sense of the metric $\delta(d_1, d_2; \Omega)$. For any finite set of positive integers $a_1, \cdots, a_k$ and for any element D^* of C_D, let $Q(a_1, \cdots, a_k, D^*)$ be the set of all decision functions $d(x)$ which satisfy the following two conditions: (1) If $x^1 = a_1, x^2 = a_2, \cdots, x^k = a_k$, then $n(x) = k$; (2) If $x^1 = a_1, \cdots, x^k = a_k$, then $d(x)$ is an element of D^*. Let C_Q^* be the class of all sets $Q(a_1, \cdots, a_k, D^*)$ corresponding to all possible values of $k, a_1, \cdots, a_k$ and all possible elements D^* of C_D. The additive class C_Q is defined as the smallest additive class containing C_Q^* as a subclass. Then with any η we can associate two sequences of functions

$$\{z_m(x^1, \cdots, x^m \mid \eta)\}$$

and

$$\{\delta_{x^1 \cdots x^m}(D^* \mid \eta)\} \, (m = 1, 2, \cdots, \text{ad inf.})$$

where $0 \leqq z_m(x^1, \cdots, x^m \mid \eta) \leqq 1$ and for any $x^1, \cdots, x^m$, $\delta_{x^1 \cdots x^m}$ is a probability measure in D defined over the additive class C_D. Here

$$z_m(x^1, \cdots, x^m \mid \eta)$$

denotes the conditional probability that $n(x) > m$ under the condition that the first m observations are equal to $x^1, \cdots, x^m$ and experimentation has not been terminated for $(x^1, \cdots, x^k)$ for $(k = 1, 2, \cdots, m - 1)$, while

$$\delta_{x^1 \cdots x^m}(D^* \mid \eta)$$

is the conditional probability that the final decision d will be an element of D^* under the condition that the sample $(x^1, \cdots, x^m)$ is observed and $n(x) = m$. Thus

$$z_1(x^1 \mid \eta) z_2(x^1, x^2 \mid \eta) \cdots z_{m-1}(x^1, \cdots, x^{m-1} \mid \eta) [1 - z_m(x^1, \cdots, x^m \mid \eta)] = \eta[Q(x^1, \cdots, x^m, D] \tag{2.3}$$

and

$$\delta_{x^1 \cdots x^m}(D^* \mid \eta) = \frac{\eta[Q(x^1, \cdots, x^m, D^*)]}{\eta[Q(x^1, \cdots, x^m, D)]}. \tag{2.4}$$

We shall now consider two sequences of functions $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}(D^*)\}$, not necessarily generated by a given η. An alternative definition of a randomized decision function can be given in terms of these two sequences as follows: After the first observation x^1 has been drawn, the statistician determines whether or not experimentation be continued by a chance mechanism constructed so that the probability of continuing experimentation is equal to $z_1(x^1)$. If it is decided to terminate experimentation, the statistician uses a chance mechanism to select the final decision d constructed so that the probability distribution of the selected d is equal to $\delta_{x^1}(D^*)$. If it is decided to take a second observation and the value x^2 is obtained, again a chance mechanism is used to determine whether or not to stop experimentation such that the probability of taking a third observation is equal to $z_2(x^1, x^2)$. If it is decided to stop experimentation, a chance mechanism is used to select the final d so that the probability distribution of the selected d is equal to $\delta_{x^1 x^2}(D^*)$, and so on.

We shall denote by ζ a randomized decision function defined in terms of two sequences $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}(D^*)\}$, as described above. Clearly, any given η generates a particular ζ. Let $\zeta(\eta)$ denote the ζ generated by η. One can easily verify that two different η's may generate the same ζ, i.e., there exist two different η's, say η_1 and η_2 such that $\zeta(\eta_1) = \zeta(\eta_2)$.

We shall now show that for any ζ there exists an η such that $\zeta(\eta) = \zeta$. Let ζ be given by the two sequences $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}(D^*)\}$. Let b_j denote a sequence of r_j positive integers, i.e., $b_j = (b_{j1}, \cdots, b_{j,r_j})$ $(j = 1, 2, \cdots, k)$ subject to the restriction that no b_j is equal to an initial segment of $b_l (j \neq l)$. Let, furthermore, $D_1^*, \cdots, D_k^*$ be k elements of C_D. Finally, let $Q(b_1, \cdots, b_k, D_1^*, \cdots, D_k^*)$ denote the class of all decision functions $d(x)$ which satisfy the following condition: If $(x^1, \cdots, x^{r_j}) = b_j$ then $n(x) = r_j$ and $d(x)$ is an element of $D_j^*(j = 1, \cdots, k)$. Let η be a probability measure such that

$$\eta[Q(b_1, \cdots, b_k, D_1^*, \cdots, D_k^*)]$$

$$(2.5) \qquad = \delta_{b_1}(D_1^*) \cdots \delta_{b_k}(D_k^*) \prod_{m=1}^{\infty} \prod_{x^m=1}^{\infty} \prod_{x^{m-1}=1}^{\infty} \cdots \prod_{x^1=1}^{\infty} \cdot \{z_m(x^1, \cdots, x^m)^{g_m(x^1, \cdots, x^m)}[1 - z_m(x^1, \cdots, x^m)]^{g_m^*(x^1, \cdots x^m)}\}$$

holds for all values of $k, b_1, \cdots, b_k, D_1^*, \cdots, D_k^*$. Here $g_m(x^1, \cdots, x^m) = 1$ if $(x^1, \cdots, x^m)$ is equal to an initial segment of at least of one of the samples $b_1, \cdots, b_k$, but is not equal to any of the samples $b_1, \cdots, b_k$. In all other cases $g_m(x^1, \cdots, x^m) = 0$. The function $g^*_m(x^1, \cdots, x^m)$ is equal to 1 if $(x^1, \cdots, x^m)$ is equal to one of the samples $b_1, \cdots, b_k$, and zero otherwise. Clearly, for any η which satisfies (2.5) we have $\zeta(\eta) = \zeta$. The existence of such an η can be shown as follows. With any finite set of positive integers $i_1, \cdots, i_r$ we associate an elementary event, say $A_r(i_1, \cdots, i_r)$. Let $\bar{A}_r(i_1, \cdots, i_r)$ denote the negation of the event $A_r(i_1, \cdots, i_r)$. Thus, we have a denumerable system of elementary events by letting $r, i_1, \cdots, i_r$ take any positive integral values. We shall assume that the events $A_1(1), A_1(2), \cdots$, ad inf. are independent and the probability that $A_1(i)$ happens is equal $z_1(i)$. We shall now define the conditional probability of $A_2(i, j)$ knowing for any k whether $A_1(k)$ or $\bar{A}_1(k)$ happened. If $A_1(i)$ happened, the conditional probability of $A_2(i, j) = z_2(i, j)$ and 0 otherwise. The conditional probability of the joint event that $A_2(i_1, j_1), A_2(i_2, j_2), \cdots, A_2(i_r, j_r), \bar{A}_2(i_{r+1}, j_{r+1}), \cdots$, and $\bar{A}_2(i_{r+s}, j_{r+s})$ will happen is the product of the conditional probabilities of each of these events (knowing for each i whether $A_1(i)$ or $\bar{A}_1(i)$ happened). Similarly, the conditional probability (knowing for any i and for any (i, j), whether the corresponding event $A_2(i, j)$ happened or not) that $A_3(i_1, j_1, k_1)$ and $A_3(i_2, j_2, k_2)$ and $\cdots A_3(i_r, j_r, k_r)$ and $\bar{A}_3(i_{r+1}, j_{r+1}, k_{r+1})$ and $\cdots$ and $\bar{A}_3(i_{r+s}; j_{r+s}, k_{r+s})$ will simultaneously happen is equal to the product of the conditional probabilities of each of them. The conditional probability of $A_3(i, j, k)$ is equal to $z_3(i, j, k)$ if $A_1(i)$ and $A_2(i, j)$ happened, and zero otherwise; and so on. Clearly, this system of probabilities is consistent.

If we interpret $A_r(i_1, \cdots, i_r)$ as the event that the decision function $\mathfrak{D} = d(x)$ selected by the statistician has the property that $n(x; \mathfrak{D}) > r$ when $x^1 = i_1, \cdots, x^r = i_r$, the above defined system of probabilities for the denumerable sequence $\{A_r(i_1, \cdots, i_r)\}$ of events implies the validity of (2.5) for $D_j^* = D(j = 1, \cdots, k)$. The consistency of the formula (2.5) for $D_j^* = D$ implies, as can easily be verified, the consistency of (2.5) also in the general case when $D_j^* \neq D$.

Let ζ_i be given by the sequences of $\{z_{mi}(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m, i}\}$ $(m = 1, 2, \cdots$, ad inf.). Let, furthermore, ζ be given by $\{z_m(x^1, \cdots, x^m)\}$ and $\{\delta_{x^1 \cdots x^m}\}$. We shall say that

$$(2.6) \qquad \lim_{i=\infty} \zeta_i = \zeta$$

if for any m, $x^1, \cdots, x^m$ we have

$$\lim_{i=\infty} z_{mi}(x^1, \cdots, x^m) = z_m(x^1, \cdots, x^m) \tag{2.7}$$

and

$$\lim_{i=\infty} \delta_{x^1 \cdots x^m, i}(D^*) = \delta_{x^1 \cdots x^m}(D^*) \tag{2.8}$$

for any open subset D^* of D whose boundary has probability measure zero according to the limit probability measure $\delta_{x^1 \cdots x^m}$.

In addition to Condition 2.4, we shall impose the following continuity condition on the cost function.

CONDITION 2.5. *If*

$$\lim_{i=\infty} \zeta(\eta_i) = \zeta(\eta),$$

then

$$\lim_{i=\infty} \int_{Q(x^1,\cdots,x^m)} c(x^1, \cdots, x^m, \mathfrak{D})\, d\eta_i = \int_{Q(x^1,\cdots,x^m)} c(x^1, \cdots, x^m, \mathfrak{D})\, d\eta.$$

where $Q(x^1, \cdots, x^m)$ is the class of all decision functions $\mathfrak{D}$ for which $n(y, \mathfrak{D}) = m$ if $y^1 = x^1, \cdots, y^m = x^m$.

2.4. *The main theorem.* In this section we shall show that the statistical decision problem, viewed as a zero sum two person game, is strictly determined. It will be shown in subsequent sections that this basic theorem has many important consequences for the theory of statistical decision functions. A precise formulation of the theorem is as follows:

THEOREM 2.1. *If Conditions 2.1–2.5 are fulfilled, the decision problem, viewed as a zero sum two person game, is strictly determined, i.e.,*

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r(\xi, \eta). \tag{2.9}$$

To prove the above theorem, we shall first derive several lemmas.

LEMMA 2.1. *For any $\epsilon > 0$, there exists a positive integer m_ϵ, depending only on ϵ, such that the value of $\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta)$, is not changed by more than ϵ if we restrict the choice of the statistician to decision functions $d(x)$ for which $n(x) \leqq m_\epsilon$ for all x.*

PROOF: Put $W_0 = \operatorname*{Sup}_{F,D} W(F, d)$ and choose m_ϵ so that

$$c(x^1, \cdots, x^m, \mathfrak{D}) > \frac{W_0^2}{\epsilon} \tag{2.10}$$

identically in $x^1, \cdots, x^m$ and $\mathfrak{D}$ for all $m \geqq m_\epsilon$. The existence of such a value m_ϵ follows from Condition 2.4. Consider the function $\operatorname*{Inf}_{\mathfrak{D}} r(\xi, \mathfrak{D})$. Our lemma is proved, if we can show that for any ξ, the value of $\operatorname*{Inf}_{\mathfrak{D}} r(\xi, \mathfrak{D})$ is not increased by more than ϵ if we restrict $\mathfrak{D}$ to be such that $n(x, \mathfrak{D}) \leq m_\epsilon$ for all x. The latter statement is proved, if we can show that for any decision function $\mathfrak{D}_1 = d_1(x)$ we can find another decision function $\mathfrak{D}_2 = d_2(x)$ such that $n(x, \mathfrak{D}_2) \leq m_\epsilon$ for all x and $r(\xi, \mathfrak{D}_2) \leq r(\xi, \mathfrak{D}_1) + \epsilon$. There are two cases to be considered: (a) prob $\{n(X, \mathfrak{D}_1) > m_\epsilon \mid \xi\} \geqq \epsilon/W_0$ and (b) prob $\{n(X, \mathfrak{D}_1) > m_\epsilon \mid \xi\} < \epsilon/W_0$. In case (a) we have $r(\xi, \mathfrak{D}_1) \geqq W_0$. In this case we can choose $\mathfrak{D}_2$ to be the rule that we decide for some element d_0 of D without taking any observations. Clearly, for this choice of $\mathfrak{D}_2$ we shall have $r(\xi, \mathfrak{D}_2) \leq r(\xi, \mathfrak{D}_1)$. In case (b) we choose $\mathfrak{D}_2$ as follows:

$$d_2(x) = d_1(x) \text{ whenever } n(x, \mathfrak{D}_1) \leqq m_\epsilon ;$$

$$d_2(x) = d_0 \text{ whenever } n(x, \mathfrak{D}_1) > m_\epsilon ,$$

where d_0 is an arbitrary element of D. Thus, $n(x, \mathfrak{D}_2) \leqq m_\epsilon$ for all x. Since prob $\{n(x, \mathfrak{D}_1) > m_\epsilon \mid \xi) < \epsilon/W_0$, it is clear that $r(\xi, \mathfrak{D}_2) \leqq r(\xi, \mathfrak{D}_1) + \epsilon$. Hence our lemma is proved.

