

AN EXTENSION OF WALD'S THEORY OF STATISTICAL DECISION FUNCTIONS¹

BY L. LECAM

University of California, Berkeley

1. Introduction. The material of the present paper was developed, during the spring of 1953, primarily to meet pedagogical needs. It is similar to the contents of Chapters 2 and 3 of Wald's book [1]. The results are an extension of Wald's theory in the sense that some requirements of boundedness or even finiteness of the loss function are removed. Moreover, Wald's requirements of equicontinuity are replaced by a requirement of lower semicontinuity of the loss function.

In the first part of the paper it is shown that, under suitable assumptions, the set \mathfrak{D} of all decision functions can be identified with a convex subset of a certain topological vector space. If further assumptions are made on the loss function, the risk functions become lower semicontinuous linear functions defined on \mathfrak{D} . It is then easy to give conditions under which \mathfrak{D} , or some subset D of \mathfrak{D} , is compact.

The next section is devoted to proofs that convexity and compactness of the space of decision functions, together with lower semicontinuity of the risk functions, imply completeness of the intersection of the class of Bayes' solutions in the wide sense with the closure of the class of Bayes' solutions.

The methods of proof differ very little from the methods used by Wald [1], though it has been necessary to use slightly more general topological methods, for instance, to prove compactness instead of sequential compactness. Although it might be possible to extend the proofs given by Wald [1] or Karlin [2], [3] to the case considered here, it is on the whole simpler and shorter to start from the basic elementary lemmas.

2. Assumptions on the decision problem. In this section weakened forms of Assumptions 3.1 to 3.6 of [1] are stated. Let $X = \{X_i\}$ for $i = 1, 2, \dots$ be a set of random elements, not necessarily real or even vector valued. Let \mathfrak{X} be the space of values of X and let Ω be an arbitrary set of indices. We will suppose that there is given on \mathfrak{X} a σ -field \mathcal{G} with respect to which all the X_i 's are measurable, and that to each $\omega \in \Omega$ corresponds a probability distribution on \mathfrak{X} .

If the variables $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ are observed in this order, we will say that $\lambda = \{i_1, i_2, \dots, i_k\}$ is observed and restrict the notation λ to ordered sets of indices which can be observed in the order given by λ , the first variable observed being X_{i_1} , the second X_{i_2} , and so on. The variables $\{X_{i_1}, \dots, X_{i_k}\}$ determine on \mathfrak{X} a smallest σ -field $\mathcal{G}_\lambda \subset \mathcal{G}$ with respect to which they are measurable. For all practical purposes it is equivalent to say that an \mathcal{G} -measurable function $f(x)$ is \mathcal{G}_λ -measurable or that it is a function of $\{X_{i_1}, \dots, X_{i_k}\}$ only.

Received May 25, 1953, revised May 3, 1954.

¹ This paper was prepared with the partial support of the Office of Naval Research.

At a certain stage λ of experimentation (including the start), the statistician has to choose an element among a set of available actions. This set Δ_λ includes the available terminal decisions T_λ as well as the decisions on how to continue experimentation, say J_λ . The set J_λ will be identified with the set of indices of variables which could be observed in the next step. We assume the following.

ASSUMPTION 1. Whatever may be λ , the set J_λ is finite.

This condition implies that the set of variables which can be observed in not more than n steps is finite. Let Λ_n be the corresponding set of λ 's and let $\Lambda = \bigcup_n \Lambda_n$. Let \mathcal{G}_n be the smallest σ -field containing all the \mathcal{G}_λ for $\lambda \in \Lambda_n$. To simplify further formulas, we denote by P_ω^n and P_ω^λ the contractions of P_ω to \mathcal{G}_n and \mathcal{G}_λ , respectively.

ASSUMPTION 2. For every n there exists on \mathcal{G}_n a σ -finite measure μ_n such that, whatever may be $\omega \in \Omega$, the probability measure P_ω^n is absolutely continuous with respect to μ_n .

Assumption 2 will be the only restriction placed upon the strategies of nature. It is implied by (3.1) or (3.2) of [1], but is strictly weaker than either (3.1) or (3.2). Discussion of Assumption 2 is deferred until after the statement of Assumption 6.

ASSUMPTION 3. For every $\lambda \in \Lambda$, the terminal space T_λ is a metrisable space, locally compact and the union of a denumerable family of compact spaces.

The space Δ_λ will be considered as topological sum of T_λ and J_λ , this last set being supplied with a discrete topology.

If δ is a certain decision function, for a given n , a given $\lambda \in \Lambda_n$, and an $x \in \mathcal{X}$, we can consider the probability that λ be observed and that the next decision belongs to a specified subset S of Δ_λ . This defines a measure denoted below by $\varphi(\lambda, x, \delta)$ on a σ -field of subsets of Δ_λ . It will always be assumed that this σ -field is the σ -field of Borel subsets of Δ_λ . If u is a numerical function defined on Δ_λ , its integral with respect to the measure $\varphi(\lambda, x, \delta)$ will be denoted by $\varphi(\lambda, x, \delta) \circ u$.

Thus if u is the indicator of a Borel set $\bar{D}^t \subset T_\lambda$, then $\varphi(\lambda, x, \delta) \circ u$ corresponds exactly to the $p(d_1^e, d_2^e, \dots, d_k^e, \bar{D}^t | x, \delta)$ defined in Formula (1.3) of [1]. In the present notation, λ replaces the ordered set of d_i^e and u replaces \bar{D}^t . It will be convenient to denote by a_λ the function identically equal to unity on T_λ and zero on J_λ , and by b_λ the function identically equal to unity on $\Delta_\lambda = T_\lambda \cup J_\lambda$. Furthermore, let K_λ denote the linear space of bounded continuous functions defined on Δ_λ and vanishing outside a compact subset of Δ_λ .

