INVARIANCE THEORY AND A MODIFIED MINIMAX PRINCIPLE

By Oscar Wesler[1]

Stanford University and *University of Michigan*

TABLE OF CONTENTS

1. Introduction and summary. One of the unpleasant facts about statistical decision problems is that they are generally much too big or too difficult to admit of practical solutions, a fact that is threatening to widen even further the gap between the theory and application of this brave new discipline. Briefly, the situation is this. For each possible decision procedure φ, the statistician is concerned only with the values $\rho(\omega, \varphi)$ of the risk function as ω ranges over the set Ω of all possible states of nature, so that a choice of a decision procedure amounts to a choice of a risk function. The obvious difficulty of comparing functions in the search for a best procedure now arises, constituting a major problem for the statistician. The Bayes and minimax principles, it should be noted, represent but two extremes to which the statistician can go to get around this difficulty rather than to meet it, the one assuming complete knowledge of an a priori probability distribution λ of the possible states ω, the other assuming least favorable circumstances about ω, so that in either case one considers only a single number per procedure rather than the entire function—the Bayesian the average risk with respect to λ, the minimax-man in all timidity the supremum of the risk function—comparisons thus becoming trivial in principle and obliging one to look simply for that procedure which minimizes these numbers. Inasmuch as the situations occurring in practice with regard to prior knowledge about ω usually lie between the two extremes just described, to this extent at least both principles are open to criticism. In view of the nature of the difficulty in choosing among procedures, the notion of admissible or complete classes of strategies is generally felt to provide the most satisfactory solution. Whereas it may be difficult to say what to do in a statistical decision problem, it is generally easier

Received March 7, 1958.

[1] This paper is the author's doctoral dissertation, accepted by Stanford University and published as Technical Report No. 35, Statistics Laboratory, Stanford University, October 6, 1955; research carried out under the sponsorship of the Office of Naval Research.

to say what not to do, so that the statistician separates out from consideration all inadmissible strategies and presents the practical man with what is at best a minimal essentially complete class of procedures. The choice of one from among these admissible procedures is then left to the best judgment, intuition, and past experience of the practical man. If the class is a small one, we have then achieved everything one can ask for, for the actual choice will then easily be made. The difficulty, however, is that the classes are usually much too big to be of real help.

The trouble, in a word, lies in the fact that it is the space Ω itself of possible states that is generally too big: one simply cannot look at and assess all the values of ω for each decision function φ. Now when problems turn out to be too big for practical purposes, it is natural to look for ways of cutting them down to size by methods of simplification or approximation in which very little of the original problem is lost. It is precisely such a cutting down or slicing up of the problem that we propose to treat in this paper, in the hope that it may help bring the theory and practice of statistical decision functions somewhat closer together.

The Modified Minimax Principle. The minimax principle, which looks only at the single value $\sup_\omega \rho(\omega, \varphi)$, suffers from the defect of being an over-simplification. Yet it suggests, by means of a simple modification, a natural way of approximating to the problem. This is to cut Ω up, to partition it into sets or "slices" Ω_s, s running over an index set S, and then look at $\sup_{\omega \varepsilon \Omega_s} \rho(\omega, \varphi) = \alpha(s, \varphi)$ for each s in S, so that corresponding to the partition $\Omega = \bigcup_{s \varepsilon S} \Omega_s$ we look only at the values $\alpha(s, \varphi)$ for $s \varepsilon S$ instead of at $\rho(\omega, \varphi)$ for all ω. The range of ρ_φ is thus replaced by the smaller range of α_φ, making comparisons of procedures that much easier. It is this slicing up of Ω (or its replacement by the smaller class S) and consequent simplification of the risk functions in the above way that we call the modified minimax principle. (By way of analogy, one might think of upper Darboux sums approximating Riemann integrals.) The reduced game can then be treated as any other game: one can play Bayes or minimax in it, or attempt to delineate its admissible strategies.

If the "slicing principle" used is such that the supremum of ρ_φ over each slice is not much different from the values within the slice, or has some other reasonable property, then very little is lost. The question of what are reasonable or natural slicing principles is clearly of primary importance here, and we shall present what we believe are several of them.

The theory of invariance provides us with the most powerful of these slicing principles and will play a central role in our considerations. The natural slicing of Ω into its orbits under the group leads us to what appears to be the best possible generalization of the Hunt-Stein theorem, and to its most natural setting: namely to the theorem that under certain regularity conditions which are often met in practice the invariant procedures form a complete class in the sense of the sliced up risk functions. The theory of composite hypotheses is also discussed in this light, and an example of theoretical interest is given illustrating these concepts, in which a difficult problem undergoes a striking simplification.

Finally, the use of previous experience as a slicing principle is discussed, and related to a purely game-theoretic model which we have constructed for the modified minimax principle and which we have called a *mixed game.*

Though these slicing principles appear to be among the most important, the search for others continues. A method of approximation and simplification having been established, it remains to be seen whether these principles and available numerical methods can be combined to make an effective instrument in practice.

2. The use of invariance as a slicing principle. We begin with a brief description of the statistical and group-theoretic setup. Background material and further remarks on statistical games can be found in Blackwell and Girshick [1] and Wald [2]. For measure-theoretic and related notions, we refer to Halmos [3]. The formidability of notation, while unavoidable in this subject, is more apparent than real, and we shall try to clarify matters with a running commentary.

DEFINITION 1. By a statistical problem $(Z, \mathcal{B}, \Omega, P, A, \mathcal{A}, L)$ is meant

i) A set Z of all possible experimental outcomes, a Borel field $\mathcal{B}$ of subsets of Z, a set Ω of states of nature, and a function P defined on $\mathcal{B} \times \Omega$ such that for each $\omega \varepsilon \Omega$, P_ω is a probability measure on $(Z, \mathcal{B})$. $(Z, \mathcal{B}, \Omega, P)$ is called the sample space

ii) A set of actions A available to the statistician, together with a Borel field $\mathcal{A}$ of subsets of A

iii) A loss function L defined on $\Omega \times A$ such that $L(\omega, a)$ is the loss to the statistician when he takes action a and ω is the true state of nature. For each $\omega \varepsilon \Omega$, L_ω is assumed to be a non-negative $\mathcal{A}$-measurable function. (A mapping $f{:}M \rightarrow N$ of one measurable space $(M, \mathfrak{M})$ into another $(N, \mathfrak{N})$ is said to be $\mathfrak{N}$-$\mathfrak{M}$ measurable if the inverse image of every $\mathfrak{N}$-set is an $\mathfrak{M}$-set. If $(N, \mathfrak{N})$ is Euclidean, with the usual Borel sets, we call the mapping simply $\mathfrak{M}$-measurable. If E is a subset of M, we shall write $f''E$ for the image of E under f.)

DEFINITION 2. $(\mathcal{G}, \gamma_Z, \gamma_\Omega, \gamma_A)$ is said to be an admissible group on the statistical problem, or the problem is said to be invariant under the group $\mathcal{G}$ (for short) if

i) $\mathcal{G}$ is a group

ii) γ_Z is a homomorphism of $\mathcal{G}$ into the class of all $\mathcal{B}$-$\mathcal{B}$ measurable 1-1 transformations of Z onto itself, γ_Ω is a homomorphism of $\mathcal{G}$ into the class of all 1-1 transformations of Ω onto itself, γ_A is a homomorphism of $\mathcal{G}$ into the class of all $\mathcal{A}$-$\mathcal{A}$ measurable 1-1 transformations of A onto itself, where for each $g \varepsilon \mathcal{G}$ we shall write g_Z for $\gamma_Z(g)$, g_Ω for $\gamma_\Omega(g)$, and g_A for $\gamma_A(g)$

iii) for each $g \varepsilon \mathcal{G}$, $B \varepsilon \mathcal{B}$, $\omega \varepsilon \Omega$, we have $P(g_Z'' B \mid g_\Omega(\omega)) = P(B \mid \omega)$ and

iv) for each $g \varepsilon \mathcal{G}$, $\omega \varepsilon \Omega$, $a \varepsilon A$, we have $L(g_\Omega(\omega), g_A(a)) = L(\omega, a)$.

