AN ESSENTIALLY COMPLETE CLASS OF ADMISSIBLE DECISION FUNCTIONS

By Abraham Wald

Columbia University

Summary. With any statistical decision procedure (function) there will be associated a risk function $r(\theta)$ where $r(\theta)$ denotes the risk due to possible wrong decisions when θ is the true parameter point. If an a priori probability distribution of θ is given, a decision procedure which minimizes the expected value of $r(\theta)$ is called the Bayes solution of the problem. The main result in this note may be stated as follows: Consider the class C of decision procedures consisting of all Bayes solutions corresponding to all possible a priori distributions of θ. Under some weak conditions, for any decision procedure T not in C there exists a decision procedure T^* in C such that $r^*(\theta) \leqq r(\theta)$ identically in θ. Here $r(\theta)$ is the risk function associated with T, and $r^*(\theta)$ is the risk function associated with T^*. Applications of this result to the problem of testing a hypothesis are made.

1. Introduction. In some previous publications [1], [2] the author has considered the following general problem of statistical inference: Let $X = (X_1, \cdots, X_n)$ be a set of chance variables. Suppose that the only information we have concerning the joint distribution function F of these chance variables is that F is an element of a given class Ω of distribution functions. Suppose, furthermore, that a class D of possible decisions d is given one of which is to be made on the basis of an observation $x = (x_1, \cdots, x_n)$ on the chance vector X. The problem is then to construct a function $d(x)$, called statistical decision function, which associates with each sample point x an element $d(x)$ of D so that the decision $d(x)$ is made when the sample point x is observed. A statistical decision function $d(x)$ is defined over all possible points x of the sample space and for each sample point x the value of the function is an element of D. Each element d of D will usually be interpreted as a decision to accept the hypothesis that the unknown distribution F of X belongs to a certain subclass ω of Ω. Different elements d of D correspond to different subclasses ω of Ω.

The problem of testing the hypothesis H that the unknown distribution function F belongs to a given subclass ω of Ω, is contained as a special case in the above general problem. The space D will then contain only two elements, d_1 and d_2, where d_1 denotes the decision of accepting H and d_2 denotes the decision of rejecting H.

As in [1] and [2], we shall assume also here that Ω is a k-parameter family of distribution functions. Then each element of Ω may be represented by a point $\theta = (\theta_1, \cdots, \theta_k)$, called parameter point, in the k-dimensional Cartesian space. The class Ω is then represented by a subset of the k-dimensional Cartesian space, called parameter space. We shall, therefore, refer to Ω as the parameter space and to its elements as parameter points.

The merits of any particular decision function $d(x)$ will usually depend on the relative importance of the various possible errors caused by not selecting the proper element d of D. The relative importance of such errors has been described in [1] and [2] by a weight function $W(\theta, d)$ defined over the product of Ω and D. For any pair (θ, d) the value of $W(\theta, d)$ is non-negative and expresses the loss caused by taking the decision d when θ is the true parameter point. For any given decision function $d(x)$ the expected value of the loss is given by

$$r(\theta) = \int_M W[\theta, d(x)]\, dF(x) \tag{1.1}$$

where M denotes the sample space and $F(x)$ is the joint cumulative distribution of $X = (X_1, \cdots, X_n)$ corresponding to the parameter point θ.

The function $r(\theta)$ is defined over the parameter space Ω and is called the risk function. The shape of the risk function $r(\theta)$ will, in general, be affected by the decision function $d(x)$ used. To put this dependence in evidence, we shall use the symbol $r[\theta \mid d(x)]$ to denote the risk function $r(\theta)$ associated with the decision function $d(x)$.

A decision function $d(x)$ is said to be uniformly better than the decision function $d^*(x)$ if

$$r[\theta \mid d(x)] \leqq r[\theta \mid d^*(x)] \tag{1.2}$$

for all θ and if there exists at least one point θ for which the inequality sign holds in (1.2). A decision function $d(x)$ is said to be admissible if no other uniformly better decision function exists.

