Consistent Decision Functions and Bayes Solutions[1,2]

Roy Radner

April 30, 1952

1. Introduction.

A central problem of statistical decision theory is the formulation of general criteria by which to judge the value of particular decision rules. The purpose of this paper is to consider a particular criterion, consistency,[3] as suggested by the notions of "consistent estimate" and "consistent test," to give a general definition of it, and to relate it to Bayes solutions of decision problems. Since in such problems the decisions may take on a wide variety of forms, it seems appropriate to aim at a fairly high degree of mathematical generality. It will become clear that the idea of consistency as considered in this paper is a nonsequential concept; thus we do not consider the cost of observation.

In Section 4 various suggestions as to the interpretation of the results are made, corresponding to the various current schools of thought about the foundations of statistics.

---

1. Research undertaken by the Cowles Commission for Research in Economics under contract Nonr-358(01), NR 047-006 with the Office of Naval Research.

2. My thanks to J. Marschak, C. M. Stein, E. L. Lehmann, and I. N. Herstein for their comments and stimulating discussions on several points in this paper.

3. Many of the ideas in this paper were outlined for the finite case in CCDP: Statistics No. 371, under the same title.

2.1. Formulation of the Nonsequential Decision Problem.

Given a sequence of statistical decision problems Q_n characterized by:

Ω, the set of possible states of nature ω.

D, the set of possible decisions d.

$K(\omega, d)$ the gain function (real valued).

$x_1, \ldots, x_n$, in X, a set of observations whose joint distribution is determined for each ω.

ω_o, the true state of nature.

Assumption Ia. $|K(\omega, d)| \leq M$, for all ω, d.

As in [1], we define metrics on both Ω and D by:

$$(1) \quad \begin{cases} \rho(\omega_1, \omega_2) = \sup_d |K(\omega_1, d) - K(\omega_2, d)| \\ \rho(d_1, d_2) = \sup_\omega |K(\omega, d_1) - K(\omega, d_2)| \end{cases}$$

Let $\mathcal{U}$ and $\mathcal{D}$ be the sigma-algebras generated by the open sets of Ω and D, respectively,[4] and $\mathcal{U} \times \mathcal{D}$ be the sigma-algebra generated by the class of sets of the form $U \times E$, with U in $\mathcal{U}$, E in $\mathcal{D}$.[5]

Assumption Ib.

$K(\omega, d)$ is measurable $(\mathcal{U} \times \mathcal{D})$. Let $\mathcal{X}$ be a sigma-algebra of subsets of X, let X^n be the n-fold Cartesian product of X with itself, let $\mathcal{X}^n$ be the corresponding sigma-algebra of subsets of X^n and, for every ω, let $F_\omega^{(n)}$ be the corresponding probability measure on X^n.

Assumption II.

There is a measure μ defined in $\mathcal{X}$ such that for every ω, and every n,

---

4,5 Cf. [2], p. 22 and p. 140.

$F_\omega^{(n)}$ is absolutely continuous[6] with respect to the product measure $\mu^{(n)}$. We denote the density of $F_\omega^{(n)}$ by $f_\omega^{(n)}$. Further, we assume that, for every n, $f_\omega^{(n)}(x_1, \ldots, x_n)$ is measurable $(\mathcal{H} \times \chi^n)$.

By a <u>decision function</u> is meant a <u>measurable</u> function δ_n from X^n to D.

2.2. The Concept of Consistency.

<u>Definition 1</u>. A sequence ($\delta_1, \delta_2, \ldots$ etc) of decision function is <u>consistent</u> ≡

$$\operatorname*{plim}_{n \to \infty} K[\omega_o, \delta_n(x_1, \ldots, x_n)] = \sup_{d \in D} K(\omega_o, d)$$ [7]

<u>Example.</u> <u>Estimation:</u>

Suppose one is estimating the parameter ω of a probability distribution $F_\omega(x)$, where ω lies in some interval Ω of the real line. The decision space D here consists of all estimates d of ω, i.e., $D = \Omega$. Let K be given by

$$K(\omega, d) = |\omega - d|.$$

Then the metrics on Ω and D given by (1) are both just the ordinary distance on the real line. If $\delta_n(x_1, \ldots, x_n)$ is what is known as a <u>consistent</u> estimator of ω then

$$\operatorname*{plim}_{n \to \infty} \delta_n(x_1, \ldots, x_n) = \omega_o$$

and δ_n is also a consistent sequence of decision functions for this gain function K. As a matter of fact we could have taken for K any monotone increasing continuous function of $|\omega - d|$.