What this definition says is that each element g of group $\mathcal{G}$ is in effect three simultaneous permutations or relabelings, g_Z, g_Ω, and g_A, of the elements of Z, Ω, and A, respectively, together with their respective Borel fields, under which probabilities of sets and losses due to actions are invariant. The homomorphisms further imply that we have three groups $\mathcal{G}_Z$, $\mathcal{G}_\Omega$, and $\mathcal{G}_A$ of such permutations.

As a simple example illustrating these concepts, we may think of the following: Let Z be the real line, $\mathcal{B}$ the ordinary Borel sets, Ω the real line, P_ω the normal distribution with mean ω and variance 1, $A = Z$, $\mathcal{A} = \mathcal{B}$, and $L(\omega, a) = \psi(\omega - a)$, i.e., the loss is some function depending only on the difference $\omega - a$. The problem is easily seen to be invariant under the full translation group $\mathcal{G}$ each element of which gives rise to three identical translations of the three real lines Z, Ω, A by the same real number. The reader may generalize this example at once to the n-dimensional case, with P_ω the multivariate normal distribution $N(\omega, I)$, vector $\omega = (\omega_1, \cdots, \omega_n)$ of means, covariance matrix I, and translations $z \to z + g = (z_1 + g, \cdots, z_n + g)$.

DEFINITION 3. A randomized decision procedure is a function φ on $\mathcal{A} \times Z$ to the unit interval of reals [0, 1] such that for each z, φ_z is a probability measure on $(A, \mathcal{A})$, i.e., $\varphi(T \mid z)$ is the probability of taking an action in $T \varepsilon \mathcal{A}$, given z, and such that for fixed T, φ_T is a $\mathcal{B}$-measurable function of z. We write Φ for the class of all randomized decision functions.

DEFINITION 4. For any state of nature ω and decision procedure φ, the value of the risk function ρ is given by

$$\rho(\omega, \varphi) = \int\int L(\omega, a)\, d\varphi(a \mid z)\, dP_\omega(z).$$

DEFINITION 5. For each $g \varepsilon \mathcal{G}$, we make correspond to each procedure φ a new procedure $g_\Phi \varphi$ given by

$$(g_\Phi\varphi)(T \mid z) = \varphi(g''_A T \mid g_Z(z)).$$

It is an easy but very important consequence of our definitions that the risk function ρ is invariant, or as we say, transforms "correctly," under the group by means of the formula

$$\rho(\omega, g_\Phi\varphi) = \rho(g_\Omega(\omega), \varphi) \qquad \text{for every } g \varepsilon \mathcal{G}.$$

The simple verification of this is as follows:

$$\rho(\omega, g_\Phi \varphi) = \int\int L(\omega, a)\, d\varphi(g_A(a) \mid g_Z(z))\, dP(z \mid \omega).$$

Writing $L(\omega, a) = L(g_\Omega(\omega), g_A(a))$ and $dP(z \mid \omega) = dP(g_Z(z) \mid g_\Omega(\omega))$, then relabeling $g_Z z = z'$ and $g_A(a) = a'$, we get

$$= \int\int L(g_\Omega(\omega), a')\, d\varphi(a' \mid z')\, dP(z' \mid g_\Omega(\omega))$$

$$= \rho(g_\Omega(\omega), \varphi).$$

DEFINITION 6. A decision procedure φ is said to be invariant under the group if $g_\Phi\varphi = \varphi$ for every $g \varepsilon \mathcal{G}$, i.e., if

$$\varphi(g''_A T \mid g_Z(z)) = \varphi(T \mid z) \qquad \text{for all } g, z, T.$$

Note, then, that if φ is invariant, $\rho(\omega, \varphi) = \rho(g_\Omega(\omega), \varphi)$ for every $g \varepsilon \mathcal{G}$.

A brief word at this point on the effect in general of an arbitrary group of permutations on an arbitrary set will give some direction to these considerations and will enable us to convey the essence and spirit of what is known as *the invariance principle.*

Let G be a group of permutations of the elements of a set X. The existence of a permutation sending one element of X into another is readily seen to establish an equivalence relation in X and gives rise to a useful partition of X into equivalence classes, known as "orbits," under the group. Formally, given $x \varepsilon X$, the orbit V_x to which it belongs is the set of elements y of X given by

$$V_x = \{y : y = gx \text{ for at least one } g \varepsilon G\}.$$

Now so far as the effect of G in its bearing on X is concerned, there is no distinguishing between elements of the same orbit. Taking an anthropomorphic view of the group, from the point of view of "the man in the group," all elements in an orbit look alike, the group in its dealings with the set displaying blindness to a mere matter of a difference in labels. To put it another way, when the man in the group looks at X, he sees only orbits.

The invariance principle. Returning to our statistical setup, the space Z of experimental outcomes breaks up into orbits under the influence of $\mathcal{G}$, more specifically under the permutation group $\mathcal{G}_Z$. Ω also breaks up into orbits under $\mathcal{G}_\Omega$. By the invariance principle is meant the adoption by the statistician of the viewpoint of the man in the group $\mathcal{G}$: he becomes himself the man in the group in that he too sees the problem in terms of orbits only. Specifically, this means that the only decision procedures he will employ are the invariant ones, that is to say, those procedures which, while free to vary at will from orbit to orbit of Z, must exhibit within each orbit complete consistency with respect to the groups $\mathcal{G}_Z$ and $\mathcal{G}_A$ as prescribed in Definition 6.

The main reason for restricting ourselves to invariant procedures has always been their undeniable plausibility. We are further encouraged by the fact that in many problems (under a finite group, say, and in other cases) an invariant decision procedure admissible within the class of all invariant procedures is known to be admissible among all procedures as well. Moreover, if we should be looking for minimax strategies, the Hunt-Stein theorem (to be described in the next section in terms of testing hypotheses) tells us that under certain weak assumptions a minimax invariant procedure is minimax among all procedures too. For us, however, there is now another and more compelling reason. It turns out that the use of invariance as a slicing principle, so that orbits rather than points become the basic elements under consideration, leads in our modified minimax sense to the best possible generalization of the Hunt-Stein result, namely, that the invariant procedures form a complete class in the sense of the sliced up risk functions. In the modified minimax sense, then, the use of invariance provides a striking simplification of the original problem, particularly when the orbits, hence the groups, are rather large.

It should be noted in this connection, by the remark in Definition 6 above, that risk functions ρ_φ for invariant procedures φ are constant over each orbit of Ω, as is to be expected, so that invariant procedures lose nothing under the modified minimax principle of taking suprema. Before presenting our main result in some detail, it will be convenient and appropriate to give a brief description of the invariance principle in the language of testing hypotheses, and to state the classical Hunt-Stein theorem in these terms.

3. The principle of invariance in testing hypotheses. Let $(Z, \mathcal{B}, \Omega, P)$ with $\Omega = \Omega_0 \cup \Omega_1$ a disjoint union, be a sample space, and suppose we want to test $H_0 : \omega \in \Omega_0$ against $H_1 : \omega \in \Omega_1$.

Let there exist a σ-finite measure ν on $(Z, \mathcal{B})$ such that there is a real-valued function p on $Z \times \Omega$ with

$$P(B \mid \omega) = \int_B p(z \mid \omega)\, d\nu\,(z) \qquad \text{for every } \omega \in \Omega \text{ and } B \in \mathcal{B},$$

i.e., we are assuming the existence of a σ-finite measure ν which dominates all our probability distributions P_ω, and p_ω is the Radon-Nikodym derivative $dP_\omega/d\nu$.

Definition 1. The group $\mathcal{G}$ is said to keep the testing problem invariant if $\omega \in \Omega_i$ implies $g_\Omega(\omega) \in \Omega_i$, $i = 0, 1$, for every $g \in \mathcal{G}$, that is to say, if each permutation g_Ω of the elements of Ω permutes the elements of Ω_0 and of Ω_1 separately. (Note that this implies for the testing problem that we are specializing the general statistical problem down to a two-action setup $A = \{0, 1\}$ with constant losses due to a wrong decision and $g_A = I$, the identity mapping, for every g.)