A class C of admissible decision functions will be said to be essentially complete if for any decision function $d(x)$ not in C there exists a decision function $d^*(x)$ in C such that

$$r[\theta \mid d^*(x)] \leqq r[\theta \mid d(x)]$$

for all θ.

In section 2 we shall formulate certain assumptions which will then be used in section 3 to derive an essentially complete class of admissible decision functions. In section 4 applications are made to the problem of testing a hypothesis.

In a recent paper Lehmann [3] obtained an essentially complete class of admissible tests for each hypothesis H of a certain restricted class of simple hypotheses. The restrictions imposed on Ω in Lehmann's paper are essentially those formulated by Neyman [4], [5] to insure the existence of the type A_1 (uniformly most powerful unbiassed) test. Our definition of an essentially complete class of admissible decision functions agrees with that given by Lehmann when the problem is to test a hypothesis and the weight function $W(\theta, d)$ can take only the values 0 and 1.

2. Assumptions. Throughout this paper we shall make the following assumptions:

Assumption 1: The parameter space Ω is a bounded and closed subset of a finite dimensional, say k-dimensional, Cartesian space.

We shall introduce the following convergence definition in the space D: a sequence $\{d_m\}$, $(m = 1, 2, \cdots, \text{ad inf.})$, of elements of D is said to converge to the element d of D if

$$\lim_{m=\infty} W(\theta, d_m) = W(\theta, d)$$

uniformly in θ.

Assumption 2: The space D is compact and, for any d, $W(\theta, d)$ is a continuous function of θ.

Assumption 3: For any point θ of Ω the joint distribution function of $X = (X_1, \cdots, X_n)$ admits a density function $p(x, \theta)$ for all points x of the n-dimensional Cartesian space M (sample space). The density function $p(x, \theta)$ is assumed to be continuous in x and θ jointly.

In what follows we shall mean by a distribution function $f(\theta)$ of θ a cumulative distribution function for which $\int_\Omega df(\theta) = 1$ and for which $\int_\Omega W(\theta, d)df(\theta)$ is not zero identically in d.

Assumption 4: For any point x of M, except perhaps for a set of measure zero, and for any cumulative distribution function $f(\theta)$ there exists one and only one element of D for which the expression

$$\int_\Omega W(\theta, d)p(x, \theta)\, df(\theta) \tag{2.1}$$

takes its minimum value with respect to d.

Assumptions 1 and 3 in this paper are exactly the same as Assumptions 1 and 3 in [2]. The formulation of Assumptions 2 and 4 is somewhat different from that given in [2]. This is mainly due to the fact that in [2] the space D has the same elements as Ω, while here this is not necessarily so. It can be verified without difficulty that this slight modification of the assumptions does not affect in any way the validity of the results obtained in [2]. Thus, we shall be able to make use of any theorems proved in [2] for the purposes of the present paper.

3. Derivation of an essentially complete class of admissible decision functions. For any distribution function $f(\theta)$ defined over Ω and for any sample point x let $d(x, f)$ denote the element of D for which the expression (2.1) takes its minimum value. It follows easily from the definition of $r(\theta)$ and $d(x, f)$ that

$$\int_\Omega r[\theta \mid d(x, f)]df(\theta) \leqq \int_\Omega r[\theta \mid d^*(x)]\, df(\theta) \tag{3.1}$$

for any decision function $d^*(x)$. If we interpret $f(\theta)$ as an a priori probability distribution of θ, inequality (3.1) says that the expected value of $r(\theta)$ takes its minimum value for the decision function $d(x, f)$. We shall refer to $d(x, f)$ as the Bayes' solution of the problem corresponding to the a priori probability distribution $f(\theta)$.

We shall now prove the following theorem.