In other words, using a consistent decision function enables one to get arbitrarily close to the maximum possible gain, with probability arbitrarily

---

6. Cf. [2], p. 124.

7. "plim" denotes "limit in probability." Cf. [2], p. 91, for convergence in measure.

close to 1, provided he takes enough observations.

2.3. Bayes Decisions.

We shall be interested in conditions under which so-called Bayes solutions of decision problems are consistent, so let us digress for a few pages to discuss some properties of Bayes solutions.

Definition 2. A decision function $\hat{\delta}_n$ is Bayes with respect to a probability measure G_o on Ω, if it maximizes the expected value of $K[\omega, \delta_n(x_1, \ldots, x_n)]$. I.e.,

$$\hat{\delta}_n = \underset{\delta_n}{\text{max'er}} \int_{\Omega} \int_{X^n} K[\omega, \delta_n(x_1, \ldots, x_n)]\, f_{\omega}(x_1, \ldots, x_n)\, d\mu^n\, dG_o$$

Definition 3. $G_n(W|x_1, \ldots, x_n)$ is the a posteriori probability measure relative to the a priori probability measure G_o = for all W in $\mathcal{A}$:

$$G_n(W|x_1, \ldots, x_n) = \frac{\int_W f_{\omega}(x_1, \ldots, x_n) dG_o}{\int_{\Omega} f_{\omega}(x_1, \ldots, x_n) dG_o} .$$

In the classical theory of probability the measures G_n arise as conditional probabilities in Bayes theorem, but there will be no need to use Bayes theorem in this paper. In fact, for reasons which will be discussed in Section 4, G_o and G_n can be thought of as not necessarily being "probabilities" at all, but just nonnegative measures which take the value 1 on Ω, and through out the rest of the paper the measures discussed will always be of this kind.

There is an intimate connection between the concepts of Bayes decision and a posteriori measure, which is expressed in the following:

Theorem 1. If D is compact (in the metric defined by (1)), G_o is an a priori measure on Ω, and $f_{\omega}^{(n)}$ is integrable on $\Omega \times X^n$, then:

(i) there exists a decision function δ_n satisfying

$$(2) \qquad \delta_n(x_1, \ldots, x_n) = \underset{d \in D}{\text{max'er}} \int_\Omega K(\omega,d)dG_n(W|x_1, \ldots, x_n), \; ^{8/}$$

for all $(x_1, \ldots, x_n)$

which is Bayes with respect to G_o, where G_n is the a posteriori measure relative to G_o; and

(ii) if δ_n' is Bayes with respect to G_o, then (2) holds for δ_n' except possibly on a set $Y \subseteq X^n$ such that:

$$\int_\Omega \int_Y f_\omega^{(n)}(x_1, \ldots, x_n)d\mu^n \, dG_o = 0 .$$

<u>Proof:</u> We shall see that any (χ^n) measurable function δ_n satisfying (2) is a Bayes solution; hence the first job is to show that there exists such a measurable function.

For the rest of the proof we drop the index n and denote $(x_1, \ldots, x_n)$ by $\bar{x}$.

Let

$$(3) \qquad u(\bar{x},d) = \int_\Omega K(\omega,d) \, f_\omega(\bar{x})dG_o .$$

Equation (2) is equivalent to:

$$(4) \qquad \delta(\bar{x}) = \underset{d \in D}{\text{max'er}} \; u(\bar{x},d) .$$

From Assumption II it follows that $u(\bar{x},d)$ is measurable (χ^n). By the definition of the metric on D, $u(\bar{x},d)$ is (uniformly) continuous in d for every $\bar{x}$, and hence, since D is compact, takes on its maximum over D. Let D^* be a countable set which is dense in D (compactness).[9/] To see that

---

8. If $u(\hat{y}) = \max_y u(y)$, we say $\hat{y} = \underset{y}{\text{max'er}}\, u(y)$.

9. A compact metric space is separable.

$\max_d u(\bar{x},d)$ is measurable χ^n we observe that

$$\max_d u(\bar{x},d) = \lim_{m\to\infty} \max_{i\leq m} u(\bar{x},d_i)$$

where d_i, $(i = 1, 2, \ldots)$ are the elements of D^*.

We proceed to construct the required function as the limit of a sequence of measurable functions.