ASSUMPTION 4. Whatever $\lambda \in \Lambda$ and whatever $u \in K_\lambda$, the integral $\varphi(\lambda, x, \delta) \circ u$ is an \mathcal{G}_λ -measurable function of x .

The purpose of this assumption is to give a meaning to integrals used below in the definition of a risk function. It also expresses the fact that $\varphi(\lambda, x, \delta)$ does not depend on variables which have not yet been observed.

ASSUMPTION 5. The cost of observing λ is a nonnegative, \mathcal{G}_λ -measurable function $C(\lambda, x, \omega)$. If λ' is an initial segment of λ then $C(\lambda', x, \omega) \leq C(\lambda, x, \omega)$. Moreover there exists a sequence $C_n^*(\omega)$ of nonnegative functions of ω such that $\lim_{n \rightarrow \infty} C_n^*(\omega) = \infty$ and $C_n^*(\omega) \leq C(\lambda, x, \omega)$ for every λ having at least n elements.

ASSUMPTION 6. If the statistician does not reach a terminal decision in a finite

number of steps, he pays an infinite amount. If the statistician reaches a decision t in n steps by observing λ he pays a total amount

$$C(\lambda, x, \omega) + W(\omega, t; \lambda)h(\lambda, x, \omega),$$

where C is the cost function defined above and W and h are extended numerical functions satisfying the following conditions:

- (i) $\inf_{\lambda \in \Lambda, t \in T_\lambda} W(\omega, t; \lambda) > -\infty$;
- (ii) h is a nonnegative \mathfrak{G}_λ -measurable function such that $\sup_x h(\lambda, x, \omega) \leq 1$ for every (λ, ω) .
- (iii) For each (λ, ω) , the function W is a lower semicontinuous function of t on T_λ .

REMARK. It is possible to make W depend directly on x by introducing conditions of strong measurability. For instance $W(\omega, t, \lambda)h(\lambda, x, \omega)$ could be replaced by $\sum_{j=1}^{\infty} W_j(\omega, t, \lambda)h_j(\lambda, x, \omega)$, with W_j and h_j nonnegative.

Before entering further developments we will discuss Assumption 2 and introduce some simplifying notation. We notice first that in Assumption 2, μ_n could as well be taken finite instead of σ -finite. Since μ_n will not enter by itself in the next section, this is irrelevant. The properties actually used in the proofs are much weaker, and Assumption 2 could be relaxed as indicated below. Assume that μ_n has been chosen finite. For each $\lambda \in \Lambda_n$ let \mathfrak{L}_λ be the space of equivalence classes of numerical functions defined on \mathfrak{X} , measurable with respect to \mathfrak{G}_λ and μ_n integrable. If the norm of $f \in \mathfrak{L}_\lambda$ is defined by $\|f\| = \int |f| d\mu_n$, then \mathfrak{L}_λ is a Banach space. The adjoint of \mathfrak{L}_λ is, according to (2), identical with the space of μ_n -equivalence classes of bounded \mathfrak{G}_λ -measurable functions on \mathfrak{X} .

It is convenient to reduce the arbitrariness of μ_n and \mathfrak{L}_λ , and take instead of \mathfrak{L}_λ the smallest L -space, say L_λ , containing the family $\{P_\omega^\lambda\}$. For the definition of this space and the properties used below, see [4] and [5]. Two \mathfrak{G}_λ -measurable functions φ_1 and φ_2 can be called P -equivalent if $\int |\varphi_1 - \varphi_2| dP_\omega^\lambda = 0$ for every $\omega \in \Omega$. Assumption 2 could then be replaced by the considerably weaker assumption that, whatever may be $\lambda \in \Lambda$, the space L_λ has for adjoint the space M_λ of P -equivalence classes of bounded \mathfrak{G}_λ -measurable functions. Thus such nonconventional families of distributions as the family of all discrete distributions on the interval $[0, 1]$ could conceivably be introduced in the theory.

Even under its restricted form, Assumption 2 is more general than, for instance, (3.1) of [1] since it does not place such a strong restriction on the dependence between successive observations. This allows the consideration of different groups of random variables as vector variables which might be very strongly related. Since the cost of observing a vector might very well be different from the total cost of observing the components separately, and since moreover some of the spaces T_λ might be empty, the scheme considered here is at least as general as the scheme considered in [1].

The spaces L_λ and M_λ just defined will be used in the rest of the paper. Let L_λ^+ denote the positive cone of L_λ , and let $L_\lambda(a)$ be the sphere $L_\lambda(a) =$

$\{f: \|f\| \leq a\}$. Furthermore, let $L_\lambda^+(a)$ be the common part of L_λ^+ and $L_\lambda(a)$. Similar notations will be used for the spaces M_λ and K_λ previously mentioned.

If $s \in M_\lambda$ and $f \in L_\lambda$, let $f \circ s$ denote the integral $f \circ s = \int s(x) df$. Let \mathfrak{M}_λ be the space of P -equivalence classes of \mathcal{G}_λ -measurable finite or infinite numerical functions on \mathfrak{X} .

Let S be a subset of \mathfrak{M}_λ such that, if $s_1, s_2 \in S$, then there exists $s_3 \in S$ satisfying $s_3 \geq s_1$ and $s_3 \geq s_2$. Assume moreover that the elements of S are all larger than a given $s_0 \in M_\lambda$. The assumption that M_λ is the adjoint of L_λ is known to be equivalent to the following: for each S with the precited properties, there exists an $\bar{s} \in \mathfrak{M}_\lambda$ such that $f \in L_\lambda^+$ implies $f \circ \bar{s} = \sup_{s \in S} f \circ s$. If $S \in M_\lambda(a)$, then $\bar{s} \in M_\lambda(a)$ and is uniquely determined. This \bar{s} will be called the *supremum* of S and denoted by $\bar{s} = \sup_{s \in S} s = \sup S$. By convention if λ is the empty set we will take M_λ and L_λ to be the real line.