Let Q^m denote the class of decision functions $\mathfrak{D}$ for which $n(x; \mathfrak{D}) \leqq m$ for all x. For any positive ϵ, let $Q^{m,\epsilon}$ denote the class of all decision functions which satisfy the following two conditions simultaneously: (1) $n(x, \mathfrak{D}) \leqq m$ for all x; (2) $d(x)$ is an element of $D^*_{x^1,\epsilon}$ where $D^*_{x^1,\epsilon}$ denotes the subset of D having the properties stated in Condition 2.3. Clearly, $Q^{m,\epsilon} \subset Q^m$. A probability measure η will be denoted by η^m if $\eta(Q^m) = 1$, and by $\eta^{m,\epsilon}$ if $\eta(Q^{m,\epsilon}) = 1$.

LEMMA 2.2. *The following inequality holds:*

$$\text{(2.11)} \quad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r(\xi, \eta^m) \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m,\epsilon}} r(\xi, \eta^{m,\epsilon}) \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r(\xi, \eta^m) + \epsilon W_0 ,$$

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

PROOF: The first half of (2.11) is obvious. If we replace the subscript x^1 by the chance variable X^1, the set $\omega_{x^1,\epsilon}$ defined in Condition 2.3 will be a random subset of Ω. It follows easily from the definition of $\omega_{x^1,\epsilon}$ that

$$\text{(2.12)} \quad \text{prob } \{F \epsilon \omega_{X^1,\epsilon} \mid F\} \geqq 1 - \epsilon.$$

With any decision function $\mathfrak{D} = d(x)$ we shall associate another decision function $\mathfrak{D}^* = d^*(x)$ such that $n(x, \mathfrak{D}) = n(x, \mathfrak{D}^*)$; $d^*(x) = d(x)$ whenever $d(x) \epsilon D^*_{x^1,\epsilon}$; and $d^*(x)$ is an element of $D^*_{x^1,\epsilon}$ that is uniformly better than $d(x)$ relative to $\omega_{x^1,\epsilon}$ whenever $d(x) \notin D^*_{x^1,\epsilon}$. It follows from (2.12) and the fact that W_0 is an upper bound of $W(F, d)$ that

$$\text{(2.13)} \quad r(F, \mathfrak{D}^*) \leqq r(F, \mathfrak{D}) + \epsilon W_0.$$

The second half of (2.11) is an immediate consequence of (2.13) and our lemma is proved.

LEMMA 2.3. *The equation*

$$\text{(2.14)} \quad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m,\epsilon}} r(\xi, \eta^{m,\epsilon}) = \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi} r(\xi, \eta^{m,\epsilon})$$

holds for all m and ϵ.

PROOF: For any positive integral values m, k and for any $\rho > 0$, let $\Omega^{m,k,\rho}$ be the class of all elements F of Ω for which

$$\text{prob } \{x^1 \leqq k \text{ and } x^2 \leqq k \text{ and } \cdots x^m \leqq k\} \geqq 1 - \rho.$$

A probability measure ξ for which $\xi(\Omega^{m,k,\rho}) = 1$ will be denoted by $\xi^{m,k,\rho}$. To prove (2.14), we shall first prove the inequality

$$(2.15) \quad \left| \underset{\xi^{m,k,\rho}}{\text{Sup}} \; \underset{\eta^{m,\epsilon}}{\text{Inf}} \; r(\xi^{m,k,\rho}, \eta^{m,\epsilon}) - \underset{\eta^{m,\epsilon}}{\text{Inf}} \; \underset{\xi^{m,k,\rho}}{\text{Sup}} \; r(\xi^{m,k,\rho}, \eta^{m,\epsilon}) \right| \leqq \rho(W_0 + C_m)$$

where C_m is an upper bound of $C(x^1, \cdots, x^r, \mathfrak{D})$ for all $r \leqq m, x^1, \cdots, x^r$ and $\mathfrak{D}$.

Since for any $d(x)$ in $Q^{m,\epsilon}$, $d(x)$ must be an element of $D^*_{x^1,\epsilon}$ and since $D^*_{x^1,\epsilon}$ is compact, it is sufficient to prove the validity of (2.14) in the case when $D^*_{x^1,\epsilon}$ is a finite set. Thus, we shall assume in the remainder of the proof that $D^*_{x^1,\epsilon}$ is finite.

Let δ be a given positive number and let $Q^{m,k,\epsilon}$ be a finite subset of $Q^{m,\epsilon}$ satisfying the following condition: for any element $\mathfrak{D} = d(x)$ in $Q^{m,\epsilon}$ there exists an element $\mathfrak{D}^* = d^*(x)$ in $Q^{m,k,\epsilon}$ such that

$$d^*(x) = d(x) \qquad \text{and} \qquad | C(x, \mathfrak{D}^*) - C(x, \mathfrak{D}) | \leqq \delta$$

for all x for which $x^1 \leqq k, x^2 \leqq k, \cdots$, and $x^m \leqq k$. Clearly, for any choice of δ there exists a finite subset $Q^{m,k,\epsilon}$ of $Q^{m,\epsilon}$ with the desired property. For any $\mathfrak{D}$ in $Q^{m,\epsilon}$, we can then find an element $\mathfrak{D}^*$ in $Q^{m,k,\epsilon}$ such that

$$r(F, \mathfrak{D}^*) \leqq r(F, \mathfrak{D}) + \rho(W_0 + C_m) + \delta,$$

for all F in $\Omega^{m,k,\rho}$. From this it follows that

$$(2.16) \quad \underset{\xi^{m,k,\rho}}{\text{Sup}} \; \underset{\eta^{m,\epsilon}}{\text{Inf}} \; r \leqq \underset{\xi^{m,k,\rho}}{\text{Sup}} \; \underset{\eta^{m,k,\epsilon}}{\text{Inf}} \; r \leqq \underset{\xi^{m,k,\rho}}{\text{Sup}} \; \underset{\eta^{m,\epsilon}}{\text{Inf}} \; r + \rho(W_0 + C_m) + \delta$$

$$(2.17) \quad \underset{\eta^{m,\epsilon}}{\text{Inf}} \; \underset{\xi^{m,k,\rho}}{\text{Sup}} \; r \leqq \underset{\eta^{m,k,\epsilon}}{\text{Inf}} \; \underset{\xi^{m,k,\rho}}{\text{Sup}} \; r \leqq \underset{\eta^{m,\epsilon}}{\text{Inf}} \; \underset{\xi^{m,k,\rho}}{\text{Sup}} \; r + \rho(W_0 + C_m) + \delta$$

where $\eta^{m,k,\epsilon}(Q^{m,k,\epsilon}) = 1$. Since $Q^{m,k,\epsilon}$ is finite, we have

$$(2.18) \quad \underset{\xi^{m,k,\rho}}{\text{Sup}} \; \underset{\eta^{m,k,\epsilon}}{\text{Inf}} \; r = \underset{\eta^{m,k,\epsilon}}{\text{Inf}} \; \underset{\xi^{m,k,\rho}}{\text{Sup}} \; r.$$

Inequality (2.15) follows from (2.16), (2.17) and (2.18) and the fact that δ can be chosen arbitrarily small.

Lemmas 1.1., 1.3 and the inequality (2.15) imply that Lemma 2.3 must hold if

$$(2.19) \quad \lim_{k=\infty} \underset{\eta^{m,\epsilon}}{\text{Inf}} \; \underset{\xi^{m,k,\rho}}{\text{Sup}} \; r = \underset{\eta^{m,\epsilon}}{\text{Inf}} \; \underset{\xi}{\text{Sup}} \; r$$

holds. Thus, the proof of Lemma 2.3 is completed if we can show the validity of (2.19).

Let $\{\eta_k^{m,\epsilon}\}$ ($k = 1, 2, \cdots$, ad inf.) be a sequence of randomized decision functions such that

$$(2.20) \quad \lim_{k=\infty} [\underset{\xi^{m,k,\rho}}{\text{Sup}} \; r(\xi^{m,k,\rho}, \eta_k^{m,\epsilon}) - \underset{\eta^{m,\epsilon}}{\text{Inf}} \; \underset{\xi^{m,k,\rho}}{\text{Sup}} \; r(\xi^{m,k,\rho}, \eta^{m,\epsilon})] = 0.$$

Let $\zeta_k = \zeta(\eta_k^{m,\epsilon})$ (see definition in Section 3.2) and let ζ_k be given by the two sequences of functions $\{z_{rk}(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1 \cdots x^r, k}\}$ $(r = 1, 2, \cdots, m)$. Since there are only countably many samples $(x^1, \cdots, x^r)$ $(r \leqq m)$, there exists a subsequence $\{k^1\}$ of the sequence $\{k\}$ such that

$$\lim_{k=\infty} z_{r,k^1}(x^1, \cdots, x^r) = z_r(x^1, \cdots, x^r) \tag{2.21}$$

and

$$\lim_{k=\infty} \delta_{x^1 \cdots x^r, k^1} = \delta_{x^1 x^2 \cdots x^r} \tag{2.22}$$

for all r and all samples $(x^1, \cdots, x^r)$. Let $\eta_0^{m,\epsilon}$ be a randomized decision function such that $\zeta(\eta_0^{m,\epsilon})$ is equal to the ζ defined by $\{z_r(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1 \cdots x^r}\}$ $(r = 1, 2, \cdots, m)$.

For any element F of Ω and for any $\nu > 0$, there exists a finite subset M_ν of the m-dimensional sample space such that the probability (under F) that the sample $(x^1, \cdots, x^m)$ will fall in M_ν is $\geqq 1 - \nu$. From this and the continuity of the cost function (Condition 2.5) it follows that

$$\lim_{k=\infty} r(F, \eta_{k^1}^{m,\epsilon}) = r(F, \eta_0^{m,\epsilon}) \text{ for all } F. \tag{2.23}$$

Clearly,

$$\operatorname*{Sup}_{\xi^{m,k,\rho}} r(\xi^{m,k,\rho}, \eta) = \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta) \tag{2.24}$$

where $F^{m,k,\rho}$ is an element of $\Omega^{m,k,\rho}$. Hence

$$\operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi^{m,k,\rho}} r(\xi^{m,k,\rho}, \eta^{m,\epsilon}) = \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,\epsilon}). \tag{2.25}$$

Since any F in Ω is contained in $\Omega^{m,k,\rho}$ for sufficiently large k, it follows from (2.20) and (2.25) that

$$\lim_{k=\infty} r(F, \eta_{k^1}^{m,\epsilon}) \leqq \lim_{k=\infty} \{\operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,\epsilon})\}. \tag{2.26}$$

Hence, because of (2.23),

$$r(F, \eta_0^{m,\epsilon}) \leqq \lim_{k=\infty} \{\operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,\epsilon})\}. \tag{2.27}$$

Thus,

$$\operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{F} r(F, \eta^{m,\epsilon}) \leqq \lim_{k=\infty} \{\operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{F^{m,k,\rho}} r(F^{m,k,\rho}, \eta^{m,\epsilon})\}. \tag{2.28}$$

Since the left hand member of (2.28) cannot be smaller than the right hand member, the equality sigh must hold. This concludes the proof of Lemma 2.3.