Definition 2. By a randomized test is meant a $\mathcal{B}$-measurable real-valued function φ on Z with values in the closed interval $[0, 1]$, where $\varphi(z)$ is the probability of rejecting H_0 when z is the observed sample point.

Definition 3. The test φ is said to be invariant under $\mathcal{G}$ if for all g and z $\varphi(g_Z(z)) = \varphi(z)$ except at most on a set of ν-measure 0, i.e., if φ is essentially constant on the orbits of Z. If the exceptional set of ν-measure 0 depends on g, then φ is called almost invariant.

Let $\mathcal{C}$ be a Borel field of subsets of $\mathcal{G}$. Two assumptions will always be made:

i) $(g, z) \to g_Z(z)$ is $\mathcal{B} - \mathcal{C} \times \mathcal{B}$ measurable

ii) $(g_1, g_2) \to g_1 g_2$ is $\mathcal{C} - \mathcal{C}^2$ measurable.

Definition 4. A measure μ on $(\mathcal{G}, \mathcal{C})$ is said to be right invariant if $\mu(Cg) = \mu(C)$ for every $g \in \mathcal{G}$ and every $C \in \mathcal{C}$. Left invariance is defined using $\mu(gC) = \mu(C)$.

As an example of an invariant measure we may take for $\mathcal{G}$ the additive group of reals, $\mathcal{C}$ the ordinary Borel sets, and μ Lebesgue measure on the reals. No invariant probability measure, however, exists for this group, which gives rise to the following useful limiting notion.

Definition 5. A sequence $\{\mu_n\}$ of probability measures on $(\mathcal{G}, \mathcal{C})$ is said to be asymptotically right invariant if $\lim_{n\to\infty}(\mu_n(Cg) - \mu_n(C)) = 0$ for every $g \in \mathcal{G}$ and $C \in \mathcal{C}$.

As an example we take the same group just given, and for each $C \in \mathcal{C}$ and $n = 1, 2, \cdots$, define the probability measure μ_n on $(\mathcal{G}, \mathcal{C})$ by

$$\mu_n(C) = (1/2n)\mu(C \cap [-n, n]),$$

i.e., we approximate Lebesgue measure μ by the conditional probabilities μ_n given that g belongs to the closed interval $[-n, n]$. Clearly the μ_n are all probability measures and $1/2n$ is the normalizing factor. To prove asymptotic invariance, we have, since clearly $Cg \cap [-n, n] = C \cap [-n - g, n - g]$,

$$| \mu_n(Cg) - \mu_n(C) | = \left| \frac{1}{2n} \mu(C \cap [-n - g, n - g]) - \frac{1}{2n} \mu(C \cap [-n, n]) \right|$$

$$\leq \frac{|g|}{2n} \to 0 \qquad \text{as } n \to \infty,$$

showing in fact that the approach to zero for each g is here even uniform in C, that is,

$$\lim_{n\to\infty} \sup_C | \mu_n(Cg) - \mu_n(C) | = 0 \qquad \text{for every } g \in \mathcal{G}.$$

(This example generalizes at once to the additive group $\mathcal{G}$ (real p-tuples) of a p-dimensional real linear space. One takes conditional probabilities of Lebesgue measure given the cubes $[(-n, \cdots, -n); (n, \cdots, n)]$, normalizing by $1/(2n)^p$, and, writing $g = (g_1, \cdots, g_p)$, showing that for n sufficiently large, in fact for $2n > \max_{i=1,\cdots,p} | g_i |$,

$$| \mu_n(Cg) - \mu_n(C) | \leq 1 - \prod_{i=1}^{p} \left(1 - \frac{|g_i|}{2n}\right) \to 0 \qquad \text{as } n \to \infty.$$

The Hunt-Stein theorem [4], in the language of testing problems, then reads:

THEOREM. *Let there exist, for the testing problem as just defined, an asymptotically (right) invariant sequence of probability measures* $\{\mu_n\}$ *on* $(\mathcal{G}, \mathcal{C})$. *Then there exists an invariant test* φ_0 *among all tests* φ *for which* $\int \varphi(z)p(z \mid \omega)\, d\nu(z) \leq \alpha$ *for all* $\omega \in \Omega_0$, *and which maximizes* $\inf_{\omega \in \Omega_1} \int \varphi(z)p(z \mid \omega)\, d\nu(z)$; *i.e., there exists an invariant minimax test* φ_0. (The probability with which a test φ rejects H_0 when ω is the true state of nature is given by $\beta_\varphi(\omega) = \int \varphi(z)p(z \mid \omega)\, d\nu(z)$. β_φ is called the *power function* of the test and is the only property of the test that interests us, bearing an obvious relation to the risk function. The original version of the Hunt-Stein theorem was in terms of most stringent tests. The connection with this version is easily seen if we subtract one minus the size α envelope power function $\beta_\alpha^*(\omega)$ from the loss function for $\omega \in \Omega_1$. The maximum risk for $\omega \in \Omega_1$ is then the stringency.)

It will be very helpful for our later understanding to present a brief sketch of the proof for this simple case, as it contains the essential ideas without the technical difficulties of the generalization. It depends rather heavily on the fact that the space of all randomized tests φ is compact in the weak -* topology, or more simply upon one version of this which is the following well-known lemma due to F. Riesz (see Banach [5] page 130).

LEMMA OF F. RIESZ. *If* $\{\varphi_n\}$ *is a sequence of* $\mathcal{B}$*-measurable functions on* Z *to the closed interval* [0, 1]—*a sequence of randomized tests in our terminology—then there* *exists a subsequence* $\{\varphi_{n_i}\}$ *and a test* φ' *such that for every function* f *integrable with respect to* ν

$$\lim_{i\to\infty} \int \varphi_{n_i}(z)f(z)\,d\nu(z) = \int \varphi'(z)f(z)\,d\nu(z).$$

PROOF OF THE HUNT-STEIN THEOREM: The first step of the proof consists in showing, by a simple application of Riesz's lemma, that a minimax solution φ_0 exists. (It is in fact the φ' of the lemma associated with a sequence of tests approaching the minimax condition.) For every $g \varepsilon \mathcal{G}$, then, we have that the test $g_\Phi\varphi_0$ given by $(g_\Phi\varphi_0)(z) = \varphi_0(g_Z(z))$ is also minimax, since the invariance of the risk function simply means here that the values of the power function of $g_\Phi\varphi_0$ are just those of the power function of φ_0 permuted separately over Ω_0 and over Ω_1. The minimax tests $g_\Phi\varphi_0$ are then averaged out by the probability measures μ_n on $(\mathcal{G}, \mathcal{C})$, giving rise to the sequence of tests

$$\varphi_n(z) = \int_{\mathcal{G}} \varphi_0(g_Z(z))\,d\mu_n(g).$$

These tests φ_n are again minimax, as a simple examination of their power functions shows (interchanging orders of integration by Fubini's theorem, we get, for each ω, permuted values of power functions averaged out over $\mathcal{G}$ by μ_n). Now take $\{\varphi_{n_i}\}$ and φ_0' in the sense of Riesz's lemma. Then just as for φ_0 above, φ_0' is also a minimax test. The one remaining difficulty is to show that φ_0' is an invariant test, more precisely an almost invariant test, and this is accomplished in two steps by a straightforward integration argument:

1) To show, for each $g \varepsilon \mathcal{G}$, that $\varphi_0'(g_Z(z)) = \varphi_0'(z)$ for almost all $z[\nu]$, it suffices to show that $\int\varphi_0'(g_Z(z))f(z)\,d\nu(z) = \int\varphi_0'(z)f(z)\,d\nu(z)$ for every integrable function f,

2) and this readily reduces to showing that a consequence of asymptotic invariance of a sequence $\{\mu_n\}$ of probability measures is that

$$\lim_{n\to\infty}\left[\int_{g'\varepsilon\mathcal{G}} \psi(g')\,d\mu_n(g'\cdot g^{-1}) - \int_{g'\varepsilon\mathcal{G}} \psi(g')\,d\mu_n(g')\right] = 0$$

for every bounded measurable function ψ of g'.