It will be necessary to distinguish between the equivalence classes belonging to M_λ and functions belonging to these equivalence classes. Thus, if $s \in M$, a function of the equivalence class of s will be denoted by \bar{s} or by $s(x)$. Similar notation will be used for M_λ and \bar{M}_λ .

3. Representation of decision functions by families of linear mappings. If

Assumptions 1 to 6 are satisfied, as far as the value of the risk is concerned a decision function is adequately described if the measures $\varphi(\lambda, x, \delta)$ are given up to a P -equivalence. It will therefore be convenient to identify decision functions which differ only on P -null sets. For a given $\lambda \in \Lambda$ the measures $\varphi(\lambda, x, \delta)$ define linear mappings from K_λ to M_λ . These linear mappings satisfy the following conditions (the subscript λ has been omitted for easier reading).

- (a) $\varphi \circ u_1 + \varphi \circ u_2 = \varphi \circ (u_1 + u_2)$,
- (b) $\varphi \circ \alpha u = \alpha \varphi \circ u$ for every real α ,
- (c) $\|\varphi \circ u\| \leq \|u\|$,
- (d) if $u \in K^+$, then $\varphi \circ u \in M^+$.

For any element $s \in M$, let $|s| = \sup(0, s) - \inf(0, s)$ and for any mapping satisfying properties (a), (b), (c) let $\nu_\varphi = \sup_{u \in K(1)} |\varphi \circ u|$. It is easily seen that for a φ satisfying (a), (b), (c), and (d), we have

$$\nu_{\varphi_\lambda} = \varphi_\lambda \circ b_\lambda = \sup_{u \in K_\lambda^+(1)} \varphi_\lambda \circ u.$$

In these formulas the supremum is taken with the meaning as previously defined.

Let Δ_0 be the space of decisions available before experimentation starts. Let $\lambda = \{i_1, \dots, i_k\}$ and $(\lambda, j) = \{i_1, \dots, i_k, j\}$ with $j \in J_\lambda$, and let u_j be the indicator of j in Δ_λ . The decision procedures also satisfy:

- (e) $\varphi_0 \circ \Delta_0 = 1$,
- (f) $\varphi_\lambda \circ u_j = \varphi_{(\lambda, j)} \circ b_{(\lambda, j)}$,
- (g) $\varphi(\lambda, x, \delta) \circ u = [\varphi(\lambda, x, \delta) \circ b_\lambda] \int_{\Delta_\lambda} u(t) dF_x(t)$,

where F_x is for each x a probability distribution on Δ_λ . It is not difficult to see that every system of measures satisfying conditions (a) to (g) defines an essentially unique decision procedure. The only condition not directly expressible in terms of equivalence classes is condition (g). The purpose of this section is to show that under reasonable conditions it is automatically satisfied, and thus eliminate it. To this end consider the following system of mappings. For each n and each $\lambda \in \Lambda_n$ let φ_λ be a mapping (not depending on n) from K_λ to M_λ . Assume that each of the φ 's is a normed linear mapping from K_λ to M_λ , that is, φ_λ satisfied (a), (b), and

$$(c') \quad \|\varphi \circ u\| \leq m \|u\| \quad \text{for some } m.$$

Such a system will be denoted by ψ_n^* , and the sequence $\{\psi_n^*\}$ for $n = 0, 1, \dots$ by ψ^* . The set of all ψ_n^* 's will be denoted by Ψ_n^* and the set of ψ^* 's by Ψ^* . A system ψ_n^* in which the φ_λ 's satisfy (a) through (f) will be denoted by ψ_n ; similarly for ψ and the corresponding capital letters. The proof that every $\psi \in \Psi$ satisfies (g), so that it defines a decision function, is an immediate consequence of the following lemma. Let \mathfrak{X} , L_λ , M_λ and Δ_λ be as before, and for simplicity omit the subscript λ .

LEMMA 1. *Let K be a vector lattice of numerical functions defined on a set Δ . Assume that*

- (h) K is separable for the topology defined by the norm $\|u\| = \sup_{t \in \Delta} |u(t)|$, and
- (k) there exists a σ -field \mathfrak{B} on Δ such that every positive normed linear functional θ on K can be represented by $\theta \circ u = \|\theta\| \int_\Delta u(t) dF(t)$, where F is a σ -additive probability distribution on (Δ, \mathfrak{B}) .

Then every linear mapping φ from K to M satisfying (a), (b), (c), and (d) can be represented in an essentially unique manner by

$$\varphi \circ u = \nu_\varphi \text{ equivalence class of } \int_\Delta u(t) dF_x(t),$$

where F_x is for each $x \in \mathfrak{X}$ a probability measure on (Δ, \mathfrak{B}) .

This lemma is a modified version of Theorem 3 of Doob [6]. (See also Gelfand [7], part 2, par. 7, Thm. 1.) If Δ is locally compact, metrisable, and denumerable at infinity, then the space K of continuous functions with compact nucleus satisfies both (h) and (k). The separability of K is well known and (k) follows immediately from the Riesz representation theorem. For other spaces to which the conclusion of the lemma applies see [6] and the theory of the Daniell integral (for instance Saks [8], p. 328).

