ELIMINATION OF RANDOMIZATION IN CERTAIN STATISTICAL DECISION PROCEDURES AND ZERO-SUM TWO-PERSON GAMES[1]

By A. Dvoretzky,[2] A. Wald,[3] and J. Wolfowitz[3]

Institute for Numerical Analysis and Columbia University

Summary. The general existence of minimax strategies and other important properties proved in the theory of statistical decision functions (e.g., [3]) and the theory of games (e.g., [5]) depends upon the convexity of the space of decision functions and the convexity of the space of strategies. This convexity can be obtained by the use of randomized decision functions and mixed (randomized) strategies. In Section 2 of the present paper the authors state the extension (first announced in [1]) of a measure theoretical result known as Lyapunov's theorem [2]. This result is applied in Section 3 to the statistical decision problem where the number of distributions and decisions is finite. It is proved that when the distributions are continuous (more generally, "atomless," see footnote 7 below) randomization is unnecessary in the sense that every randomized decision function can be replaced by an equivalent nonrandomized decision function. Section 4 extends this result to the case when the decision space is compact. Section 5 extends the results of Section 3 to the sequential case. Sections 6 and 7 show, by counterexamples, that the results of Section 3 cannot be extended to the case of infinitely many distributions without new restrictions.[4] Section 8 gives sufficient conditions for the elimination of randomization under maintenance of ϵ-equivalence. Section 9 concludes with a restatement of the results in the language of the theory of games.

1. Introduction.

We shall consider the following statistical decision problem: Let x be the generic point in an n-dimensional Euclidean[5] space R, and let Ω be a given class of cumulative distribution functions $F(x)$ in R. The cumulative distribution function $F(x)$ of the vector chance variable $X = (X_1, \cdots, X_n)$ with range in R is not known. It is known, however, that F is an element of the given class Ω. There is also given a space D whose elements d represent the possible decisions that can be made by the statistician in the problem under consideration. Let $W(F, d, x)$ denote the "loss" when F is the true distribution of

[1] The main results of this paper were announced without proof in an earlier publication [1] of the authors.

[2] On leave of absence from the Hebrew University, Jerusalem, Israel.

[3] Research under a contract with the Office of Naval Research.

[4] The impossibility of such an extension is related to the failure of Lyapunov's theorem when infinitely many measures are considered. (cf. A. Lyapunov, "Sur les fonctions-vecteurs complètement additives," *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, Vol. 10 (1946), pp. 277–279.)

[5] The restriction to a Euclidean space is not essential (see [1]).

X, the decision d is made and x is the observed value of X. We shall define the distance between two elements d_1 and d_2 of D by

$$\rho(d_1, d_2) = \operatorname*{Sup}_{F,x} | W(F, d_1, x) - W(F, d_2, x) |. \tag{1.1}$$

Let B be the smallest Borel field of subsets of D which contains all open subsets of D as elements. Let B_0 be the totality of Borel sets of R. We shall assume that $W(F, d, x)$ is bounded[6] and, for every F, a function of d and x which is measurable $(B \times B_0)$. By a decision function $\delta(x)$ we mean a function which associates with each x a probability measure on D defined for all elements of B. We shall occasionally use the symbol δ_x instead of $\delta(x)$ when we want to emphasize that x is kept fixed. A decision function $\delta(x)$ is said to be nonrandomized if for every x the probability measure $\delta(x)$ assigns the probability one to a single point d of D. For any measurable subset D^* of D (D^* an element of B), the symbol $\delta(D^* \mid x)$ will denote the probability measure of D^* according to the set function $\delta(x)$. It will be assumed throughout this paper that for any given D^* the function $\delta(D^* \mid x)$ is a Borel measurable function of x. The adoption of a decision function $\delta(x)$ by the statistician means that he proceeds according to the following rule: Let x be the observed value of X. The element d of the space D is selected by an independent chance mechanism constructed in such a way that for any measurable subset D^* of D the probability that the selected element d will be included in D^* is equal to $\delta(D^* \mid x)$.

Given the sample point x and given that $\delta(x)$ is the decision function adopted, the expected value of the loss $W(F, d, x)$ is given by

$$W^*(F, \delta, x) = \int_D W(F, d, x)\, d\delta_x. \tag{1.2}$$

The expected value of the loss $W(F, d, x)$ when F is the true distribution of X and $\delta(x)$ is the decision function adopted (but x is not known) is obviously equal to

$$r(F, \delta) = \int_R W^*(F, \delta, x)\, dF(x). \tag{1.3}$$

The above expression is called the risk when F is true and δ is adopted.

We shall say that the decision functions $\delta(x)$ and $\delta^*(x)$ are equivalent if

$$r(F, \delta^*) = r(F, \delta) \qquad \text{for all } F \text{ in } \Omega. \tag{1.4}$$

We shall say that $\delta(x)$ and $\delta^*(x)$ are strongly equivalent if for every measurable subset D^* of D we have

$$\int_R \delta(D^* \mid x)\, dF(x) = \int_R \delta^*(D^* \mid x)\, dF(x) \qquad \text{for all } F \text{ in } \Omega. \tag{1.5}$$

[6] The restriction of boundedness is not essential (see [1]).

If δ and δ^* are strongly equivalent, they are equivalent for any loss function which is a function of F and d only.

For any positive ϵ, we shall say that $\delta(x)$ and $\delta^*(x)$ are ϵ-equivalent if

$$(1.6) \qquad | r(F, \delta) - r(F, \delta^*) | \leqq \epsilon \qquad \text{for all } F \text{ in } \Omega,$$

and strongly ϵ-equivalent if

$$(1.7) \qquad \left| \int_R \delta(D^* | x) \, dF(x) - \int_R \delta^*(D^* | x) \, dF(x) \right| \leqq \epsilon$$

for all measurable D^* and for all F in Ω.

In Section 2 we state a measure-theoretical result first announced in [1] and proved in [6]. This result is then used in Section 3 to prove that for every decision function there exists an equivalent, as well as a strongly equivalent, nonrandomized decision function δ^*, if Ω and D are finite and if each element $F(x)$ of Ω is atomless.[7] This result is extended in Section 4 to the case where D is compact. Section 5 deals with the sequential case for which similar results are proved. A precise definition of a sequential decision function is given in Section 5.

The finiteness of Ω is essential for the validity of the results given in Sections 2–5. The examples given in Section 6 show that even when Ω is such a simple class as the class of all univariate normal distributions with unit variance, there exist decision functions δ such that no equivalent nonrandomized decision functions exist. In Section 7, an example is given where a decision function δ and a positive ϵ exist such that no nonrandomized decision function δ^* is ϵ-equivalent to δ.

In Section 8, sufficient conditions are given which guarantee that for every δ and for every $\epsilon > 0$ there exists a nonrandomized decision function δ^* which is ϵ-equivalent to δ.

2. A measure-theoretical result. Let $\{y\} = Y$ be any space and let $\{S\} = \mathfrak{S}$ be a Borel field of subsets of Y. Let $\mu_k(S) (k = 1, \cdots, q)$ be a finite number of real-valued, σ-finite and countably additive set functions defined for all $S \in \mathfrak{S}$. The following theorem was stated by the authors [1]:

THEOREM 2.1. *Let $\delta_j(y)$ $(j = 1, 2, \cdots, m)$ be real non-negative $\mathfrak{S}$-measurable functions satisfying*

$$(2.1) \qquad \sum_{j=1}^{m} \delta_j(y) = 1$$

for all $y \in Y$. Then if the set functions $\mu_k(S)$ are atomless there exists a decomposition of Y into m disjoint subsets $S_1, \cdots, S_m$ belonging to $\mathfrak{S}$ having the property

[7] A set function μ defined on a Borel field $\mathfrak{S}$ is called atomless if it has the following property: If for some $S \in \mathfrak{S}$, $\mu(S) \neq 0$, then there exists an $S' \subset S$ such that $S' \in \mathfrak{S}$ and such that $\mu(S') \neq \mu(S)$ and $\mu(S') \neq 0$. A cumulative distribution function is called atomless if its associated set function is atomless.

that

$$\int_Y \delta_j(y)\, d\mu_k(y) = \mu_k(S_j) \qquad (j = 1, \cdots, m; k = 1, \cdots, q). \tag{2.2}$$

If $\delta_j^*(y) = 1$ *for all* $y \in S_j$ *and* $= 0$ *for any other* $y (j = 1, \cdots, m)$, *then the above equation can be written as*

$$\int_Y \delta_j(y)\, d\mu_k(y) = \int_Y \delta_j^*(y)\, d\mu_k(y) \qquad (j = 1, \cdots, m; k = 1, \cdots, q). \tag{2.3}$$

This theorem is an extension of a result of A. Lyapunov [2] and is basic for deriving most of the results of the present paper.

