On the Theory of Statistical Decision Functions

by Kameo MATUSITA

1 Introduction

The theory of statistical decision functions has been developed since 1939 principally by A. Wald, and the results thus far obtained are presented in his book *Statistical Decision Functions (1950)*. He founded his theory on that of zero-sum two-person games with infinitely many strategies, an extension of the theory of J. von Neumann. In the present paper we shall give a new establishment of this theory on the basis of simpler assumptions. Our treatment differs from that of Wald in two principal points: Firstly, while Wald considered the randomized decision function i. e., the probability distribution on the space of all decisions, which is determined by the sample point, we shall consider the probability distribution on the space of all decision functions, which we call the *mixed* decision function. Secondly, while he took into account of the cost function from the beginning, we shall do that only after the main results are obtained. This way of treatment has brought about a considerable simplification of the theory, and as we hope it, it will be not without interest also from the viewpoint of the practical application. In the last section we shall treat the case where the number of possible distributions is finite, and show how we can make the risk in decision as small as desired.

2 Basic Assumptions and Some Easy Consequences

In this section we shall formulate the underlying assumptions of our theory.

Let R denote the sample space. Then

$(A.\ R)$ There are defined a Borel field $\mathfrak{K}$ of subsets of R and a measure m on $\mathfrak{K}$.

The sample space R may be a finite dimensional space or an infinite dimensional one and the measure m may be a Lebesgue-measure or a discrete one, i. e., a measure which gives the value 1 to each point of a certain subset consisting of (discrete) countably many points in R. In the following exposition we limit ourselves to the "continuous case", but it is clear that the "discrete case" can also be treated in a quite analogous way.

Let Ω be the set of all admissible distribution functions on $\mathfrak{K}$. The first assumption concerning Ω is the following

(A, Ω, I) Each distribution function F of Ω is absolutely continuous with regard to m, i.e. each F admits an expression $F(E) = \int_E p(x)\, dm$.

$p(x)$ is the density function of F; it is uniquely determined up to a set of m-measure zero.

Now, taking into account of the fact that $\sqrt{p(x)}$ belongs to L^2-class with respect to the measure m, as $\int_R p(x)\, dm = 1$, we introduce a metric in Ω as follows.

Definition The distance between any pair of elements F_1 and F_2 of Ω is given by $\|F_1 - F_2\| = \left(\int_R (\sqrt{p_1(x)} - \sqrt{p_2(x)})^2\, dm\right)^{\frac{1}{2}}$, where $p_1(x)$ and $p_2(x)$ denote the density functions of F_1 and F_2, respectively.

Ω becomes thus a metric space. We have then the following corollary to this definition.

Corollary 1 If a sequence of distribution $\{F_n\}$ converges to a distribution F in the sense of the metric $\| \ \|$, then the sequence $\{F_n(E)\}$ also converges to $F(E)$ uniformly with respect to E, where E denotes an arbitrary subset of $\mathfrak{R}$.

Proof As

$$|F_n(E) - F(E)| = \left| \int_E p_n(x)\, dm - \int_E p(x)\, dm \right|$$

$$\leqq \int_E |p_n(x) - p(x)|\, dm \leqq \int_R |p_n(x) - p(x)|\, dm$$

where $p_n(x)$, $p(x)$ denote the density functions of F_n and F respectively, it is sufficient to show that $\int_R |p_n(x) - p(x)|\, dm \to 0$ when $\int_R \left(\sqrt{p_n(x)} - \sqrt{p(x)}\right)^2 dm \to 0$.

Now, we have

$$\int_R |p_n(x) - p(x)|\, dm = \int_R \left|\sqrt{p_n(x)} + \sqrt{p(x)}\right| \left|\sqrt{p_n(x)} - \sqrt{p(x)}\right| dm$$

$$\leqq \left\{\int_R \left(\sqrt{p_n(x)} + \sqrt{p(x)}\right)^2 dm \int_R \left(\sqrt{p_n(x)} - \sqrt{p(x)}\right)^2 dm\right\}^{\frac{1}{2}}$$

$$\leqq 2\|F_n - F\|$$

The last term tends to zero as $n \to \infty$. Therefore, $\int_R |p_n(x) - p(x)|\, dm$ also tends to zero as $n \to \infty$, which completes the proof.