By the compactness of D there exist a finite number of closed spheres S_i with centers d_i and radii 1, which cover D. Let

$$D_{\bar{x}} = \left\{ d \middle| d = \text{max'er}_{d'}\, u(\bar{x},d') \right\}$$

$$X_k(d) = \left\{ \bar{x} \middle| \max_{d'} u(\bar{x},d') - u(\bar{x},d) < \frac{1}{2^k} \right\}$$

and $$Z_i = \left\{ \bar{x} \middle| S_i \cap D_{\bar{x}} \neq 0 \right\} .$$

<u>Lemma:</u> $$Z_i = \bigcap_k \bigcup_{d^* \in S_i} X_k(d^*) .$$

<u>Proof:</u> If $\bar{x}_o$ is in Z_i, there is a d_o in S_i such that $d_o = \text{max'er}_d\, u(\bar{x}_o,d)$. From the continuity of $u(\bar{x}_o,d)$ in d, and the denseness of D^* in D, it follows that for any k there is a d^* in $D^* \cap S_i$ such that

$$u(\bar{x}_o,d_o) - u(\bar{x}_o,d^*) < \frac{1}{2^k} .$$ [10]

That is to say $\bar{x}_o$ is in $\bigcap_k \bigcup_{d^* \in S_i} X_k(d^*)$.

---

10. Actually $D^* \cap S_i$ may not be dense in S_i, but since S_i is compact, and therefore separable, we could add (at most) a countable number of elements to D^* so that $D^* \cap S_i$ would become dense in S_i. There are only a finite number of S_i, so this presents no difficulty.

On the other hand, if $\bar{x}_o$ is in this last mentioned set, then there is a sequence d_j^* of elements of $D^* \cap S_i$ such that

$$\lim_{j \to \infty} u(\bar{x}_o, d_j^*) = \max_d u(\bar{x}_o, d) .$$

By compactness there is a subsequence of the d_j^*'s which converges to some point d_o. Since S_i is closed, d_o is in S_i, and $u(\bar{x}_o, d_o) = \max_d u(\bar{x}_o, d)$, because of the continuity of u. Hence $\bar{x}_o$ is in Z_i, and the lemma is proved.

We now define

$$X_i = Z_i - \bigcup_{m<i} Z_m$$

and

(5) $$\delta^{(1)}(\bar{x}) = d_i \quad \text{if} \quad \bar{x} \in X_i \quad .$$

Since the $Y_k(d^*)$ are measurable, the lemma tells us that the Z_i are also measurable; hence $\delta^{(1)}$ is measurable χ^n.

Consider one of the sets

$$T_i = S_i - \bigcup_{m<i} S_m \quad .$$

Since the closure of T_i is compact there exist a finite number of closed spheres with centers d_j' and radii $\frac{1}{2}$, which cover T_i. Let S_j' be the intersection of T_i with the j'th such sphere.

Let:

$$Y_k'(d) = \left\{ \bar{x} \middle| \bar{x} \in X_i \cdot \max_{d'} u(\bar{x}, d') - u(\bar{x}, d) < \frac{1}{2^k} \right\}$$

$$Z_j' = \left\{ \bar{x} \middle| x \in X_i \cdot S_j' \cap D_{\bar{x}} \neq 0 \right\} .$$

As in the previous lemma we can show that

(6) $$Z_j' = \bigcap_k \bigcup_{d^* \in S_j'} Y_k'(d^*) .$$

The proof that Z_j' is contained in the right side of (6) is the same as in the lemma. To prove the inclusion in the other direction we proceed as in the lemma up to where we have a sequence d_m^* in S_j' which converges to some d_o and such that

$$\lim_{m \to \infty} u(\bar{x}_o, d_m^*) = u(\bar{x}_o, d_o) = \max_d u(\bar{x}_o, d) \quad .$$

Since S_j' is not (necessarily) closed we cannot argue as before. However, an examination of the definition of T_i will show that d_o, being a limit point of S_j', is either in T_i, and hence in S_j', or it is in S_m, for some $m < i$, in which case $D_{\bar{x}_o} \cap S_m \neq 0$, $\bar{x}_o$ is not in X_i, and hence not in any X_k'. Contradiction!

Hence we can define (on T_i):

$$\delta^{(2)}(\bar{x}) = d_j' \quad \text{if} \quad \bar{x} \in Z_j' - \bigcup_{m<j} Z_j' \; .$$

This can be done for all the sets T_i, and the arguments for $\delta^{(1)}$ will apply here to show that $\delta^{(2)}$ is measurable ($\mathcal{X}^n$).

By successive repetitions of the above procedure we can then construct a sequence of functions $\delta^{(m)}$, ($m = 1, 2, \ldots$), with the properties:

(a) For every m, $\delta^{(m)}$ is measurable ($\mathcal{X}^n$).

(b) For every $\bar{x}$, the values $\delta^{(m)}(\bar{x})$ lie in a sequence of nested sets, whose diameters are $\leq \frac{2}{m}$.