3. Elimination of randomization when Ω and D are finite and each element $F(x)$ of Ω is atomless. In this section we shall assume that Ω consists of the elements $F_1(x), \cdots, F_p(x)$ and D of the elements $d_1, \cdots, d_m$. Moreover, we assume that $F_i(x)$ is atomless for $i = 1, \cdots, p$. A decision function $\delta(x)$ is now given by a vector function $\delta(x) = [\delta_1(x), \cdots, \delta_m(x)]$ such that

$$\delta_j(x) \geqq 0, \qquad \sum_{j=1}^{m} \delta_j(x) = 1 \tag{3.1}$$

for all $x \in R$. Here $\delta_j(x)$ is the probability that the decision d_j will be made when x is the observed value of X. The risk when F_i is true and the decision function $\delta(x)$ is adopted is now given by

$$r(F_i, \delta) = \sum_{j=1}^{m} \int_R W(F_i, d_j, x)\delta_j(x)\, dF_i(x). \tag{3.2}$$

A nonrandomized decision function $\delta^*(x)$ is a vector function whose components $\delta_j^*(x)$ can take only the values 0 and 1 for all x.

For any measurable subset S of R let

$$\nu_{ij}(S) = \int_S W(F_i, d_j, x)\, dF_i(x) \qquad (i = 1, \cdots p; j = 1, \cdots, m). \tag{3.3}$$

Then the measures $\nu_{ij}(S)$ are finite, atomless and countably additive. Using these set functions, equation (3.2) can be written as

$$r(F_i, \delta) = \sum_{j=1}^{m} \int_R \delta_j(x)\, d\nu_{ij}(x). \tag{3.4}$$

Replacing in Theorem 2.1 the space Y by R, the set of measures $\{\mu_1, \cdots, \mu_q\}$ by the set $\{\nu_{ij}\} (i = 1, \cdots, p; j = 1, \cdots, m)$, it follows from Theorem 2.1 that there exists a nonrandomized decision function $\delta^*(x)$ such that

$$\int_R \delta_j(x)\, d\nu_{ij}(x) = \int_R \delta_j^*(x)\, d\nu_{ij}(x) \qquad (i = 1, \cdots, p; j = 1, \cdots, m). \tag{3.5}$$

This immediately yields the following theorems:

THEOREM 3.1. *If Ω and D are finite and if each element $F(x)$ of Ω is atomless, then for any decision function $\delta(x)$ there exists an equivalent nonrandomized decision function $\delta^*(x)$.*

Putting $W(F, d, x) = 1$ identically in F, d and x, equation (3.5) immediately yields the following theorem:

THEOREM 3.2. *If Ω and D are finite and if each element $F(x)$ of Ω is atomless, then for any decision function $\delta(x)$ there exists a strongly equivalent nonrandomized decision function $\delta^*(x)$.*

4. Elimination of randomization when Ω is finite, D is compact and each element $F(x)$ of Ω is atomless. Again, let $\Omega = \{F_1, \cdots, F_p\}$ where the distributions F are atomless. If the loss $W(F, d, x)$ does not depend on x, the finiteness of Ω implies that D is at least conditionally compact with respect to the metric (1.1) (see Theorem 3.1 in [3]). We postulate that D is compact (but permit the loss to depend on x), and shall prove that if $\delta(x)$ is any decision function, there exists a nonrandomized decision function $\delta^*(x)$ such that $\delta^*(x)$ is equivalent to $\delta(x)$, i.e.,

$$r_i(\delta) = r_i(\delta^*) \qquad (i = 1, \cdots, p), \tag{4.1}$$

where $r_i(\delta)$ stands for $r(F_i, \delta)$.

Since D is compact there exists an infinite sequence of decompositions of the space D into a finite number of disjoint nonempty measurable sets, the l^{th} decomposition to be $C(1, 1, \cdots, 1), \cdots, C(k_1, \cdots, k_l)$ with the properties:

(a) Any two sets C which have the same number of indices not all identical, are disjoint.

(b) The sum of all sets with the same number l of indices is D ($l = 1, 2, \cdots$ ad inf.).

(c) If the sequence of indices of one set C constitutes a proper initial part of the sequence of indices of another set C, the first set includes the second.

(d) The diameters of all sets with l indices are bounded above by $h(l)$ and

$$\lim_{l \to \infty} h(l) = 0.$$

Let l be fixed and define

$$\Delta_{m_1, \cdots, m_l}(x) = \delta[C(m_1, \cdots, m_l \mid x]. \tag{4.2}$$

Define, furthermore,

$$\begin{aligned} W_i[x, C(m_1, \cdots, m_l)] &= \frac{1}{\Delta_{m_1 \cdots m_l}(x)} \int_{C(m_1, \cdots, m_l)} W(F_i, d, x)\, d\delta_x \\ &\qquad \text{if } \Delta_{m_1 \cdots m_l}(x) > 0, \\ &= 0 \qquad \text{if } \Delta_{m_1 \cdots m_l}(x) = 0. \end{aligned} \tag{4.3}$$

Clearly,

$$(4.4) \qquad r_i(\delta) = \sum_{m_l=1}^{k_l} \cdots \sum_{m_1=1}^{k_1} \int_R W_i[x, C(m_1, \cdots, m_l)]\Delta_{m_1\cdots m_l}(x)\, dF_i(x).$$

Considering a decision space D_l with elements $d_{m_1\cdots m_l}$ $(m_i = 1, \cdots, k_i;$ $i = 1, \cdots, l)$ and putting the loss $W(F_i, d_{m_1\cdots m_l}, x) = W_i[x, C(m_1, \cdots, m_l)]$, equations (3.3) and (3.5) imply that there exists a finite sequence of measurable functions $\bar{\Delta}_{m_1\cdots m_l}(x)$ $(m_1 = 1, \cdots, k_1; \cdots; m_l = 1, \cdots, k_l)$ such that

$$(4.5) \qquad \bar{\Delta}_{m_1\cdots m_l}(x) = 0 \text{ or } 1 \qquad \text{for all } x,$$

$$(4.6) \qquad \sum_{m_l} \cdots \sum_{m_1} \bar{\Delta}_{m_1\cdots m_l}(x) = 1 \qquad \text{for all } x,$$

$$(4.7) \qquad \bar{\Delta}_{m_1\cdots m_l}(x) = 0 \qquad \text{whenever } \Delta_{m_1\cdots m_l}(x) = 0,$$

and

$$(4.8) \qquad \begin{aligned} &\int_R W_i[x, C(m_1, \cdots, m_l)]\bar{\Delta}_{m_1\cdots m_l}(x)\, dF_i(x) \\ &\qquad = \int_R W_i[x, C(m_1, \cdots, m_l)\Delta_{m_1\cdots m_l}(x)\, dF_i(x). \end{aligned}$$

Let now $\bar{\delta}(x)$ be the decision function for which

$$(4.9) \qquad \bar{\delta}[C(m_1, \cdots, m_l) \mid x] = \bar{\Delta}_{m_1\cdots m_l}(x)$$

and for any measurable subset $D_{m_1\cdots m_l}$ of $C(m_1, \cdots, m_l)$

$$(4.10) \qquad \bar{\delta}[D_{m_1\cdots m_l} \mid x]\bar{\Delta}_{m_1\cdots m_l}(x) = \frac{\delta(D_{m_1\cdots m_l} \mid x)}{\delta[C(m_1, \cdots, m_l) \mid x]},$$

where $\dfrac{\delta(D_{m_1\cdots m_l} \mid x)}{\delta[C(m_1, \cdots, m_l) \mid x]}$ is defined to be $= 0$ when $\delta[C(m_1, \cdots, m_l) \mid x] = 0$.