Then we make a further assumption concerning Ω.

($A. \Omega. II$) Ω is a compact metric space with respect to the metric $\| \ \|$.

From this assumption we have clearly the following

Corollary 2 Ω is separable.

Let D denote the set of all decisions, which can be made by the experimenter. Further, let $W(F, d)$ denote the weight function, which represents the loss by the decision d of D when F is the true distribution function. We make the following assumptions concerning the weight function.

$(A. W. I)$ $W(F, d)$ is a bounded, non-negative function defined on the product space $\Omega \times D$. We assume

$$0 \leqq W(F, d) \leqq 1 \qquad \text{for every } F \text{ and } d.$$

$(A. W. II)$ For every d of D $W(F, d)$ is a continuous function with respect to F.

Now we introduce a metric in D as follows.

Definition. The distance between two elements d_1, d_2 of D is given by $\rho_D(d_1, d_2) = \max_{F \in \Omega} | W(F, d_1) - W(F, d_2) |$.

D becomes thus a metric space. Then we make the following assumption.

$(A. D)$ D is a compact with respect to ρ_D.

From $(A. D)$ we have clearly the

Corollary 3 D is separable.

Definition A mapping from R onto D is called a (statistical) decision function.

As the class of all decision functions is too broad, we must restrict ourselves to a narrower class, which is easy to deal. For this purpose we introduce a notion of measurability of decision functions.

Definition A statistical decision function $d = \varphi(x)$ is called $\mathfrak{R}$-measurable, when for any open subset e of D the set $\{x; \varphi(x) \epsilon e\}$ belongs to $\mathfrak{R}$.

Then we have the following

Corollary 4 When a sequence of $\mathfrak{R}$-measurable decision functions $\{\varphi_n(x)\}$ converges to a decision function $\varphi(x)$ at every point x in R, then $\varphi(x)$ is also $\mathfrak{R}$-measurable.

Proof For an arbitrary open set e in D, we have the following logical equivalence:

$$\varphi(x) \epsilon e \rightleftarrows \varphi_n(x) \epsilon e, \ n \geqq \text{some } n_0 \qquad \left(\to x \epsilon \{x; \varphi_n(x) \epsilon e\}, n \geqq \text{some } n_0\right)$$

Hence

$$\{x; d(x) \epsilon e\} = \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{x; d_n(x) \epsilon e\} \ \epsilon \mathfrak{R}$$

This proves the corollary.

Now, we assume

$(A. \mathfrak{D})$ The class of decision functions at the disposal of the exprimenter is the set $\mathfrak{D}$ of all the $\mathfrak{K}$-measurable decision functions.

When a sequence of decision functions $\{\varphi_n(x)\}$ converges to a decision function $\varphi(x)$ at every point x, we say that $\{\varphi_n\}$ converges to φ. Then a topology can be defined in the class of all decision functions by this convergence and we have from Corollary 4 the following

Corollary 5 $\mathfrak{D}$ is a closed set.

3 Some Lemmas

In this section we shall prove some lemmas for the later use.

Lemma 1 $W(F, d)$ *is a continuous function on* $\Omega \times D$.

Proof Let $F_i \to F$, $d_i \to d$. Then we have

$$|W(F_i, d_i) - W(F, d)| \leqq |W(F_i, d_i) - W(F_i, d)| + |W(F_i, d) - W(F, d)|$$
$$\leqq \rho_D(d_i, d) + |W(F_i, d) - W(F, d)|$$

The last term tends to zero on account of $(A. W. II)$.

Lemma 2 *For any* F *of* Ω $W(F, \varphi(x))$ *is a* $\mathfrak{K}$*-measurable function of* x.

Proof The lemma follows clearly from the fact that

$$\{x; W(F, \varphi(x)) > \alpha\} = \{x; \varphi(x) \varepsilon e, e = \{d; W(F, d) > \alpha\}\}$$

for any real α.

As a consequence of this lemma we can consider the risk function

$$r(F, \varphi) = \int_R W(F, \varphi(x)) dF$$

for any F and $\varphi(x)$.