THEOREM 3.1. *The class C of all Bayes' solutions $d(x, f)$ corresponding to all possible a priori distributions $f(\theta)$ is an essentially complete class of admissible decision functions.*

PROOF. First we show that for any distribution $f(\theta)$ the decision function $d(x,f)$ is admissible. Let $d(x)$ be a decision function such that

$$r[\theta \mid d(x)] \leqq r[\theta \mid d(x,f)]$$

for all θ. Then

$$\int_\Omega r[\theta \mid d(x)]\, df(\theta) \leqq \int_\Omega r[\theta \mid d(x,f)]\, df(\theta). \tag{3.2}$$

From the definition of $d(x, f)$ it follows that the equality sign must hold in (3.2), i.e.,

$$\int_\Omega r[\theta \mid d(x)]\, df(\theta) = \int_\Omega r[\theta \mid d(x,f)]\, df(\theta). \tag{3.3}$$

From the second half of Theorem 4.2 in [2] it then follows that

$$r[\theta \mid d(x)] = r[\theta \mid d(x,f)]$$

for all θ. Hence $d(x,f)$ is an admissible decision function.

We shall now show that the class C of decision functions $d(x,f)$ corresponding to all possible a priori distributions $f(\theta)$ is essentially complete. Let $d_0(x)$ be any decision function not in the class C. The essential completeness of the class C is proved if we can show that there exists a distribution $f(\theta)$ such that

$$r[\theta \mid d(x,f)] \leqq r[\theta \mid d_0(x)] \tag{3.4}$$

for all θ.

To prove (3.4) we shall consider the weight function

$$W^*(\theta, d) = W(\theta, d) - r[\theta \mid d_0(x)] + \operatorname*{Max}_\theta r[\theta \mid d_0(x)] \tag{3.5}$$

The maximum of $r[\theta \mid d_0(x)]$ exists, since according to Theorem 4.1 in [2] $r[\theta \mid d_0(x)]$ is a continuous function of θ. Clearly, Assumptions 1–4 remain valid if we replace $W(\theta, d)$ by $W^*(\theta, d)$. Let $r^*[\theta \mid d(x)]$ denote the risk function associated with the decision function $d(x)$ if the weight function is given by $W^*(\theta, d)$. According to Theorem 5.2 in [2] there exists a decision function $d^*(x)$ such that

$$\operatorname*{Max}_\theta r^*[\theta \mid d^*(x)] \leqq \operatorname*{Max}_\theta r^*[\theta \mid d(x)] \tag{3.6}$$

for any decision function $d(x)$. Since

$$\underset{\theta}{\text{Max}}\, r^*[\theta \mid d_0(x)] = \underset{\theta}{\text{Max}}\, r[\theta \mid d_0(x)]$$

it follows from (3.6) that

$$\underset{\theta}{\text{Max}}\, r^*[\theta \mid d^*(x)] \leqq \underset{\theta}{\text{Max}}\, r[\theta \mid d_0(x)]. \tag{3.7}$$

Inequalities (3.5) and (3.7) imply

$$r[\theta \mid d^*(x)] \leqq r[\theta \mid d_0(x)] \tag{3.8}$$

for all θ.

For any distribution $f(\theta)$ we shall denote by $d^*(x, f)$ the Bayes solution of the problem corresponding to the a priori distribution $f(\theta)$ when the weight function is given by $W^*(\theta, d)$. Since $W^*(\theta, d) - W(\theta, d)$ depends only on θ but not on d, one can easily verify that $d^*(x, f) = d(x, f)$. It follows from Theorems 4.4 and 5.1 in [2] that there exists a distribution $\bar{f}(\theta)$, the so-called least favorable distribution, such that (3.6) remains valid if we replace $d^*(x)$ by $d^*(x, \bar{f})$. Thus we can put

$$d^*(x) = d^*(x, \bar{f}) = d(x, \bar{f}). \tag{3.9}$$

Hence, from (3.8) we obtain

$$r[\theta \mid d(x, \bar{f})] \leqq r[\theta \mid d_0(x)]$$

for all θ. This completes the proof of Theorem 3.1.

4. Applications to the problem of testing a hypothesis. In this section we shall apply the results of the preceding section to the problem of testing the hypothesis H that the true parameter point is included in a given subset ω of Ω. We shall assume that ω is an open subset of Ω. The space D consists now only of two elements, d_1 and d_2, where d_1 denotes the decision of accepting H and d_2 denotes the decision of rejecting H.