It then follows from (4.4) and (4.8) that

$$(4.11) \qquad r_i(\delta) = r_i(\bar{\delta}).$$

Applying the above result for $l = 1$, we conclude that there exists a decision function $\delta^1(x)$ with the following properties: The choice among the C's with one index is nonrandom. The decision, once given the C (with one index) chosen, is made according to $\delta(x)$. We have $\delta^1[C(m_1) \mid x] = 0$ whenever $\delta[C(m_1) \mid x] = 0$ and

$$r_i(\delta) = r_i(\delta^1) \qquad (i = 1, \cdots, p).$$

Repeat the above procedure for every C with two indices, using $W_i\{x, C(m_1, m_2)\}$ as weight function and $\delta^1(x)$ as the decision function. We conclude that there exists a decision function $\delta^2(x)$ with the following properties: The choice among the C's with two indices is nonrandom. $\delta^2[C(m_1, m_2) \mid x] = 0$ whenever $\delta^1[C(m_1, m_2) \mid x] = 0$. The decision, once given the C (with two indices) chosen, is made according to $\delta^1(x)$ and, therefore, in accordance with $\delta(x)$. We have

$$\int_R \int_{C(m_1)} W(F_i, d, x)\, d\delta_x^1\, dF_i(x) = \int_R \int_{C(m_1)} W(F_i, d, x)\, d\delta_x^2\, dF_i(x) \quad \begin{matrix}(m_1 = 1, 2, \cdots, k_1) \\ (i = 1, \cdots, p).\end{matrix}$$

Repeat the above procedure for all C's with l indices, $l = 3, 4, \cdots$ ad inf. At the l^{th} stage we obtain a decision function $\delta^l(x)$ with the following properties: The decision among the C's with l indices is nonrandom. $\delta^l[C(m_1, \cdots, m_l) \mid x] = 0$ whenever $\delta^{l-1}[C(m_1, \cdots, m_l) \mid x] = 0$. The decision, once given the chosen C with l indices, is made according to $\delta(x)$. We have

$$\int_R \int_{C(m_1, \cdots, m_{l-1})} W(F_i, d, x)\, d\delta_x^{l-1}\, dF_i(x) = \int_R \int_{C(m_1, \cdots, m_{l-1})} W(F_i, d, x)\, d\delta_x^{l}\, dF_i(x)$$

$$\begin{pmatrix} i = 1, \cdots, p \\ m_1 = 1, \cdots, k_1 \\ m_{l-1} = 1, \cdots, k_{l-1} \end{pmatrix}.$$

Hold x fixed and let $C(x; l)$ be that C with l indices for which

$$\int_{C(x;l)} d\delta_x^l = 1.$$

Then $C(x; l + 1)$ is a proper subset of $C(x; l)$ for every positive l. The sequence $C(x; l)$, $l = 1, 2, \cdots$, determines, because D is compact, a unique limit point $c(x)$ such that any neighborhood of $c(x)$ contains almost all sets $C(x; l)$. Hence the sequence of probability measures $\delta_x^l (l = 1, 2, \cdots$, ad inf.) converges to a limit probability measure δ_x^* which assigns probability one to any measurable set which contains the point $c(x)$. Since $W(F_i, d, x)$ is continuous in d, we have

$$\lim_{l=\infty} \int_D W(F_i, d, x)\, d\delta_x^l = \int_D W(F_i, d, x)\, d\delta_x^* \tag{4.12}$$

for any x.

Now let x vary over R. It follows from (4.12) and the boundedness of $W(F, d, x)$ that $\lim_{l=\infty} r_i(\delta^l) = r_i(\delta^*)$. Since $r_i(\delta^l) = r_i(\delta)$, also $r_i(\delta^*) = r_i(\delta)$ $(i = 1, \cdots, p)$. Thus the probability measures $\delta^*(x)$ constitute the desired nonrandomized decision function.

It remains to show that for any measurable subset D^* of D, the function $\delta^*(D^* \mid x)$ is a measurable function of x. The measurability of $\delta^*(D^* \mid x)$ can easily be shown for any D^*, if it is shown for all closed sets D^*, since every measurable set can be attained by a denumerable number of Borel operations (denumerably infinite sums and complements) starting with closed sets. Thus we shall assume that D^* is closed. For any positive ρ let D_ρ^* be the sum of all open spheres with center in D^* and radius ρ. It is easy to see that

$$\delta^*(D_{2\rho}^* \mid x) \geqq \liminf_{l=\infty} \delta^l(D_\rho^* \mid x) \geqq \delta^*(D^* \mid x).$$

Since $\lim_{\rho=0} \delta^*(D_{2\rho}^* \mid x) = \delta^*(D^* \mid x)$, it follows from the above relation that

$$\lim_{\rho=0} \liminf_{l} \delta^l(D_\rho^* \mid x) = \delta^*(D^* \mid x).$$

Since $\delta^l(D_\rho^* \mid x)$ is a measurable function of x, the measurability of $\delta^*(D^* \mid x)$ is proved.

5. Elimination of randomization in the sequential case. In this section we shall consider the following sequential decision problem: Let $X = \{X_n\}$ $(n = 1, 2, \cdots, \text{ad inf.})$ be a sequence of chance variables. Let x be the generic point in the space $\bar{R}$ of all infinite sequences of real numbers, i.e., $x = \{x_n\}$ $(n = 1, 2, \cdots, \text{ad inf.})$ where each x_n is a real number. It is known that the distribution function $F(x)$ of X is an element of Ω, where Ω consists of a finite number of distribution functions $F_1(x), \cdots, F_p(x)$, and that the distribution function of X_1 is continuous according to $F_i(x)$, $i = 1, \cdots, p$. The statistician is assumed to have a choice of a finite number of (terminal) decisions $d_1, \cdots, d_m$, i.e., the space D consists of the elements $d_1, d_2, \cdots, d_m$. A decision rule δ is now given by a sequence of nonnegative, measurable functions $\delta_{\nu t}(x_1, \cdots, x_t)$ $(\nu = 0, 1, \cdots, m; t = 1, 2, \cdots, \text{ad inf.})$ satisfying

$$\sum_{\nu=0}^{m} \delta_{\nu t}(x_1, \cdots, x_t) = 1 \tag{5.1}$$

for $-\infty < x_1, \cdots, x_t < \infty$. The decision rule δ is defined in terms of the functions $\delta_{\nu t}$ as follows: After the value x_1 of X_1 has been observed, the statistician decides either to continue experimentation and take another observation, or to stop further experimentation and adopt a terminal decision $d_j (j = 1, \cdots, m)$ with the respective probabilities $\delta_{01}(x_1)$ and $\delta_{j1}(x_1)$ $(j = 1, \cdots, m)$. If it is decided to continue experimentation, a value x_2 of X_2 is observed and it is again decided either to take a further observation or adopt a terminal decision $d_j (j = 1, \cdots, m)$ with the respective probabilities $\delta_{02}(x_1, x_2)$ and $\delta_{j2}(x_1, x_2) (j = 1, \cdots, m)$, etc. The decision rule is called nonrandomized if each $\delta_{\nu t}$ can take only the values 0 and 1.

Let $v_{i\nu t}(x_1, \cdots, x_t)$ represent the sum of the loss and the cost of experimentation when F_i is true, the terminal decision d_ν is made and experimentation is terminated with the t^{th} observation

$$(\nu = 1, 2, \cdots, m; i = 1, \cdots, p; t = 1, 2, \cdots, \text{ad inf.}).$$

The functions $v_{i\nu t}(x_1, \cdots, x_t)$ are assumed to be finite, nonnegative and measurable. We shall consider only decision rules δ for which the probability is one that experimentation will be terminated at some finite stage. The risk (expected loss plus expected cost of experimentation) when F_i is true and the rule δ is adopted is then given by

$$r_i(\delta) = \sum_{t=1}^{\infty}\sum_{\nu=1}^{m}\int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01}(x_1)\delta_{02}(x_1, x_2) \cdots \delta_{0(t-1)}(x_1, \cdots, x_{t-1}) \cdot \delta_{\nu t}(x_1, \cdots, x_t)\, dF_{it}(x_1, \cdots, x_t), \tag{5.2}$$

where R_t is the t-dimensional space of $x_1, \cdots, x_t$ and $F_{it}(x_1, \cdots, x_t)$ is the cumulative distribution function of $X_1, \cdots, X_t$ when F_i is the distribution function of X.