Lemma 3 $r(F, \varphi)$ *is a continuous function on* $\Omega \times \mathfrak{D}$, *and satisfies the inequality* $0 \leqq r(F, \varphi) \leqq 1$.

Proof Let $(F_i, \varphi_i) \to (F, \varphi)$, where F, $F_i \varepsilon \Omega$, φ, $\varphi_i \varepsilon \mathfrak{D}$. Then, we have

$$\left| r(F_i, \varphi_i) - r(F, \varphi) \right| = \left| \int_R W(F_i, \varphi_i(x)) p_i(x) dm - \int_R W(F, \varphi(x)) p(x) dm \right|$$
$$\leqq \left| \int_R W\big(F_i, \varphi_i(x)\big)\big(p_i(x) - p(x)\big) dm \right|$$
$$+ \left| \int_R \Big(W(F_i, \varphi_i(x)) - W(F, \varphi(x)) \Big) p(x) dm \right|$$

Now

$$\left| \int_R W(F_i, \varphi_i(x)) (p_i(x) - p(x)) dm \right|$$

(*) $$\leqq \int_R |p_i(x) - p(x)| dm \to 0 \qquad (i \to \infty)$$

On the other hand, as

$$\left| W(F_i, \varphi_i(x)) - W(F, \varphi(x)) \right| p(x) \leqq 2p(x),$$

$p(x)$ is integrable and for every x

$$\big(W(F_i, \varphi_i(x)) - W(F, \varphi(x))\big)p(x) \to 0 \qquad (i \to \infty),$$

it holds therefore

$$(\ast\!\ast) \qquad \int_R \big(W(F_i, \varphi_i(x)) - W(F, \varphi(x))\big)p(x)dm \to 0 \qquad (i \to \infty).$$

From ($*$) and ($\ast\!\ast$) we have

$$r(F_i, \varphi_i) \to r(F, \varphi) \qquad (i \to \infty),$$

which proves the first part of the lemma.

The second part is clear.

Lemma 4 $\mathfrak{D}$ *is a compact space.*

Proof As $\varphi(x)$ assumes a value in D for every x in R, $\mathfrak{D}$ is embedded in the (infinite) product space D^R as a topological space. Now, D is compact, therefore, D^R also is compact by Tchychonoff's theorem. Then it follows from Corollary 5 that $\mathfrak{D}$ is compact.

From this lemma we see that $\Omega \times \mathfrak{D}$ is a compact space. Hence we have the

Lemma 5 *When* $\varphi_n(x) \to \varphi(x)$ $(n \to \infty)$,

$$r(F, \varphi_n) \to r(F, \varphi) \qquad \textit{uniformly in } F.$$

Now we introduce a metric in $\mathfrak{D}$.

Definition For two elements φ, φ' of $\mathfrak{D}$ the distance $\rho_{\mathfrak{D}}$ between them is given

$$\rho_{\mathfrak{D}}(\varphi, \varphi') = \max_{F \in \Omega} |r(F, \varphi) - r(F, \varphi')|$$

The topology in $\mathfrak{D}$ by this metric $\rho_{\mathfrak{D}}$ is weaker than the former one, as can easily be seen from Lemma 5. Therefore, if we identify the elements, between which the distance is zero, and denote by $\mathfrak{D}^*$ the space obtained in this way from $\mathfrak{D}$, then we have the

Lemma 6 $\mathfrak{D}^*$ *is a compact, metric space.*

Consequently, $\mathfrak{D}^*$ *is separable.*

It is clear that $r(F, \varphi)$ is a continuous function on the product space $\Omega \times \mathfrak{D}^*$, which is a compact, metric and separable space.