Having done this, φ_0' is almost invariant. Finally, a standard argument allows us to replace φ_0' by an invariant test φ_0'' such that $\varphi_0'' = \varphi_0'$ almost everywhere with respect to ν, provided only that another regularity condition is imposed on the group $\mathcal{G}$.

4. The modified minimax principle, completeness, and the generalized Hunt-Stein theorem. We return in this section to the general setup of a statistical decision problem invariant under a group of transformations, as described in Section 2, and assume in addition that all our probability distributions P_ω are dominated by σ-finite measure ν on $(Z, \mathcal{B})$. Without loss of generality, see Halmos and Savage [6], ν may be assumed equivalent to the family of all P_ω, that is, if $P_\omega(B) = 0$ for all ω, then $\nu(B) = 0$.

From Definitions 3 and 4 of Section 2, it is clear that a decision procedure φ can be changed on a set of ν-measure 0 without changing its risk function, so that two procedures differing on such a set may henceforth always be taken as equivalent. If in Definition 6 we have $g_\Phi \varphi = \varphi$ for each $g \varepsilon \mathcal{G}$, where the exceptional set of ν-measure 0 depends on g, we shall call φ almost invariant. We denote the class of almost invariant procedures by Φ^* and the class of invariant procedures by Φ^{**}. Further, we write $\Omega = \bigcup_{s \varepsilon S} \Omega_s$ where the Ω_s, s running over an index set S, are the orbits of Ω under the group $\mathcal{G}_\Omega$.

In strict accordance with our modified minimax principle as described in Sections 1 and 2, we have the following definition:

DEFINITION 1. A procedure φ is said to be at least as good in the modified minimax sense as a procedure ψ if

$$\sup_{\omega \varepsilon \Omega_s} \rho(\omega, \varphi) \leqq \sup_{\omega \varepsilon \Omega_s} \rho(\omega, \psi) \qquad \text{for every } s \varepsilon S.$$

The related notions of admissibility, complete class, etc., in the modified minimax sense are all similarly defined, in the obvious way. Henceforth in this section, such phrases as "at least as good as" are always to be understood in the modified minimax sense.

In generalizing upon the Hunt-Stein theorem, we run into the complicated notions of weak -* compactness, convergence and cluster point for a set of measures on a topological space. As these concepts are by no means obvious, it will be necessary to spell them out in some detail. Readers who wish to avoid topological difficulties may pass lightly over this part of the proof, pausing only long enough to note that in substance it is designed to show how a cluster point of a sequence of procedures may be used in constructing a new decision procedure capable of playing the central role analogous to the one in Riesz's lemma.

With the general plan of the proof of the Hunt-Stein theorem as outlined in Section 3 firmly in mind, we now proceed to our generalization. It will be useful to make here the obvious generalization of Definition 5, Section 3, to an asymptotically right invariant net $\{\mu_\alpha\}$ of probability measures on $(\mathcal{G}, \mathcal{C})$. Readers who are unfamiliar with the notion of nets may continue to think of sequences. For a discussion of nets we refer to Kelley [7]. The modified minimax principle, it should be remembered, requires us to pay strict attention to the fact that the various permutations mentioned there never take us out of an orbit, leading to the sharper result of completeness.

THE GENERALIZED HUNT-STEIN THEOREM. *Let* $(Z, \mathcal{B}, \Omega, P, A, \mathcal{A}, L)$ *be a statistical problem invariant under a group* $\mathcal{G}$ *as previously described. If the following regularity conditions are satisfied,*

i) *there exists a* σ*-finite measure* ν *equivalent to the set of all* P_ω,

ii) *there is a topology on the action set* A *such that* A *is a separable metric space and such that* $\mathcal{A}$ *is the Borel field generated by the compact subsets of* A, *and the loss function* L *is such that for each* $\omega \varepsilon \Omega$, $L(\omega, a)$ *is non-negative and continuous in* a *and is such that for every real number* τ, *the set* $\{a : L(\omega, a) \leqq \tau\}$ *is compact,*

iii) $\mathcal{C}$ *is a Borel field of subsets of* $\mathcal{G}$ *and there is an asymptotically right invariant net* $\{\mu_\alpha\}$ *of probability measures on* $(\mathcal{G}, \mathcal{C})$,
then the class Φ^* *of almost invariant decision procedures is essentially complete in the modified minimax sense. If in addition*

iv) $\mathcal{G}$ *is a locally compact* σ*-compact topological group with* $\mathcal{C}$ *generated by the compact subsets of* $\mathcal{G}$,
then the class Φ^{**} *itself is essentially complete in this sense.*

PROOF. Let $\varphi \varepsilon \Phi$ be given. We are required to find a $\varphi^* \varepsilon \Phi^*$ at least as good as φ. We may assume without loss of generality that the risks of φ are bounded on each orbit Ω_s by finite numbers m_s, for otherwise we may clearly ignore such orbits when comparing a possible φ^* and have only to carry out the following argument for the remaining orbits.

A) We observe first of all that for each $g \varepsilon \mathcal{G}$, the procedure $g_\Phi \varphi$ is equivalent to φ in the modified minimax sense. This follows at once from the invariance of the risk function, expressed by the formula $\rho(\omega, g_\Phi \varphi) = \rho(g_\Omega(\omega), \varphi)$, which tells us that the values of one risk function are simply those of the other permuted within each orbit of Ω separately.

We define for each α,

$$\varphi_\alpha(T \mid z) = \int_{\mathcal{G}} \varphi(g''_A T \mid g_Z(z)) \, d\mu_\alpha(g),$$

a net of randomized procedures got by averaging the procedures $g_\Phi \varphi$ with respect to the measures μ_α on $(\mathcal{G}, \mathcal{C})$, and shall show by a simple calculation of their risk functions that the φ_α are all at least as good as φ.

$$\rho(\omega, \varphi_\alpha) = \int\int L(\omega, a) \, d\varphi_\alpha(a \mid z) \, dP_\omega(z).$$

The Lebesgue integral transformation comes to our aid here, enabling us to compute such an expression by writing it as

$$= \int\int_0^\infty \varphi_\alpha(\{a : L(\omega, a) > h\} \mid z) \, dh \, dP_\omega(z)$$

$$= \int\int\int \varphi(g''_A\{a : L(\omega, a) > h\} \mid g_Z(z)) \, d\mu_\alpha(g) \, dh \, dP_\omega(z).$$

Now interchanging orders of integration (the integrand is positive) we integrate out with respect to h and the Lebesgue integral transformation in reverse gives us

$$= \int\int\int L(\omega, a) \, d\varphi(g_A(a) \mid g_Z(z)) \, dP_\omega(z) \, d\mu_\alpha(g)$$

$$= \int \rho(\omega, g_\Phi \varphi) \, d\mu_\alpha(g)$$

$$= \int \rho(g_\Omega(\omega), \varphi) \, d\mu_\alpha(g)$$

$$\leqq \sup_{\omega' \varepsilon \Omega_s} \rho(\omega', \varphi),$$

where Ω_s is the particular orbit to which ω belongs; hence,

$$\sup_{\omega \varepsilon \Omega_s} \rho(\omega, \varphi_\alpha) \leqq \sup_{\omega \varepsilon \Omega_s} \rho(\omega, \varphi) \qquad \text{for all } s \varepsilon S,$$

so that the φ_α are all at least as good as φ.