We shall say that the decision rules δ^1 and δ^2 are equivalent if $r_i(\delta^1) = r_i(\delta^2)$ for $i = 1, \cdots, p$. We shall say that δ^1 and δ^2 are strongly equivalent if

$$\begin{aligned}&\int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta^1_{01}(x_1) \cdots \delta^1_{0(t-1)}(x_1, \cdots, x_{t-1})\delta^1_{\nu t}(x_1, \cdots, x_t)\, dF_{it}\\ &\quad = \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta^2_{01}(x_1) \cdots \delta^2_{0(t-1)}(x_1, \cdots, x_{t-1})\delta^2_{\nu t}(x_1, \cdots, x_t)\, dF_{it}\end{aligned} \tag{5.3}$$

for $i = 1, 2, \cdots, p$; $\nu = 1, \cdots, m$ and $t = 1, 2, \cdots$, ad inf.

Clearly, if δ^1 and δ^2 are strongly equivalent and if the functions $v_{i\nu t}(x_1, \cdots, x_t)$ reduce to constants $v_{i\nu t}$, then δ^1 and δ^2 are equivalent for all possible choices of the constants $v_{i\nu t}$.

Let

$$\varphi_i(x, \delta) = \sum_{t=1}^{\infty}\sum_{\nu=1}^{m} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01}(x_1) \cdots \delta_{0(t-1)}(x_1, \cdots, x_{t-1})\delta_{\nu t}(x_1, \cdots, x_t). \tag{5.4}$$

We shall prove the following lemma:

LEMMA 5.1. *Let δ be a decision rule for which $\varphi_i(x, \delta) < \infty$ for all x, except perhaps on a set of x's whose probability is zero according to every distribution function $F_i(x)(i = 1, \cdots, p)$. Let τ and T be given positive integers. Then there exists a decision function $\bar\delta$ with the following properties:*

$$\bar\delta_{\nu\tau}(x_1, \cdots, x_\tau) = 0 \quad\text{or}\quad 1, \qquad \sum_{\nu=0}^{m}\bar\delta_{\nu\tau}(x_1, \cdots, x_\tau) = 1, \tag{5.5}$$

for every point in $R_\tau(\nu = 0, 1, \cdots, m)$,

$$\bar\delta_{\nu t}(x_1, \cdots, x_t) = \delta_{\nu t}(x_1, \cdots, x_t) \qquad (\nu = 0, 1, \cdots, m; t \neq \tau), \tag{5.6}$$

$$r_i(\delta) = r_i(\bar\delta) \qquad (i = 1, \cdots, p), \tag{5.7}$$

$$\int_{R_t} v_{i\nu t}\delta_{01} \cdots \delta_{0(t-1)}\delta_{\nu t}\, dF_{it} = \int_{R_t} v_{i\nu t}\bar\delta_{01} \cdots \bar\delta_{0(t-1)}\bar\delta_{\nu t}\, dF_{it} \qquad (\nu = 1, \cdots, m; t = 1, \cdots, T), \tag{5.8}$$

$$\varphi_i(x, \bar\delta) < \infty, \tag{5.9}$$

for all x except perhaps on a set whose probability is zero according to every distribution $F_i(x)(i = 1, \cdots, p)$.

Proof. We can write $\varphi_i(x, \delta)$ as follows:

$$\begin{aligned}\varphi_i(x, \delta) = &\sum_{t=1}^{\tau-1} \sum_{\nu=1}^{m} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01} \cdots \delta_{0(t-1)}\delta_{\nu t} \\ &+ \sum_{t=\tau}^{\infty} \sum_{\nu=0}^{m} g_{i\nu\tau t}(x_1, \cdots, x_t)\delta_{\nu\tau},\end{aligned} \tag{5.10}$$

where $g_{i\nu\tau t}(x_1, \cdots, x_t)$ does not depend on $\delta_{0\tau}, \delta_{1\tau}, \cdots, \delta_{m\tau}$. The first double sum reduces to zero when $\tau = 1$. Clearly, if a $\bar{\delta}$ with the desired properties exists, then

$$\begin{aligned}\varphi_i(x, \bar{\delta}) = &\sum_{t=1}^{\tau-1} \sum_{\nu=1}^{m} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01} \cdots \delta_{0(t-1)}\delta_{\nu t} \\ &+ \sum_{t=\tau}^{\infty} \sum_{\nu=0}^{m} g_{i\nu\tau t}(x_1, \cdots, x_t)\bar{\delta}_{\nu\tau}.\end{aligned} \tag{5.11}$$

For any subset S of R, let

$$\mu_{i\nu\tau t}(S) = \int_S g_{i\nu\tau t}(x_1, \cdots, x_t)\, dF_i \qquad (t = \tau, \tau + 1, \cdots, T), \tag{5.12}$$

and

$$\mu_{i\nu\tau}(S) = \int_S \left[\sum_{t=T+1}^{\infty} g_{i\nu\tau t}(x_1, \cdots x_t) \right] dF_i. \tag{5.13}$$

The measures $\mu_{i\nu\tau t}$ are not defined if $\tau > T$. Clearly, the measures

$$\mu_{i\nu\tau t}(\nu = 0, 1, \cdots, m; t = \tau, \tau + 1, \cdots, T)$$

and the measures $\mu_{i\nu\tau}(\nu = 1, \cdots, m)$ are nonnegative, countably additive and σ-finite. Since for any x for which $\varphi_i(x, \delta) < \infty$ and $\delta_{0\tau} > 0$, the sum

$$\sum_{t=T+1}^{\infty} g_{i0\tau t}(x_1, \cdots, x_t) < \infty,$$

it follows from the assumptions of Lemma 5.1 that $\mu_{i0\tau}$ is σ-finite over the space R' consisting of all x for which $\delta_{0\tau} > 0$. Of course, $\mu_{i0\tau}$ is nonnegative and countably additive. Let R'' be the set of all points x for which $\delta_{0\tau} = 0$. We put

$$\bar{\delta}_{0\tau}(x_1, \cdots, x_\tau) = 0 \quad \text{for all} \quad x \text{ in } R''. \tag{5.14}$$

Application of Theorem 2.1 to each of the spaces R' and R'' shows that there exist measurable functions $\bar{\delta}_{\nu\tau}(x_1, \cdots, x_\tau)(\nu = 0, 1, \cdots, m)$ such that in addiion to (5.14) the following conditions hold:

$$\bar{\delta}_{\nu\tau} = 0 \quad \text{or} \quad 1(\nu = 0, 1, \cdots m) \quad \text{and} \quad \sum_{\nu=0}^{m} \bar{\delta}_{\nu\tau} = 1 \quad \text{for all } x, \tag{5.15}$$

$$\int_R \delta_{\nu\tau}\, d\mu_{i\nu\tau t} = \int_R \bar{\delta}_{\nu\tau}\, d\mu_{i\nu\tau t} \tag{5.16}$$

$$(i = 1, \cdots, p; \nu = 0, 1, \cdots m; t = \tau, \tau + 1, \cdots, T),$$

$$\int_R \delta_{\nu\tau}\, d\mu_{i\nu\tau} = \int_R \bar{\delta}_{\nu\tau}\, d\mu_{i\nu\tau} \qquad (i = 1, \cdots, p; \nu = 0, 1, \cdots m). \tag{5.17}$$

Lemma 5.1 is a simple consequence of the equations (5.14)–(5.17).

For any positive integer u, we shall say that a decision rule δ is truncated at the u^{th} stage if $\delta_{0u'} = 0$ for $u' \geqq u$ identically in x.

