ON THE PROPERTY (W) OF THE CLASS OF STATISTICAL DECISION FUNCTIONS[1]

BY HIROKICHI KUDŌ

Osaka City University and University of California, Berkeley

0. Summary. The property (W) of the class of decision functions, which corresponds to the concept of weak compactness in the intrinsic sense in [6], is discussed and several sufficient conditions for it are given in this article. Some examples concerning the non-sequential case are discussed.

1. Introduction. LeCam proved some complete class theorems under the assumption that the class D of decision functions is compact in some sense [3], and in the same paper he mentioned that the compactness of D can be replaced by the property (W). This property is an extension of Wald's concept of weak compactness in the intrinsic sense, which the reader would find in Wald's book [6], page 77. This paper will be devoted to giving a precise description of this property and sufficient conditions for it, some of which were previously sketched in miscellaneous remarks (5), (6) and (8) of LeCam's paper [3], and also in [4].

The property (W) of D is essentially a geometrical concept of the subset $R = \{r(\cdot, \delta): \delta \varepsilon D\}$ of the function space $\mathfrak{F}$ on the parameter space Θ, where $r(\theta, \delta)$ represents a risk imposed on a statistician who adopts $\delta \varepsilon D$ when θ is true value of the parameter. We shall refer to the corresponding property of R as half-closedness. We shall give the definitions of half-closedness and of the property (W) in Section 2. To see how these properties work in the complete class theorems, Wald-LeCam's complete class theorems are restated in Section 3. The form of Wald-LeCam's theorem we describe here can be proved by the same way as that done in [3], and is also a very geometrical statement, in the sense that any structure of the risk function $r(\theta, \delta)$ will not be needed in the proof. In Section 4 we give two theorems concerning the geometrical property of a function of two variables. These theorems could be used as a criterion of a loss function $L(\theta, a)$ being half-closed and of the class D having the property (W). To obtain more precise criteria the risk function $r(\theta, \delta)$ is specialized in the usual way in Sections 6 and 7. We will give the definitions of decision functions and risk function in non-sequential case, according to LeCam [3], for the completeness of descriptions (Section 5). In Section 6 we give a sufficient condition for the class $\mathfrak{D}$ of all the decision functions defined in Section 5 having the property (W). Roughly speaking, the condition in Theorem 4 (and Theorem 4$'$) is the half-closedness of the loss function $L(\theta, a)$. For a subclass of $\mathfrak{D}$, it happens that D does not have the property (W) while $L(\theta, a)$ satisfies the assumptions of Theorem 4. Theorem 5 of Section 7 says that if the loss function tends to ∞ at the infinity point of the action space and if the sample distribution has positive density everywhere, every closed subclass of $\mathfrak{D}$ has the property (W).

Received 28 June 1965; revised 23 May 1966.

[1] This research was supported by the National Science Foundation.

The author wishes to express his sincere gratitude to Professor Lucien LeCam for very valuable conversations. He is also grateful to Mr. Tokitake Kusama and the referees for their advice.

2. The property (W) **and half-closedness.** Let Θ be an arbitrary set, and $\mathfrak{F}$ the set of all nonnegative real extended functions defined on Θ. We shall assign each element $f \varepsilon \mathfrak{F}$ with a family of neighborhoods $V(f: \theta_1, \cdots, \theta_k, \epsilon)$ consisting of all elements g of $\mathfrak{F}$ such that

$$|g(\theta_i) - f(\theta_i)| < \epsilon \quad \text{if} \quad f(\theta_i) < \infty,$$
$$g(\theta_i) > 1/\epsilon \quad \text{if} \quad f(\theta_i) = \infty,$$

where k is an arbitrary positive integer, $\{\theta_1, \cdots, \theta_k\}$ a finite subset of Θ and $\epsilon > 0$. Such a system of neighborhoods of every f in $\mathfrak{F}$ defines a topology $\mathfrak{J}$ in $\mathfrak{F}$, which we shall refer as a pointwise convergence topology. $\mathfrak{F}$ is compact with respect to this topology $\mathfrak{J}$.

DEFINITION 1. A subset F of $\mathfrak{F}$ is said to be *half-closed* if, for any element f^* of the closure F^* of F with respect to $\mathfrak{J}$, there exists an element $f \varepsilon F$ such that

$$f(\theta) \leqq f^*(\theta) \qquad \text{for every} \quad \theta \varepsilon \Theta.$$

Let us consider a statistical decision function problem (Θ, D, r), where Θ is the space of the parameter θ, D the class of decision functions δ to which the choice of a statistician is restricted, and $r(\theta, \delta)$ the risk function imposed on him when δ is chosen and θ is the true value of parameter. Since for each $\delta \varepsilon D$ the risk function $r(\cdot, \delta)$ is regarded as an element of $\mathfrak{F}$, we shall denote by R the subset of $\mathfrak{F}$ whose elements are all $r(\cdot, \delta)$, $\delta \varepsilon D$.

DEFINITION 2. A class D of decision functions is said to have *the property* (W), if the corresponding R is half-closed.

3. Wald-LeCam's theorem. For the sake of understanding the role of the property (W), we shall restate the general complete class theorem, which is initiated by Wald [6] and then extended by LeCam [3].