(c) For every $\bar{x}$ there is a $d_{\bar{x}}$ in $D_{\bar{x}}$ such that $\rho(\delta^{(m)}(\bar{x}), d_{\bar{x}}) \leq \frac{2}{m}$.

Property (b) implies that for every $\bar{x}$,

(7) $$\delta(\bar{x}) = \lim_{m \to \infty} \delta^{(m)}(\bar{x}) \quad \text{exists.}$$

For every $\bar{x}$, the uniform continuity of $u(\bar{x}, d)$ in d and property (c) together imply that:

(8) $$u(\bar{x}, \hat{\delta}(\bar{x})) = \lim_{m \to \infty} u(\bar{x}, \delta^{(m)}(\bar{x})) = \max_d u(\bar{x}, d) \quad .$$

Property (a) and equation (7) imply that $\hat{\delta}$ is measurable ($\mathcal{X}^n$).

The worst of the proof is done, and it remains to show that $\hat{\delta}$ as defined by (7) is Bayes with respect to G_o. We first note that by Assumption Ib and the ($\mathcal{X}^n$) measurability of $\hat{\delta}$, $K(\omega, \hat{\delta}(\bar{x}))$ is measurable ($\mathcal{W} \times \mathcal{X}^n$).

The rest follows from Fubini's theorem;[11] for by equations (3) and (8):

$$\int_{\Omega}\int_{X^n} K(\omega, \hat{\delta}(\bar{x})\; f_{\omega}(\bar{x})\; d\mu^n\; dG_o$$

$$= \int_{X^n}\int_{\Omega} K(\omega, \hat{\delta}(\bar{x})\; f_{\omega}(\bar{x})\; dG_o\; d\mu^n$$

$$= \int_{X^n} \max_{d} \int_{\Omega} K(\omega, d)\; f_{\omega}(\bar{x})\; dG_o\; d\mu^n$$

$$= \max_{\delta} \int_{X^n}\int_{\Omega} K(\omega, \delta(\bar{x}))f_{\omega}(\bar{x})\; dG_o\; d\mu^n$$

$$= \max_{\delta} \int_{\Omega}\int_{X^n} K(\omega, \delta(\bar{x}))f_{\omega}(\bar{x})\; d\mu^n\; dG_o$$

and hence $\hat{\delta}$ is Bayes with respect to G_o. Further, this same kind of argument gives us a proof of (ii). Q.E.D.

If the double integral of Definition 2 does not take on a maximum value, we cannot speak of a Bayes decision function; hence the following:

<u>Definition 4</u>. For fixed n, a sequence of decision functions $\hat{\delta}_n^{(m)}(m=1,2,\ldots)$ is <u>Bayes</u> with respect to a probability measure G_o on Ω if (using the notation of Theorem 1),

$$\lim_{m\to\infty}\int_{\Omega}\int_{X^n} K(\omega, \hat{\delta}_n^{(m)}(\bar{x}))f_{\omega}(\bar{x})\; d\mu^n\; dG_o$$

$$= \sup_{\delta_n}\int_{\Omega}\int_{X^n} K(\omega, \delta_n(\bar{x}))f_{\omega}(\bar{x})\; d\mu^n\; dG_o$$

<u>Theorem 2</u>. If D is separable (in the metric defined by (1)), G_o is an

11. cf. [2], p. 145

a priori measure on Ω, and $f_\omega^{(n)}$ is integrable on $\Omega \times X^n$, then:

(i) there exists a sequence of decision functions $\delta_n^{(m)}$ satisfying:

$$(9) \qquad \lim_m \int_\Omega K(\omega, \delta_n^{(m)}(\bar{x})) \, dG_n(W \mid \bar{x})$$

$$= \sup_d \int_\Omega K(\omega, d) \, dG_n(W \mid \bar{x}), \text{ for all } \bar{x} = (x_1, \ldots, x_n)$$

which is Bayes with respect to G_o, where G_n is the a posteriori measure relative to G_o; and

(ii) If $\delta_n^{(m)}$ is any Bayes sequence with respect to G_o, then (9) holds for $\delta_n^{(m)}$ except possibly on a set $Y \subseteq X^n$ such that

$$\int_\Omega \int_Y f_\omega(x_1, \ldots, x_n) \, d\mu^n \, dG_o = 0$$

<u>Proof</u>: The proof of this theorem is simpler than that of Theorem 1 since we are not concerned with proving convergence in the space D.