Since conditions (a) to (f) are preserved under convex combinations, we have

THEOREM 1. *Under Assumptions 1 to 4, every system $\psi \in \Psi$ of mappings satisfying (a) through (f) can be obtained in an essentially unique way by a decision procedure, and conversely. The space \mathfrak{D} of decision procedures satisfying Assumptions 1, 3, and 4 can thus be identified with the convex subset Ψ of the linear system Ψ^* .*

Since Ψ^* is a linear space, it is possible to define on it a number of interesting topologies. One such topology \mathfrak{J} is defined below. Under the conditions (3.1)–(3.6) of [1] the topology \mathfrak{J} coincides with the topology of regular convergence of [1]. Consider mappings φ_λ satisfying (a), (b), and (c'). On the set of such mappings we can define a topology \mathfrak{J}_λ by the generic neighborhoods

$$V\{\varphi_\lambda^0 \mid \lambda, u, f, \epsilon\} = \{\varphi_\lambda: |f \circ \varphi_\lambda \circ u - f \circ \varphi_\lambda^0 \circ u| < \epsilon, |f \circ \varphi_\lambda \circ a - f \circ \varphi_\lambda^0 \circ a| < \epsilon\},$$

where u ranges through K_λ and f through L_λ , and ϵ is a positive number. On Ψ_n^* define the topology \mathfrak{J}_n by replacing φ_λ by ψ_n^* in the preceding formula and letting λ run through Λ_n . For $m < n$ consider the projection $\psi_n^* \rightarrow \Pi_{m,n}\psi_n^*$ defined in the following way. If $\lambda \in \Lambda_m$, then to φ_λ corresponds φ_λ . If $\lambda = \{i_1, \dots, i_m, i_{m+1}, \dots, i_r\}$, with $m < r \leq n$, then to φ_λ corresponds $\varphi_{\lambda'}$ with $\lambda' = \{i_1, \dots, i_m\}$.

It is clear that $\Pi_{mn} = \Pi_{m\nu}\Pi_{\nu n}$ for $m < \nu < n$ and that the projections Π_{mn} are continuous for the topologies \mathfrak{J}_m and \mathfrak{J}_n . Let \mathfrak{J} be the topology defined on Ψ^* as projective limit of the \mathfrak{J}_n 's with system of projection the $\Pi_{m,n}$'s. (See, for instance, Lefschetz [9], p. 31.) For \mathfrak{J} , the linear system Ψ^* becomes a locally convex, Hausdorff vector space. Moreover, conditions (a) through (f) are preserved by passages to the limit under \mathfrak{J} ; hence Ψ is a closed convex subset of Ψ^* .

4. Compactness of the space of decision functions and lower semicontinuity of the risk functions. This section gives a necessary and sufficient condition that a subset D of \mathfrak{D} be compact in the topology \mathfrak{J} defined above. The word compact is used in the sense that every covering by open sets has a finite subcovering, *not* in the sequential sense. The conditions for compactness of D result from

LEMMA 2. *Let λ be fixed and let Φ be a family of mappings from K_λ to M_λ which satisfy conditions (a) through (d). Let Assumptions 1 to 4 be satisfied.*

Then a necessary condition that Φ be relatively compact for \mathfrak{J}_λ is that, whatever $f \in L_\lambda^+$ and whatever $\epsilon > 0$, there exists $u(f \circ \epsilon) \in K_\lambda^+(1)$ such that

$$f \circ [\nu_\varphi - \varphi \circ u(f, \epsilon)] \leq \epsilon \|f\| \quad \text{for every } \varphi \in \Phi.$$

A sufficient condition is that the preceding condition holds for every f in a subfamily of $\{P_\omega^\lambda\}$ generating the same L -space.

PROOF. If instead of the smallest L -space containing the $\{P_\omega^\lambda\}$ we had taken another space \mathfrak{L}_λ , then in the statement of the lemma P_ω^λ would have to be replaced by $f \in \mathfrak{L}_\lambda^+(1)$. Lemma 2 is of the same general nature as the well known Helly compactness theorem. The proof sketched below differs from the usual proofs of Helly's theorem in that the denumerable Helly selection principle has been replaced by the topological tool known variously as "ultrafilter" ([10], pp. 25, 59) or "universal net" [11]. The use of this tool is necessary to avoid separability assumptions.

The necessity of the condition is quite obvious. The proof of sufficiency follows. Let U be an ultrafilter on Φ (or a universal net of φ 's). Then $\lim_U f \circ \varphi \circ u = \alpha(f, u)$ exists for every $f \in L_\lambda$ and $u \in K_\lambda$. Conditions (a) through (d) are preserved when taking limits so that $\alpha(f, u)$ can be written $f \circ \varphi^0 \circ u$ where φ^0 is a mapping having

the properties (a) through (d). Moreover $\lim_U \nu_\varphi = \nu \geq \nu_{\varphi^0}$. For a given $f \in \{P_\omega\}$ we have

$$f \circ [\nu_\varphi - \varphi \circ u(f, \epsilon)] < \epsilon, \quad f \circ [\nu - \varphi^0 \circ u(f, \epsilon)] < \epsilon.$$

This implies $f \circ [\nu - \nu_{\varphi^0}] \leq \epsilon$, hence $f \circ [\nu - \nu_{\varphi^0}] = 0$ for every $f \in \{P_\omega^\lambda\}$. Consequently $g \circ [\nu - \nu_{\varphi^0}] > 0$ only if g has a nonzero part disjoint from every $f \in \{P_\omega^\lambda\}$. But then g does not belong to the minimal L -space containing $\{P_\omega^\lambda\}$. This implies $\nu = \lim_U \nu_\varphi = \nu_{\varphi^0} = \nu(\lim_U \varphi)$ so that U converges to φ^0 for the topology \mathfrak{J}_λ . This leads us to the following

ASSUMPTION 7. A set D of decision functions satisfies Assumption 7 if, for each $\epsilon > 0$, each $\lambda \in \Lambda$, and each $f \in \{P_\omega^\lambda\}$, there exists a $u \in K_\lambda$ satisfying

(i) $0 \leq u \leq 1$ and $u(t) = 0$ for $t \in J_\lambda$,

(ii) $f \circ \varphi_\lambda(\delta) \circ u \geq f \circ \varphi_\lambda(\delta) \circ a_\lambda - \epsilon$,

this last condition being fulfilled uniformly for every $\delta \in D$.