DEFINITION 3. A class D of decision functions is said to be *subconvex* if for any two elements δ_1 and δ_2 of D and a real number α $(0 \leqq \alpha \leqq 1)$ there is a $\delta_3 \varepsilon D$ such that

$$r(\theta, \delta_3) \leqq \alpha r(\theta, \delta_1) + (1 - \alpha) r(\theta, \delta_2) \qquad \text{for every} \quad \theta \varepsilon \Theta.$$

A function $\xi(\theta)$ on Θ is called a probability function when it is a nonnegative function vanishing everywhere on Θ except for a finite number of θ's and $\sum_\theta \xi(\theta) = 1$. Here we shall understand that $\sum_\theta$ stands for the sum of non-zero values. Denote by Ξ the set of all such probability functions on Θ and define a function $F(\delta)$ of δ as

$$F(\delta) = \inf_{\xi \varepsilon \Xi} \{ \textstyle\sum_\theta r(\theta, \delta)\xi(\theta) - \inf_{\delta' \varepsilon D} \sum_\theta r(\theta, \delta')\xi(\theta) \}.$$

DEFINITION 4. A decision function $\delta \varepsilon D$ is called a *Bayes solution* if

$\sum_\theta r(\theta, \delta)\xi(\theta) = \inf_{\delta' \varepsilon D} \sum_\theta r(\theta, \delta')\xi(\theta)$ for some $\xi \varepsilon \Xi$. By B we denote the set of all Bayes solutions.

DEFINITION 5. A decision function $\delta \varepsilon D$ satisfying $F(\delta) = 0$ is called a *Bayes solution in the wide sense*, and the set of such δ's will be denoted by W.

Let us denote by B^* the set of the δ's, $r(\cdot, \delta)$ of which belongs to the closure of $R_B = \{r(\cdot, \delta'): \delta' \varepsilon B\}$.

DEFINITION 6. A decision function $\delta \varepsilon D$ is said to be *improvable uniformly*, if there are a positive number ϵ and another decision function $\delta' \varepsilon D$ such that

$$r(\theta, \delta') \leqq r(\theta, \delta) - \epsilon \qquad \text{for every} \quad \theta \varepsilon \Theta.$$

LeCam extends Wald's complete class theorems as follows:

If D is subconvex and has the property (W), then

(i) there is the minimal complete class in D,

(ii) W, B^* and $W \cap B^*$ are all complete classes in D,

(iii) $F(\delta) > 0$ if and only if δ is improvable uniformly.

((ii) and (iii) are slight modifications of LeCam's original work, but there is no essential difference. These are proved in a quite similar way as LeCam's proof.)

From Definition 2 we have very easily

THEOREM 1. *If D has the property (W) and C is an esentially complete class of D, then C has the property (W).*

4. Sufficient conditions for the half-closedness.

THEOREM 2. *Let Θ and T be two arbitrary spaces, and suppose that $f(\theta, t)$ is a nonnegative real extended function defined on the cartesian product $\Theta \times T$. If for any $\theta \varepsilon \Theta$ and for any real number $k < \sup_{t\varepsilon T} f(\theta, t)$ there exists a proper subset C of T such that* (i) *$G = \{f(\cdot, t): t \varepsilon C\}$ is half-closed and* (ii) *$\inf_{t \notin C} f(\theta, t) > k$, then $F = \{f(\cdot, t): t \varepsilon T\}$ is half-closed.*

PROOF. Let $h(\theta)$ be an element of the closure F^* of F relative to $\mathfrak{J}$ and put $S(\theta) = \sup_{t\varepsilon T} f(\theta, t)$. Without any loss of generality we may assume that there is a point $\theta_0 \varepsilon \Theta$ such that $h(\theta_0) < S(\theta_0)$. Suppose that k be a given real number such that $h(\theta_0) < k < S(\theta_0)$. For such θ_0 and k we can take C such that $G = \{f(\cdot, t): t \varepsilon C\}$ is half-closed and $\inf_{t \notin C} f(\theta_0, t) > k$. This shows that h belongs to the closure G^* of G. Since G is half-closed, there is an element $t_0 \varepsilon C$ such that $h(\theta) \geqq f(\theta, t_0)$ for all $\theta \varepsilon \Theta$.

THEOREM 3. *Let Θ be an arbitrary space and T a compact Hausdorff space. Suppose that $f(\theta, t)$ is a nonnegative real extended function on $\Theta \times T$ and lower semicontinuous on T for any fixed $\theta \varepsilon \Theta$. Then $F = \{f(\cdot, t): t \varepsilon T\}$ is half-closed.*

PROOF. Let h be an element of the closure F^* of F. For any finite subset N of Θ and any $\epsilon > 0$ we shall denote by $U_{N,\epsilon}$ the set of all points t for which $f(\cdot, t)$ belongs to the neighborhood $V = V(h:N, \epsilon)$. Since $\{U_{N,\epsilon}\}$ has the finite intersection property, and T is compact, the intersection of the closures of all $U_{n,\epsilon}$'s has at least one point $t^* \varepsilon T$. From the lower semicontinuity of $f(\theta, t)$ on T, we have

$$f(\theta, t^*) \leqq \lim_{V\to h} \inf_{t\varepsilon U_{N,\epsilon}} f(\theta, t) = h(\theta).$$

Combining Theorems 2 and 3, we have

COROLLARY 1. *Let* Θ *be an arbitrary space and* T *a Hausdorff space. Suppose that* $F(\theta, t)$ *is a nonnegative real extended function* $f(\theta, t)$ *on* $\Theta \times T$ *and that* $f(\theta, t)$ *is lower semicontinuous on* T *for each* $\theta \varepsilon \Theta$. *If for any* $\theta \varepsilon \Theta$ *and for any real number* $k < \sup_{t \varepsilon T} f(\theta, t)$ *there is a compact proper subset* C *of* T *such that* $\inf_{t \notin C} f(\theta, t) > k$, *then* $F = \{f(\cdot, t) : t \varepsilon T\}$ *is half-closed.*

If we read T and $f(\theta, t)$ in Theorems 2, 3 and Corollary 1 as D and $r(\theta, \delta)$, respectively, and "half-closed of F" as "the property (W) of D," we have the corresponding statement about the property (W).