B) Let $\mathcal{K}$ be the collection of compact subsets K of A. For fixed K and any $\varphi \varepsilon \Phi$, φ_K is a $\mathcal{B}$-measurable function on Z to the closed interval [0, 1]. By Φ_K we mean the collection of all these functions φ_K with $\varphi \varepsilon \Phi$, i.e., the set of all randomized tests. Now the Banach space $L^\infty(\nu)$ of all bounded $\mathcal{B}$-measurable functions is the adjoint of the Banach space $L^1(\nu)$ of all extended real-valued functions integrable with respect to ν, and Alaoglu's theorem tells us that the solid unit sphere of the adjoint of a Banach space is compact in the weak -* topology. Hence Φ_K, as the intersection of a closed set with the solid unit sphere, is also compact in the weak -* topology on $L^\infty(\nu)$. By Tychonoff's compactness theorem, the cartesian product of the compact sets Φ_K for all $K \varepsilon \mathcal{K}$ is a compact set in the product topology of these Banach spaces $L^\infty(\nu)$. Therefore our net of functions φ_α defined above has a cluster point $\bar{\varphi}$ in the product space. A consequence of this, called the cluster point condition, is that for any finite number of sets K, any finite number of functions f integrable with respect to ν and any number $\epsilon > 0$, there is an arbitrarily large α such that

$$\left| \int \varphi_\alpha(K \mid z) f(z) \, d\nu(z) - \int \bar{\varphi}(K \mid z) f(z) \, d\nu(z) \right| < \epsilon$$

Further, since $\bar{\varphi}$ is a cluster point of the net of functions φ_α, there is a subnet $\varphi_{\alpha\beta}$ converging to it. To simplify the argument, we shall henceforth confine ourselves to this subnet and drop the second subscript β, so that $\bar{\varphi}$ is now a limit point of the $\{\varphi_\alpha\}$, i.e., for any compact set K and any integrable f we have

$$\int \varphi_\alpha(K \mid z) f(z) \, d\nu(z) \rightarrow \int \bar{\varphi}(K \mid z) f(z) \, d\nu(z).$$

Note that nothing is said about $\bar{\varphi}$ being a procedure, for it most certainly need not be. However, we have the following. By definition, every procedure φ has the properties

1) If $K_1 \subset K_2$, then $\varphi_{K_1} \leqq \varphi_{K_2}$;
2) If $K_1 \cap K_2 = \phi$, then $\varphi_{K_1 \cup K_2} = \varphi_{K_1} + \varphi_{K_2}$;
3) $\varphi_K \leqq 1$;

and it is easily seen that the limit point $\bar{\varphi}$ must have the same three properties almost everywhere with respect to ν (in fact has the third property everywhere). For example, we show property 2).

Let f be a function everywhere positive and integrable with respect to ν (such an f clearly exists since ν is σ-finite). Define a function g by

$$g(z) = f(z)[\bar{\varphi}(K_1 \cup K_2 \mid z) - \bar{\varphi}(K_1 \mid z) - \bar{\varphi}(K_2 \mid z)].$$

g is still integrable, as the product of f by a bounded $\mathcal{B}$-measurable function. Applying the limit point condition to the function g and the three sets $K_1 \cup K_2$,

K_1, K_2 we have

$$\int g(z)[\bar{\varphi}(K_1 \cup K_2 \mid z) - \bar{\varphi}(K_1 \mid z) - \bar{\varphi}(K_2 \mid z)]\, d\nu(z)$$
$$- \int g(z)[\varphi_\alpha(K_1 \cup K_2 \mid z) - \varphi_\alpha(K_1 \mid z) - \varphi_\alpha(K_2 \mid z)]\, d\nu(z) \rightarrow 0.$$

But the second square bracket is zero, hence rewriting g

$$\int f(z)[\bar{\varphi}(K_1 \cup K_2 \mid z) - \bar{\varphi}(K_1 \mid z) - \bar{\varphi}(K_2 \mid z)]^2\, d\nu(z) = 0.$$

But f was chosen everywhere positive, hence

$$\bar{\varphi}_{K_1 \cup K_2} = \bar{\varphi}_{K_1} + \bar{\varphi}_{K_2}[\nu].$$

We now show how $\bar{\varphi}$ may be used to construct a suitable procedure φ^*.

Let $\mathcal{R}$ be a countable subset of $\mathcal{K}$ such that $\mathcal{R}$ is closed with respect to finite unions and finite intersections and such that every open subset U of A is a countable union of interiors of elements R of $\mathcal{R}$ which are themselves subsets of U (such an $\mathcal{R}$ is known to exist by virtue of our hypothesis (ii) on A). Consider the function $\bar{\varphi}$ restricted to the set $\mathcal{R}$. Thus restricted, the limit point $\bar{\varphi}$ has the three mentioned properties of a φ with only countably many exceptional sets of ν-measure 0, which we may hereby disregard and assume the properties hold everywhere for $\bar{\varphi}$.

We define an outer measure φ^* on the open sets U in A by the formula, for each z,

$$\varphi^*(U \mid z) = \sup_{R \subset U} \bar{\varphi}(R \mid z).$$

φ_U^* is clearly a $\mathcal{B}$-measurable function of z, and for each z we do have an outer measure by the subadditivity of the supremum. Hence, we can extend φ^* to be a Caratheodory outer measure by defining for every subset W of A,

$$\varphi^*(W \mid z) = \inf_{U \supset W} \varphi^*(U \mid z).$$

φ_W^* is clearly $\mathcal{B}$-measurable, and for each z all the Borel sets $T \varepsilon \mathcal{A}$ are measurable. Restricting φ^* to the set $\mathcal{A}$, it is thus a measure on $(A, \mathcal{A})$ for each z. If we can show, therefore, that $\varphi^*(A \mid z) = 1$ $[\nu]$, φ^* will then be a randomized decision procedure. To do this, we first show that for every compact set K, $\varphi_K^* \geqq \bar{\varphi}_K[\nu]$:

Let $U \supset K$ be any open set. Then we have seen U may be written as $U = \bigcup_{R_i \subset U}$ (interior of R_i) which is then an open covering for the compact set K, so that for some n, $K \subset \bigcup_{i=1}^n R_i = R \varepsilon \mathcal{R}$. Hence $\bar{\varphi}_K \leqq \bar{\varphi}_R$ $[\nu]$. But $R \subset U$, so that $\bar{\varphi}_K \leqq \varphi_U^*[\nu]$. Now write $K = \bigcap_j U_j \downarrow$ the intersection of a descending sequence of open sets, which can be done since our space is metric. Since φ^* is a finite measure on the Borel sets, we have $\varphi_K^* = \lim \varphi_{U_j}^*$ everywhere. Combining this with the last inequality gives $\varphi_K^* \geqq \bar{\varphi}_K[\nu]$ since there are only countably many j involved.

Now consider any $\epsilon > 0$ and any $\omega \,\varepsilon\, \Omega$. We have for every α, $\rho(\omega, \varphi_\alpha) \leqq m_\epsilon$, where Ω_s is the particular orbit to which ω belongs. A simple consequence of this inequality is that for the compact set $K = \{a:L(\omega, a) \leqq m_s/\epsilon\}$ and for every α there holds

$$\int \varphi_\alpha(K \mid z)\, dP_\omega(z) > 1 - \epsilon.$$

Hence by the limit point condition, since $dP_\omega(z) = p(z \mid \omega)\, d\nu(z)$, we have

$$\int \bar{\varphi}(K \mid z)\, dP_\omega(z) \geqq 1 - \epsilon,$$

so that by the inequality just proved,

$$\int \varphi^*(K \mid z)\, dP_\omega(z) \geqq 1 - \epsilon,$$

therefore,

$$\int \varphi^*(A \mid z)\, dP_\omega(z) \geqq 1 - \epsilon.$$

But ϵ is arbitrary, and as A is open, we know $\varphi^*(A \mid z) \leqq 1$ everywhere; hence, $\varphi^*(A \mid z) = 1$ almost everywhere with respect to P_ω, for each $\omega \varepsilon \Omega$. From the equivalence of ν to the family of probability measures P_ω, there clearly follows $\varphi^*(A \mid z) = 1\ [\nu]$.