According to Lemma 2, in order that a set $D \subset \mathfrak{D}$ be compact for \mathfrak{J} it is necessary and sufficient that it be closed and satisfy Assumption 7.

We can now obtain a relation between \mathfrak{J} and the risk function. Let λ be a particular element of Δ_λ and let v be a lower semicontinuous function defined on Δ_λ and such that $v \geq 0$. If φ is a mapping having the properties (a) through (d), it can be extended to a class of Baire functions by the Daniell-Bourbaki procedure giving

$$\varphi \circ v = \sup_{u \in K_\lambda; u \leq v} \varphi \circ u,$$

the supremum being taken in the sense previously defined for measurable functions in \mathfrak{M}_λ . Under conditions (a) through (d), and assuming $f \in L_\lambda^+$, the following equality holds

$$f \circ \varphi \circ v = \sup_{u \in K_\lambda; u \leq v} f \circ \varphi \circ u = f \circ \left[\sup_{u \in K_\lambda; u \leq v} \varphi \circ u \right].$$

Thus for the topology \mathfrak{J}_λ the function $f \circ \varphi \circ v$ is a lower-semicontinuous function of φ . The same property still holds if v is bounded from below (that is $\inf_{t \in \Delta_\lambda} v(t) > -\infty$) instead of being nonnegative.

The risk function $R(\omega, \delta)$ is a sum of two terms, $R(\omega, \delta) = R_1(\omega, \delta) + R_2(\omega, \delta)$, where $R_1(\omega, \delta)$ is equal to zero if the probability that the process terminates is unity, and to infinity otherwise, and $R_2(\omega, \delta)$ is the amount paid when the process terminates. Since $R_2(\omega, \delta)$ is a sum of terms of the form $f \circ \varphi_\lambda \circ v$, and since by Assumptions 5 and 6 only a finite number of these terms can be negative, it is clear that $R_2(\omega, \delta)$ is lower semicontinuous on \mathfrak{D} for the topology \mathfrak{J} . Moreover, the probability of taking at least n steps is a continuous function of δ so that, by Assumption 5, $R(\omega, \delta)$ is also lower semicontinuous on \mathfrak{D} . It is also clear that the assumption of lower semicontinuity in (iii) of Assumption 6 cannot be weakened if $R(\omega, \delta)$ is to be lower-semicontinuous on the whole of \mathfrak{D} .

The results just established are collected in

THEOREM 2. *Let \mathfrak{D} be the family of decision functions satisfying Assumptions 1 to 4. If Assumptions 5 and 6 are also satisfied, and D is a subset of \mathfrak{D} , then*

(1) *\mathfrak{D} is a closed convex subset of the locally convex Hausdorff vector space Ψ^* supplied with the topology \mathfrak{J} . The risk function $R(\omega, \delta)$ is a lower semicontinuous linear function of $\delta \in \mathfrak{D}$ for each $\omega \in \Omega$. Furthermore,*

$$\inf_{\delta \in \mathfrak{D}} R(\omega, \delta) > -\infty \quad \text{for each } \omega \in \Omega.$$

(2) *In order that D be relatively compact in $(\mathfrak{D}, \mathfrak{J})$, it is necessary and sufficient that D satisfy also Assumption 7. In this case the closure \bar{D} and the closed convex hull \bar{D} of D in \mathfrak{D} are compact in $(\mathfrak{D}, \mathfrak{J})$.*

In view of Theorem 2, the problem of finding complete classes of solutions can be reduced to similar problems considered by Karlin [2] in an abstract setting. There is, however, some difficulty due to the fact that the risk function $R(\omega, \delta)$ can take strictly infinite values.

5. Complete classes of decision functions. Let A and B be two nonempty sets. In this section A will represent the "states of nature" and B the set of decision functions available to the statistician. Let R be an extended numerical function defined on $A \times B$. Let $H = \{x_1, \dots, x_k\}$ be a finite subset of A , and let $\{\beta_i\}$ for $i = 1, 2, \dots, k$ be k strictly positive real numbers such that $\sum_{i=1}^k \beta_i = 1$. Let $g = \{(x_i, \beta_i)\}$ for $i = 1, \dots, k$. Denote by $K(g, y)$ the function $K(g, y) = \sum_{i=1}^k \beta_i R(x_i, y)$. For any subset A' of A let $G_{A'}$ be the set of g 's corresponding to finite subsets of A' . If H is a finite subset of A , let G_H^+ denote the subset of G_H for which all the β_i 's are strictly positive numbers.

The theorems on the existence of complete classes given in the present paper are proved in two steps. It is first shown that a result is correct when the set of "states of nature" is reduced to a finite subset of A or G_A . The result is then extended to an arbitrary A by a limiting process, using the fact that the family Σ of finite subsets of A is directed and covers A . That is, if S_1 and S_2 belong to Σ , then there exists an $S_3 \in \Sigma$ such that $S_1 \cup S_2 \subset S_3$ and moreover $A \subset \bigcup_{S \in \Sigma} S$. The required tool for such a passage to the limit is given by the following entirely obvious proposition.