Using the notation of Theorem 1, let

$$S(\bar{x}) = \sup_d u(\bar{x}, d)$$

$$Y_m(d) = \left\{ \bar{x} \,\middle|\, (S(\bar{x})) - u(\bar{x}, d) < \frac{1}{m} \right\}$$

Set $D^* = \left\{ d_1^*, d_2^*, \ldots \right\}$ be a countable set dense in D (separability). As in Theorem 1, it is easily seen that $S(\bar{x})$ is measurable (χ^n) and $Y_m(d)$ is in χ^n, for every d and m. Define:

$$\delta^{(m)}(\bar{x}) = d_j^* \text{ if } \bar{x} \in Y_m(d_j^*) - \bigcup_{i<j} Y_m(d_i^*)$$

Then $\delta^{(m)}$ is measurable (χ^n), for each m, and

$$\lim_m \delta^{(m)}(\bar{x}) = S(\bar{x}), \text{ for all } \bar{x}.$$

An argument similar to that of Theorem 1 now shows that the sequence $\delta^{(m)}$ is Bayes with respect to G_o, and also proves statement (ii). Q.E.D.

2.4. The Consistency of Bayes Decisions

We can now begin to consider the question: "When are Bayes decision functions consistent?"

Let $G_n(W | x_1, \ldots, x_n)$ be a sequence of a posteriori probability measures.

Definition 5: G_n is said to converge upon ω_o in Ω if:

for every open U with $\omega_o \in U$, $\operatorname{plim}_{n \to \infty} G_n(U | x_1, \ldots, x_n) = 1$

for every closed W with $\omega_o \notin W$, $\operatorname{plim}_{n \to \infty} G_n(W | x_1, \ldots, x_n) = 0$

Theorem 3. Suppose D is separable and, the a posteriori measures $G_n(W | x_1, \ldots, x_n)$ converge upon ω_o (the true state of nature); then for any K satisfying Assumption I, if $\delta_n^{(m)}$, $(m = 1, 2, \ldots)$, is, for every n, a Bayes sequence of decision functions with respect to G_o, then there is a consistent sequence of decision functions γ_n such that, for every n, γ_n is an element of $\{\delta_n^{(m)}\}$.

Corollary. If D is compact and G_n converges upon ω_o, then for any K satisfying Assumption I the sequence of Bayes decisions with respect to G_o is consistent.

Proof. Let $S_o = \sup_d K(\omega_o, d)$

$$S_n(x_1, \ldots, x_n) = \sup_d \int_\Omega K(\omega, d) \, dG_n(W | x_1, \ldots, x_n)$$

By Theorem 2, for every n and almost every $(x_1, \ldots, x_n)$,[12/] there is an m such that:

(10) $$S_n(x_1, \ldots, x_n) - \int_\Omega K(\omega, \delta_n^{(m)}(x_1, \ldots, x_n)) \, dG_n(W | x_1, \ldots, x_n) < \frac{1}{n}$$

Let $\gamma_n = \delta_n^{(m)}$ for that m.

---

12. Meaning: except possibly on a set $Y^n \subseteq X^n$ such that
$$\int_\Omega \int_{Y^n} f_\omega^{(n)}(x_1, \ldots, x_n) \, d\mu^n \, dG_o = 0$$

For the remainder of the proof we will leave out explicit reference to $(x_1,\ldots,x_n)$ in all the equations, these being understood to hold for all, or almost all, $(x_1,\ldots,x_n)$. Thus we can write:

(11)
$$S_o - K(\omega_o, \gamma_n) \leq \left| S_o - \int_\Omega K(\omega, \gamma_n)\, dG_n \right| + \left| \int_\Omega K(\omega, \gamma_n)\, dG_n - K(\omega_o, \gamma_n) \right|$$

By equation (1), for any $\varepsilon > 0$ there is a neighborhood U of ω_o such that for every ω in U and every d in D,

$$\left| K(\omega, d) - K(\omega_o, d) \right| < \varepsilon$$

and since G_n converges upon ω_o, for any $\beta > 0$ there is an integer $N(\varepsilon)$ such that for all $n \geq N(\varepsilon)$

(12) $$P_r\left\{ C_n(U, \varepsilon) \right\} > 1 - \beta$$

where $C_n(U, \varepsilon)$ denotes the condition:

$$1 - G_n(U) < \varepsilon$$

Focusing attention on the second term of the right side of (11), $C_n(U, \varepsilon)$ implies:

(13)
$$\left| \int_\Omega K(\omega, \gamma_n)\, dG_n - K(\omega_o, \gamma_n) \right|$$
$$\leq \left| \int_U (K(\omega, \gamma_n) - K(\omega_o, \gamma_n))\, dG_n \right| + \left| \int_{\bar{U}} (K(\omega, \gamma_n) - K(\omega_o, \gamma_n))\, dG_n \right|$$
$$\leq \varepsilon + 2M\varepsilon \text{ (Assumption I).}$$