Theorem 2.1 can easily be proved with the help of lemmas 2.1, 2.2 and 2.3. From Lemma 2.2 it follows that

$$\lim_{\epsilon=0} \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^{m,\epsilon}} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r. \tag{2.29}$$

From this and Lemma 2.3 we obtain

$$\lim_{\epsilon=0} \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r. \tag{2.30}$$

But

$$\lim_{\epsilon=0} \operatorname*{Inf}_{\eta^{m,\epsilon}} \operatorname*{Sup}_{\xi} r \geqq \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r.$$

Hence

$$\operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r. \tag{2.31}$$

Hence, because of Lemma 1.3, we then must have

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r. \tag{2.32}$$

It follows from Lemma 2.1 that

$$\lim_{m=\infty} \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r. \tag{2.33}$$

Hence, because of (2.32), we have

$$\lim_{m=\infty} \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r. \tag{2.34}$$

But

$$\lim_{m=\infty} \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r \geqq \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r. \tag{2.35}$$

Hence

$$\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r \tag{2.36}$$

Theorem 2.1 is an immediate consequence of (2.36) and Lemma 1.3.

2.5. *Theorems on complete classes of decision functions and minimax solutions.* For any positive ϵ we shall say that the randomized decision function η_0 is an ϵ-Bayes solution relative to the a priori distribution ξ if

$$r(\xi, \eta_0) \leqq \operatorname*{Inf}_{\eta} r(\xi, \eta) + \epsilon. \tag{2.37}$$

If η_0 satisfies (2.37) for $\epsilon = 0$, we shall say that η_0 is a Bayes solution relative to ξ.

A randomized decision rule η_1 is said to be uniformly better than η_2 if

$$r(F, \eta_1) \leqq r(F, \eta_2) \text{ for all } F \tag{2.38}$$

and if

$$r(F, \eta_1) < r(F, \eta_2) \text{ at least for one } F. \tag{2.39}$$

A class C of randomized decision functions η is said to be complete if for any η not in C we can find an element η^* in C such that η^* is uniformly better than η.

THEOREM 2.2. *If Conditions* 2.1–2.5 *are fulfilled, then for any* $\epsilon > 0$ *the class* C_ϵ *of all* ϵ*-Bayes solutions corresponding to all possible a priori distributions* ξ *is a complete class.*

PROOF: Let η_0 be a randomized decision function that is not an ϵ-Bayes solution relative to any ξ. That is,

$$r(\xi, \eta_0) > \operatorname*{Inf}_{\eta} r(\xi, \eta) + \epsilon \text{ for all } \xi. \tag{2.40}$$

If $r(F, \eta_0) = \infty$ for all F, then there is evidently an element of C_ϵ that is uniformly better than η_0. Thus, we can restrict ourselves to the case where

$$r(F, \eta_0) < \infty \text{ at least for one } F. \tag{2.41}$$

Put

$$W^*(F, d) = W(F, d) - r(F, \eta_0) \tag{2.42}$$

and let r^* (ξ, η) denote the risk when $W(F, d)$ is replaced by $W^*(F, d)$. Then

$$r^*(\xi, \eta) = r(\xi, \eta) - r(\xi, \eta_0). \tag{2.43}$$

Let Q^m denote the class of all decision functions $d(x)$ for which $n(x) \leqq m$ identically in x. Furthermore, denote any η for which $\eta(Q^m) = 1$ by η^m. We shall first prove the following relation.

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r^*(\xi, \eta^m) = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^*(\xi, \eta^m) \tag{2.44}$$

for any positive integral value m. For any positive constant c, let Ω_c denote the class of all elements F for which $r(F, \eta_0) \leqq c$.

Clearly, Conditions 2.1-2.5 remain valid if we replace $W(F, d)$ by $W^*(F, d)$ and Ω by Ω_c where c is restricted to values for which Ω_c is not empty. Hence, Theorem 2.1 can be applied and we obtain

$$\operatorname*{Sup}_{\xi^c} \operatorname*{Inf}_{\eta^m} r^*(\xi^c, \eta^m) = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi^c} r^*(\xi^c, \eta^m), \tag{2.45}$$

where ξ^c denotes any ξ for which $\xi(\Omega_c) = 1$. Let h and w be two positive values for which

$$\operatorname*{Sup}_{\xi^c} \operatorname*{Inf}_{\eta^m} r^*(\xi^c, \eta^m) \geqq -h \quad \text{for all} \quad c \tag{2.46}$$

and

$$r(F, \eta^m) \leqq w \text{ for all } F \text{ and all } \eta^m. \tag{2.47}$$

Clearly, such two constants h and ω exist. From (2.46) and Lemma 1.3 we obtain

$$\operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^* (\xi, \eta^m) \geqq -h. \tag{2.48}$$

Since

$$r^*(F, \eta^m) < -(h + \delta) \text{ for any } F \text{ not in } \Omega_{h+\delta+w} (\delta > 0), \tag{2.49}$$

it follows from (2.48) that

(2.50) $$\operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi^c} r^* = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^* \quad \text{for all} \quad c > h + w.$$

From (2.45) and (2.50) we obtain

(2.51) $$\operatorname*{Sup}_{\xi^c} \operatorname*{Inf}_{\eta^m} r^* = \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^* \quad \text{for all} \quad c > h + w.$$

Hence,

(2.51a) $$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta^m} r^* \geqq \operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^*.$$

Because of Lemma 1.3, the equality sign must hold and (2.44) is proved.

Since η_0 is not an element of C_ϵ, we must have

(2.52) $$\operatorname*{Inf}_{\eta} r(\xi, \eta) < r(\xi, \eta_0) - \epsilon.$$

From this it follows that

(2.53) $$\operatorname*{Inf}_{\eta} r^*(\xi, \eta) \leqq -\epsilon.$$

Hence

(2.54) $$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r^*(\xi, \eta) \leqq -\epsilon.$$

It was shown in the proof of Lemma 2.1 that for any $\rho > 0$ there exists a positive integer m_ρ, depending only on ρ, such that

(2.55) $$\operatorname*{Inf}_{\eta^{m_\rho}} r(\xi, \eta^{m_\rho}) \leqq \operatorname*{Inf}_{\eta} r(\xi, \eta) + \rho \text{ for all } \xi.$$

From (2.44), (2.54) and (2.55) it follows that there exists a positive integer m_0, namely $m_0 = m_{\epsilon/2}$, such that

(2.56) $$\operatorname*{Inf}_{\eta^m} \operatorname*{Sup}_{\xi} r^*(\xi, \eta^m) \leq -\frac{\epsilon}{2} \text{ for any } m \geqq m_0.$$

From (2.44) and (2.56) it follows that there exists an a priori distribution ξ_1 and an ϵ-Bayes solution η_1^m relative to ξ_1 such that

(2.57) $$r^*(F, \eta_1^m) \leqq -\frac{\epsilon}{4} \text{ for all } F.$$

Hence, because of (2.43),

(2.58) $$r(F, \eta_1^m) \leqq r(F, \eta_0) - \frac{\epsilon}{4} \text{ for all } F.$$

and Theorem 2.2 is proved.

THEOREM 2.3. *If D is compact, and if Conditions 2.1, 2.2, 2.4, 2.5 are fulfilled, then there exists a minimax solution, i.e., a decision rule η_0 for which*

(2.59) $$\operatorname*{Sup}_F r(F, \eta_0) \leqq \operatorname*{Sup}_F r(F, \eta) \text{ for all } \eta.$$

To prove the above theorem, we shall first prove the following lemma.

LEMMA 2.4. *If D is compact and if Conditions 2.1, 2.2, 2.4, 2.5 are fulfilled, then for any sequence $\{\eta_i\}$ ($i = 1, 2, \cdots$, ad inf.) of randomized decision functions for which $r(F, \eta_i)$ is a bounded function of F and i, there exists a subsequence $\{\eta_{i_j}\}$ ($j = 1, 2, \cdots$, ad inf.) and a randomized decision function η_0 such that*

(2.60) $$\liminf_{j=0} r(\xi, \eta_{i_j}) \geqq r(\xi, \eta_0) \text{ for all } \xi.$$

PROOF: Let $\zeta_i = \zeta(\eta_i)$ (defined in Section 2.3) be given by $\{z_{r,i}(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1x^2\cdots x^r,i}\}$ ($r = 1, 2, \cdots$, ad inf.). Thus, $z_{r,i}(x^1, \cdots, x^r)$ is the conditional probability that we shall take an observation on X^{r+1} using the rule η_i and knowing that the first r observations are given by $x^1, \cdots, x^r$ and that experimentation was not terminated for $(x^1, \cdots, x^k)$ $(k < r)$. As stated in section 2.3, for any $r, x^1, \cdots, x^r$ the symbol $\delta_{x^1\cdots x^r,i}$ denotes the conditional probability distribution of the selected d when η_i is used and is known that the first r observations are equal to $x^1, \cdots, x^r$ and that $n(x) = r$. Since there are only countably many finite samples $(x^1, \cdots, x^r)$, it is possible to find a subsequence $\{i_j\}$ of $\{i\}$ such that $\lim_{j=\infty} z_{r,i_j}(x^1, \cdots, x^r)$ and $\lim_{j=\infty} \delta_{x^1\cdots x^r,i_j}$ exist.[5] Put

(2.61) $$\lim_{j=\infty} z_{r,i_j}(x^1, \cdots, x^r) = z_{r,0}(x^1, \cdots, x^r)$$

and

(2.62) $$\lim_{j=\infty} \delta_{x^1\cdots x^r,i_j} = \delta_{x^1\cdots x^r,0}\,.$$

As shown in section 2.3, there exists a randomized decision function η_0 such $\zeta_0 = \zeta(\eta_0)$ is given by $\{z_{r,0}(x^1, \cdots, x^2)\}$ and $\{\delta_{x^1\cdots x^r,0}\}$. Let $q_{r,i}(x^1, \cdots, x^r \mid \xi)$ denote the probability that the sample $(x^1, \cdots, x^r)$ will be obtained and that experimentation will be stopped at the r-th observation when ξ is the a priori distribution and η_i is the decision rule used by the statistician. For any sample $(x^1, \cdots, x^r)$ let $R_i(x^1, \cdots, x^r)$ denote the expected value of $W(F, d)$ when the distribution of F is equal to the a posteriori distribution of F as implied by ξ and $(x^1, \cdots, x^r)$ and where d is a chance variable independent of F with the probability distribution $\delta_{x^1\cdots x^r,i}$. Since, $r(\xi, \eta_i)$ is bounded by assumption, the probability that experimentation will go on indefinitely is equal to zero. From this it follows that

(2.63) $$\sum_{r,x^1,\cdots,x^r} q_{r,i}(x^1, \cdots, x^r \mid \xi) = 1 \text{ for all } \xi.$$

---

[5] The existence of $\lim_{j=\infty} \delta_{x^1\cdots x^r,i_j}$ follows from the compactness of D (see Theorem 3.6 in [3]).

Then $r(\xi, \eta_i)$ is given by

$$r(\xi, \eta_i) \tag{2.64}$$
$$= \sum_{r, x^1, \cdots, x^r} q_{r,i}(x^1, \cdots, x^r \mid \xi) \left[R_i(x^1, \cdots, x^r) + \frac{\int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r, \mathfrak{D})\, d\eta_i}{\int_{Q_{x^1 \cdots x^r}} d\eta_i} \right]$$

where $Q_{x^1 \cdots x^r}$ is the totality of all decision functions $d(x)$ for which $n(y) = r$ whenever $y^1 = x^1, \cdots, y^r = x^r$. Clearly,

$$\lim_{j=\infty} q_{r,i_j}(x^1, \cdots, x^r \mid \xi) = q_{r,0}(x^1, \cdots, x^r \mid \xi). \tag{2.65}$$

Since D is compact and since $W(F, d)$ is a continuous function of d uniformly in F (in the sense of the metric defined in D), we have

$$\lim_{j=\infty} R_{i_j}(x^1, \cdots, x^r) = R_0(x^1, \cdots, x^r). \tag{2.66}$$

From Condition 2.5 it follows that

$$\lim_{j=\infty} \frac{\int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r, \mathfrak{D})\, d\eta_{i_j}}{\int_{Q_{x^1 \cdots x^r}} d\eta_{i_j}} = \frac{\int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r \mathfrak{D}_{,0})\, d\eta_0}{\int_{Q_{x^1 \cdots x^r}} d\eta_0}. \tag{2.67}$$

Lemma 2.4 is an immediate consequence of the equations (2.64) − (2.67).