LEMMA 3. *Let u be an extended numerical function defined on G_A . Let G be a subset of G_A such that, for every $y \in B$, the inequality $K(g, y) \leq u(g)$ for every $g \in G$ implies $K(g, y) \leq u(g)$ for every $g \in G_A$. Let Σ be a directed family of subsets of G_A covering G . Finally let C be a subset of B having properties:*

(C₁) *C is supplied with a topology for which it is a compact topological space (in the Borel-Lebesgue sense);*

(C₂) *for every fixed $g \in G$ the function defined by $y \rightarrow K(g, y)$ is lower semicontinuous on C ;*

(C₃) *for every $S \in \Sigma$ and every $\epsilon > 0$, there exists a $y_* \in C$ such that $K(g, y_*) \leq u(g) + \epsilon$ for every $g \in S$.*

Then, there exists $y \in C$ such that $K(g, y) \leq u(g)$ for every $g \in G_A$.

A particular family Σ used below is obtained as follows. For a finite subset H

of A , let S be a convex subset of G_H^+ spanned by a finite number of points of G_H^+ . Let G_H be supplied with the usual Euclidean metric, and assume moreover that S is closed in G_H for this metric. It is clear that the family Σ_H of such subsets of G_H^+ is directed and covers G_H^+ .

Let y_0 be a particular point of B and assume that $K(g, y_0)$ is finite for every $g \in G_H^+$. It is also clear that $K(g, y_1) \leq K(g, y_0)$ for every $g \in G_H^+$ implies $K(g, y_1) \leq K(g, y_0)$ for every $g \in G_H$, provided that $K(g, y_1)$ is bounded from below on G_H .

In order to obtain completeness theorems, it will be sufficient to add, to the preceding lemma, the following result.

LEMMA 4. *Let H be a finite subset of A , say $H = \{x_1, \dots, x_k\}$. Assume that B is convex with respect to (H, R) , that is, whatever may be $y_1, y_2 \in B$ and $0 \leq \alpha \leq 1$, there exists $y \in B$ such that*

$$R(x, y) \leq \alpha R(x, y_1) + (1 - \alpha)R(x, y_2), \quad \text{for every } x \in H.$$

Assume moreover that $\inf_{y \in B} R(x_i, y) > -\infty$ for each $x_i \in H$. Let $a = \inf_{y \in B} \sup_{x \in H} R(x, y)$.

Then, whatever $b < a$, there exists $g \in G_H^+$ such that $K(g, y) > b$ for every $y \in B$.

Or, equivalently, if b' is such that whatever $g \in G_H^+$ there exist $y_0 \in B$ such that $K(g, y_0) \leq b'$, then whatever $a' > b'$ there exists $y \in B$ such that $K(g, y) < a'$ for every $g \in G_H$.

This lemma is well known for the case where $R(x, y)$ takes only finite values (see for instance Karlin [2]). For a proof readily adaptable to our case see Ville [12] and Kneser [13]. The lemma is obvious if H contains only two elements, and an extension by induction does not present any difficulties. We will make use of the following assumptions.

ASSUMPTION 8. For every $x \in A$, $\inf_{y \in B} R(x, y) > -\infty$.

ASSUMPTION 9. Whatever may be $y_1, y_2 \in B$ and $0 \leq \alpha \leq 1$, there exists $y \in B$ such that $R(x, y) \leq \alpha R(x, y_1) + (1 - \alpha)R(x, y_2)$.

ASSUMPTION 10. B is a compact topological space and $R(x, y)$ for each x is lower semicontinuous on B .

It is well known and obvious that these assumptions imply the completeness of the class of admissible solutions.

Let $h(g) = \inf_{y \in B} R(g, y)$ and let G_A^* be the subset of G_A for which $h(g) < +\infty$.

THEOREM 3. *Let Assumptions 8, 9, and 10 be satisfied. If $y_0 \in B$ is such that $\inf_{g \in G_A^*} K(g, y_0) - h(g) \geq \epsilon$ then there exists $y_1 \in B$ such that $K(g, y_1) + \frac{1}{2}\epsilon \leq K(g, y_0)$ for every $g \in G_A^*$. Hence the class of Bayes' solutions in the wide sense is complete.*

PROOF. It results from Lemma 4 that such a y exists for every finite subset of G_A^* . Consequently by Lemma 3 there exists a y_1 such that the condition be satisfied for every $g \in G_A^*$. This result is related to a result of Kiefer [14].

If Assumptions 1 to 6 are satisfied and the set D of decision functions considered is compact, Theorem 3 implies that the class of Bayes' solutions in the wide sense is complete. To prove this result, Kiefer used the boundedness of the loss functions.

If G is a convex subset of G_A supplied with some topology for which it is compact, and $K(g, y)$ is lower-semicontinuous for every $y \in B$, and moreover such that $h(g)$ is upper-semicontinuous on G , then $K(g, y) - h(g)$ must reach its minimum on G . For such subsets G , Bayes' solutions in the wide sense coincide with Bayes' solutions. This applies in particular to the sets $S \in \Sigma_H$ introduced after the statement of Lemma 3. Take then for C in Lemma 3 the closure of the class of Bayes' solutions and apply Lemma 3 twice, first with $\Sigma = \Sigma_H$ for each H and then for $\Sigma = \{G_H^+\}$, with H finite. This gives

THEOREM 4. *If assumptions (8), (9), and (10) are satisfied, the closure of the class of Bayes' solutions is a complete class.*

Let G_Ω be the set of a priori distributions with finite support on Ω . Let $g = \{\{\omega_i\}, i = 1 \cdots k; \{\alpha_i\}, i = 1 \cdots k\}$ be an element of G_Ω . Denote by $K(g, \delta)$ the corresponding risk function $K(g, \delta) = \sum_1^k \alpha_i R(\omega_i, \delta)$. From Theorem 4 we obtain at once