5. Decision function and its linearly structured risk. In the following sections we shall consider statistical problems (Θ, D, r) with a *linearly structured risk.* In such a problem, there are given three factors on which the problem is based:

(1) a σ-finite measure space $(X, \mathcal{B}, \lambda)$ named the sample space,

(2) a locally compact space $\mathbf{A}$, named the action space, and

(3) a real nonnegative function $L(\theta, a)$, named the loss function of $\theta \varepsilon \Theta$ and $a \varepsilon \mathbf{A}$, which is Borel measurable on $\mathbf{A}$ for any fixed $\theta \varepsilon \Theta$.

Let $\mathbf{L}_1$ be the Banach space of all integrable functions p on the sample space $(X, \mathcal{B}, \lambda)$ with norm $\|p\|_1 = \int |p(x)| \, d\lambda$. The distribution space Π of this problem is a subset of $\mathbf{L}_1$, consisting of nonnegative functions p of norm $\|p\|_1 = 1$, whose elements are labelled by the parameter $\theta \varepsilon \Theta$, and will be denoted by $p_\theta(x)$. Denote by $\mathcal{P}$ the linear subspace of $\mathbf{L}_1$ spanned by Π. Let us consider the space $C_0(\mathbf{A})$ of all continuous functions with compact carrier, and denote by $\|c\|$ the norm $(\max_{a \varepsilon \mathbf{A}} |c(a)|)$ of $c \varepsilon C_0(\mathbf{A})$.

Consider the linear space $\Phi = \Phi(C_0(\mathbf{A}), \mathcal{P})$ of all bilinear functionals φ on $C_0(\mathbf{A})$ and $\mathcal{P}$, which is bounded in the following sense: there is a positive number k such that $|\varphi(c, p)| \leqq k \|c\| \cdot \|p\|_1$ for every $c \varepsilon C_0(\mathbf{A})$ and $p \varepsilon \mathcal{P}$. The norm $\|\varphi\|$ of φ is defined as the infimum of such k's. An element φ of Φ is said to be positive if $c \geqq 0$ and $p \geqq 0$ implies $\varphi(c, p) \geqq 0$. According to LeCam [3], if $\mathbf{A}$ is separable, locally compact and metrizable, then every positive bilinear functional φ of norm 1 can be represented by an integral

$$\varphi(c, p) = \int_X [\int_{\mathbf{A}} c(a)\delta(da : x)] p(x)\lambda(dx)$$

by using a measure-function version $\delta(A : x)$, which is (1) a probability measure defined on the σ-field $\mathcal{A}$ of all Borel subsets of $\mathbf{A}$ for all $x \varepsilon X$, and (2) an essentially (λ) bounded measurable function of x for any $A \varepsilon \mathcal{A}$. Here δ and δ' are called equivalent if $\int \delta(A : x) p_\theta(x)\lambda(dx) = \int \delta'(A : x) p_\theta(x)\lambda(dx)$ holds for every $\theta \varepsilon \Theta$ and $A \varepsilon \mathcal{A}$. In this meaning an equivalent class of such δ's, or in other words, a positive bilinear functional on $C_0(\mathbf{A})$ and $\mathcal{P}$ of norm 1 is called a decision function. Throughout the sequel, we shall use the notation $\mathfrak{D}$ for the set of all decision functions thus defined. The risk function associated to a decision function δ is defined as

$$r(\theta, \delta) = \int_X [\int_{\mathbf{A}} L(\theta, a)\delta(da : x)] p_\theta(x)\lambda(dx) \tag{5.1}$$

as far as $L(\theta, a)$ is Borel measurable on $\mathbf{A}$ for every $\theta \varepsilon \Theta$.

The topology of $\mathfrak{D}$, which we shall follow to LeCam [3], is a relative topology of the weak topology of Φ, i.e. a topology of Φ generated by a system of neighborhoods

$$N(\varphi: c_1, \cdots, c_k, p_1, \cdots, p_k, \epsilon) = \{\varphi' \varepsilon \Phi: |\varphi(c_i, p_i) - \varphi'(c_i, p_i)| < \epsilon, i = 1, 2, \cdots, k\},$$

where k is an arbitrary positive integer, $c_i \varepsilon C_0(\mathbf{A})$, $p_i \varepsilon \mathcal{P}$ $(i = 1, 2, \cdots, k)$ and ϵ a positive number. When we are concerned only with a subset D of $\mathfrak{D}$, we shall refer to the relative topology of D induced by the above topology of Φ as the *regular* topology, after the Wald's terminology "the regular convergence" [6]. LeCam gave a condition for D being compact in the regular topology. As a special case of this condition, if $\mathbf{A}$ is compact, so is $\mathfrak{D}$ in the regular topology.

If for each $\theta \varepsilon \Theta$, $L(\theta, a)$ is lower semicontinuous of a, then $r(\theta, \delta)$ is lower semicontinuous in the regular topology (see [3], page 75), and so, for any nonnegative extended function $f(\theta)$ on Θ, the set $\mathfrak{D}_f = \{\delta \varepsilon \mathfrak{D}: r(\theta, \delta) \leqq f(\theta)\}$ is closed in $\mathfrak{D}$. In this case the compactness of $\mathfrak{D}$ implies that of $\mathfrak{D}_f$. From this fact we get

EXAMPLE 5.1. In the problem of testing hypothesis, the action space $\mathbf{A}$ is finite and so compact. Hence $\mathfrak{D}$ is compact. The set of all tests of level α is also compact in the regular topology, so that it has the property (W).