We are now in a position to prove Theorem 2.3. Because of Theorem 2.1 there exists a sequence $\{\eta_i\}$ such that

$$\lim_{i=\infty} \operatorname*{Sup}_F r(F, \eta_i) = \operatorname*{Inf}_\eta \operatorname*{Sup}_F r(F, \eta). \tag{2.68}$$

According to Lemma 2.4 there exists a subsequence $\{\eta_{i_j}\}$ ($j = 1, 2, \cdots$, ad inf.) and a randomized decision function η_0 such that

$$\liminf_{j=\infty} r(F, \eta_{i_j}) \geqq r(F, \eta_0) \text{ for all } F. \tag{2.69}$$

It follows from (2.68) and (2.69) that η_0 is a minimax solution and Theorem 2.3 is proved.

THEOREM 2.4. *If D is compact and if Conditions* 2.1, 2.2, 2.4, 2.5 *are fulfilled, then for any ξ there exists a Bayes solution relative to ξ.*

This theorem is an immediate consequence of Lemma 2.4.

We shall say that η_0 is a Bayes solution in the wide sense, if there exists a sequence $\{\xi_i\}$ ($i = 1, 2, \cdots$, ad inf.) such that

$$\lim_{i=\infty} [r(\xi_i, \eta_0) - \operatorname*{Inf}_\eta r(\xi_i, \eta)] = 0. \tag{2.70}$$

We shall say that η_0 is a Bayes solution in the strict sense, if there exists a ξ such that η_0 is a Bayes solution relative to ξ.

THEOREM 2.5. *If D is compact and Conditions* 2.1–2.5 *hold, then the class of all Bayes solutions in the wide sense is a complete class.*

PROOF: Let η_0 be a decision rule that is not a Bayes solution in the wide sense. Consider the weight function $W^*(F, d) = W(F, d) - r(F, \eta_0)$. We may assume that $r(F, \eta_0) < \infty$ for at least some F, since otherwise there obviously exists a Bayes solution in the wide sense that is uniformly better than η_0. Then it follows easily from (2.44) and Lemmas 2.1 and 1.3 that

$$\text{(2.71)} \qquad \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r^*(\xi, \eta) = \operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r^*(\xi, \eta) = v^* \quad \text{(say)},$$

where $r^*(\xi, \eta)$ is the risk corresponding to $W^*(F, d)$, i.e.,

$$\text{(2.72)} \qquad r^*(\xi, \eta) = r(\xi, \eta) - r(\xi, \eta_0).$$

Theorem 2.3 is clearly applicable to the risk function $r^*(\xi, \eta)$. Then, there exists a minimax solution η_1 for the problem corresponding to the new weight function $W^*(F, d)$. Since, because of 2.72, $v^* \leqq 0$, we have

$$\text{(2.73)} \qquad r^*(\xi, \eta_1) = r(\xi, \eta_1) - r(\xi, \eta_0) \leqq 0 \text{ for all } \xi.$$

Our theorem is proved, if we can show that η_1 is a Bayes solution in the wide sense. Let $\{\xi_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of a priori distributions such that

$$\text{(2.74)} \qquad \lim_{i=\infty} \operatorname*{Inf}_{\eta} r^*(\xi_i, \eta) = v^*.$$

Since η_1 is a minimax solution, we must have

$$\text{(2.75)} \qquad r^*(\xi_i, \eta_1) \leqq v^*.$$

It follows from (2.74) and (2.75) that η_1 is a Bayes solution in the wide sense and our theorem is proved.

We shall now formulate an additional condition which will permit the derivation of some stronger theorems. First, we shall give a convergence definition in the space Ω. We shall say that F_i converges to F in the ordinary sense if

$$\text{(2.76)} \qquad \lim_{i=\infty} p_r(x^1, \cdots, x^r | F_i) = p_r(x^1, \cdots, x^r | F) \qquad (r = 1, 2, \cdots, \text{ad inf.}).$$

Here $p_r(x^1, \cdots, x^r \mid F)$ denotes the probability, under F, that the first r observations will be equal to $x^1, \cdots, x^r$, respectively. We shall say that a subset ω of Ω is compact in the ordinary sense, if ω is compact in the sense of the convergence definition (2.76).

CONDITION 2.6. *The space Ω is compact in the ordinary sense. If F_i converges to F, as $i \to \infty$, in the ordinary sense, then*

$$\lim_{i=\infty} W(F_i, d) = W(F, d)$$

uniformly in d.

THEOREM 2.6. *If D is compact and if Conditions* 2.1, 2.2, 2.4, 2.5, 2.6 *hold, then:*

(i) *there exists a least favorable a priori distribution, i.e., an a priori distribution* ξ_0 *for which*

$$\operatorname*{Inf}_{\eta} r(\xi_0, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta).$$

(ii) *A minimax solution exists and any minimax solution is a Bayes solution in the strict sense.*

(iii) *If* η_0 *is a decision rule which is not a Bayes solution in the strict sense and for which* $r(F, \eta_0)$ *is a bounded function of* F, *then there exists a decision rule* η_1 *which is a Bayes solution in the strict sense and is uniformly better than* η_0.

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

$$\lim_{i=\infty} \operatorname*{Inf}_{\eta} r(\xi_i, \eta) = \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r(\xi, \eta). \tag{2.77}$$

Since Ω is compact in the ordinary sense, there exists an a priori distribution ξ_0 and a subsequence $\{\xi_{i_j}\}$ or $\{\xi_i\}$ such that

$$\lim_{j=\infty} \xi_{i_j}(\omega) = \xi_0(\omega) \tag{2.78}$$

for any subset ω of Ω which is open (in the sense of the ordinary convergence definition in Ω) and for which $\xi_0(\omega^*) = 0$, where ω^* denotes the set of all boundary points of ω. We shall show that ξ_0 is a least favorable distribution. Assume that it is not. Then there exists a decision function $\mathfrak{D}_0 = d_0(x)$ such that

$$r(\xi_0, \mathfrak{D}_0) \leqq v - \delta, \tag{2.79}$$

where $\delta > 0$ and v denotes the common value of $\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\eta} r$ and $\operatorname*{Inf}_{\eta} \operatorname*{Sup}_{\xi} r$. It was shown in the proof of Lemma 2.1 that (2.79) implies the existence of a decision function $\mathfrak{D}_1 = d_1(x)$ and that of a positive integer m such that

$$n(x; \mathfrak{D}_1) \leqq m \text{ for all } x \tag{2.80}$$

and

$$r(\xi_0, \mathfrak{D}_1) \leqq v - \frac{\delta}{2}. \tag{2.81}$$

Since $c(x^1, \cdots, x^n, \mathfrak{D}_1)$ and $W(F, d)$ are uniformly bounded and $W(F, d)$ is continuous in F uniformly in d, we have

$$\lim_{i=\infty} r \quad (F_i, \mathfrak{D}_1) = r(F, \mathfrak{D}_1) \tag{2.82}$$

for any sequence $\{F_i\}$ for which $F_i \to F$ in the ordinary sense. From (2.78), (2.82) and the compactness of Ω (in the ordinary sense) it follows that

$$\lim_{j=\infty} r \quad (\xi_{i_j}, \mathfrak{D}_1) = r(\xi_0, \mathfrak{D}_1) \leqq v - \frac{\delta}{2}. \tag{2.83}$$

But this is in contradiction to (2.77) and, therefore, ξ_0 must be a least favorable distribution. Hence, statement (i) of our theorem is proved.

Statement (ii) is an immediate consequence of Theorems (2.1), (2.3) and statement (i) of Theorem (2.6).

To prove (iii), replace the weight function $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \eta_0)$ where η_0 satisfies the conditions imposed on it in (iii).

We shall show that (i) remains valid also when $W(F, d)$ is replaced by $W^*(F, d)$. This is not clear, since $W^*(F, d)$ may not be continuous in F. First we shall prove that

$$\liminf_{i=\infty} r(\xi_i', \eta_0) \geqq r(\xi_0', \eta_0) \tag{2.84}$$

for any sequence $\{\xi_i'\}$ for which $\xi_i' \to \xi_0'$ in the ordinary sense, i.e., for which

$$\lim_{i=\infty} \xi_i'(\omega) = \xi_0'(\omega) \tag{2.85}$$

for any open subset ω (open in the sense of ordinary convergence defined in Ω) whose boundary has probability measure zero according to ξ_0'. For any sample $x^1, \cdots, x^r$ let $q_{ri}(x^1, \cdots, x^r)$ denote the probability that the first r observations will be equal to $x^1, \cdots, x^r$, respectively, when ξ_i' is the a priori distribution. Clearly,

$$q_{ri}(x^1, \cdots, x^r) = \int_\Omega p_r(x^1, \cdots, x^r \mid F)\, d\xi_i'. \tag{2.86}$$

Since $p_r(x^1, \cdots, x^r \mid F)$ is a continuous function of F, we have

$$\lim_{i=\infty} q_{ri}(x^1, \cdots, x^r) = q_{r0}(x^1, \cdots, x^r). \tag{2.87}$$

The function $r(\xi, \eta_0)$ can be split into two parts, i.e., $r(\xi, \eta_0) = r_1(\xi, \eta_0) + r_2(\xi, \eta_0)$ where r_1 is the expected value of the loss $W(F, d)$ and r_2 is the expected cost of experimentation. Since $W(F, d)$ is a bounded function of F and d, and since $W(F, d)$ is continuous in F uniformly in d, we have

$$\lim r_1(\xi_i', \eta_0) = r_1(\xi_0', \eta_0) \tag{2.88}$$

for any sequence $\{\xi_i'\}$ which satisfies (2.85). To prove (2.84), we merely have to show that

$$\liminf_{i=\infty} r_2(\xi_i', \eta_0) \geqq r_2(\xi_0', \eta_0). \tag{2.89}$$

But

$$r_2(\xi_i', \eta_0) = \sum_{r, x^1, \cdots, x^r} q_{ri}(x^1, \cdots, x^r) \int_{Q_{x^1 \cdots x^r}} c(x^1, \cdots, x^r; \mathfrak{D})\, d\eta_0 \tag{2.90}$$

where $Q_{x^1 \cdots x^r}$ is the totality of all decision functions $d(x)$ with the property that $d(y) = r$ for any y whose first r coordinates are equal to $x^1, \cdots, x^r$, respectively. Equation (2.89) is an immediate consequence of (2.87) and (2.90). Hence, (2.84) is proved.

Let $r^*(\xi, \eta)$ be the risk function when $W(F, d)$ is replaced by $W^*(F, d)$, i.e., $r^*(\xi, \eta) = r(\xi, \eta) - r(\xi, \eta_0)$. Let, furthermore, $\{\xi_i^*\}$ be a sequence of a priori distributions such that

$$\lim_{i=\infty} \operatorname{Inf}_{\eta} r^*(\xi_i^*, \eta) = \operatorname{Sup}_{\xi} \operatorname{Inf}_{\eta} r^*(\xi, \eta). \tag{2.91}$$

There exists a subsequent $\{\xi_{i_j}^*\}$ of the sequence $\{\xi_i^*\}$ such that $\xi_{i_j}^*$ converges (in the ordinary sense) to a limit distribution ξ_0^* as $j \to \infty$. We shall show that ξ_0^* is a least favorable distribution. For suppose that ξ_0^* is not a least favorable distribution. Then there exists a decision function $\mathfrak{D}_0^* = d_0^*(x)$ such that

$$r^*(\xi_0^*, \mathfrak{D}_0^*) \leqq v^* - \delta \tag{2.92}$$

where $\delta > 0$ and $v^* = \operatorname{Sup}_{\xi} \operatorname{Inf}_{\eta} r^* = \operatorname{Inf}_{\eta} \operatorname{Sup}_{\xi} r^*$. But then there exists a decision function $\mathfrak{D}_1^* = d_1^*(x)$ and a positive integer m such that

$$n(x; \mathfrak{D}_1^*) \leqq m \text{ for all } x \tag{2.93}$$

and

$$r^*(\xi_0^*, \mathfrak{D}_1^*) \leqq v^* - \frac{\delta}{2}. \tag{2.94}$$

Since $r^*(\xi, \mathfrak{D}_1^*) = r(\xi, \mathfrak{D}_1^*) - r(\xi, \eta_0)$, and since

$$\lim_{j=\infty} r(\xi_{i_j}^*, \mathfrak{D}_1^*) = r(\xi_0^*, \mathfrak{D}_1^*),$$

it follows from (2.84) and (2.94) that

$$\lim_{j=\infty} \sup r^*(\xi_{i_j}^*, \mathfrak{D}_1^*) \leqq v^* - \frac{\delta}{2} \tag{2.95}$$

which is in contradiction to (2.91). Hence, the validity of (i) is proved also when $W(F, d)$ is replaced by $W^*(F, d)$. Clearly, also (ii) remains valid when $W(F, d)$ is replaced by $W^*(F, d)$.