Considering now the first term of the right side of (11), $C_n(U, \varepsilon)$ implies:

(14)
$$\int_\Omega K(\omega, \gamma_n)\, dG_n - S_o$$
$$\leq \int_\Omega K(\omega, \gamma_n)\, dG_n - K(\omega_o, \gamma_n)$$
$$\leq \varepsilon + 2M\varepsilon \qquad \text{(by (13))}$$

Let d_o be such that:

$$S_o - K(\omega_o, d_o) < \mathcal{E} .$$

Then $C_n(U, \mathcal{E})$ implies:

$$(15) \qquad \begin{aligned} & S_o - \int_{\Omega} K(\omega, \gamma_n)\, dG_n \\ & \leqslant K(\omega_o, d_o) + \mathcal{E} - \left(\int_{\Omega} K(\omega, d_o)\, dG_n - \frac{1}{n} \right) \\ & \leqslant \int_U (K(\omega_o, d_o) - K(\omega, d_o))\, dG_n \\ & \quad + \int_{CU} (K(\omega_o, d_o) - K(\omega, d_o))\, dG_n + \mathcal{E} + \frac{1}{n} \\ & \leqslant \mathcal{E} + 2M\,\mathcal{E} + \mathcal{E} + \frac{1}{n} \end{aligned}$$

Thus, by (11), (13), (14) and (15), $C_n(U, \mathcal{E})$ implies:

$$(16) \qquad \left| S_o - K(\omega_o, \gamma_n) \right| \leqslant 3\,\mathcal{E} + 4M\,\mathcal{E} + \frac{1}{n}$$

We wish to show that for any $\alpha, \beta > 0$ there is an integer N' such that for all $n \geqslant N'$

$$P_r \left\{ \left| S_o - K(\omega_o, \gamma_n) \right| \leqslant \alpha \right\} > 1 - \beta$$

and this is satisfied by taking

$$3\,\mathcal{E} + 4M\mathcal{E} < \frac{\alpha}{2} \quad \text{and}$$

$$N' \geqslant \max \left\{ N(\mathcal{E}), \frac{\alpha}{2} \right\} \qquad \text{Q.E.D.}$$

Theorem 3, together with its corollary, is the best of its kind in the sense that if the sequence does not converge upon ω_o for any true ω_o, then in the compact case, for example, there is a gain function K such that the sequence of Bayes solutions will not be consistent. However, given K, the following more restrictive statement can be made:

<u>Definition 5</u>. ω_1 is <u>related</u> to ω_2 ≡

$$d_o = \underset{d}{\text{max'er}}\ K(\omega_1, d) \Rightarrow d_o = \underset{d}{\text{max'er}}\ K(\omega_2, d)$$

<u>Theorem 4</u>. Suppose G_n is a sequence of a posteriori measures such that for every ω, G_n converges upon $\gamma(\omega)$, given that ω is "true", where $\gamma(\omega)$ is some function from Ω into itself. Suppose further that D is compact and $\delta_n(x_1,\dots,x_n)$ is a sequence of Bayes decision functions with respect to G_o. Then a sufficient condition that $\delta = (\delta_1, \delta_2,\dots)$ be consistent is that $\gamma(\omega)$ be related to ω, for every ω.

<u>Lemma 1</u>. Let

$$\begin{cases} D_o = \{d \mid d \text{ ; } \underset{d'}{\text{max'er}}\ K(\omega_o, d'). \\ S_o = \max\limits_{d} K(\omega_o, d) \end{cases}$$

Given any $\varepsilon > 0$ there exists $\eta > 0$ such that if $\rho(d, D_o) > \varepsilon$ then $S_o - K(\omega_o, d) > \eta$.

<u>Proof</u>. Suppose the contrary. Then there exists $\varepsilon > 0$ such that for every positive integer n there exists d_n in D such that:

$$\begin{cases} \rho(d_n, D_o) > \varepsilon \\ S_o - K(\omega_o, d_n) \leq \frac{1}{n} \end{cases}$$

By the compactness of D, the sequence $\{d_n\}$ has a limit point d_o, and by the continuity of K, $K(\omega_o, d_o) = S_o$, i.e. d_o is in D_o. But also $\rho(d_o, D_o) \geq \varepsilon$. Contradiction!