6. A criterion for $\mathfrak{D}$ having the property (W). In Section 4, Theorems 2 and 3, we gave general criteria for the space D of decision functions available to a statistician having the property (W), without any specialization of the structure of the risk function. Now we have a precise structure (5.1) of the risk function which gives us a criterion for $\mathfrak{D}$ having the property (W).

Before we proceed to our theorem we should give a preparatory lemma.

LEMMA. *Let T and S be σ-compact, locally compact metrizable spaces, $\mathcal{P}$ a linear subspace of $\mathbf{L}_1$ space on a measure space $(X, \mathcal{B}, \lambda)$, where λ is a σ-finite measure on $\mathcal{B}$. Suppose that u be a mapping of T onto S such that for any Borel subset B of S the inverse image $u^{-1}(B)$ is also a Borel subset of T. Then for any positive bilinear functional φ on $(C_0(S), \mathcal{P})$ there exists a positive bilinear functional ψ on $(C_0(T), \mathcal{P})$ such that $\|\psi\|_T = \|\varphi\|_S$ and*

$$\varphi(c, p) = \int_X [\int_T c(u(t))\eta(dt: x)]p(x)\lambda(dx) \tag{6.1}$$

for every $c \varepsilon C_0(S)$ and $p \varepsilon \mathcal{P}$, where $\|\cdot\|_T$ and $\|\cdot\|_S$ are the norms of bilinear functionals on $(C_0(T), \mathcal{P})$ and $(C_0(S), \mathcal{P})$, respectively, and η is a measure-function version of the bilinear functional ψ on T.

PROOF. Let $\mathfrak{S}$ be the collection of linear subspaces of $\mathcal{P}$ for which our lemma holds. For any $p \varepsilon \Pi$, the integral $\delta \circ p(S^*) = \int \delta(S^* | x)p(x)\lambda(dx)$ for every Borel set $S^* \subset S$ defines a probability measure on the σ-field of the Borel subsets of S, where δ is a measure-function version of φ. By virtue of Varadarajan's lemma [5], Lemma 2.2, there is a probability measure q_p on T such that

$$\delta \circ p(S^*) = q_p(u^{-1}S^*) \tag{6.2}$$

holds for any Borel subset S^* of S. Taking $q_{\alpha p} = \alpha q_p$ for any real α, we can see that q_p is a bilinear functional on $(C_0(T), \{p\})$, where $\{p\}$ is a one-dimensional linear subspace through p. Thus $\mathfrak{S}$ contains all one-dimensional subspaces of $\mathcal{P}$ and hence it is nonempty. Since $\mathfrak{S}$ is of the finite property, it follows from Zorn's lemma that there is a maximal element $\mathcal{P}_0$ in $\mathfrak{S}$. Suppose that $\mathcal{P}_0$ does not coincide with $\mathcal{P}$ and let $p' \varepsilon \mathcal{P} - \mathcal{P}_0$. For this p' we can define the measure $q_{p'}$ on T and put $q_{\alpha p' + p} = \alpha q_{p'} + q_p$ for $p \varepsilon \mathcal{P}_0$. This q_p is well defined for all $p \varepsilon$ the subspace $\{p', \mathcal{P}_0\}$ spanned by $\mathcal{P}_0$ and p' and satisfies (6.2) for all $p \varepsilon \{p', \mathcal{P}_0\}$. Therefore $\psi(c, p) = \int c\, dq_p$ is a linear functional satisfying (6.1), which shows that the linear subspace $\{p', \mathcal{P}_0\}$ belongs to $\mathfrak{S}$ again. This is a contradiction with the maximality of $\mathcal{P}_0$ in $\mathfrak{S}$. Thus we have $\mathcal{P}_0 = \mathcal{P}$, or in other words, $\mathfrak{S}$ contains $\mathcal{P}$ itself.

THEOREM 4. *Suppose that*

(i) $\mathbf{A}$ *is a σ-compact, locally compact and metrizable space;*

(ii) $L(\theta, a)$ *is a Borel measurable nonnegative real function of* $a \varepsilon \mathbf{A}$ *for any fixed* $\theta \varepsilon \Theta$;

(iii) $\mathfrak{L} = \{L(\cdot, a) : a \varepsilon \mathbf{A}\}$ *is a subset of* $\mathfrak{F}$ *which is homeomorphic to a σ-compact, locally compact metric space, in the relative topology of the pointwise convergence topology* $\mathfrak{J}$ *of* $\mathfrak{F}$;

(iv) *there is a mapping* τ *of* $\mathfrak{L}^*$, *the closure of* $\mathfrak{L}$, *into* $\mathfrak{L}$ *such that*

(a) *for any* $\theta \varepsilon \Theta$ *and any positive* α, *the set* $\{f \varepsilon \mathfrak{L}^* : L(\theta, \tau f) \leqq \alpha\}$ *is the common part of* $\mathfrak{L}^*$ *and a Baire set in the topology* $\mathfrak{J}$ *in* $\mathfrak{F}$,

(b) $L(\theta, \tau f) \leqq f(\theta)$ *for all* $\theta \varepsilon \Theta$ *and* $f \varepsilon \mathfrak{L}^*$, *where* $L(\theta, \tau f)$ *means the value of* $L(\cdot, a)$ *at* θ *for* $\tau f = L(\cdot, a)$.