We might mention in passing that this newly constructed procedure φ^* is a limit point for the net of procedures φ_α when the appropriate (weak -*) topology is put on the space Φ of randomized decision procedures, so that we have shown that the set of all procedures having risk $\leqq m_\omega$ for each ω, where the m_ω are preassigned finite numbers, is compact in the weak -* topology. [The appropriate weak -* topology for procedures is characterized as follows. A procedure φ is a limit point of a net $\{\varphi_\alpha\}$ of procedures if for every compact set K and every non-negative integrable function f, there holds the inequality $\lim\sup_\alpha \int \varphi_\alpha(K \mid z)f(z)\, d\nu(z) \leqq \int \varphi(K \mid z)f(z)\, d\nu(z)$. Since $\lim_\alpha \int \varphi_{\alpha K} f\, d\nu = \int \bar{\varphi}_K f\, d\nu$ and $\bar{\varphi}_K \leqq \varphi^*_K[\nu]$ and $f \geqq 0$, the inequality $\lim_\alpha \int \varphi_{\alpha K} f\, d\nu \leqq \int \varphi^*_K f\, d\nu$ follows.]

C) We now show that this new procedure φ^* just constructed satisfies the conclusion of the theorem.

We consider its risk function, making the Lebesgue integral transformation and interchanging orders of integration when necessary.

$$\begin{aligned}
\rho(\omega, \varphi^*) &= \int\!\!\int \varphi^*(\{a:L(\omega, a) > h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh \\
&= \int\!\!\int [1 - \varphi^*(\{a:L(\omega, a) \leqq h\} \mid z)]p(z \mid \omega)\, d\nu(z)\, dh \\
&= \lim_{H \to \infty} \left(H - \int_0^H \int \varphi^*(\{a:L(\omega, a) \leqq h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh\right)
\end{aligned}$$

$$\leqq \lim_{H\to\infty} \left(H - \int_0^H \lim_\alpha \int \varphi_\alpha(\{a:L(\omega, a) \leqq h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh\right)$$

$$= \lim_{H\to\infty} \lim_\alpha \left(H - \int_0^H \int \varphi_\alpha(\{a:L(\omega, a) \leqq h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh\right)$$

$$= \lim_{H\to\infty} \lim_\alpha \int_0^H \int [1 - \varphi_\alpha(\{a:L(\omega, a) \leqq h\} \mid z)]p(z \mid \omega)\, d\nu(z)\, dh$$

$$\leqq \lim_\alpha \int\int \varphi_\alpha(\{a:L(\omega, a) > h\} \mid z)p(z \mid \omega)\, d\nu(z)\, dh$$

$$= \lim_\alpha \rho(\omega, \varphi_\alpha)$$

$$\leqq \sup_{\omega' \varepsilon \Omega_s} \rho(\omega', \varphi),$$

where Ω_s is the particular orbit to which ω belongs, hence φ^* is at least as good as φ in the modified minimax sense.

It remains to show that φ^* is an almost invariant procedure. To do this, we shall require the following lemma.

LEMMA. *The asymptotic right invariance of* $\{\mu_\alpha\}$ *implies for fixed* $g \; \varepsilon \; \mathcal{G}$ *that*

$$\lim_\alpha \int_{g' \varepsilon \mathcal{G}} \psi(g')[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')] = 0$$

for every bounded measurable function ψ *of* g'.

PROOF. This follows directly from the definition of integral. Without loss of generality, we assume $|\psi| \leqq 1$. Partitioning the interval $[-1, 1]$ into equal subintervals of length $1/M$, we have at once for each positive integer M the inequality.

$$\Bigg| \int \psi(g')[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')] - \sum_{k=-M}^{M} \frac{k}{M} \left[\mu_\alpha\left\{\left(g' : \frac{k}{M} \leqq \psi(g') < \frac{k+1}{M}\right) g^{-1}\right\} - \mu_\alpha\left\{g' : \frac{k}{M} \leqq \psi(g') < \frac{k+1}{M}\right\}\right] \Bigg| \leqq \frac{2}{M}.$$

For fixed M, the second square bracket approaches zero as α approaches infinity, hence $\limsup_\alpha | \int \psi(g')[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')] | \leqq 2/M$. As this is true for every M, the lemma follows.

Let, now, $g \; \varepsilon \; \mathcal{G}$ and a compact set K be given. We shall first show that $\bar{\varphi}(g_A'' K \mid g_Z(z)) = \bar{\varphi}(K \mid z)[\nu]$. To do this, it suffices to show that

$$\int [\bar{\varphi}(g_A'' K \mid g_Z(z)) - \bar{\varphi}(K \mid z)]f(z)\, d\nu(z) = 0$$

for every integrable function f. Since $\bar{\varphi}$ is a limit point of the φ_α in the weak -*

sense, the formula for φ_α in terms of φ makes this last integral

$$= \lim_{\alpha} \int\int \varphi(g_A''' K \mid g_Z'(z))f(z)\, d\nu(z)[d\mu_\alpha(g' \cdot g^{-1}) - d\mu_\alpha(g')].$$

But $\psi(g') = \int \varphi(g_A''' K \mid g_Z'(z))f(z)\, d\nu(z)$ is a measurable function of g' and is bounded by $\int |f|\, d\nu$, so our integral is indeed zero by our lemma.

We now show that $\varphi^*(g_A'' K \mid g_Z(z)) = \varphi^*(K \mid z)[\nu]$. Let U, an open set, and L, a compact set, be given such that $K \subset U \subset L$. φ_U^* is by definition the supremum of the $\bar{\varphi}_R$, all of which are less than or equal to $\bar{\varphi}_L$, so that $\varphi_U^* \leqq \bar{\varphi}_L[\nu]$, thus permitting us to write $\varphi_K^* \leqq \varphi_U^* \leqq \bar{\varphi}_L \leqq \varphi_L^*[\nu]$. Applying the group element g to these inequalities, we may extract the following:

$$\varphi^*(g_A'' K \mid g_Z(z)) \leqq \bar{\varphi}(g_A'' L \mid g_Z(z)) = \bar{\varphi}(L \mid z) \leqq \varphi^*(L \mid z)\ [\nu].$$

We may also extract

$$\varphi^*(K \mid z) \leqq \bar{\varphi}(L \mid z) = \bar{\varphi}(g_A'' L \mid g_Z(z)) \leqq \varphi^*(g_A'' L \mid g_Z(z))\ [\nu].$$

Taking a sequence $\{L_n\}$ of such compact sets with $K = \bigcap L_n$ and passing to the limit we get $\varphi^*(g_A'' K \mid g_Z(z)) \leqq \varphi^*(K \mid z) \leqq \varphi^*(g_A'' K \mid g_Z(z))\ [\nu]$, thus establishing the equality $[\nu]$.

From the nature of our topology on A and the way $\mathfrak{A}$ is generated by the compact sets K, this last equality extends by a standard transfinite induction argument to every set $T \in \mathfrak{A}$, thus establishing the almost invariance of φ^*. The argument goes as follows.

Note first that $\mathfrak{A}$ is equal to the smallest monotone class $\mathfrak{M}(\mathfrak{K})$ containing the compact sets: A is σ-compact (the union of a sequence of compact sets). The σ-ring of subsets of A generated by $\mathfrak{K}$, containing the whole space A, is therefore a Borel field, thus equal to the Borel field $\mathfrak{A}$ generated by $\mathfrak{K}$. Obviously $\mathfrak{M}(\mathfrak{K}) \subset \mathfrak{A}$. The proper difference $K - L$ of two compact sets belongs to $\mathfrak{M}(\mathfrak{K})$ since, writing $L = \bigcap_n U_n$ the intersection of open sets, we have $K - L = \bigcup_n (K - U_n)$ which is σ-compact. Therefore the class of all finite unions of these proper differences belongs to $\mathfrak{M}(\mathfrak{K})$. Thus (see Halmos [3] Theorem F, p. 223) $\mathfrak{M}(\mathfrak{K})$ contains a ring containing $\mathfrak{K}$, so that (Halmos again, Theorem B, p. 27) $\mathfrak{M}(\mathfrak{K}) \supset \mathfrak{A}$, hence $\mathfrak{M}(\mathfrak{K}) = \mathfrak{A}$.