THEOREM 5.1. *If δ is truncated at the u^{th} stage there exists a nonrandomized decision rule δ^* that is strongly equivalent to δ.*

PROOF. It is sufficient to prove Theorem 5.1 in the case where $\delta_{\nu t} = 0$ for $t > u$ and $\nu \neq 1$ and $\delta_{1t} = 1$ for $t > u$. Clearly, $\varphi_i(x, \delta) < \infty$ for all x. Putting $\tau = 1$ and $T = u$ in Lemma 5.1, this lemma implies the existence of a decision rule δ^1 with the following properties: (a) δ^1 is strongly equivalent to δ; (b) $\delta^1_{\nu 1} = 0$ or 1 ($\nu = 0, 1, \cdots, m$); (c) $\delta^1_{\nu t} = \delta_{\nu t}$ for $\nu = 0, 1, \cdots, m$ and $t > 1$. Applying Lemma 5.1 to δ^1 and putting $\tau = 2$ and $T = u$, we see that there exists a decision rule δ^2 with the following properties: (a) δ^2 is strongly equivalent to δ^1; (b) $\delta^2_{\nu 2} = 0$ or 1 ($\nu = 0, 1, \cdots, m$); (c) $\delta^2_{\nu t} = \delta^1_{\nu t}$ for $\nu = 0, 1, \cdots, m$ and $t \neq 2$. Continuing this procedure, at the u^{th} step we obtain a decision rule δ^u that is nonrandomized and is strongly equivalent to all the preceding ones. This proves our theorem.

We shall say that two decision rules δ^1 and δ^2 are strongly equivalent up to the T^{th} stage if

$$\begin{aligned} &\int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta^1_{01} \cdots \delta^1_{0(t-1)}\delta^1_{\nu t}\, dF_{it} \\ &\qquad = \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta^2_{01} \cdots \delta^2_{0(t-1)}\delta^2_{\nu t}\, dF_{it} \end{aligned} \tag{5.18}$$

$$\text{for} \quad i = 1, \cdots, p; \nu = 1, \cdots, m \quad \text{and} \quad t = 1, \cdots, T.$$

Furthermore, we shall say that a decision rule δ is nonrandomized up to the stage T if $\delta_{\nu t} = 0$ or 1 for $\nu = 0, 1, \cdots, m$ and $t = 1, \cdots, T$.

We now prove the following theorem.

THEOREM 5.2. *If δ is a decision rule for which $\varphi_i(x, \delta) < \infty$, except perhaps on a set of x's of probability zero according to every $F_i(x) (i = 1, \cdots, p)$, then there exists a nonrandomized decision rule δ^* that is equivalent to δ.*

PROOF. Let $\{\epsilon_i\}$ and $\{\eta_i\} (i = 1, 2, \cdots,$ ad inf.) be two sequences of positive numbers such that $\lim_{i=\infty} \epsilon_i = 0$ and $\lim_{i=\infty} \eta_i = \infty$. Let T_1 be a positive integer such that

$$r_i(\delta) - \sum_{t=1}^{T_1}\sum_{\nu=1}^{m} \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01} \cdots \delta_{0(t-1)}\delta_{\nu t}\, dF_{it} < \epsilon_1 \quad \text{if} \quad r_i(\delta) < \infty, \tag{5.19}$$

and

$$(5.20) \qquad \sum_{t=1}^{T_1} \sum_{\nu=1}^{m} \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01} \cdots \delta_{0(t-1)} \delta_{\nu t} \, dF_{it} > \eta_1 \quad \text{if} \quad r_i(\delta) = \infty.$$

Let δ^1 be a decision rule such that $\varphi_i(x, \delta^1) < \infty$ (except perhaps on a set of probability measure zero); δ^1 is equivalent to δ; δ^1 is strongly equivalent to δ up to the T_1th stage; δ^1 is nonrandomized up to the T_1th stage and $\delta^1_{\nu t} = \delta_{\nu t}$ for $t > T_1$. The existence of such a decision rule follows from a repeated application of Lemma 5.1. In general, after $\delta^1, \cdots, \delta^j$ and $T_1, \cdots, T_j$ are given, let δ^{j+1} be a decision rule such that $\varphi_i(x, \delta^{j+1}) < \infty$ (except perhaps on a set of probability measure zero); δ^{j+1} is equivalent to δ^j; δ^{j+1} is strongly equivalent to δ^j up to the T_{j+1}th stage, where T_{j+1} is a positive integer chosen so that $T_{j+1} > T_j$ and (5.19) and (5.20) hold with δ replaced by δ^j, ϵ_1 replaced by ϵ_{j+1} and η_1 replaced by η_{j+1}; δ^{j+1} is nonrandomized up to the stage T_{j+1}; $\delta^{j+1}_{\nu t} = \delta^j_{\nu t}$ for $t \leqq T_j$ and $\delta^{j+1}_{\nu t} = \delta_{\nu t}$ for $t > T_{j+1}$. The existence of such a decision rule δ^{j+1} follows again from a repeated application of Lemma 5.1.

Let δ^* be the decision rule given by the equations

$$(5.21) \qquad \delta^*_{\nu t} = \delta^t_{\nu t} \qquad (\nu = 0, 1, \cdots, m; t = 1, 2, \cdots, \text{ ad inf.}).$$

It follows easily from the above stated properties of the decision rules δ^j ($j = 1, 2, \cdots$, ad inf.) that δ^* is nonrandomized and $r_i(\delta^*) = r_i(\delta) (i = 1, \cdots, p)$. This completes the proof of Theorem 5.2.

6. Examples where admissible[8] decision functions do not admit equivalent nonrandomized decision functions. In this section we shall construct examples which show that there exist admissible decision functions $\delta(x)$ which do not admit equivalent nonrandomized decision functions $\delta^*(x)$.

Example 1. Let X be a normally distributed chance variable with unknown mean θ and variance unity. This means that Ω is the totality of all univariate normal distributions with unit variance. Suppose we wish to test the hypothesis H_0 that the true mean θ is rational on the basis of a single observation x on X. Thus, D consists of two elements d_1 and d_2 where d_1 is the decision to accept H_0 and d_2 is the decision to reject H_0. For any decision function $\delta(x)$, let $\delta_1(x)$ denote the value of $\delta(d_1 \mid x)$. Let the loss be zero when a correct decision is made, and the loss be one when a wrong decision is made. Then the risk when θ is the true mean and the decision function $\delta(x)$ is adopted is given by

$$(6.1) \qquad r(\theta, \delta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} \delta_1(x) \, dx \qquad \text{when } \theta \text{ is irrational,}$$

$$(6.2) \qquad r(\theta, \delta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} (1 - \delta_1(x)) \, dx \qquad \text{when } \theta \text{ is rational.}$$

---

[8] A decision function with risk function $r(F)$ is called admissible if there exists no other decision function with risk function $r'(F)$ such that $r'(F) \leqq r(F)$ for every $F \in \Omega$, and the inequality sign holds for at least one $F \in \Omega$.

Let $\delta_1^0(x) = \frac{1}{2}$ for all x. Clearly,

$$r(\theta, \delta^0) = \tfrac{1}{2} \tag{6.3}$$

for all θ. We shall now show that $\delta^0(x)$ is an admissible decision function. For suppose that there exists a decision function $\delta'(x)$ such that

$$r(\theta, \delta') \leqq r(\theta, \delta^0) = \tfrac{1}{2} \tag{6.4}$$

for all θ, and

$$r(\theta_1, \delta') < r(\theta_1, \delta^0) = \tfrac{1}{2} \tag{6.5}$$

for some value θ_1. Suppose first that θ_1 is rational. Since the integrals in (6.1) and (6.2) are continuous functions of θ, for an irrational value θ_2 sufficiently near to θ_1 we shall have $r(\theta_2, \delta') > \frac{1}{2}$ which contradicts (6.4). Thus, θ_1 cannot be rational. In a similar way, one can show that θ_1 cannot be irrational. Hence, the assumption that a decision function $\delta'(x)$ satisfying (6.4) and (6.5) exists leads to a contradiction and the admissibility of $\delta^0(x)$ is proved.