Then $R = \{r(\cdot, \delta) : \delta \varepsilon \mathfrak{D}\}$ *is half-closed and so* $\mathfrak{D}$ *has the property* (W).

PROOF. For any decision function $\delta \varepsilon \mathfrak{D}$ we shall associate a positive measure on $\mathfrak{A}$:

$$\delta \circ p(A) = \int \delta(A : x) p(x) \lambda(dx),$$

for each $p \varepsilon \mathcal{P}$. Let V be a basic neighborhood in $\mathfrak{F}$ given by the manner in Section 2. From the assumption (ii), $\{a: L(\cdot, a) \varepsilon V\}$ is Borel measurable, and so the set $\{a: L(\cdot, a) \varepsilon M\}$ is Borel measurable for any Baire subste M of $\mathfrak{F}$. Let $\pi(M : \delta \circ p) = \delta \circ p(\{a: L(\cdot, a) \varepsilon M\})$. This is a signed measure defined on the σ-field of all Baire subsets M of $\mathfrak{F}$, but vanishes for M disjoint of the set $\mathfrak{L}$. Therefore the closure $\mathfrak{L}^*$ of $\mathfrak{L}$ in the topology $\mathfrak{J}$ is a thick set (for definition, see [4], p. 74) relative to $\pi(\cdot : \delta \circ p)$. Noticing that every Borel subset of $\mathfrak{L}^*$ can be regarded as an intersection of a Baire set and $\mathfrak{L}^*$ because of the separability of $\mathfrak{L}^*$, $\pi(\cdot : \delta \circ p)$ may be regarded as a signed measure on the σ-field of Borel subsets of $\mathfrak{L}^*$. We shall denote by f an element of $\mathfrak{L}^*$ and by $\theta(f)$ the value $f(\theta)$ of f at the point $\theta \varepsilon \Theta$. Then $\theta(f)$ is a continuous function of f with respect to the relative topology of $\mathfrak{J}$ in $\mathfrak{F}$, and

$$r(\theta, \delta) = \int_{\mathfrak{L}^*} \theta(f) \pi(df : \delta \circ p_\theta). \tag{6.3}$$

Let h be an element of R^*, the closure of R in $\mathfrak{J}$. For any neighborhood

$V = V(h: \theta_1, \cdots, \theta_k, \epsilon)$ of h, there exists an element $\delta \varepsilon \mathfrak{D}$ such that $r(\cdot, \delta) \varepsilon R \cap V$. Let us keep an element $p \varepsilon \mathcal{P}$ fix for a while and then consider a class

$$\Delta_V = \{\pi(\cdot : \delta \circ p) : r(\cdot, \delta) \varepsilon R \cap V\}$$

of signed measures on $\mathcal{L}^*$. Since $\mathcal{L}^*$ is compact, so is the set of the signed measures, bounded by $\|p\|_1$, in the weak topology. Take the closure Δ_V^* of Δ_V in the weak topology. Then the intersection $\bigcap_V \Delta_V^*$ is not empty, because $\{\Delta_V : V\}$ has the finite intersection property. Thus we have a signed measure $\pi^*(\cdot : p)$ on $\mathcal{L}^*$ belonging commonly in Δ_V^*. By the condition (a) of (iv), a signed measure $\bar{\pi}(\cdot : p)$ on $\mathcal{L}$ will be induced by $\pi^*(\cdot : p)$ through τ as follows: $\bar{\pi}(M : p) = \pi^*(\tau^{-1}M : p)$ for every Borel subset M of $\mathcal{L}$. Obviously we have

$$\int_{\mathcal{L}} \theta(f)\bar{\pi}(df: p) = \int_{\mathcal{L}} \theta(f)\pi^*(\tau^{-1}\, df: p) = \int_{\mathcal{L}^*} \theta(\tau f)\pi^*(df: p). \tag{6.4}$$

On the other hand, we can observe that $\int_{\mathcal{L}} u(f)\bar{\pi}(df: p)$, $u \varepsilon C_0(\mathcal{L})$, is a positive bilinear functional on $(C_0(\mathcal{L}), \mathcal{P})$ of norm 1. In fact, the integral $\int_{\mathcal{L}^*} u(f)\pi^*(df: p)$, $u \varepsilon C_0(\mathcal{L}^*)$, is a cluster point, in the weak topology, of the set of bilinear functionals $\int u(f)\pi(df: \delta \circ p)$, and hence it is bilinear. The positivity and the norm of functionals are preserved invariantly. Consequently $\int_{\mathcal{L}} u(f)\bar{\pi}(df: p)$ has also the same property.

From the assumptions (i), (ii), (iii) and the lemma, there is a decision function $\delta_0 \varepsilon \mathfrak{D}$ such that

$$\int_{\mathcal{L}} u(f)\bar{\pi}(df: p) = \int_X \int_{\mathbf{A}} u(L(\cdot, a))\delta_0(da: x)p(x)\lambda(dx) \tag{6.5}$$

for every $u \varepsilon C_0(\mathcal{L})$. This means that $\bar{\pi}(\cdot : p)$ is an induced measure on $\mathcal{L}$ by the mapping $a \rightarrow f(\cdot) = L(\cdot, a) \varepsilon \mathcal{L}$. Therefore this equation holds for a continuous function $\theta(f)$. Since $\theta(f) = L(\theta, a)$, we have

$$\int_{\mathcal{L}} \theta(f)\bar{\pi}(df: p) = \int_X \int_{\mathbf{A}} L(\theta, a)\delta_0(da: x)p(x)\lambda(dx). \tag{6.6}$$

Especially for $p = p_\theta \varepsilon \Pi$, we have, by (b) of the assumption (iv),

$$\int_{\mathcal{L}^*} \theta(\tau f)\pi^*(df: p_\theta) \leqq \int_{\mathcal{L}^*} \theta(f)\pi^*(df: p_\theta) \tag{6.7}$$

and, by the continuity of $\theta(f)$,

$$\int_{\mathcal{L}^*} \theta(f)\pi^*(df: p_\theta) \leqq \lim_{V \rightarrow h} \inf_{r(\cdot, \delta) \varepsilon V} \int_{\mathcal{L}^*} \theta(f)\pi(df: \delta \circ p_\theta) = h(\theta). \tag{6.8}$$

From (6.4) and (6.6)–(6.8), we have $r(\theta, \delta_0) \leqq h(\theta)$.