<u>Lemma 2</u>. If δ_n is a sequence of decision functions such that

$$\operatorname*{plim}_{n\to\infty}\left(S_o - K(\omega_o, \delta_n)\right) = 0, \qquad \text{then}$$

$$\operatorname*{plim}_{n\to\infty} \rho(D_o, \delta_n) = 0.$$

<u>Proof</u>. If not then there exist ε and $\beta > 0$, and a subsequence δ_{n_i} such that for every i,

$$P_r \left\{ \rho(\delta_{n_i}, D_o) > \varepsilon \right\} > \beta.$$

Taking η as given by Lemma 1, we then have:

$$P_r \left\{ \left| S_o - K(\omega_o, \delta_{n_i}) \right| > \eta \right\} > \beta$$, for every i, which contradicts the hypothesis.

To complete the proof of the theorem we first note that the proof of Theorem 3 shows that given any ω,

$$\operatorname*{plim}_{n\to\infty} (\max_d K(\gamma (\omega), d) - K(\gamma(\omega), \delta_n)) = 0$$

Let $D_\omega = \left\{ d \,\middle|\, d = \operatorname*{max'er}_{d'} K(\omega, d) \right\}$. Then by Lemma 2, and the fact that $\gamma(\omega)$ is related to ω,

$$\operatorname*{plim}_{n\to\infty} \rho(D_\omega, \delta_n) = 0$$

which in turn implies that

$$\operatorname*{plim}_{n\to\infty} (\max_d K(\omega, d) - K(\omega, \delta_n) = 0. \qquad \text{Q.E.D.}$$

The question of the necessity of the condition given in Theorem 4 is complicated by a certain amount of indeterminacy in case there exist decisions which are optimal under more than one state of nature. However a partial converse is given in the following:

<u>Theorem 5</u> Let $E(\omega) \equiv \left\{ d \,\middle|\, d = \operatorname*{max'er}_{d'} K(\omega, d') \right.$

If $\omega_1 \neq \omega_2$ implies $E(\omega_1) \cap E(\omega_2) = 0$, then under the conditions of Theorem 4, a necessary condition that $\delta = (\delta_{1,2 \cdots})$ be consistent is that $\gamma(\omega)$ be related to ω for, every ω.

<u>Proof</u>: Suppose $\gamma (\omega_o)$ is not related to ω_o. Then there exists $\Delta > 0$ such that for every d in $E(\gamma (\omega_o))$

$$K(\omega_o, d) + \Delta < \max_d K(\omega_o, d).$$

For if not then there is a sequence d_i in $E(\gamma (\omega_o))$ such that

$$K(\omega_o, d_i) + \frac{1}{i} \geq \max_d K(\omega_o, d)$$

and hence there is a limit point d_o of that sequence such that

$$K(\omega_o, d_o) = \max_d K(\omega_o, d).$$

i.e. d_o is in $E(\omega_o)$. But $E(\gamma(\omega_o))$ is easily seen to be closed; hence d_o is also in $E(\gamma(\omega_o))$. Contradiction!

Having made this point, the rest of the proof follows quickly from Lemma 2 of Theorem 4. Q.E.D.

3. The Convergence of a Posteriori Measures

It is clear from the definition of the a posteriori measures that if the a priori measure G_o assigns zero measure to a neighborhood of the true ω_o, then the corresponding sequence of G_n's cannot converge upon ω_o.

Definition: G_o is unprejudiced if $G_o(W) > 0$ for every open set $W \subseteq \Omega$.

For the case in which the random variables x_i are independent and identically distributed, Wald, in [3], proved the consistency of the maximum likelihood estimate of the parameter ω, under certain assumptions about Ω and about the density functions $f_\omega(x)$. Wald's proof can be modified in a minor way to show that under the same assumptions, the a posteriori measures corresponding to an unprejudiced a priori measure will converge upon the true ω_o, for any ω_o in Ω. The assumptions are too long to make it worth while to in clude them here ; they involve, among other things, the continuity of $f_\omega(x)$ in ω for almost all x, and are satisfied for most of the commonly used density functions. This whole question, including the case of dependent random variables, deserves further study.

4. Interpretation of Results

By definition, if one uses a consistent decision function, one can get arbitrarily close to the maximum possible gain, with probability arbitrarily close to 1, provided enough observations are taken. Theorem 3 and the considerations of section 3 tell us that, under regular enough conditions, any sequence of decision functions which are Bayes with respect to an unprejudiced a priori measure on Ω will be consistent.

Naturally, the interpretation of these last two statements rests upon the interpretation given to the word <u>probability</u>. Note that their are two sets of probability measures floating about here:

(1) The set of measures $F_{\omega}^{(n)}$ on the spaces of observations; it is a sequence $F_{\omega_0}^{(n)}$ of these which determine the "probability" occurring in the definition of consistency.

(2) The set of possible a priori measures on Ω, and their corresponding a posteriori measures.