Write $\mathfrak{B}_0 = \mathfrak{K} = \{$the compact sets$\}$ and for every ordinal number α let $\mathfrak{B}_\alpha = \{$all monotone limits of sets S_n, where $S_n \in \mathfrak{B}_{\beta_n}$ and $\beta_n < \alpha\}$. For every α, $\mathfrak{B}_\alpha$ is well-defined. Then $\mathfrak{A} = \mathfrak{M}(\mathfrak{K})$ means that $\mathfrak{A} = \bigcup_{\alpha < \omega_1} \mathfrak{B}_\alpha$ where ω_1 is the first uncountable ordinal. Having thus characterized $\mathfrak{A}$, the proof proceeds by transfinite induction on α: Suppose our equality $[\nu]$ holds for all $\mathfrak{B}_\beta$ with $\beta < \alpha$. Consider any set $T \in \mathfrak{B}_\alpha$. Then $T = \lim T_n$ where $T_n \in \mathfrak{B}_{\beta_n}$ and $\beta_n < \alpha$. The equality $[\nu]$ holds for these T_n by the induction hypothesis, therefore it holds for the limit T since there are only a countable number of sets of ν-measure 0 involved. The equality $[\nu]$ holding for all $T \in \mathfrak{B}_\alpha$, for all α, it holds for every $T \in \mathfrak{A}$, and φ^* is almost invariant.

D) With our assumptions on $(\mathcal{G}, \mathcal{C})$ we now proceed to replace this almost invariant procedure φ^* by an invariant procedure φ^{**} such that $\varphi^{**} = \varphi^*$ $[\nu]$, thereby completing the proof of the theorem.

Let us return to the class $\mathcal{R}$ of part (B). Since $\mathcal{G}$ is a locally compact topological group, there is a right invariant Haar measure λ on $(\mathcal{G}, \mathcal{C})$. We define a function φ^{**} as follows: For each $R \varepsilon \mathcal{R}$

i) let $\varphi^{**}(R \mid z) = \varphi^*(R \mid z)$, for all z such that

$$\varphi^*(g''_A R \mid g_Z(z)) = \varphi^*(R \mid z) \quad [\lambda];$$

ii) for all such z in (i), and each g, let

$$\varphi^{**}(g''_A R \mid g_Z(z)) = \varphi^*(R \mid z);$$

iii) let $\varphi^{**}(R \mid z) = 0$ elsewhere.

To show that (i) and (ii) are not contradictory, suppose there existed R_1, z_1, g_1 with $R_2 = g''_{1A} R_1$, $z_2 = g_{1Z}(z_1)$, and

$$\varphi^*(g''_A R_1 \mid g_Z(z_1)) = \varphi^*(R_1 \mid z_1) \qquad [\lambda]$$
$$\varphi^*(g''_A R_2 \mid g_Z(z_2)) = \varphi^*(R_2 \mid z_2) \qquad [\lambda],$$

but with $\varphi^*(R_1 \mid z_1) \neq \varphi^*(R_2 \mid z_2)$. Then it would follow that

$$\begin{aligned} 0 &= \lambda\{g:\varphi^*(g''_A R_2 \mid g_Z(z_2)) \neq \varphi^*(R_2 \mid z_2)\} \\ &\geqq \lambda\{g:\varphi^*(g''_A R_2 \mid g_Z(z_2)) = \varphi^*(R_1 \mid z_1)\} \\ &= \lambda\{g:\varphi^*(g''_A g''_{1A} R_1 \mid g_Z(g_{1Z}(z_1))) = \varphi^*(R_1 \mid z_1)\}, \end{aligned}$$

which by the right invariance of λ

$$\begin{aligned} &= \lambda\{g:\varphi^*(g''_A R_1 \mid g_Z(z_1)) = \varphi^*(R_1 \mid z_1)\} \\ &= \lambda(\mathcal{G}). \end{aligned}$$

But $0 \geqq \lambda(\mathcal{G})$ contradicts the fact that λ is a Haar measure, hence φ^{**} is not contradictorily defined.

We now show, for each $R \varepsilon \mathcal{R}$, that

$$\varphi^{**}(R \mid z) = \varphi^*(R \mid z) \quad [\nu]$$

Let

$$\begin{aligned} V &= \{(z, g):\varphi^*(g''_A R \mid g_Z(z)) \neq \varphi^*(R \mid z)\} \\ V_z &= \{g:\varphi^*(g''_A R \mid g_Z(z)) \neq \varphi^*(R \mid z)\} \\ V_g &= \{z:\varphi^*(g''_A R \mid g_Z(z)) \neq \varphi^*(R \mid z)\}. \end{aligned}$$

The almost invariance of φ^* implies that $\nu(V_g) = 0$ for each g. Moreover, λ is σ-finite since $\mathcal{G}$ was assumed σ-compact, hence by Fubini's theorem (Halmos [3] Theorem A, p. 147) $(\nu \times \lambda)(V) = 0$. By the same reference this implies $\lambda(V_z) = 0$ $[\nu]$, hence $\varphi^{**}(R \mid z) = \varphi^*(R \mid z)[\nu]$.

Since $\mathcal{R}$ is countable, we may disregard sets of ν-measure 0, so that for all z,

z and g we have

$$\varphi^{**}(g''_A R \mid g_Z(z)) = \varphi^*(R \mid z) = \varphi^{**}(R \mid z).$$

Extending this equality in the obvious way to the open sets U, then to the compact sets K, a transfinite induction on α exactly as in part (C) above extends it to all the Borel sets $T \varepsilon \mathfrak{A}$, thus completing the proof of the generalized Hunt-Stein theorem.

Remarks. It may prove useful in certain applications to point out an easy extension of the above result. Clearly, everything we have said goes through if Φ is any invariant, convex and closed set of randomized decision procedures rather than the class of all of them. That is, under our regularity conditions, if Φ is closed in the weak -* topology, and if for every probability measure λ on $(\mathcal{G}, \mathcal{C})$ and every $\varphi \varepsilon \Phi$, the procedure φ_λ given by $\varphi_\lambda(T \mid z) = \int \varphi(g''_A T \mid g_Z(z)) \, d\lambda(g)$ belongs to Φ, then for each $\varphi \varepsilon \Phi$ there is an almost invariant $\varphi^* \varepsilon \Phi$ which is at least as good as φ in the modified minimax sense, and so on. For example, the classical Hunt-Stein theorem follows from this by taking Φ to be the set of all procedures φ satisfying the inequality $\sup_{\omega \varepsilon \Omega_0} \rho(\omega, \varphi) \leqq \alpha$.

The condition in hypothesis (ii) that $L(\omega, a)$ be non-negative is not essential. We only require it to be bounded from below, and this is automatically assured by the compactness condition.

An example of theoretical interest. Let $\{P_\omega\}$ be the set of all probability distributions on the positive integers, E the set of even integers, $A = \{0, 1\}$, $L(\omega, 0) = 0$ if $P_\omega(E) > 1/2$ and $= 1$ if $P_\omega(E) \leqq 1/2$, $L(\omega, 1) = 1$ if $P_\omega(E) > 1/2$ and $= 0$ if $P_\omega(E) \leqq 1/2$, an action to be taken after observing one positive integer.

Let φ be a strategy, where $\varphi(n)$ is the probability of taking action 0 (saying "even") when integer n is observed. It can be shown, but only after a complicated argument, that φ is a Bayes solution if and only if it does "the right thing" for at least one integer n, i.e., if $\varphi(2n) = 1$ or $\varphi(2n - 1) = 0$ for at least one integer n. Further, all these strategies are admissible, but the class of all admissible strategies is not known. A plausible conjecture seems to be that φ is admissible if and only if $\inf_n \min \{\varphi(2n - 1), 1 - \varphi(2n)\} = 0$. Thus we see that even in the simple case of one observation, the classes are extremely large.