Remark. If we assume, in addition to the other conditions of Theorem 4, that

(v) *the class of sets* $\{a: L(\theta, a) < \alpha\}$, $\theta \varepsilon \Theta$, $0 < \alpha < \infty$, *generates the* σ*-field* $\mathcal{A}$ *of the Borel subsets of* $\mathbf{A}$,

we can prove Theorem 4 directly without using the lemma preceding Theorem 4.

In fact, (6.5) is equivalent to the coincidence of $\bar{\pi}(\cdot : p)$ and $\int_X \delta_0(\cdot : x)p(x)\lambda(dx)$ on $\mathfrak{A}$, and hence (6.6) is implied directly by the condition (ii) of the theorem. For the same reason, the following theorem can be proved in a similar way as Theorem 4 except for using the lemma.

THEOREM 4'. *Suppose that*

(1) $\mathbf{A}$ *is equipped with a topology* $\mathfrak{J}_1$, *relative to which* $\mathbf{A}$ *is* σ*-compact, locally compact and metrizable and the loss function* $L(\theta, a)$ *is Borel measurable on* $\mathbf{A}$ *for any* $\theta \varepsilon \Theta$;

(2) *there is a compact metric space* $\mathbf{A}^*$, *whose induced topology is denoted by* $\mathfrak{J}_2$, *having the following property:*

(a) $\mathbf{A}$ *is embedded in* $\mathbf{A}^*$ *as a dense subset in* $\mathfrak{J}_2$*-sense,*

(b) *for every* $\theta \varepsilon \Theta$, $L(\theta, a)$ *is continuous on* $(\mathbf{A}, \mathfrak{J}_2)$ *and has a continuous extension* $L^*(\theta, a)$ *onto* $(\mathbf{A}^*, \mathfrak{J}_2)$,

(c) *every Borel set in* $(\mathbf{A}, \mathfrak{J}_2)$ *is a Borel set in* $(\mathbf{A}, \mathfrak{J}_1)$;

(3) *there is a mapping* τ *of* $\mathbf{A}^*$ *into* $\mathbf{A}$ *such that*

(a) *for any Borel set* A *in* $(\mathbf{A}, \mathfrak{J}_1)$, *the inverse image* $\tau^{-1}(A)$ *is Baire measurable in* $(\mathbf{A}, \mathfrak{J}_2)$,

(b) $L(\theta, \tau a) \leqq L^*(\theta, a)$ *for every* $a \in \mathbf{A}^*$, *and* $\theta \varepsilon \Theta$.

Then $\mathfrak{D}$ *has the property* (W).

Theorem 4' covers the case of the statement [3], p. 80, Miscellaneous remark (6), due to LeCam.

EXAMPLE 6.1. The quadratic loss estimation of a real valued parameter is one of the cases of Theorems 4 and 4', and so the class of all estimates has the property (W). Furthermore the class D^* of all nonrandomized estimates has the property (W), since D^* is an essentially complete class in $\mathfrak{D}$ [1], page 294. In a later section (Example 7.3) we shall discuss this problem again.

EXAMPLE 6.2. Consider an interval estimation problem of a real valued parameter θ with the loss function $L(\theta, (\underline{\theta}, \bar{\theta})) = u(\bar{\theta} - \underline{\theta}) + \alpha v(\theta, \underline{\theta}, \bar{\theta})$, where $(\underline{\theta}, \bar{\theta})$, $\underline{\theta} < \bar{\theta}$, is an estimated interval, $u(t)$ a monotone nondecreasing left-continuous nonnegative function of $t > 0$, α a positive real and

$$v(\theta, \underline{\theta}, \bar{\theta}) = 1, \qquad \text{when } \theta < \underline{\theta} \text{ or } \bar{\theta} < \theta,$$

$$= 0, \qquad \text{when } \underline{\theta} \leqq \theta \leqq \bar{\theta}.$$

In the case of $u(t) = t$, the class $\mathfrak{D}$ does not have the property (W) and there is no minimal complete class in $\mathfrak{D}$. However if we assume that $u(t) = 0$ for $t < t_0$ for some $t_0 > 0$, $\mathfrak{D}$ has the property (W). We shall show this fact.

Let us denote by $\bar{U}$ the closure of the range U of the function $u(t)$. Clearly $\bar{U}$ consists of an at most countable number of closed intervals. As easily seen, the closure $\mathfrak{L}^*$ of the class $\mathfrak{L}$ of the loss functions in $\mathfrak{F}$ is the set of the functions of the following four forms:

(a) $f(\theta) = u + \alpha$, for some $u \varepsilon \bar{U}$;

(b) $f(\theta) = u(t) + \alpha v(\theta, \theta', \theta' + t)$ for some θ' and some $t > 0$;

(c) $f(\theta) = u(t+) + \alpha v(\theta, \theta', \theta' + t)$ for some θ' and some discontinuity point t of $u(t)$;

(d) $f(\theta) = \alpha, \quad$ for $\theta \neq \theta'$,

$= 0, \quad$ for $\theta = \theta'$ for some θ'.