Let η_1 be a minimax solution relative to the problem corresponding to $W^*(F, d)$. Then because of (ii), η_1 is a Bayes solution in the strict sense. Since η_0 is not a Bayes solution in the strict sense, $\eta_1 \neq \eta_0$ and $v^* < 0$. Hence η_1 is uniformly better than η_0. This completes the proof of Theorem 2.6.

We shall now replace Condition 2.6 by the following weaker one.

CONDITION 2.6*. *There exists a sequence* $\{\Omega_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ *of subsets of* Ω *such that Condition* 2.6 *is fulfilled when* Ω *is replaced by* Ω_i, $\Omega_{i+1} \supset \Omega_i$ *and* $\lim_{i=\infty} \Omega_i = \Omega$.

We shall say that η_i converges weakly to η as $i \to \infty$, if $\lim_{i=\infty} \zeta(\eta_i) = \zeta(\eta)$. We shall also say that η is a weak limit of η_i. This limit definition seems to be natural, since $r(\xi, \eta_1) = r(\xi, \eta_2)$ if $\zeta(\eta_2) = \zeta(\eta_1)$. We shall now prove the following theorem:

THEOREM 2.7. *If D is compact and if Conditions* 2.1, 2.2, 2.4, 2.5 *and* 2.6* *are fulfilled, then:*

(i) *A minimax solution exists that is a weak limit of a sequence of Bayes solutions in the strict sense.*

(ii) *Let η_0 be a decision rule for which $r(F, \eta_0)$ is a bounded function of F. Then there exists a decision rule η_1 that is a weak limit of a sequence of Bayes solutions in the strict sense and such that $r(F, \eta_1) \leqq r(F, \eta_0)$ for all F in Ω.*

PROOF: According to theorem 2.6, there exists a decision rule η_i that is a Bayes solution in the strict sense and a minimax solution if Ω is replaced by Ω_i. There exists a subsequence $\{\eta_{i_j}\}$ ($j = 1, 2, \cdots$, ad inf.) of the sequence $\{\eta_i\}$ such that $\{\eta_{i_j}\}$ admits a weak limit. Let η_0 be a weak limit of $\{\eta_{i_j}\}$. Then, as shown in the proof of Lemma 2.4, equation (2.60) holds and η_0 is a minimax solution relative to the original space Ω. Thus, statement (i) is proved.

To prove (ii), replace $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \eta_0)$. According to Theorem 2.6 there exists a decision rule η_{1i} such that η_{1i} is a minimax solution and a Bayes solution in the strict sense when Ω is replaced by Ω_i and $W(F, d)$ by $W^*(F, d)$. Clearly, η_{1i} remains to be a Bayes solution in the strict sense also relative to Ω and $W(F, d)$. Since η_{1i} is a minimax solution relative to Ω_i and $W^*(F, d)$, we have

$$r(F, \eta_{1i}) \leqq r(F, \eta_0) \text{ for all } F \text{ in } \Omega_i. \tag{2.96}$$

Let $\{\eta_{1i_j}\}$ be a subsequence of the sequence $\{\eta_{1i}\}$ such that $\{\eta_{1i_j}\}$ admits a weak limit η_1. Then, (2.60) holds for $\{\eta_{1i_j}\}$ and η_1, and

$$r(F, \eta_1) \leqq r(F, \eta_0) \text{ for all } F \text{ in } \Omega. \tag{2.97}$$

Since η_1 is a weak limit of strict Bayes solution, statement (ii) is proved.

3. Statistical decision functions: the case of continuous chance variables.

3.1. *Introductory remarks.* In this section we shall be concerned with the case where the probability distribution F of X is absolutely continuous, i.e., for any element F of Ω and for any positive integer r there exists a joint density function $p_r(x^1, \cdots, x^r \mid F)$ of the first r chance variables $X^1, \cdots, X^r$.

The continuous case can immediately be reduced to the discrete case discussed in section 2 if the observations are not given exactly but only up to a finite number of decimal places. More precisely, we mean this: For each i, let the real axis R be subdivided into a denumerable number of disjoint sets $R_{i1}, R_{i2}, \cdots$, ad inf. Suppose that the observed value x^i of X^i is not given exactly; it is merely known which element of the sequence $\{R_{ij}\}$ ($j = 1, 2, \cdots$, ad inf.) contains x^i. This is the situation, for example, if the value of x^i is given merely up to a finite number, say r, decimal places (r fixed, independent of i). This case can be reduced to the previously discussed discrete case, since we can regard the sets R_{ij} as our points, i.e., we can replace the chance variable X^i by Y^i where Y^i can take only the values $R_{i1}, R_{i2}, \cdots$, ad inf. (Y^i takes the value R_{ij} if X^i falls in R_{ij}). If $W(F_1, d) = W(F_2, d)$ whenever the distribution of Y under

F_1 is identical with that under F_2, only the chance variables Y^1, Y^2, $\cdots$, etc. play a role in the decision problem and we have the discrete case. If, the latter condition on the weight function is not fulfilled, i.e., if there exists a pair (F_1, F_2) such that $W(F_1, d) \neq W(F_2, d)$ for some d and the distribution of Y is the same under F_1 as under F_2, we can still reduce the problem to the discrete case, if in the discrete case we permit the weight W to depend also on a third extraneous variable G, i.e., if we put $W = W(F, G, d)$, where G is a variable about whose value the sample does not give any information. The results obtained in the discrete case can easily be generalized to include the situation where $W = W(F, G, d)$.

In practical applications the observed value x^i of X^i will usually be given only up to a certain number of decimal places and, thus, the problem can be reduced to the discrete case. Nevertheless, it seems desirable from the theoretical point of view to develop the theory of the continuous case, assuming that the observed value x^i of X^i is given precisely.

In section 2.3 an alternative definition of a randomized decision rule was given in terms of two sequences of functions $\{z_r(x^1, \cdots, x^r)\}$ and $\{\delta_{x^1 \cdots x^r}\}$ $(r = 1, 2, \cdots$, ad. inf.). We used the symbol ζ to denote a randomized decision rule given by two such sequences. It was shown in the discrete case that the use of a randomized decision function η generates a certain $\zeta = \zeta(\eta)$, and that for any given ζ there exists an η such that $\zeta = \zeta(\eta)$. Furthermore, because of Condition 2.5, in the discrete case we had $r(F, \eta_1) = r(F, \eta_2)$ if $\zeta(\eta_1) = \zeta(\eta_2)$. It would be possible to develop a similar theory as to the relation between ζ and η also in the continuous case. However, a somewhat different procedure will be followed for the sake of simplicity. Instead of the decision functions $d(x)$, we shall regard the ζ' s as the pure strategies of the statistician, i.e., we replace the space Q of all decision functions $d(x)$ by the space Z of all randomized decisions rules ζ. It will then be necessary to consider probability measures η defined over an additive class of subsets of Z. It will be sufficient, as will be seen later, to consider only discrete probability measures η. A probability measure η is said to be discrete, if it assigns the probability 1 to some denumerable subset of Z. Any discrete η will clearly generate a certain $\zeta = \zeta(\eta)$. In the next section we shall formulate some conditions which will imply that $r(F, \eta_1) = r(F, \eta_2)$ if $\zeta(\eta_1) = \zeta(\eta_2)$. Thus, it will be possible to restrict ourselves to consideration of pure strategies ζ which will cause considerable simplifications.

The definitions of various notions given in the discrete case, such as minimax solution, Bayes solution, a priori distribution ξ in Ω, least favorable a priori distribution, complete class of decision functions, etc. can immediately be extended to the continuous case and will, therefore, not be restated here.

3.2 *Conditions on Ω, D, $W(F, d)$ and the cost function.* In this section we shall formulate conditions similar to those given in the discrete case.

CONDITION 3.1. *Each element F of Ω is absolutely continuous.*

CONDITION 3.2. *$W(F, d)$ is a bounded function of F and d.*

CONDITION 3.3. *The space D is compact in the sense of its intrinsic metric $\delta(d_1, d_2; \Omega)$ (see equation 2.2).*

This condition is somewhat stronger than the corresponding Condition 2.3. While it may be possible to weaken this condition, it would make the proofs of certain theorems considerably more involved.

CONDITION 3.4. *The cost of experimentation* $c(x^1, \cdots, x^m)$ *does not depend on* ζ. *It is non-negative and* $\lim_{m=\infty} c(x^1, \cdots, x^m) = \infty$ *uniformly in* $x^1, \cdots, x^m$. *For each positive integral value* m, $c(x^1, \cdots, x^m)$ *is a bounded function of* $x^1, \cdots, x^m$.

This condition is stronger than Conditions 2.4 and 2.5 postulated in the discrete case. The reason for formulating a stronger condition here is that we wish the relation $r(F, \eta_1) = r(F, \eta_2)$ to be fulfilled whenever $\zeta(\eta_1) = \zeta(\eta_2)$ which will make it possible for us to eliminate the consideration of η's altogether. Since the ζ's are regarded here as the pure strategies of the statistician, it is not clear what kind of dependence of the cost on ζ would be consistent with the requirement that $r(F, \eta_1) = r(F, \eta_2)$ whenever $\zeta(\eta_1) = \zeta(\eta_2)$.

We shall say that $F_1 \rightarrow F$ in the ordinary sense, if for any positive integral value m

$$\lim_{i=\infty} \int_{S_m} p_m(x^1, \cdots, x^m \mid F_i)\, dx^1 \cdots dx^m = \int_{S_m} p_m(x^1, \cdots, x^m \mid F)\, dx^1 \cdots dx^m$$

uniformly in S_m where S_m is a subset of the m-dimensional sample space.

CONDITION 3.5. *The space* Ω *is separable in the sense of the above convergence definition.*[6]

No such condition was formulated in the discrete case for the simple reason that in the discrete case Ω is always separable in the sense of the convergence definition given in (2.76).

3.3. *Some lemmas.* We shall first give a convergence definition in the space Z of all ζ's which is somewhat different from the one given in the discrete case. Let $h_r(x^1, x^2, \cdots, x^r, D^*)$ denote the probability that experimentation will be terminated with the rth observation and that the final decision d selected will be an element of D^*, knowing that the first r observations are equal to $x^1, \cdots, x^r$, respectively. That is,

$$\begin{aligned} h_r(x^1, \cdots, x^r, D^*) &= z_1(x^1) z_2(x^1, x^2) \\ &\quad \cdots z_{r-1}(x^1, \cdots, x^{r-1})(1 - z_r(x^1, \cdots, x^r))\delta_{x^1\cdots x^r}(D^*). \end{aligned} \tag{3.1}$$

Clearly, the functions $h_r(x^1, \cdots, x^r, D^*)$ are non-negative and satisfy the following conditions:

$$\sum_{r=1}^{m} h_r(x^1, \cdots, x^r, D^*) \leqq 1 \text{ for any } D^* \text{ and for any sample } (x^1, \cdots, x^m). \tag{3.2}$$

$$\sum_{j=1}^{\infty} h_r(x^1, \cdots, x^r, D_j^*) = h_r(x^1, \cdots, x^r, D^*), \tag{3.3}$$

if $\sum_{j=1}^{\infty} D_j^* = D^*$ and $D_1^*, D_2^*, \cdots$, etc. are disjoint.

[6] For a definition of a separable space, see F. Hausdorff, *Mengenlehre* (3rd edition), p. 125.