THEOREM 5. *Let Assumptions 1 to 6 be satisfied and let S be a convex subset of G_Ω . Let D be a class of decision functions satisfying Assumption 7 and let \bar{D} be its closed convex hull. Let $B \subset \bar{D}$ be the class of decision functions δ satisfying $\delta \in \bar{D}$ and such that $K(g, \delta) = \min_{\delta' \in \bar{D}} K(g, \delta')$ for some $g \in S$. Let \bar{B} be the closure of B . Let C be the class of δ 's satisfying*

$$\delta \in \bar{D}, \quad \inf_{g \in S} [K(g, \delta) - \min_{\delta' \in \bar{D}} K(g, \delta')] = 0.$$

Then, whatever may be $\delta_0 \in \bar{D}$, there exists $\delta_1 \in \bar{B} \cap C$ such that

$$K(g, \delta_1) \leq K(g, \delta_0) \quad \text{for every } g \in S.$$

REMARK. The preceding theorem is an extension of Wald's Theorems 3.17 and 3.18 ([1], p. 100). Usually the class S can be replaced by a convex class of a priori distributions which do not necessarily have finite support. Moreover, an inequality of the type $K(g, \delta_1) \leq K(g, \delta_0)$ for every $g \in S$ usually implies that the same inequality holds for a larger class.

This brings us close to Wald's Theorem 3.19 ([1], p. 101). It is, however, clear that Wald meant to assume the convexity of the class ζ of a priori distributions used in this theorem. Wald's proof uses the convexity of ζ ; it is easy to show on examples that without this assumption the conclusion of the theorem does not generally hold. This correction being made, it is possible to use our Theorem 5 to obtain extensions of Wald's Theorem 3.19.

6. Miscellaneous remarks.

(1) The theorems given in this paper are generalizations of Wald's [1] Theorems 3.17 and 3.18. Assumptions on Ω are required to obtain the equivalent of his Theorem 3.20. The same is true of Wolfowitz's [15] theorems on ϵ -complete classes.

(2) Wald [1] assumed that the spaces L_n of equivalence classes of integrable functions on $(\mathfrak{X}, \mathfrak{A}_n, \mu_n)$ are separable. In this case our Assumption 2 is certainly satisfied. Moreover, the topology \mathfrak{J} on \mathfrak{D} admits at each point a countable basis,

so that a compact subset of \mathfrak{D} is also sequentially compact. Consequently every admissible decision function is a limit of a *sequence* of Bayes' decision procedures.

In this case it is possible to use general a priori distributions on Ω instead of the discrete distributions considered in the present paper since, according to Fatou's lemma, $\int R(\omega, \delta) d\xi(\omega)$ is lower semicontinuous in δ .

(3) In many problems the spaces $(\mathfrak{X}, \mathfrak{A})$ are countable discrete spaces. In such cases the method used here is unnecessarily involved. The topologies \mathfrak{J}_λ become topologies of pointwise convergence of the measures $\varphi(\lambda, x, \delta)$.

Thus the separability of K_λ is no longer necessary. For instance, it is possible to choose for T_λ any compact space, or even any completely regular space, and then to let K_λ be, for instance, the space of all continuous functions on Δ_λ .

(4) We will now give another example of application of Lemma 3. Suppose that Assumptions 1 to 6 are satisfied and assume that T_λ and $W(\omega, t; \lambda)$ do not depend on λ . It frequently happens that under such conditions the statistician would be able to find probability measures $\delta(x, \lambda)$ defined on T such that

$$r(\omega, \delta, \lambda) = E\{\delta(x, \lambda) \circ W(\omega, t) \mid \omega\}$$

is finite for each $\omega \in \Omega$, provided only that λ is large enough, say provided that the number of elements in λ be larger than some $n(\omega)$. It is clear that under circumstances, whatever may be the finite set $\{\omega_1, \dots, \omega_k\}$ and whatever $\epsilon > 0$ and whatever may be the decision function δ_0 , there exists a finite N and a decision function δ_N terminating surely in not more than N steps such that

$$R(\omega_i, \delta_N) \leq R(\omega_i, \delta_0) + \epsilon \quad \text{for every } \omega_i \in \{\omega_1, \dots, \omega_k\}.$$

If we assume that T is compact this implies that there exists a decision function δ_1 which is Bayes' among those terminating in not more than N steps and which satisfies

$$R(\omega_i, \delta_1) \leq R(\omega_i, \delta_0) + 2\epsilon, \quad i = 1, 2, \dots, k.$$

According to Lemma 3 the closure of the class of solutions which for some N are Bayes' among those requiring less than N steps is then complete. This has immediate applications to sequential analysis.

(5) The assumption of compactness used in Lemma 3, Theorem 3, and Theorem 4 is obviously too strong. With the notation of Lemma 3, let \mathfrak{F} be the space of all extended numerical functions on G_A topologized by the topology of pointwise convergence on G_A . The space \mathfrak{F} is a compact Hausdorff space for this topology. For each $y \in C$ let W_y be the function of g defined by $g \rightarrow K(g, y)$ and let F be the family $F = \{W_y; y \in C\} \subset \mathfrak{F}$. Furthermore, let \bar{F} be the closure of F in \mathfrak{F} . The only property actually used in the proofs of Lemma 3, Theorem 3 and Theorem 4 is that for every $v \in \bar{F}$ there exists a $W \in F$ such that $W(y) \leq v(g)$ for every $g \in G_A$. This property is the analogue of the property of "weak compactness" used by Wald ([1] p. 53); it will be referred to as property (W).

Similarly in Theorem 5 the closure \bar{B} of the class of Bayes' procedures could

be replaced by any class B^* containing B and having the property (W) . It is easy to see that whenever \tilde{D} satisfies (W) the same is also true of the class of admissible solutions of \tilde{D} , although these classes might not be compact.