Let now $\delta^*(x)$ be any decision function for which

$$r(\theta, \delta^*) = r(\theta, \delta^0) \tag{6.6}$$

for all θ. Now (6.6) implies that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} (\delta_1(x) - \delta_1^*(x))\, dx = 0 \tag{6.7}$$

identically in θ. Since $\delta_1(x) - \delta_1^*(x)$ is a bounded function of x, it follows from the uniqueness properties of the Laplace transform that (6.7) can hold only if $\delta_1(x) - \delta_1^*(x) = 0$ except perhaps on a set of measure zero. Hence, no nonrandomized decision function $\delta^*(x)$ can satisfy (6.6).

In the above example, the distributions consistent with the hypothesis H_0 which is to be tested (normal distributions with rational means) are not well separated from the alternative distributions (normal distributions with irrational means). One might think that this is perhaps the reason for the existence of an admissible decision function δ^0 such that no nonrandomized decision function δ^* can have as good a risk function as δ^0 has. That this need not be so, is shown by the following:

EXAMPLE 2. Suppose that X is a normally distributed chance variable with mean θ and variance unity. The value of θ is unknown. It is known, however, that the true value of θ is contained in the union of the two intervals $[-2, -1]$ and $[1, 2]$. Suppose that we want to test the hypothesis that θ is contained in the interval $[-2, -1]$ on the basis of a single observation x on X. Suppose, furthermore, that the chance variable X itself is not observable and only the chance variable $Y = f(X)$ can be observed where $f(x) = x$ when $|x| < 1$, and $= |x|$ when $|x| \geqq 1$. Let the loss be zero when a correct decision is made, and one when a wrong decision is made. For any decision function $\delta(y)$, let

$\delta_1(y)$ denote the value of $\delta(d_1 \mid y)$ where d_1 denotes the decision to accept H_0. Let $\delta^0(y)$ be the following decision function:

$$\begin{aligned} \delta_1^0(y) &= 1 \quad \text{when } -1 < y < 0 \\ &= 0 \quad \text{when } \quad 0 \leqq y < 1 \\ &= \tfrac{1}{2} \quad \text{when } \quad y \geqq 1. \end{aligned} \tag{6.8}$$

First we shall show that $\delta^0(y)$ is an admissible decision function. For this purpose, consider the following probability density function $g(\theta)$ in the parameter space: $g(\theta) = \frac{1}{2}$ when $-2 \leqq \theta \leqq -1$ or $1 \leqq \theta \leqq 2$, $= 0$ for all other θ. If we interpret $g(\theta)$ as the a priori probability distribution of θ, the a posteriori probability of the θ-interval $[-2, -1]$ is greater (less) than the a posteriori probability of the θ-interval $[1, 2]$ when $-1 < y < 0$ $(0 < y < 1)$, and the a posteriori probabilities of the two intervals are equal to each other when $y = 0$ or $y \geqq 1$. Hence, $\delta^0(y)$ is a Bayes solution relative to the a priori distribution $g(\theta)$, i.e.,

$$\int_{-2}^{-1} r(\theta, \delta^0)\, d\theta + \int_1^2 r(\theta, \delta^0)\, d\theta \leqq \int_{-2}^{-1} r(\theta, \delta)\, d\theta + \int_1^2 r(\theta, \delta)\, d\theta \tag{6.9}$$

for any decision function δ. Suppose now that δ is a decision function for which $r(\theta, \delta) \leqq r(\theta, \delta^0)$ for all θ. It then follows from (6.9) that $r(\theta, \delta) < r(\theta, \delta^0)$ can hold at most on a set of θ's of measure zero. Since, as can easily be verified, $r(\theta, \delta)$ and $r(\theta, \delta^0)$ are continuous functions of θ, it follows that $r(\theta, \delta) = r(\theta, \delta^0)$ everywhere and the admissibility of δ^0 is proved.

Let now $\delta'(y)$ be any decision function for which $r(\theta, \delta') = r(\theta, \delta^0)$ for all θ, i.e.,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} [\delta^0(y) - \delta'(y)]\, dx = 0 \qquad \text{for all } \theta. \tag{6.10}$$

Since $\delta_1^0(y) - \delta_1'(y)$ is a bounded function of x, it follows from the uniqueness properties of the Laplace transform that (6.10) can hold only if $\delta_1^0(y) = \delta_1'(y)$ except perhaps on a set of measure zero. Thus, no nonrandomized decision function δ^* exists such that $r(\theta, \delta^*) = r(\theta, \delta^0)$ for all θ.

7. Compactness of Ω in the ordinary sense is not sufficient for the existence of ϵ-equivalent nonrandomized decision functions. Let $\Omega = \{F\}$ be the totality of density functions[9] on the interval $0 \leqq x \leqq 1$ for which $F(x) \leqq c$ for every x, where c is some positive constant greater than 2. The sample space will be the interval $0 \leqq x \leqq 1$. We shall say that the sequence $F_1, F_2, \cdots$ converges to F if

$$\lim_{n\to\infty} \int_{-\infty}^{x} F_n(y)\, dy = \int_{-\infty}^{x} F(y)\, dy$$

[9] Here $F(x)$ denotes a density function. This represents a change in notation from preceding sections.

for every real x. The set Ω is compact in the sense of the above convergence definition.[10] Let A be a fixed interval $a_1 \leqq x \leqq a_2$ where $0 < a_1 < a_2 < 1$. Let $D = \{d_1, d_2\}$ and define W as follows:

$$W(F, d_1) + W(F, d_2) \equiv 1,$$

$$W(F, d_1) = 0 \text{ or } 1$$

according as the probability of A under F is rational or not. For any decision function $\delta(x)$, let $\delta_1(x)$ denote the probability assigned to d_1 by $\delta(x)$, i.e., $\delta_1(x) = \delta(d_1 \mid x)$.

Let $\delta'(x)$ be the decision function for which $\delta_1'(x) \equiv \frac{1}{2}$. We shall prove that $\delta'(x)$ is an admissible decision function. For suppose there exists a decision function $\delta^0(x)$ such that

$$r(F, \delta^0) \leqq r(F, \delta') = \tfrac{1}{2} \tag{7.1}$$

for every F, and for F_0 we have

$$r(F_0, \delta^0) < r(F_0, \delta'). \tag{7.2}$$

Now, if $F_i \to F_0$ and $W(F_i, d_1) = W(F_0, d_1)$ for every i, then $r(F_i, \delta) \to r(F_0, \delta)$ for every decision function $\delta(x)$, and, in particular, for $\delta^0(x)$. If $F_i \to F_0$ and $W(F_i, d_1) + W(F_0, d_1) = 1$ for every i, then $r(F_i, \delta) \to 1 - r(F_0, \delta)$ for every decision function $\delta(x)$ and, in particular, for $\delta^0(x)$. Clearly, we can construct two sequences of functions F such that each sequence converges to F_0, the probability of A according to every member of the first sequence is rational, and the probability of A according to every member of the second sequence is irrational. Because of (7.2) it follows that inequality (7.1) will be violated for almost every member of one of these two sequences. Hence δ' is admissible.

Let us now prove that there cannot exist a nonrandomized decision function $\delta^*(x)$ such that

$$r(F, \delta^*) \leqq r(F, \delta') + \tfrac{1}{4} = \tfrac{3}{4} \tag{7.3}$$

for every $F \in \Omega$. Suppose there were such a decision function $\delta^*(x)$. Let H be the set of x's where $\delta_1^*(x) = 1$, and let $\bar{H}$ be the complement of H with respect to the interval [0, 1]. If H is a set of measure zero or one then obviously (7.3) is violated for some F. Thus, it is sufficient to consider the case when H is a set of positive measure $\alpha < 1$. Suppose for a moment that $\alpha > \frac{1}{2}$. Let G be the density which is zero on $\bar{H}$ and constant on H. From (7.3) it follows that $P\{A \mid G\}$ is rational. There exists a density $G' \in \Omega$ such that $P\{H \mid G'\} > \frac{3}{4}$ and $P\{A \mid G'\}$ is irrational. But then (7.3) is violated for G'. If $\alpha \leqq \frac{1}{2}$, let $\bar{G}$ be the density which is zero on H and constant on $\bar{H}$. From (7.3) it follows that $P\{A \mid \bar{G}\}$ is irrational. There exists a density $\bar{G}' \in \Omega$ such that $P\{\bar{H} \mid \bar{G}'\} >$

[10] The cumulative distribution functions are well-known to be compact in the usual convergence sense. Since the densities are bounded above the limit cumulative distribution function must be absolutely continuous.