The mapping τ of $\mathfrak{L}^*$ into $\mathfrak{L}$ is

$$\begin{aligned} \tau(f) &= u(t) + \alpha v(\theta, 0, t), && \text{when } f \text{ is of the form (a)}, \\ &= f(\theta), && \text{when } f \text{ is of the form (b)}, \\ &= u(t) + \alpha v(\theta, \theta', \theta' + t), && \text{when } f \text{ is of the form (c)}, \\ &= \alpha v(\theta, \theta' - \tfrac{1}{2}t_0, \theta' + \tfrac{1}{2}t_0), && \text{when } f \text{ is of the form (d)}. \end{aligned}$$

We can easily see that this mapping τ satisfies the conditions of Theorem 4. Thus $\mathfrak{D}$ has the property (W).

7. The property (W) of the closed subclass of $\mathfrak{D}$. Theorems 4 and 4′ are very powerful for the whole class $\mathfrak{D}$, but they do not answer the question whether a restricted subclass of $\mathfrak{D}$ satisfies the property (W). However in many practical problems statisticians' concern is about a subclass of $\mathfrak{D}$, as the class of unbiased estimates, the class of tests of level α, etc. Some of such subclasses do not have the property (W). Even closed subclasses of $\mathfrak{D}$ do not have this property.

EXAMPLE 7.1. Consider the case where the parameter space Θ, the action space $\mathbf{A}$ and the sample space X are all the real line and the loss function $L(\theta, a) = 0$ if $|\theta - a| < 1$, and $= 1$ otherwise. For this loss function $L(\theta, a)$, $\mathfrak{L} = \{L(\cdot, a) : a \varepsilon \mathbf{A}\}$ is apparently half-closed, and so, by Theorem 4, $\mathfrak{D}$ has the property (W). Let

$$\begin{aligned} a_n(x) &= x + n && \text{if } -1 < x < 1, \\ &= x - n && \text{if } n - 1 < x < n + 1, \\ &= x && \text{otherwise,} \end{aligned}$$

and $\delta_n(A : x) = 1$ if $A \ni a_n(x)$, and $= 0$ if $A \not\ni a_n(x)$. Consider the class $D = \{\delta_n\}$, $n = 1, 2, \cdots$, of decision functions and, for the simplicity of calculations, the family of Cauchy distributions $[\pi\{1 + (x - \theta)^2\}]^{-1}$ on X with the location parameter θ. Then D is not compact, but closed in $\mathfrak{D}$ in the regular topology. The linearly structured risk $r(\theta, \delta_n)$ of δ_n is as follows:

$$\begin{aligned} r(\theta, \delta_n) = \tfrac{1}{2} &+ P([-n - 1, -n + 1] \cap [\theta - 1, \theta + 1]) \\ &+ P([n - 1, n + 1] \cap [-\theta - 1, -\theta + 1]) \\ &- P([-1, 1] \cap [-\theta - 1, -\theta + 1]) \\ &- P([-1, 1] \cap [n - \theta - 1, n - \theta + 1]) \end{aligned}$$

and $f(\theta) = \lim_{n\to\infty} r(\theta, \delta_n) = \frac{1}{2} - P([-1, 1] \cap [-\theta - 1, -\theta + 1])$ is a single cluster point of $R = \{r(\cdot, \delta_n) : \delta_n \varepsilon D\}$, where $P(A) = \int_A [\pi(1 + x^2)]^{-1} dx$.

Therefore for $n > 1$ there is a $\theta < 0$ such that $r(\theta, \delta_n) > f(\theta)$, which shows that D does not have the property (W).

EXAMPLE 7.2. Consider an estimation problem of the real parameter θ in the uniform distribution $p_\theta(x) = 1$ if $\theta - \frac{1}{2} < x < \theta + \frac{1}{2}$; and $= 0$ otherwise, with the quadratic loss $(\theta - a)^2$. The class $D = \{a_n(\cdot)\}$ of nonrandomized estimates

$$a_n(x) = \text{sign}\,(x)\cdot|x|^n$$

is obviously closed in the space $\mathfrak{D}$. And a simple calculation shows us that the risk function $r(\theta, a_n(\cdot))$ has the limit:

$$\begin{aligned}\lim_{n\to\infty} r(\theta, a_n(\cdot)) &= \infty \qquad \text{if } |\theta| > \tfrac{1}{2},\\ &= \theta^2 \qquad \text{if } |\theta| \leqq \tfrac{1}{2},\end{aligned}$$

and we have, at $\theta = 0$, $r(0, a_n(\cdot)) > 0 = \lim_{n\to\infty} r(0, a_n(\cdot))$ for any positive integer n. This shows that D does not satisfy the property (W). Thus we see that the closedness of D does not imply the property (W) even if the loss function is quadratic.

The above two examples show that the property (W) is a more profound character than the topological character of D like the closedness. It seems to the author that the topological structures of D are mainly determined by the topology of the action space $\mathbf{A}$, whereas the property (W) is closely related to the property of the loss function $L(\theta, a)$ and the sample distribution family Π. We shall give a sufficient condition for the subclass D of $\mathfrak{D}$ having the property (W) in the rest of this section.