One can easily verify that for any sequence of non-negative functions $\{h_r(x^1, \cdots, x^r, D^*)\}$ $(r = 1, 2, \cdots)$ satisfying (3.2) and (3.3) there exists exactly one sequence $\{z_r(x^1, \cdots, x^r)\}$ and one sequence $\{\delta_{x^1\cdots x^r}(D^*)\}$ such that (3.1) is fulfilled. Thus, a randomized decision rule ζ can be given by a sequence $\{h_r(x^1, \cdots, x^r, D^*)\}$ satisfying (3.2) and (3.3). The functions $z_r(x^1, \cdots, x^r)$ and $\delta_{x^1\cdots x^r}$ need be defined only for samples $x^1, \cdots, x^r$ for which $z_i(x^1, \cdots, x^i) > 0$ for $i = 1, \cdots, r - 1$. The above mentioned uniqueness of $z_r(x^1, \cdots, x^r)$ and $\delta_{x^1\cdots x^r}$ was meant to hold if the definition of these functions is restricted to such samples $x^1, \cdots, x^r$.

For any bounded subset S_r of the r-dimensional sample space, let

$$H_r(S_r, D^*) = \int_{S_r} h_r(x^1, \cdots, x^r, D^*)\, dx^1 \cdots dx^r. \tag{3.4}$$

Let $\{\zeta_i\}(i = 0, 1, 2, \cdots, \text{ad inf.})$ be a sequence of decision rules, and $H_{r,i}(S_r, D^*)$ be the function $H_r(S_r, D^*)$ corresponding to ζ_i. We shall say that

$$\lim_{i=\infty} \zeta_i = \zeta_0 \tag{3.5}$$

if

$$\lim_{i=\infty} H_{r,i}(S_r, D^*) = H_{r,0}(S_r, D^*) \tag{3.6}$$

for any r, any bounded set S_r and for any D^* that is an element of a sequence $\{D_{k_1,\cdots k_l}\}$ $(k_j = 1, \cdots, r_j; j = 1, \cdots, l; l = 1, 2, \cdots, \text{ad inf.})$ of subsets of D satisfying the following conditions:

$$\sum_{k_1=1}^{r_1} D_{k_1} = D; \sum_{k_l} D_{k_1\cdots k_l} = D_{k_1k_2\cdots k_{l-1}}, \tag{3.7}$$

$$D_{k_1\cdots k_{l-1}1}, \cdots, D_{k_1\cdots k_{l-1}r_l} \text{ are disjoint,} \tag{3.8}$$

and

(3.9) the diameter of $D_{k_1\cdots k_l}$ converges to zero as $l \to \infty$ uniformly in $k_1, \cdots, k_l$.

LEMMA 3.1. *For any sequence* $\{\zeta_i\}(i = 1, 2, \cdots, \textit{ad inf.})$ *of decision rules there exists a subsequence* $\{\zeta_{i_j}\}$ $(j = 1, 2, \cdots, \textit{ad inf.})$ *and a decision rule* ζ_0 *such that* $\lim_{j=\infty} \zeta_{i_j} = \zeta_0$.

PROOF: Let $H_{r,i}(S_r, D^*)$ $(r = 1, 2, \cdots, \text{ad inf.})$ be the sequence of functions associated with ζ_i. Let, furthermore, $\{D^*_{k_1\cdots k_l}\}$ be a sequence of subsets of D satisfying the relations (3.7), (3.8) and (3.9). Clearly, for any fixed r and any fixed element $D^*_{k_1\cdots k_l}$ of the sequence $\{D^*_{k_1\cdots k_l}\}$, it is possible to find a subsequence $\{i_j\}$ $(j = 1, 2, \cdots, \text{ad inf.})$ of the sequence $\{i\}$ (the subsequence $\{i_j\}$ may depend on r and $D^*_{k_1\cdots k_l}$) and a set function $H_{r,0}(S_r)$ such that

$$\lim_{j=\infty} H_{r,i_j}(S_r, D^*_{k_1\cdots k_l}) = H_{r,0}(S_r). \tag{3.10}$$

Using the well known diagonal procedure, it is therefore possible to find a fixed subsequence $\{i_j\}$ (independent of r and D^*) and a sequence of set functions $\{H_{r,0}(S_r, D^*_{k_1 \cdots k_l})\}$ such that

$$\lim_{j=\infty} H_{r,i_j}(S_r, D^*_{k_1 \cdots k_l}) = H_{r,0}(S_r, D^*_{k_1 \cdots k_l}) \tag{3.11}$$

for all values of $r, k_1, \cdots, k_l$ and l.

To complete the proof of Lemma 3.1, it remains to be shown that there exists a decision rule ζ_0 such that the associated function $H_r(S_r, D^*)$ is equal to $H_{r,0}(S_r, D^*)$ for any D^* that is an element of $\{D^*_{k_1 \cdots k_l}\}$. Since $h_{r,i}(x^1, \cdots, x^r, D^*)$ is uniformly bounded, the set function $H_{r,0}(S_r, D^*_{k_1 \cdots k_l})$ is absolutely continuous. Hence for any values of $k_1, \cdots, k_l$ there exists a function $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ such that

$$\int_{S_r} h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l}) dx^1 \cdots dx^r = H_{r,0}(S_r, D^*_{k_1 \cdots k_l}). \tag{3.12}$$

The existence of a ζ_0 with the desired property is proved, if we show that the functions $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ satisfy the relations (3.2) and (3.3). Let $h_r(x^1, \cdots, x^m, D^*) = h_r(x^1, \cdots, x^r, D^*)$ for any $m > r$. Then, since the functions $h_{r,i}$ satisfy (3.2), we have

$$\sum_{r=1}^{m} H_{r,i}(S_m, D^*) \leqq V(S_m) \tag{3.13}$$

where $V(S_m)$ denotes the m-dimensional Lebesgue measure of S_m. From (3.13) it follows that

$$\sum_{r=1}^{m} H_{r,0}(S_m, D^*_{k_1 \cdots k_l}) \leqq V(S_m). \tag{3.14}$$

Hence, the functions $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ must satisfy (3.2) except perhaps on a set of Lebesgue measure zero. Since the functions $h_{r,i}(x^1, \cdots, x^r, D^*)$ satisfy (3.3), we must have

$$H_{r,i}(S_r, D^*_{k_1 \cdots k_{l-1}}) = \sum_{k_l=1}^{r_l} H_{r,0}(S_r, D^*_{k_1 \cdots k_l}). \tag{3.15}$$

Hence, the same relation must hold also for $H_{r,0}(S_r, D^*_{k_1 \cdots k_l})$. But this implies that the functions $h_{r,0}(x^1, \cdots, x^r, D^*_{k_1 \cdots k_l})$ satisfy (3.3) except perhaps on a set of Lebesgue measure zero, and the proof of Lemma 3.1 is completed.

LEMMA 3.2. *Let $T_i(S)$ ($i = 0, 1, 2, \cdots$) be a non-negative, completely additive set function defined for all measurable subsets S of the r-dimensional sample space M_r. Assume that*

$$T_i(S) \leqq V(S) \tag{3.16}$$

for all S ($i = 0, 1, 2, \cdots$, ad inf.) where $V(S)$ denotes the Lebesgue measure of S. Let, furthermore, $g(x^1, \cdots, x^r)$ be a non-negative function such that

$$\int_{M_r} g(x^1, \cdots, x^r)\, dx^1 \cdots dx^r < \infty. \tag{3.17}$$

If

(3.18) $$\lim_{i=\infty} T_i(S) = T_0(S)$$

then

(3.19) $$\lim_{i=\infty} \int_{M_r} g(x^1, \cdots, x^r)\, dT_i = \int_{M_r} g(x^1, \cdots, x^r)\, dT_0 .$$

PROOF: Let $M_{r,c}$ be the sphere in M_r with center at the origin and radius c. Clearly,

(3.21) $$\lim_{c=\infty} \int_{M_{r,c}} g(x^1, \cdots, x^r)\, dx^1 \cdots dx^r = \int_{M_r} g(x^1, \cdots, x^r)\, dx^1 \cdots dx^r .$$

Hence, because of (3.16), we have

(3.21) $$\lim_{c=\infty} \left[\int_{M_{r,c}} g(x^1, \cdots, x^r)\, dT_i - \int_{M_r} g(x^1, \cdots, x^r)\, dT_i \right] = 0$$

uniformly in i. Hence our lemma is proved if we show that

(3.22) $$\lim_{i=\infty} \int_{M_{r,c}} g(x^1, \cdots, x^r)\, dT_i = \int_{M_{r,c}} g(x^1, \cdots, x^r)\, dT_0$$

for any finite c. Let $g_A(x^1, \cdots, x^r) = g(x^1, \cdots, x^r)$ when $g(x^1, \cdots, x^r) \leqq A$, and $= 0$ otherwise. Since

$$\lim_{A=\infty} \int_{M_{r,c}} (g - g_A)\, dx^1 \cdots dx^r = 0$$

it follows from (3.16) that

(3.23) $$\lim_{A=\infty} \int_{M_{r,c}} (g - g_A)\, dT_i = 0$$

uniformly in i. Hence, our lemma is proved if we can show that

(3.24) $$\lim_{i=\infty} \int_{M_{r,c}} g_A\, dT_i = \int_{M_{r,c}} g_A\, dT_0$$

for any $c > 0$ and any $A > 0$. Let S_j denote the set of all points in $M_{r,c}$ for which

(3.25) $$(j - 1)\, \epsilon \leqq g_A < j\, \epsilon$$

where ϵ is a given positive number. We have

(3.26) $$\sum_j (j - 1)\, \epsilon \int_{S_j} dT_i \leqq \int_{M_{r,c}} g_A\, dT_i \leqq \sum_j j\, \epsilon \int_{S_j} dT_i, \quad (i = 0, 1, 2, \cdots).$$

Since for any ϵ, j can take only a finite number of values, and since ϵ can be chosen arbitrarily small, our lemma follows easily from (3.18) and (3.26).

LEMMA 3.3. *Let* $\{\zeta_1\}$ *be a sequence of decision rules such that* $\lim_{i=\infty} \zeta_i = \zeta_0$ *and*

$r(F, \zeta_i)$ is a bounded function of F and i ($i \geqq 1$). Then

$$\liminf_{i=\infty} r(F, \zeta_i) \geqq r(F, \zeta_0). \tag{3.27}$$

PROOF: First we shall show that it is sufficient to prove Lemma 3.3 for any finite space D. For this purpose, assume that Lemma 3.3 is true for any finite decision space, but there exists a non-finite compact decision space D and a sequence $\{\zeta_i\}$ such that $\lim_{i=\infty} \zeta_i = \zeta_0$ and

$$\liminf_{i=\infty} r(F, \zeta_i) = r(F, \zeta_0) - \delta \text{ for some } F (\delta > 0). \tag{3.28}$$

Since $\zeta_i \rightarrow \zeta_0$, there exists a sequence $\{D_{k_1 \cdots k_l}\}$ of subsets of D satisfying the conditions (3.7)–(3.9) and such that

$$\lim_{i=\infty} H_{r,i}(S_r, D^*_{k_1 \cdots k_l}) = H_{r,0}(S_r, D^*_{k_1 \cdots k_l}) \tag{3.29}$$

where $H_{r,i}(S_r, D^*)$ is the function H_r associated with $\zeta_i (i = 0, 1, 2, \cdots)$. Let λ be a fixed value of l and consider the corresponding finite sequence $\{D_{k_1 \cdots k_\lambda}\}$ of subsets of D. Let k be the number of elements in this finite sequence. We select one point from each element of the finite sequence $\{D_{k_1 \cdots k_\lambda}\}$. Let the points selected be $d_1, d_2, \cdots, d_k$ and let $\bar{D}$ denote the set consisting of the points $d_1, \cdots, d_k$. Let $\bar{\zeta}_i$ be the decision rule defined as follows: the function $h_r(x^1, \cdots, x^r, d_j)$ associated with $\bar{\zeta}_i$ is equal to $h_{r,i}(x^1, \cdots, x^r, D_j^*)$ where D_j^* is equal to the element of the finite sequence $\{D_{k_1 \cdots k_\lambda}\}$ which contains the point $d_j (j = 1, \cdots, k)$. Clearly, because of (3.29),

$$\lim_{i=\infty} \bar{\zeta}_i = \bar{\zeta}_0 . \tag{3.30}$$

Furthermore, for sufficiently large λ we obviously have

$$| r(F, \zeta_i) - r(F, \bar{\zeta}_i) | \leqq \epsilon \text{ for } i = 0, 1, 2, \cdots, \text{ ad inf.} \tag{3.31}$$

Since for finite D our lemma is assumed to be true, we have

$$\liminf_{i=\infty} r(F, \bar{\zeta}_i) \geqq r(F, \bar{\zeta}_0). \tag{3.32}$$

Choosing $\epsilon \leqq \frac{\delta}{3}$, we obtain a contradiction from (3.28), (3.31) and (3.32). Thus, it is sufficient to prove Lemma 3.3 for finite D. In the remainder of the proof we shall assume that D consists of the points $d_1, \cdots, d_k$.