If one regards probability as a logical or subjective concept, e.g. measure of implication, degree of confirmation, degree of belief, etc. (as in the work of Keynes, Jeffreys, Ramsey, de Finneti, Savage, Koopman , Carnap, etc.), then the a posteriori probability measures arise naturally as a consequence of Bayes theorem. Provided our decision maker is assumed to want to maximize the expectation of something or other, he will naturally use "Bayes decision functions" and the first paragraph of this section becomes a statement about conditional probabilities in his probability structure.

If on the other hand, probability is regarded as something objective, i.e. a mathematical description of something to be found in nature, like relative frequencies of certain kinds of events, then it is only the measures described in (1) which can reasonably be described as "probabilities." The measures of (2), however, though not probabilities in this sense, may arise in another way. Wald [1], p. 57, has shown that if Ω and D are both compact, then every <u>admissible</u> decision function is Bayes with respect to some a priori measure on Ω.13/ Thus the a priori measures are means of characterising admissible

13/ More precisely, the class of Bayes decision functions is <u>complete</u> in Wald's sense.

decisions. It should be emphasised, though, that for any given decision rule the characteristic a priori measure may depend upon the number of observations to be taken. For example the minimax estimator of the binomial parameter p , when the loss function is the square of the error, is:

$$\frac{x}{n}\left(\frac{\sqrt{n}}{\sqrt{n}+1}\right)+\frac{1}{2(\sqrt{n}+1)}$$

where x is the number of successes in n trials (cf [4]). This is a consistent estimator, and for every n is Bayes with respect to the a priori density $C[\,p\,(1-p\,)]^{\left(\frac{\sqrt{n}}{2}-1\right)}$. This density depends upon n, and the theorems of this paper, as stated, do not apply to this case.

However, Theorem 4 would remain true if the words "a posteriori" were struck out, and the δ_n were any sequence of decision functions satisfying:

$$\delta_n\,(x_1,\ldots,x_n) = \underset{d}{\text{max'er}} \int_{\Omega} K(\omega,\,d)\; dG_n\,(W \mid x_1,\ldots,x_n)$$

This form of the theorem will apply to the above example, but not in a very natural way.

At any rate, it is clear that the class of admissible consistent decision functions will be quite a large one in most problems, and within the framework of the "objective probability" viewpoint, this requirement goes only a part of the way towards characterising optimal procedures.

Finally, one can take the position that the word "probability" refers to <u>both</u> objective and subjective (or logical) phenomena. In this case the first paragraph of this section provides a link between the two, saying "If I obey certain laws of inductive inference, then I can get arbitrarily close to the maximum possible gain, with (objective) probability arbitrarily close to one, provided I take enough observations."

5. Historical Note

Theorem 1 was stated by Wald in [1], (p. 124), for the case in which D and Ω are finite. The result for more general spaces is to some extent implicit in theorems by Wald and Wolfowitz [5], and Arrow, Blackwell and Girshick [6], although these theorems were concerned with sequential rather than non-sequential problems. However, it appears that these authors did not demonstrate the measurability of the decision function defined by equation (2) (though such a thing is not likely to bother anyone in practice!)

Various authors have been concerned with the problem of the convergence of a posteriori measures. J. L. Doob [7] demonstrated this convergence under very simple and general conditions, but unfortunately the convergence may fail if the true ω_o lies in some exceptional set of measure zero. Unlike most sets of measure zero (which are usually innocuous) this one might cause difficulty, and it seems to me preferable to stick to stronger assumptions which rule out such a possibility.

References

[1] A. Wald, Statistical Decision Functions, Wiley, New York, 1950.

[2] P. R. Halmos, Measure Theory, Van Nostrand, New York, 1950.

[3] A. Wald, "Note on the Consistency of the Maximum Likelihood Estimate", Annals of Mathematical Statistics, Volume 20, (1949) pp 595-601.

[4] J. L. Hodges, Jr., and E. L. Lehmann, "Some Problems in Minimax Point Estimation", Annals of Mathematical Statistics, Volume 21, (1950), p. 190.

[5] A. Wald and J. Wolfowitz, "Bayes Solutions of Sequential Decision Problems", Annals of Mathematical Statistics, Volume 21 (1950), pp 82-99, esp. Theorem 3.6.

[6] K. J. Arrow, D. Blackwell and M. A. Girshick, "Bayes and Minimax Solutions of Sequential Decision Problems", Econometrica, Volume 17 (1949), pp. 213-244, esp. pp 216,217.

[7] J. L. Doob, "Application of the Theory of Martingales", Colloques Internationaux du Centre National de la Recherche Scientifique, XIII, Le Calcul des Probabilites et ses Applications, (1949) pp 23-27.