The family $\mathfrak{M}$ of the probability measures on the locally compact space $\mathbf{A}$ is topologized by the convergence of the integral $\int u(a)m(da)$, $m \in \mathfrak{M}$, for every $u \in C_0(\mathbf{A})$. This topology is usually called the weak topology of $\mathfrak{M}$. A subset M of $\mathfrak{M}$ is relatively compact in this topology if and only if for any positive ϵ there corresponds a compact subset C such that $m(C) > 1 - \epsilon$ for every $m \in M$. For any subset D of $\mathfrak{D}$ we shall denote by $D \circ p$ the set $\{\delta \circ p : \delta \in D\}$ of the probability measures on $\mathbf{A}$ for every $p \in \Pi$.

DEFINITION 7.1. A subset D of $\mathfrak{D}$ is said to be *homogeneous* relative to a distribution family Π if for any subset D^* of D the compactness of $D^* \circ p^*$ for some $p^* \in \Pi$ implies that of $D^* \circ p$ for all $p \in \Pi$.

The homogeneity is satisfied by $\mathfrak{D}$ when elements of Π are mutually absolutely continuous. In fact, let $\beta(\alpha)$ be the power of the most powerful test of level α for the hypothesis $p^*(\in \Pi)$ against $p(\in \Pi)$. Since p and p^* are mutually absolutely continuous, for any positive $\epsilon > 0$ there is a uniquely determined positive number α_0 such that $\beta(\alpha_0) = 1 - \epsilon$. Suppose that there is a compact subset C_0 of $\mathbf{A}$ such that $\delta \circ p^*(C_0^c) < \alpha_0$ for every $\delta \in D^* \subset \mathfrak{D}$, where C_0^c is the complementary of C_0. For such a $\delta \in D^*$, $\varphi(x) = \delta(C_0^c : x)$ can be regarded as a test function of level α_0, so that $\delta \circ p(C_0^c) = E_p[\varphi] < \beta(\alpha_0) = 1 - \epsilon$, which shows the homogeneity of $\mathfrak{D}$ for Π.

THEOREM 5. *Suppose that*

(i) *the action space* **A** *is a* σ-*compact, locally compact and metrizable space;*

(ii) *for any fixed* θ *the loss function* $L(\theta, a)$ *is a lower semicontinuous function of* a;

(iii) D *is a closed subset of* $\mathfrak{D}$ *being homogeneous relative to* Π;

(iv) *for any* $\theta \varepsilon \Theta$ *and for any positive number* n *there exists a compact proper subset* $C_{n,\theta}$ *of* **A** *such that*

$$n \leqq \inf_{a \notin C_{n,\theta}} L(\theta, a).$$

Then D *has the property* (W).

PROOF. It follows from the assumption (ii) that, for $\theta \varepsilon \Theta$, $r(\theta, \delta)$ is a lower semicontinuous function of δ (see [3], p. 75). Therefore by Corollary 1 of Theorem 3 it is sufficient to prove that for any $\theta \varepsilon \Theta$ and nonnegative k there is a compact subset $D_{k,\theta} \subset D$ such that $\inf_{\delta \notin D_{k,\theta}} r(\theta, \delta) > k$ holds. Consider the set

$$D_{k,\theta} = \bigcap_{n=1}^{\infty} \{\delta \,\varepsilon\, D: \delta \circ p_\theta(C_{n,\theta}) \geqq 1 - k/n\}$$

(which may be empty) of decision functions. This $D_{k,\theta} \circ p_\theta$ is a compact subset of $\mathfrak{M}$ in the weak topology, and so from the assumption (iii) $D_{k,\theta} \circ p_\theta$ is also compact for every $p_\theta \,\varepsilon\, \Pi$. From Lemma 2 of [3], p. 74, $D_{k,\theta}$ is relatively compact in the regular topology of $\mathfrak{D}$. Since D is closed in $\mathfrak{D}$, the closure $D_{k,\theta}^*$ of $D_{k,\theta}$ is a compact subset of D.

Let $\delta \,\varepsilon\, D - D_{k,\theta}^* \subset D - D_{k,\theta}$. By the definition of $D_{k,\theta}^*$, there corresponds a positive integer n such that $\delta \circ p_\theta(C_{n,\theta}) < 1 - k/n$. Therefore we have

$$\begin{aligned} r(\theta, \delta) &= \int_{\mathbf{A}} L(\theta, a)\delta \circ p_\theta(da) \\ &\geqq \int_{C_{n,\theta}^c} L(\theta, a)\delta \circ p_\theta(da) \\ &\geqq n \cdot \delta \circ p_\theta(C_{n,\theta}^c) > n(k/n) = k. \end{aligned}$$

Thus our theorem is proved.

EXAMPLE 7.3. Consider a problem of estimating a real valued parameter with quadratic loss, and suppose that the sample distribution has a positive density function. Since the homogeneity of D is an inherited character for subsets, Theorem 5 is available for every closed subset D of $\mathfrak{D}$, and hence D has the property (W). As stated at the end of Section 5, the class D_f of decision functions $\varepsilon\, D$ whose risk function is bounded by a function $f(\theta)$ is closed. Therefore D_f has the property (W). Moreover the intersection of D_f and the class D^* of nonrandomized decision functions is essentially complete in D_f, and so by Theorem 1 the class $D_f^* = D_f \cap D^*$ has the property (W).

REFERENCES

[1] BLACKWELL, DAVID and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

[2] HALMOS, PAUL R. (1951). *Measure Theory*. Van Nostrand, New York.
[3] LECAM, LUCIEN (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.
[4] LECAM, LUCIEN (1962). Notes for 258 (April 1962). Mimeographed.
[5] VARADARAJAN, V. S. (1963). Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109** 191–220.
[6] WALD, ABRAHAM (1950). *Statistical Decision Functions*. Wiley, New York.