The probability that we shall take exactly m observations when ζ_i is used and F is true is given by

$$\text{prob. } \{n = m \mid F, \zeta_i\} = \int_{M_m} p_m(x^1, \cdots, x^m \mid F) h_{m,i}(x^1, \cdots, x^m, D) dx^1, \cdots dx^m \tag{3.33}$$

where M_m denotes the m-dimensional sample space. Since

$$\lim_{i=\infty} H_{m,i}(S_m, D) = H_{m,0}(S_m, D),$$

it follows from Lemma 3.2 that

$$\lim_{i=\infty} \text{prob}\,\{n = m \mid F, \zeta_i\} = \text{prob}\,\{n = m \mid F, \zeta_0\}. \tag{3.34}$$

Hence

$$\lim_{i=\infty} \text{prob}\,\{n \leqq m \mid F, \zeta_i\} = \text{prob}\,\{n \leqq m \mid F, \zeta_0\}. \tag{3.35}$$

Since $r(F, \zeta_i)$ is a bounded function of F and i ($i \geqq 1$), we must have

$$\lim_{m=\infty} \text{prob}\,\{n \leqq m \mid F, \zeta_i\} = 1 \qquad (i = 1, 2, \cdots) \tag{3.36}$$

uniformly in F and i. From (3.35) and (3.36) it follows that

$$\lim_{m=\infty} \text{prob}\,\{n \leqq m \mid F, \zeta_0\} = 1 \tag{3.37}$$

uniformly in F. Because of (3.36) and (3.37), we have

$$r(F, \zeta_i) = \sum_{m=1}^{\infty} r_m(F, \zeta_i) \qquad (i = 0, 1, 2, \cdots, \text{ad inf.}), \tag{3.38}$$

where

$$\begin{aligned} r_m(F, \zeta_i) = \sum_{l=1}^{k} \int_{M_m} & p_m(x^1, \cdots, x^m \mid F) W(F, d_l)\, dH_{m,i}(S_m, d_l) \\ & + \int_{M_m} p_m(x^1, \cdots, x^m \mid F) c(x^1, \cdots, x^m)\, dH_{m,i}(S_m, D). \end{aligned} \tag{3.39}$$

Since

$$\lim_{i=\infty} H_{m,i}(S_m, D^*) = H_{m,0}(S_m, D^*)$$

for any subset D^* of D, it follows from Lemma 3.2 that

$$\lim_{i=\infty} r_m(F, \zeta_i) = r_m(F, \zeta_0). \tag{3.40}$$

Lemma 3.3 is an immediate consequence of (3.38) and (3.40).

3.4. *Equality of* Sup Inf r *and* Inf Sup r, *and other theorems.* In this section we shall prove the main theorems for the continuous case, using the lemmas derived in the preceding section.

THEOREM 3.1. *If Conditions* 3.1–3.5 *are fulfilled, then*

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta} r(\xi, \zeta) = \operatorname*{Inf}_{\zeta} \operatorname*{Sup}_{\xi} r(\xi, \zeta). \tag{3.41}$$

PROOF: Let Z^m denote the class of all ζ's for which prob $\{n \leqq m \mid \zeta, F\} = 1$ for all F. We shall denote an element of Z^m by ζ^m. First we shall show that it is sufficient if for any finite m we can prove Theorem 3.1 under the restriction that ζ must be an element of Z^m. For this purpose, put $W_0 = \operatorname{Sup}_{F,d} W(F, d)$ and choose a positive integer m_ϵ so that

$$c(x^1, \cdots, x^m) > \frac{W_0^2}{\epsilon} \tag{3.42}$$

for all $m \geqq m_\epsilon$. The existence of such a value m_ϵ follows from Condition 3.4. We shall now show that for any ξ we have

$$\operatorname*{Inf}_{\zeta^m} r(\xi, \zeta^m) \leqq \operatorname*{Inf}_{\zeta} r(\xi, \zeta) + \epsilon \text{ for any } m \geqq m_\epsilon. \tag{3.43}$$

Let ζ_1 be any decision rule. There are two cases to be considered: (a) prob $\{n \geqq m_\epsilon \mid \xi, \zeta_1\} \geqq \frac{\epsilon}{W_0}$; (b) prob $\{n \geqq m_\epsilon \mid \xi, \zeta_1\} < \frac{\epsilon}{W_0}$. In case (a) we have $r(\xi, \zeta_1) \geqq W_0$. In this case, let ζ_2 be the rule that we decide for some d without taking any observations. Clearly, we shall have $r(\xi, \zeta_2) \leqq W_0$ and, therefore, $r(\xi, \zeta_2) \leqq r(\xi, \zeta_1)$. In case (b), let ζ_2 be defined as follows: $h_r(x^1, \cdots, x^r, D^*)$ for ζ_2 is the same as that for ζ_1 when $r < m_\epsilon$, and $h_r(x^1, \cdots, x^r, d_0)$ for ζ_2 is equal to $1 - \sum_{k=1}^{m_\epsilon - 1} h_k(x^1, \cdots, x^k, D)$ when $r = m_\epsilon$, and zero when $r > m_\epsilon$ where d_0 is a fixed element of D. Since prob $\{n \geqq m_\epsilon \mid \xi, \zeta_1\} < \frac{\epsilon}{W_0}$, we have

$$r(\xi, \zeta_2) \leqq r(\xi, \zeta_1) + \epsilon.$$

In both cases ζ_2 is an element of Z^{m_ϵ}. Hence (3.43) is proved. From (3.43) we obtain

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta^{m_\epsilon}} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta} r + \epsilon. \tag{3.44}$$

Assume now that

$$\operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta^m} r = \operatorname*{Inf}_{\zeta^m} \operatorname*{Sup}_{\xi} r \tag{3.45}$$

holds for any m. From (3.44) and (3.45) we obtain

$$\operatorname*{Inf}_{\zeta^{m_\epsilon}} \operatorname*{Sup}_{\xi} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta} r + \epsilon. \tag{3.46}$$

Hence

$$\operatorname*{Inf}_{\zeta} \operatorname*{Sup}_{\xi} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta} r + \epsilon. \tag{3.47}$$

Since this is true for any ϵ, we have

$$\operatorname*{Inf}_{\zeta} \operatorname*{Sup}_{\xi} r \leqq \operatorname*{Sup}_{\xi} \operatorname*{Inf}_{\zeta} r. \tag{3.48}$$

Theorem 3.1 follows from (3.48) and Lemma 1.3.

To complete the proof of Theorem 3.1, it remains to be shown that (3.45) holds for any m. Since D is compact, (3.45) is proved if we can prove it for any finite D. In the remainder of the proof we shall, therefore assume that D consists of k points $d_1, \cdots, d_k$. Let ω be a subset of Ω that is conditionally compact in the sense of the metric[7]

$$\delta_0(F_1, F_2) = \operatorname*{Sup}_{S_m} \left| \int_{S_m} dF_1 - \int_{S_m} dF_2 \right| \tag{3.49}$$

where S_m is a subset of the m-dimensional sample space. We shall show that ω is conditionally compact also in the sense of the intrinsic metric given by

$$\delta_1(F_1, F_2) = \operatorname*{Sup}_{\zeta^m} | r(F_1, \zeta^m) - r(F_2, \zeta^m) | . \tag{3.50}$$

Let

$$\delta_2(F_1, F_2) = \operatorname*{Sup}_{d} | W(F_1, d) - W(F_2, d) | \tag{3.51}$$

and

$$\delta_3(F_1, F_2) = \delta_0(F_1, F_2) + \delta_2(F_1, F_2). \tag{3.52}$$

It follows from Condition 3.3 and Theorem 3.1 in [3] that Ω, and therefore also ω, is conditionally compact in the sense of the metric $\delta_2(F_1, F_2)$. Hence ω is conditionally compact in the sense of the metric $\delta_3(F_1, F_2)$. The conditional compactness of ω relative to the metric $\delta_1(F_1, F_2)$ is proved, if we can show that any sequence $\{F_i\}$ that is a Cauchy sequence relative to the metric δ_3 is a Cauchy sequence also relative to the metric δ_1. Let $\{F_i\}$ ($i = 1, 2, \cdots$, ad inf.) be a Cauchy sequence relative to δ_3. Then there exists a distribution F_0 (not necessarily an element of Ω) and a function $W(d)$ such that

$$\lim_{i=\infty} W(F_i, d) = W(d) \text{ uniformly in } d \tag{3.53}$$

and

$$\lim_{i=\infty} \int_{S_m} dF_i = \int_{S_m} dF_0 \tag{3.54}$$

uniformly in S_m. We have

$$\begin{aligned} r(F_i, \zeta^m) = {} & \sum_{r=1}^{m} \sum_{j=1}^{k} \int_{M_r} \\ & \cdot p(x^1, \cdots, x^r \mid F_i) W(F_i, d_j) h_r(x^1, \cdots, x^r, d_j) \, dx^1 \cdots dx^r \\ & + \sum_{r=1}^{m} \int_{M_r} c\,(x^1, \cdots, x^r) p(x^1, \cdots, x^r \mid F_i) h_r(x^1, \cdots, x^r, D) \, dx^1 \cdots dx^r, \end{aligned} \tag{3.55}$$

[7] By $\int_{S_m} dF$ we mean $\int_{S_m} p_m(x^1, \cdots, x^m \mid F) dx^1 \cdots dx^m$.

where M_r denotes the r-dimensional sample space. The sequence $\{F_i\}$ is a Cauchy sequence relative to the metric δ_1 if there exists a function $r(\zeta^m)$ such that

$$\lim_{i=\infty} r(F_i, \zeta^n) = r(\zeta^m) \tag{3.56}$$

uniformly in ζ^m. Let $\bar{r}(F_i, \zeta^m)$ be the function we obtain from $r(F_i, \zeta^m)$ by replacing the factor $W(F_i, d_j)$ by $W(d_j)$ under the first integral on the right hand side of (3.55). Because of (3.53), we have

$$\lim_{i=\infty} [r(F_i, \zeta^m) - \bar{r}(F_i, \zeta^m)] = 0 \tag{3.57}$$

uniformly in ζ^m. Thus, (3.56) is proved if we can show the existence of a function $\bar{r}(\zeta^m)$ such that

$$\lim_{i=\infty} \bar{r}(F_i, \zeta^m) = \bar{r}(\zeta^m) \tag{3.58}$$

uniformly in ζ^m. Let C be a class of functions $\varphi(x^1, \cdots, x^m)$ such that

$$|\varphi(x^1, \cdots, x^m)| < A < \infty \text{ for all } \varphi \text{ in } C.$$

It then follows from (3.54) that there exists a functional $g(\varphi)$ such that

$$\lim_{i} \int_{M_m} \varphi\, dF_i = g(\varphi) \tag{3.59}$$

uniformly in φ. Application of this general result yields (3.58) immediately. Hence, $\{F_i\}$ is a Cauchy sequence relative to the metric δ_1 and, therefore ω is shown to be conditionally compact relative to the metric δ_1 if it is relative to the metric δ_0.

It then follows from Theorem 3.2 in [3] that $\operatorname{Sup}_{\xi} \operatorname{Inf}_{\zeta^m} r = \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi} r$ if we replace Ω by a subset ω that is conditionally compact relative to δ_0.[8] Since Ω is separable relative to δ_0, there exists a sequence $\{\Omega_i\}$ of subsets of Ω such that Ω_i is conditionally compact relative to δ_0, $\Omega_{i+1} \supset \Omega_i$ and $\sum_{i=\infty} \Omega_i = \Omega^*$ is dense in Ω. Let ξ^i denote an a priori distribution ξ for which $\xi(\Omega_i) = 1$. Since the left and right hand members in (3.45) remain unchanged when Ω is replaced by Ω^*, it follows from Theorem 1.3 that equation (3.45) is proved if we can show that

$$\lim_{i=\infty} \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi^i} r = \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi} r. \tag{3.60}$$

Let $\{\zeta_i^m\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of decision rules such that

$$\lim_{i=\infty} [\operatorname{Sup}_{\xi^i} r(\xi^i, \zeta_i^m) - \operatorname{Inf}_{\zeta^m} \operatorname{Sup}_{\xi^i} r] = 0. \tag{3.61}$$

---

[8] Strictly, we would have to write $\operatorname{Inf}_{\eta^m}$ instead of $\operatorname{Inf}_{\zeta^m}$ where η^m is a probability measure in the space of all ζ^m. But, since the use of any discrete probability measure is equivalent to the use of a ζ^m, and since the restriction to discrete η^m does not change $\operatorname{Sup}_{\xi} \operatorname{Inf}_{\eta^m} r$ or $\operatorname{Inf}_{\eta^m} \operatorname{Sup}_{\xi} r$ we can replace $\operatorname{Inf}_{\eta^m}$ by $\operatorname{Inf}_{\zeta^m}$.