(6) In many problems it is convenient to compact the spaces T_λ instead of restricting the set of decision functions to a compact subset of \mathfrak{D} . For instance, let T_λ be precompact and let its compacted \tilde{T}_λ be metrisable. Furthermore, assume that for every sequence $\{t_n\} \subset T_\lambda$ converging to a point $t_0 \in \tilde{T}_\lambda - T_\lambda$ the following equality holds

$$\liminf_{n \rightarrow \infty} W(\omega, t_n; \lambda) = \sup_{t \in \tilde{T}_\lambda} W(\omega, t; \lambda).$$

Take for each (ω, λ) the largest lower semicontinuous extension of $W(\omega, t, \lambda)$ on \tilde{T}_λ . These Assumptions 1 to 7 are satisfied so that the set $\tilde{\mathfrak{D}}$ of measurable decision functions obtained by replacing T_λ by \tilde{T}_λ is compact. However, if for instance T_λ is a Borel subset of \tilde{T}_λ , every $\delta_0 \in \mathfrak{D}$ is dominated by some $\delta_1 \in \mathfrak{D}$. In this case \mathfrak{D} has the property (W) without being compact.

As a simple example consider the case where T is for each λ the whole real line R . Then \tilde{T}_λ can be taken as the extended real line (compactified by the adjunction of points at infinity). Let ω be a real valued parameter and let $W(\omega, t; \lambda) = (\omega - t)^2$. Then the class \mathfrak{D} is not compact but $\tilde{\mathfrak{D}}$ is compact and every $\tilde{\delta} \in \tilde{\mathfrak{D}}$ is dominated by the δ obtained by replacing infinite values of the estimate by say zero. The preceding argument gives practical content to a remark of Karlin [3] that "every game can be completed to a game having a value."

(7) After this paper was written, an article by M. N. Ghosh [16] was brought to the author's attention. Ghosh's assumptions are stronger than those of the present paper. For instance, his Assumptions IV and VII imply Assumption 3 of the present paper, and his Assumptions III and IV are substantially stronger than our Assumption 6. Ghosh's theorems on compactness and existence of complete classes are special cases of Theorems 2 and 5 of the present paper.

(8) To give a simple illustration of the relevance of the results of the present paper, consider the problem of estimating the mean of a normal population with known variance from a sample of fixed size. Take T and Ω to be the real line. Let $W(\omega, t)$ be any positive nondecreasing function of $\tau = |\omega - t|$, continuous on the left and such that $\lim_{\tau \rightarrow \infty} W(\tau) = \infty$. For instance $W(\omega, t)$ can be taken equal to $(\omega - t)^2$. Then

- (a) Wald's intrinsic topology on T is usually the discrete topology on T .
- (b) If W is not continuous the assumptions of [16] are not necessarily satisfied.
- (c) According to Theorem 2 the subclass of decision functions satisfying $R(\omega_0, \delta) \leq M < \infty$ for one arbitrary given ω_0 is a closed convex compact subset of \mathfrak{D} .
- (d) While \mathfrak{D} itself is not compact the class of estimates $\tilde{\mathfrak{D}}$ obtained by allowing infinite values for t is compact in the corresponding topology. According to Remark (6), \mathfrak{D} satisfies also property (W) .
- (e) The class B^* obtained by taking the closure in $\tilde{\mathfrak{D}}$ of the class of Bayes'

solutions and then replacing infinite values of the estimate by zero or even by the value of any finite estimate, is a complete class for \mathfrak{D} .

Instead of estimating the mean ω by a point estimate, it might be desirable to give confidence intervals for ω . For instance, it might be desirable to obtain confidence intervals of given length $2l$. In such a case take $W(\omega, t)$ to be zero if $|\omega - t| \leq l$ and unity if $|\omega - t| > l$. Again allow t to take infinite values. The Assumptions 1 to 7 are then satisfied so that Theorem 5 applies directly to \mathfrak{D} .

REFERENCES

- [1] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.
- [2] S. KARLIN, "Operator treatment of the minimax principle," *Contribution to the theory of games*, Princeton University Press, 1950.
- [3] S. KARLIN, "The theory of infinite games," *Ann. Mathematics*, Vol. 58 (1953), pp. 371-401.
- [4] N. BOURBAKI, *Intégration*, Actualités scientifiques et industrielles, No. 1175, Hermann et Cie., Paris, 1952.
- [5] F. RIESZ, "Sur la théorie générale des opérations linéaires," *Ann. Mathematics*, Vol. 41 (1940), pp. 174-206.
- [6] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, 1953, pp. 29-31.
- [7] I. GELFAND, "Abstrakte Functionen und lineare Operatoren," *Mat. Sbornik*, Vol. 46 (1938), pp. 235-286.
- [8] S. SAKS, *Theory of the Integral*, 2nd. ed., Hafner Publishing Co., New York, 1937.
- [9] S. LEFSCHETZ, *Algebraic Topology*, Amer. Math. Soc. Colloq. Publ., Vol. 27, 1942.
- [10] N. BOURBAKI, *Topologie Générale*, Actualités scientifiques et industrielles, No. 858, Hermann et Cie., Paris, 1940.
- [11] J. L. KELLEY, "Convergence in topology," *Duke Math. J.*, Vol. 17 (1950), pp. 277-283.
- [12] J. VILLE, "Note sur la théorie des jeux où intervient l'habilité des joueurs," *Traité du calcul des probabilités*, Vol. 4, No. 2, pp. 105-113. Gauthier-Villars, Paris, 1938.
- [13] H. KNESER, "Sur un théorème fondamental de la théorie des jeux," *C. R. Acad. Sci. Paris*, Vol. 234 (1952), pp. 2418-2420.
- [14] J. C. KIEFER, "On Wald's complete class theorems," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 70-75.
- [15] J. WOLFOWITZ, "On ϵ -complete classes of decision functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 461-465.
- [16] M. N. GHOSH, "An extension of Wald's decision theory to unbounded weight functions," *Sankhyā*, Vol. 12 (1952), pp. 8-26.