$\frac{3}{4}$ and $P\{A \mid \bar{G}'\}$ is rational. But then (7.3) is violated for $\bar{G}'$. Thus (7.3) can never hold for every $F \in \Omega$ and the desired result is proved.

8. Sufficient conditions for the existence of ϵ-equivalent nonrandomized decision functions. In this section we shall consider the nonsequential decision problem (as described in the introduction), and we shall give sufficient conditions for the existence of ϵ-equivalent nonrandomized decision functions. We shall consider the following four metrics in the space Ω:

$$\rho_1(F_1, F_2) = \operatorname{Sup}_S \left| \int_S dF_1 - \int_S dF_2 \right| \tag{8.1}$$

when S is any measurable subset of R,

$$\rho_2(F_1, F_2) = \operatorname{Sup}_{d,x} | W(F_1, d, x) - W(F_2, d, x) |, \tag{8.2}$$

$$\rho_3(F_1, F_2) = \rho_1(F_1, F_2) + \rho_2(F_1, F_2), \tag{8.3}$$

$$\rho_4(F_1, F_2) = \operatorname{Sup}_{\delta} | r(F_1, \delta) - r(F_2, \delta) | . \tag{8.4}$$

First we prove the following lemma:

LEMMA 8.1. *If Ω is conditionally compact in the sense of the metric ρ_3, then it is conditionally compact in the sense of the metric ρ_4.*

PROOF. Let $\{F_i\}(i = 1, 2, \cdots$, ad inf.) be a Cauchy sequence in the sense of the metric ρ_3, i.e.,

$$\lim_{i,j=\infty} \rho_3(F_i, F_j) = 0. \tag{8.5}$$

It follows from (8.5) and (8.3) that $W(F_i, d, x)$ converges, as $i \to \infty$, to a limit function $W(d, x)$ uniformly in d and x, i.e.,

$$\lim_{i=\infty} W(F_i, d, x) = W(d, x) \tag{8.6}$$

uniformly in d and x. Hence

$$\lim_{i=\infty} \int_D W(F_i, d, x)\, d\delta_x = \int_D W(d, x)\, d\delta_x \tag{8.7}$$

uniformly in x and δ. Because of (8.5), we have

$$\lim_{i,j=\infty} \rho_1(F_i, F_j) = 0. \tag{8.8}$$

Hence there exists a distribution function $F_0(x)$ (not necessarily an element of Ω) such that

$$\lim_{i=\infty} \rho_1(F_i, F_0) = 0. \tag{8.9}$$

It follows from (8.7) and (8.9) that

$$\lim_{i=\infty} \int_R \left[\int_D W(F_i, d, x)\, d\delta_x\right] dF_i(x) = \int_R \left[\int_D W(d, x)\, d\delta_x\right] dF_0(x) \tag{8.10}$$

uniformly in δ. Hence $\{F_i\}$ is a Cauchy sequence in the sense of the metric ρ_4 and Lemma 8.1 is proved.

Next we prove

LEMMA 8.2. *If D is conditionally compact in the sense of the metric* (1.1) *and if δ is any decision function, then for any $\epsilon > 0$ there exists a finite subset D^1 of D and a decision function δ^1 such that $\delta^1(D^1 \mid x) = 1$ identically in x and δ^1 is ϵ-equivalent to δ.*

PROOF. Since D is conditionally compact, it is possible to decompose D into a finite number of disjoint subsets $D_1, \cdots, D_u$ such that the diameter of D_j is less then $\epsilon (j = 1, \cdots, u)$. Let d_j be an arbitrary but fixed point of $D_j (j = 1, \cdots, u)$ and let $\delta^1(x)$ be the decision function determined by the condition

$$\delta^1(d_j \mid x) = \delta(D_j \mid x) \qquad (j = 1, \cdots, u). \tag{8.11}$$

Clearly

$$\left| \int_D W(F, d, x)\, d\delta_x - \int_D W(F, d, x)\, d\delta_x^1 \right| \leqq \epsilon \tag{8.12}$$

for all F and x. Hence,

$$| r(F, \delta^1) - r(F, \delta) | \leqq \epsilon \tag{8.13}$$

for all F and our lemma is proved.

We are now in a position to prove the main theorem.

THEOREM 8.1. *If the elements $F(x)$ of Ω are atomless, if Ω is conditionally compact in the sense of the metrics ρ_1 and ρ_2, and if D is conditionally compact in the the sense of the metric* (1.1), *then for any $\epsilon > 0$ and for any decision function $\delta(x)$ there exists an ϵ-equivalent nonrandomized decision function $\delta^*(x)$.*

PROOF. Because of Lemma 8.2, it is sufficient to prove our theorem for finite D. Thus, we shall assume that D consists of the elements $d_1, \cdots, d_m$. It is easy to verify that conditional compactness of Ω in the sense of both metrics ρ_1 and ρ_2 implies conditional compactness in the sense of the metric ρ_3, and because of Lemma 8.1, also in the sense of the metric ρ_4. Thus, conditional compactness of Ω in the sense of the metrics ρ_1 and ρ_2 implies the existence of a finite subset $\Omega^* = \{F_1, \cdots, F_k\}$ of Ω such that Ω^* is $\epsilon/2$-dense in Ω in the sense of the metric ρ_4. Let δ^* be a nonrandomized decision function that is equivalent to δ if Ω is replaced by Ω^*. The existence of such a δ^* follows from Theorem 3.1. Since Ω^* is $\epsilon/2$-dense in Ω (in the sense of the metric ρ_4), we have

$$| r(F, \delta^*) - r(F, \delta) | \leqq \epsilon \quad \text{for all} \quad F \text{ in } \Omega \tag{8.14}$$

and our theorem is proved.

We shall now introduce some notions with the help of which we shall be able to strengthen Theorem 3.1. For any measurable subset S of R, let

$$r(F, \delta \mid S) = \int_S \left[\int_D W(F, d, x) \, d\delta_x \right] dF(x). \tag{8.15}$$

We shall refer to the above expression as the contribution of the set S to the risk. For any S we shall consider the following four metrics in Ω:

$$\rho_{1S}(F_1, F_2) = \operatorname*{Sup}_{S} \left| \int_{S^*} dF_1 - \int_{S^*} dF_2 \right| \tag{8.16}$$

where S^* is any measurable subset of S,

$$\rho_{2S}(F_1, F_2) = \operatorname*{Sup}_{d, x \epsilon S} \mid W(F_1, d, x) - W(F_2, d, x) \mid, \tag{8.17}$$

$$\rho_{3S}(F_1, F_2) = \rho_{1S}(F_1, F_2) + \rho_{2S}(F_1, F_2), \tag{8.18}$$

$$\rho_{4S}(F_1, F_2) = \operatorname*{Sup}_{\delta} \mid r(F_1, \delta \mid S) - r(F_2, \delta \mid S) \mid. \tag{8.19}$$

Finally let the metric $\rho_S(d_1, d_2)$ in D be defined by

$$\rho_S(d_1, d_2) = \operatorname*{Sup}_{F, x \epsilon S} \mid W(F, d_1, x) - W(F, d_2, x) \mid. \tag{8.20}$$

We shall now prove the following stronger theorem:

THEOREM 8.2. *Let all elements F of Ω be atomless. If there exists a decomposition of R into a sequence $\{R_i\}$ $(i = 1, 2, \cdots, ad\ inf.)$ of disjoint subsets such that Ω is conditionally compact in the sense of the metrics ρ_{1R_i} and ρ_{2R_i} for each i, and such that D is conditionally compact in the sense of the metric ρ_{R_i} for each i, then for any $\epsilon > 0$ and for any decision function δ there exists an ϵ-equivalent nonrandomized decision function δ^*.*