We shall assume that the $W(\theta, d_1)$ is equal to zero for points θ in the interior or on the boundary of ω, and positive elsewhere. Similarly, $W(\theta, d_2)$ will be assumed to be positive for points θ inside ω and zero outside ω. For any a priori distribution $f(\theta)$ the Bayes solution is given by the following test: We reject the hypothesis H if (and only if)[1]

$$\int_{\Omega-\omega} W(\theta, d_1)p(x, \theta)\, df(\theta) > \int_{\omega} W(\theta, d_2)p(x, \theta)\, df(\theta). \tag{4.1}$$

Thus, the class C of regions (4.1), corresponding to all possible distributions $f(\theta)$, is an essentially complete class of admissible critical regions.

For any critical region R we shall denote the probability that the sample x

[1] Whether the equality sign is included or not in (4.1) is of no consequence, since by Assumption 4 the measure of the set of points x for which the equality holds in (4.1) is zero.

will fall in R when θ is true by $P(\theta \mid R)$. It follows from Lemma 4.4 in [2] and Assumption 3 that $P(\theta \mid R)$ is a continuous function of θ for any region R. Since $W(\theta, d_1)$ is positive in the interior of $\Omega - \omega$, and $W(\theta, d_2)$ is positive in ω, the class C of regions defined in (4.1) will have the following properties:

(a) For any region R outside the class C there exists a region R^* in C such that

$$P(\theta \mid R^*) \leqq P(\theta \mid R) \text{ in } \omega$$

and

$$P(\theta \mid R^*) \geq P(\theta \mid R) \text{ in } \Omega - \omega.$$

(b) If R and R^* are members of C such that

$$P(\theta \mid R^*) \leqq P(\theta \mid R) \text{ in } \omega$$

and

$$P(\theta \mid R^*) \geq P(\theta \mid R) \text{ in } \Omega - \omega,$$

then

$$P(\theta \mid R^*) = P(\theta \mid R) \text{ for all } \theta.$$

For any distribution $g(\theta)$ consider the critical region consisting of all sample points x satisfying

$$\int_{\Omega-\omega} p(x, \theta)\, dg(\theta) > \int_{\omega} p(x, \theta)\, dg(\theta). \tag{4.2}$$

Let C^* be the class of regions (4.2) corresponding to all possible distributions $g(\theta)$. One can easily verify that any region in C is also a member of C^*. Thus, the following theorem holds:

THEOREM 4.1 *Suppose that Assumptions 1 and 3 are fulfilled and ω is an open subset of Ω. Suppose, furthermore, that for any distribution $g(\theta)$ the set of sample points x satisfying the equation*

$$\int_{\Omega-\omega} p(x, \theta)\, dg(\theta) = \int_{\omega} p(x, \theta)\, dg(\theta)$$

has the measure zero. Then, for any region R outside the class C^* there will be a region R^* in C^* such that

$$P(\theta \mid R^*) \leq P(\theta \mid R) \textit{ in } \omega$$

and

$$P(\theta \mid R^*) \geqq P(\theta \mid R) \textit{ in } \Omega - \omega.$$

Addition at proof reading: After this paper was sent to the printer, the author obtained a generalization of Theorem 3.1 to sequential decision functions, as well as some other results. They will appear in a forthcoming issue of *Econometrica.*

REFERENCES

[1] A. WALD, "Contributions to the theory of statistical estimation and testing hypotheses", *Annals of Math. Stat.*, Vol. 10 (1939).

[2] A. WALD, "Statistical decision functions which minimize the maximum risk", *Annals of Math.*, Vol. 46 (1945).

[3] E. L. LEHMANN, "On families of admissible tests", *Annals of Math. Stat.*, Vol. 18 (1947), pp. 97–104.

[4] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, Part II", *Statistical Research Memoirs*, Vol. II, London, 1938.

[5] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, Part I", *Statistical Research Memoirs*, Vol. I, London, 1936.