Let us consider this problem in the light of the invariance principle. Let g be the particular kind of permutation of the integers which, for some n, permutes the first n even numbers and the first n odd numbers separately and leaves the tail untouched. The set $\mathcal{G}$ of all such permutations for all n clearly forms a group, containing countably many elements g, which leaves the problem invariant. Let $\mathcal{C}$ be the set of all subsets of $\mathcal{G}$. The set Z of all integral outcomes breaks up into two orbits under the group, which we may write as E and O for "even" and "odd." The orbits Ω_s of Ω (closed in a suitable topology, a matter we need not go into here) are given by

$$\Omega_s = \{\omega : P_\omega(E) = s\}$$

for all s belonging to the closed interval [0, 1]. There is an asymptotically (right) invariant sequence of probability measures on $(\mathcal{G}, \mathcal{C})$, as follows: Let μ_n be the measure which assigns equal probabilities of $1/(n!)^2$ to each of the $(n!)^2$ elements g affecting the first $2n$ integers only, and 0 to the remaining elements of $\mathcal{G}$. To prove the asymptotic invariance of this sequence, we simply note that given $g \varepsilon \mathcal{G}$, there is an integer n_0 such that g affects the first $2n_0$ integers only, so that for every $C \varepsilon \mathcal{C}$ we necessarily have $\mu_n(Cg) = \mu_n(C)$ for all $n \geqq n_0$. By the generalized Hunt-Stein theorem, then, the problem reduces to the classical binomial case, say of tossing a coin once with unknown probability s of showing heads (even).

5. Mixed games, and the use of previous experience. In this section we present a purely game-theoretic model for the modified minimax principle, and suggest how by means of this model and our previous experience the principle might be applied in practice to bridge the gap between the Bayes and minimax approach.

Mixed Games. Let there be given a pair of games $G_1 = (X_1, Y, M_1)$ and $G_2 = (X_2, Y, M_2)$ with the same Y-space of randomized strategies available to the second player. Let p and $q = 1 - p$ be fixed probabilities with which player I is to play the respective games. We shall call such a setup a mixed game for player I. A strategy for the first player is then equivalent to an ordered couple (x_1, x_2), $x_1 \varepsilon X_1$, $x_2 \varepsilon X_2$, signifying his choice of strategy depending on the game he must eventually play (say, after a coin with probability p of heads is tossed). The payoff function M is clearly given by

$$M((x_1, x_2), y) = pM_1(x_1, y) + qM_2(x_2, y).$$

We denote the mixed game by $G = (X_1 \times X_2, Y, M)$.

More generally, given $G_i = (X_i, Y, M_i)$ $i = 1, 2, \cdots$ and probabilities p_i of playing the game $G_i (p_i \geqq 0, \sum_i p_i = 1)$, the mixed game for player I is equivalent to the game $G = (X, Y, M)$ where

$$X = X_1 \times X_2 \times \cdots, x = (x_1, x_2, \cdots) \varepsilon X$$

is a strategy for the first player, and $M(x, y) = \sum_i p_i M_i(x_i, y)$.

In the most general case, we have games $G_s = (X_s, Y, M_s)$ with $s \varepsilon S$ an index set, $\mathcal{S}$ a Borel field in S, and P a probability measure on $\mathcal{S}$. A strategy for player I is now a function f on S with $f(s) \varepsilon X_s$ for each $s \varepsilon S$ (i.e., f is a point in the function space $X = \times_{s \varepsilon S} X_s$) and

$$M(f, y) = \int_S M_s(f(s), y)\, dP(s).$$

Let us now consider the minimax principle from the point of view of the second player. For simplicity we confine ourselves to the discrete case, although the argument is general. For each strategy $y \varepsilon Y$, the second player seeks to minimize $\sup_x M(x, y)$ over y. But $\sup_x M(x, y) = \sum_i p_i \sup_{x_i \varepsilon X_i} M_i(x_i, y) = \sum_i p_i \alpha(i, y)$ is just the average with respect to the p_i of the sliced up risk function in the modified minimax sense, where each slice is the old payoff function M_i for the ith game. (The point $(\alpha(1, y), \alpha(2, y), \cdots)$ is the generalization of the

α-β set for testing composite hypotheses.) Hence playing minimax in the mixed game is equivalent to playing Bayes in the sliced up game.

The use of previous experience as a slicing principle: In a paper on the use of previous experience in reaching statistical decisions, Hodges and Lehmann [8] propose a guess at an a priori distribution λ for the states of nature on the basis of previous experience, and a safeguard in case the guess is incorrect. If c is the minimax risk, they seek to minimize $\int \rho(\omega, \varphi)\, d\lambda(\omega)$ among all procedures φ satisfying the restriction $\sup_\omega \rho(\omega, \varphi) \leqq c + k$ where k is a given positive number. They call such a minimizing procedure φ_0 a restricted Bayes solution (we might say the Bayes principle is given a minimax restriction). If λ is suspect, k is made small; as confidence in λ grows, k may be raised. Note that as $k \to \infty$ or is large enough, the restricted Bayes solution approaches the Bayes solution; as $k \to 0$, the restricted solution approaches the minimax state (for $k = 0$, we have a best minimax procedure relative to λ).

Another way of looking at the modified minimax principle, with a view to applying it in practice, suggests that we reverse this procedure. That is to say, to start with the minimax principle and then modify it gradually or greatly as previous experience is gathered—to assume less and trust it completely, a sort of Bayesian modification of the minimax principle. Specifically, to take the most simple case, we might use the best of our information to break up or stratify Ω into Ω_1 and Ω_2 with certain probabilities p and $1 - p$ attached thereto, and then play minimax in the resulting mixed game. (Say, a machine is known to produce with probability $p = .95$ coins whose probabilities of showing heads when tossed lie between .4 and .6. Ω_1 is then the interval from .4 to .6, Ω_2 its complement in the unit interval, and the corresponding games are mixed in the ratio .95 to .05. If the probability that $\omega \varepsilon \Omega_1$ is at least p, then we would mix Ω_1 and all of Ω in the proportion p to $1 - p$, and so on.) As more knowledge is acquired, we might adjust the value of p, or better still, feel confident enough to partition Ω into a larger number of sets with certain p_i attached, or even into a family $\{\Omega_s\}$ of sets with a probability measure P over it, and then play minimax in the resulting mixed game. By the above, this is equivalent to playing Bayes in the sliced up game. If each Ω_s consists of only one point, we are really postulating an a priori probability distribution P over the states of nature, so that minimax in the mixed game is exactly Bayes for the original game. Thus we see how previous information might be used in any given problem to slice Ω into subsets, and how it is possible to go from the minimax extreme to the Bayes in easy stages, by a gradual modification of the minimax principle.

Acknowledgment. I wish to express my deep personal indebtedness to the late Professor M. A. Girshick for suggesting this topic, and for his interest and guidance throughout my statistical studies. I am indebted to Professor Herman Rubin for several helpful discussions. I wish to thank the referee for pointing out references [9] and [10], as papers containing further useful remarks on the Hunt-Stein theorem and the role of invariance in Statistics; I am glad to call them to the attention of the reader. To the Office of Naval Research my thanks for its support of this work.

REFERENCES

[1] David Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley & Sons, New York, 1954.
[2] Abraham Wald, *Statistical Decision Functions*, John Wiley & Sons, New York, 1950.
[3] Paul R. Halmos, *Measure Theory*, D. Van Nostrand, New York, 1950.
[4] G. Hunt and C. Stein, "Most stringent tests of composite hypotheses," unpublished.
[5] S. Banach, *Théorie des opérations linéaires*, Warszawa, 1932.
[6] Paul R. Halmos and L. J. Savage, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 225–241.
[7] J. L. Kelley, *General Topology*, D. Van Nostrand, New York, 1955.
[8] J. L. Hodges, Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 396–407.
[9] E. L. Lehmann, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1–26.
[10] J. Kiefer, "Invariance, minimax sequential estimation, and continuous time processes," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 573–601.