According to Lemmas 3.1 and 3.3, there exists a subsequence $\{i_j\}$ of $\{i\}$ and a decision rule ζ_0^m such that

$$(3.62) \qquad \liminf_{j=\infty} r(F, \zeta_{i_j}^m) \geqq r(F, \zeta_0^m) \text{ for all } F.$$

Since Ω^* is dense in Ω, it follows from (3.61) and (3.62) that

$$(3.63) \qquad \operatorname*{Sup}_F r(F, \zeta_0^m) \leqq \lim_{i=\infty} \operatorname*{Inf}_{\zeta^m} \operatorname*{Sup}_{\xi^i} r$$

and, therefore, 3.60 holds. Thus, (3.45) is proved and the proof of Theorem 3.1 is completed.

THEOREM 3.2. *If Conditions 3.1–3.5 are fulfilled, then there exists a minimax solution, i.e., a decision rule ζ_0 for which*

$$(3.64) \qquad \operatorname*{Sup}_F r(F, \zeta_0) \leqq \operatorname*{Sup}_F r(F, \zeta) \text{ for all } \zeta.$$

PROOF: Because of Theorem 3.1 there exists a sequence $\{\zeta_i\}$ ($i = 1, 2, \cdots$, ad inf.) of decision rules such that

$$(3.65) \qquad \lim_{i=\infty} \operatorname*{Sup}_F r(F, \zeta_i) = \operatorname*{Inf}_{\zeta} \operatorname*{Sup}_F r(F, \zeta).$$

According to Lemmas 3.1 and 3.3 there exists a subsequence $\{\zeta_{i_j}\}$ of $\{\zeta_i\}$ and a decision rule ζ_0 such that

$$3.66 \qquad \liminf_{j=\infty} r(F, \zeta_{i_j}) \geqq r(F, \zeta_0) \text{ for all } F.$$

It follows from (3.65) and (3.66) that ζ_0 is a minimax solution and Theorem 3.2 is proved.

THEOREM 3.3. *If Conditions 3.1–3.5 are fulfilled, then for any ξ there exists a Bayes solution relative to ξ.*

This theorem is an immediate consequence of Lemmas 3.1 and 3.3.

THEOREM 3.4. *If Conditions 3.1–3.5 are fulfilled, then the class of all Bayes solutions in the wide sense is a complete class.*

The proof is omitted, since it is entirely analogous to that of Theorem 2.5.

3.5. *Formulation of an additional condition.* In this section we shall formulate an additional condition which will permit the derivation of some stronger theorems. Let the metric $\delta_0(F_1, F_2)$ be defined by

$$\delta_0(F_1, F_2) = \sum_{m=1}^{\infty} \frac{1}{m^2} \operatorname*{Sup}_{S_m} \left| \int_{S_m} dF_1 - \int_{S_m} dF_2 \right|$$

where S_m may be any subset of the m-dimensional sample space.

CONDITION 3.6. *The space Ω is compact relative to the metric $\delta_0(F_1, F_2)$*

$$\lim_i W(F_i, d) = W(F_0, d)$$

uniformly in d if $\lim_i \delta_0(F_i, F_0) = 0$.

THEOREM 3.5. *If Conditions 3.1–3.6 hold, then*

(i) *there exists a least favorable a priori distribution*
(ii) *any minimax solution is a Bayes solution in the strict sense*
(iii) *for any decision rule ζ_0 which is not a Bayes solution in the strict sense and for which $r(F, \zeta_0)$ is a bounded function of F there exists a decision rule ζ_1 which is a Bayes solution in the strict sense and is uniformly better than ζ_0.*

PROOF: The proofs of (i) and (ii) are entirely analogous to those of (i) and (ii) in Theorem 2.6, and will therefore be omitted here.

To prove (iii), let ζ_0 be a decision rule that is not a Bayes solution in the strict sense and for which $r(F, \zeta_0)$ is bounded. We replace the weight function $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \zeta_0)$. We shall show that (i) remains valid when $W(F, d)$ is replaced by $W^*(F, d)$. This is not obvious, since $r(F, \zeta_0)$, and therefore also $W^*(F, d)$ may not be continuous in F. First we shall prove that

$$\liminf_{i=\infty} r(\xi_i', \zeta_0) \geqq r(\xi_0', \zeta_0) \tag{3.67}$$

for any sequence $\{\xi_i'\}$ for which

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

for any open subset ω of Ω (in the sense of the metric δ_0) whose boundary has probability measure zero according to ξ_0'. Let $r_m(F, \zeta)$ denote the conditional expected value of the loss $W(F, d)$ plus the cost of experimentation when $n = m$, F is true and the rule ζ is used by the statistician (see equation (3.39)). Since $W(F, d)$ and the cost of experimentation when m observations are taken are uniformly bounded, one can easily verify that

$$\lim_{i=\infty} r_m(F_i, \zeta_0) = r_m(F_0, \zeta_0) \tag{3.68}$$

for any sequence $\{F_i\}$ for which

$$\lim_{i=\infty} \delta_0(F_i, F_0) = 0. \tag{3.69}$$

Hence, since Ω is compact (Condition 3.6),

$$\lim_{i=\infty} r_m(\xi_i', \zeta_0) = r_m(\xi_0', \zeta_0) \tag{3.70}$$

where

$$r_m(\xi, \zeta_0) = \int_\Omega r_m(F, \zeta_0)\, d\xi. \tag{3.71}$$

Since

$$r(\xi, \zeta_0) = \sum_{m=1}^{\infty} r_m(\xi, \zeta_0)$$

inequality (3.67) follows from (3.70).

The remainder of the proof of (iii) will be omitted here, since it is the same as that of (iii) in Theorem 2.6.

We shall now replace Condition 3.6 by the following weaker one.

CONDITION 3.6*. *There exists a sequence* $\{\Omega_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ *of subsets of* Ω *such that Condition* 3.6 *is fulfilled when* Ω *is replaced by* Ω_i, $\Omega_{i+1} \supset \Omega_i$ *and* $\lim_{i=\infty} \Omega_i = \Omega$.

THEOREM 3.6. *If Conditions* 3.1–3.5 *and* 3.6* *are fulfilled then*

(i) *A minimax solution* ζ_0 *and a sequence* $\{\zeta_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ *exist such that* $\lim_{i=\infty} \zeta_i = \zeta_0$ *and* ζ_i $(i = 1, 2, \cdots, ad\ inf.)$ *is a Bayes solution in the strict sense.*

(ii) *For any decision rule* ζ_0 *for which* $r(F, \zeta_0)$ *is bounded there exists another decision rule* ζ_1 *such that* ζ_1 *is a limit of a sequence of Bayes solutions in the strict sense and* $r(F, \zeta_1) \leqq r(F, \zeta_0)$ *for all* F *in* Ω.

PROOF: According to Theorem 3.5, for each i there exists a decision rule ζ_i $(i = 1, 2, \cdots, \text{ad inf.})$ such that ζ_i is a minimax solution and a Bayes solution in the strict sense when Ω is replaced by Ω_i. Let $\{\zeta_{i_j}\}$ be a subsequence of the sequence $\{\zeta_i\}$ such that $\{\zeta_{i_j}\}$ admits a limit ζ_0, i.e., $\lim_{j=\infty} \zeta_{i_j} = \zeta_0$. Because of Lemma 3.3,

$$\liminf_{j=\infty} r(F, \zeta_{i_j}) \geqq r(F, \zeta_0). \tag{3.72}$$

Hence ζ_0 is a minimax solution relative to the original space Ω and statement (i) is proved.

To prove (ii), replace $W(F, d)$ by $W^*(F, d) = W(F, d) - r(F, \zeta_0)$ where ζ_0 is a decision rule for which $r(F, \zeta_0)$ is bounded. In proving statement (iii) of Theorem 3.5, we have shown that there exists a decision rule $\zeta_{1i}(i = 1, 2, \cdots,$ ad inf.) such that ζ_{1i} is a minimax solution and a Bayes solution in the strict sense when Ω is replaced by Ω_i and $W(F, d)$ by $W^*(F, d)$. Clearly, ζ_{1i} remains to be a Bayes solution in the strict sense also relative to Ω and $W(F, d)$. Since ζ_{1i} is a minimax solution relative to Ω_i and $W^*(F, d)$, we have

$$r(F, \zeta_{1i}) \leqq r(F, \zeta_0) \text{ for all } F \text{ in } \Omega_i. \tag{3.73}$$

Let $\{\zeta_{1i_j}\}$ be a convergent subsequence of $\{\zeta_{1i}\}$ and let $\lim_{j=\infty} \zeta_{1i_j} = \zeta_1$. Then, because of Lemma 3.3, we have

$$r(F, \zeta_1) \leqq r(F, \zeta_0) \text{ for all } F \text{ in } \Omega.$$

Since ζ_1 is a limit of a sequence of Bayes solutions in the strict sense, statement (ii) is proved.

Addition at proof reading. After this paper was sent to the printer the author found that Ω is always separable (in the sense of the convergence definition in Condition 3.5) and, therefore, Condition 3.5 is unnecessary. A proof of the separability of Ω will appear in a forthcoming publication of the author.

The boundedness of $r(F, \zeta_i)$ is not necessary for the validity of Lemma 3.3. Let $\lim_{i=\infty} \zeta_i = \zeta_0$ and suppose that for some F, say F_0, $r(F_0, \zeta_i)$ is not bounded in i. If $\liminf_{i=\infty} r(F_0, \zeta_i) = \infty$, Lemma 3.3 obviously holds for $F = F_0$. If

$\liminf_{i=\infty} r(F_0, \zeta_i) = g < \infty$, let $\{i_j\}$ be a subsequence of $\{i\}$ such that $\lim_{j=\infty} r(F_0, \zeta_{i_j}) = g$. Since $r(F_0, \zeta_{i_j})$ is a bounded function of j, Lemma 3.3 is applicable and we obtain $g \geqq r(F_0, \zeta_0)$. In a similar way, one can see that also Lemma 2.4 remains valid without assuming the boundedness of $r(F, \eta_i)$.

Although not stated explicitly, several functions considered in this paper are assumed to be measurable with respect to certain additive classes of subsets. In the continuous case, for example, the precise measurability assumptions may be stated as follows: Let B be the class of all Borel subsets of the infinite dimensional sample space M. Let H be the smallest additive class of subsets of Ω which contains any subset of Ω which is open in the sense of at least one of the convergence definitions considered in this paper. Let T be the smallest additive class of subsets of D which contains all open subsets of D (in the sense of the metric $\delta(d_1, d_2, \Omega)$). By the symbolic product $H \times T$ we mean the smallest additive class of subsets of the Cartesian product $\Omega \times D$ which contains the Cartesian product of any member of H by any member of T. The symbolic product $H \times B$ is similarly defined. It is assumed that: (1) $W(F, d)$ is measurable $(H \times T)$; (2) $p_m(x^1, \cdots, x^m \mid F)$ is measurable $(B \times H)$; (3) $\delta_{x^1 \ldots x^r}(D^*)$ is measurable (B) for any member D^* of T; (4) $z_r(x^1, \cdots, x^r)$ and $c_r(x^1, \cdots, x^r)$ are measurable (B). These assumptions are sufficient to insure the measurability (H) of $r(F, \zeta)$ for any ζ.

REFERENCES

[1] J. v. Neumann and Oskar Morganstein, *Theory of Games and Eonomic Behavior*, Princeton University Press, 1944.

]2] A. Wald, "Generalizatoin of a theorem by v. Neumann concerning zero sum two-person games," *Annals of Mathematics*, Vol. 46 (April, 1945).

[3] A. Wald, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (October, 1947).