PROOF. Let $\{R_i\}$ be a decomposition of R for which the conditions of our theorem are fulfilled. Let $\{\epsilon_i\}$ be a sequence of positive numbers such that $\sum_{i=1}^{\infty} \epsilon_i = \epsilon$. Let $\delta^1(x)$ be a decision function such that $\delta_1(x) = \delta(x)$ for any x not in R_1, $\delta^1(x)$ is nonrandomized over R_1 (for any x in R_1, $\delta^1(x)$ assigns the probability one to a single point d in D) and such that

$$\mid r(F, \delta \mid R_1) - r(F, \delta^1 \mid R_1) \mid \leqq \epsilon_1 \quad \text{for all } F. \tag{8.21}$$

The existence of such a decision function δ^1 follows from Theorem 8.1 (replacing R by R_1). After $\delta^1, \cdots, \delta^{i-1}$ have been defined $(i \geqq 1)$, let δ^i be a decision function such that δ^i is nonrandomized over R^i, $\delta^i(x) = \delta^{i-1}(x)$ for all x in $\bigcup_{j=1}^{i-1} R_j$, $\delta^i(x) = \delta(x)$ for all x in $R - \bigcup_{j=1}^{i} R_j$ and such that

$$\mid r(F, \delta^i \mid R_i) - r(F, \delta \mid R_i) \mid \leqq \epsilon_i \quad \text{for all} \quad F \text{ in } \Omega. \tag{8.22}$$

The existence of such a decision function δ^i follows again from Theorem 8.1. Clearly $\delta^i(x)$ converges to a limit $\delta^*(x)$, as $i \to \infty$. This limit decision function $\delta^*(x)$ is obviously nonrandomized and satisfies the conditon

$$(8.23) \qquad | r(F, \delta \mid R_i) - r(F, \delta^* \mid R_i) | \leqq \epsilon_i$$

for all i and F. Theorem 8.2 is an immediate consequence of this.

The conditions of Theorem 8.2 will be fulfilled for a wide class of statistical decision problems. For example, this is true for the decision problems which satisfy the following six conditions:

CONDITION 1. *The sample space R is a finite dimensional Euclidean space. All elements $F(x)$ of Ω are absolutely continuous.*

CONDITION 2. *Ω admits a parametric representation, i.e., each element F of Ω is associated with a parametric point θ in a finite dimensional Euclidean space E.*

We shall denote the density function $p(x)$ corresponding to the parameter point θ by $p(x, \theta)$.

CONDITION 3. *The set of parameter points θ which correspond to all elements F of Ω is a closed subset of E.*

We shall call this set of all parameter points θ the parameter space. Since there is a one-to-one correspondence between the elements F of Ω and the points θ of the parameter space, there is no danger of confusion if we denote the parameter space also by Ω.

CONDITION 4. *The density function $p(x, \theta)$ is continuous in $\theta \in \Omega$ for every x.*

CONDITION 5. *The loss $W(\theta, d)$ when θ is true and the decision d is made does not depend on x. D is conditionally compact in the sense of the metric $\rho(d_1, d_2) = \operatorname{Sup}_{\theta} | W(\theta, d_1) - W(\theta, d_2) |$.*

CONDITION 6. *For any bounded subset M of R, we have* $\lim_{\left\{ \substack{|\theta| = \infty \\ \theta \in \Omega} \right\}} \int_M p(x, \theta)\, dx = 0.$

We shall now show that the conditions of Theorem 8.2 are fulfilled for any decision problem that satisfies Conditions 1–6. Let S_i be the sphere in R with center at the origin and radius i. Let $R_1 = S_1$ and $R_i = S_i - \bigcup_{j=1}^{i-1} R_j (i = 1, 2, \cdots$, ad inf.). Condition 5 implies that D is conditionally compact in the sense of the metric ρ_{R_i} for all i. It follows from Condition 5 and Theorem 2.1 in [3] that Ω is conditionally compact in the sense of the metric $\rho(\theta_1, \theta_2) = \operatorname{Sup}_{d} | W(\theta_1, d) - W(\theta_2, d) |$. Hence Ω is conditionally compact in the sense of the metric ρ_{2R_i} for each i. It remains to be shown that Ω is conditionally compact in the sense of the metric ρ_{1R_i} for each i. For this purpose, consider any sequence $\{\theta_j\} (j = 1, 2, \cdots$, ad inf.) of parameter points. There are 2 cases possible: (a) $\{\theta_j\}$ admits a subsequence that converges in the Euclidean sense to a finite point θ_0; (b) $\lim_{j=\infty} | \theta_j | = \infty$. Let us consider first the case (a) and let $\{\theta_j'\}$ be a subsequence of $\{\theta_j\}$ which converges to a finite point θ_0. It then follows from Condition 4 and a theorem of Robbins [4] that $\{\theta_j'\}$ is a Cauchy subsequence in the sense of the metric ρ_{1R_i} for each i. In case (b), Condition 6 implies that the sequence $\{\theta_j\}$ is a Cauchy sequence in the sense of the metric ρ_{1R_i} for each i. Thus, Ω is conditionally compact in the sense of the metric ρ_{1R_i}. This completes the proof of our assertion that a decision problem that satisfies Conditions 1–6, satisfies also the conditions of Theorem 8.2.

9. Application to the theory of games. Translation of the results of Section 2 into the language of the theory of games is immediate and we shall do this only very briefly. The function $W(F_i, d_j, x)$ $(i = 1, \cdots, p; j = 1, \cdots, m; x \in R)$, of Section 1 is now called the pay-off function of a zero-sum two-person game. The game is played as follows: Player I selects one of the integers $1, \cdots, p$, say i, without communicating his choice to player II. A random observation $x \in R$ on a chance variable whose distribution function is F_i is obtained and communicated to player II. The latter chooses one of the integers $1, \cdots, m$, say j. The game now ends with the receipt by player I and player II of the respective sums $W(F_i, d_j, x)$ and $-W(F_i, d_j, x)$. Randomized (mixed) and nonrandomized (pure) strategies are defined in the same manner as the corresponding decision functions in Section 1. When the distribution functions $F_i(x)(i = 1, \cdots, p)$ are all atomless the obvious analogues of Theorems 3.1 and 3.2 hold.

It should be remarked that the usual definition of randomized (mixed) strategy is not as general as the one given above. In the usual definition player II chooses, by a random mechanism independent of the random mechanism which yields the point x, some one of a (usually finite) number of nonrandomized (pure) strategies, and then plays the game according to the nonrandomized strategy selected. In our definition (used in [3]) the random choice is allowed to depend on x. Clearly our method of randomization includes the usual one as a special case. The relation between the two methods of randomization will be discussed by two of the authors in a forthcoming paper [7].

Suppose that the number of possible decisions is at most denumerable, and that the decision procedure consists in choosing at random and in advance of the observations, one of a finite number of nonrandomized decision functions. The sample space can be divided into an at most denumerable number of sets in each of which only a finite number of decisions is possible (the possible decisions vary from set to set). In each set our results are applicable. Since the number of sets is denumerable the resultant decision function is measurable. We conclude: It follows from our results that if a decision procedure consists of selecting with preassigned probabilities one of a finite number of nonrandomized decision functions with the number of possible decisions at most denumerably infinite, and if the possible distributions are finite in number and atomless, then there exists an equivalent nonrandomized decision function. More general results can be obtained for this case (where one chooses at random and in advance of the observations, one of a finite number of nonrandomized decision functions). By application of the methods of Sections 4 and 8 the requirement that the number of possible decisions be denumerable can be easily removed. The procedures are straightforward and we omit them.

REFERENCES

[1] A. Dvoretzky, A. Wald, and J. Wolfowitz, "Elimination of randomization in certain problems of statistics and of the theory of games," *Proc. Nat. Acad. Sci.*, Vol. 36 (1950), pp. 256–260.

[2] A. Lyapunov, "Sur les fonctions-vecteurs complètement additives," *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, Vol. 4 (1940), pp. 465–478.

[3] A. Wald, *Statistical Decision Functions*, John Wiley & Sons, 1950.

[4] Herbert Robbins, "Convergence of distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 72–75.

[5] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

[6] A. Dvoretzky, A. Wald, and J. Wolfowitz, "Relations among certain ranges of vector measures," *Pacific Journal of Mathematics* (1951).

[7] A. Wald and J. Wolfowitz, "Two methods of randomization in statistics and the theory of games," *Annals of Mathematics*, to be published.