Lemma 7 *(Helly's Theorem)* *Let* $\{\mu_n\}$ *be a sequence of distribution functions on a Borel field generated by all the open sets in a compact, separable (metric) space, which converges to* μ *in the ordinary sense.**) *Then it holds for any continuous function* $f(x)$ *and any set with* $\mu(\overline{E} - E^{\circ}) = 0$

*) The convergence of $\{\mu_n\}$ to μ in the *ordinary* sense means that for any set E with $\mu(\bar{E} - E^0) = 0$ $\mu_n(E)$ converges to $\mu(E)$, where $\bar{E}$ and E^0 denotes the closure and the open kernel of E respectively.

$$\int_E f(x)d\mu_n \rightarrow \int_E f(x)d\mu \qquad (n \rightarrow \infty)$$

Lemma 8 *The set of all the distributions on a Borel field generated by all the open sets in a compact, separable (metric) space is a compact space with respect to the ordinary topology.**)

These two lemmas are the Lemma 2 and 6 in my previous note *A Remark to the Wald's theory of Statistical Inference,* Ann. Inst. Stat. Math. Vol. I (1950) and the proofs are omitted.

Lemma 9 *Let an (m-n)-matrix* (a_{ij}) *of real numbers satisfy the following condition:*

For any system of non-negative numbers $\xi_1, \ldots\ldots, \xi_m$ *there exists at least one non-negative number among* $\sum_{i=1}^{m} \xi_i a_{ij}$, $(j = 1, \ldots\ldots, n)$.

Then, there exists a system of non-negative numbers $\eta_1, \ldots\ldots \eta_n$, *whose sum is 1, such that all* $\sum_{j=1}^{n} a_{ij}\eta_j$ $(i = 1, \ldots, m)$ *are non-negative.*

This is a theorem due to J. Ville and the proof is omitted.**)

4 Main Results

In this section we shall show the existence of an *optimum* solution to the problem of decision functions and deduce its main properties as they were obtained by Wald.

Let $\mathfrak{S}$, $\mathfrak{F}$ be the Borel fields generated by all open sets in Ω, $\mathfrak{D}^*$ respectively. In the following we shall consider the distributions μ, δ on those Borel field $\mathfrak{S}$, $\mathfrak{F}$ and call them singly the distributions on Ω, $\mathfrak{D}^*$. μ is called also an a priori distribution and δ a *mixed* decision function. As $\Omega \times \mathfrak{D}^*$ is metrizable and $r(F, \varphi)$ is continuous on $\Omega \times \mathfrak{D}^*$, $r(F, \varphi)$ is $\mathfrak{S} \times \mathfrak{F}$-measurable and we can consider the integral

$$r(\mu, \delta) = \int_\Omega \int_{\mathfrak{D}^*} r(F, \varphi) d\mu\, d\delta.$$

Let M and Δ represent the sets of all distributions on Ω and $\mathfrak{D}^*$ respectively. M and Δ are then compact with regard to the ordinary topology by Lemma 8, and we obtain from Lemma 7 the

THEOREM I $r(\mu, \delta)$ *is a continuous function on a compact space* $\mathrm{M} \times \Delta$.

As a consequence of this theorem we have

THEOREM II $\operatorname{Max}_{\mu \in M} r(\mu, \delta)$ *and* $\operatorname{Min}_{\delta \in \Delta} r(\mu, \delta)$ *are continuous functions on* Δ *and* M *respectively.*

---

*) The *ordinary* topology means the topology by the *ordinary* convergence.

**) Cf. J. Ville's note "*Sur la théorie générale des jeux où intervient l'habilité des joueurs*" in E. Borel "*Traité du calcul des probabilités et de ses applications, Tome IV, Facsc. II.*"

So, there exist μ_0 and δ_0 such that

$$\underset{\delta\epsilon\Delta}{\text{Min}}\ \underset{\mu\epsilon M}{\text{Max}}\ r(\mu, \delta) = r(\mu_0, \delta_0)$$

Such a δ_0 is called a *minimax solution.* We shall see in the following that a minimax solution is, in a sense, an optium solution to the decision problem.

It is also an immediate consequence of THEOREM I that for an arbitrary $\mu\epsilon$M there exists a $\delta_\mu\ \epsilon\Delta$, which minimizes $r(\mu, \delta)$. Such a δ_μ is called a *Bayes solution* relative to μ.

THEOREM III For any a priori distribution μ there exists a Bayes solution δ_μ relative to μ.

THEOREM IV For any Bayes solution δ_μ relative to μ we have

$$\delta_\mu\{\varphi_0; r(\mu, \varphi_0) = \underset{\varphi}{\text{Min}}\, r(\mu, \varphi)\} = 1.$$

THEOREM V Let $\{\delta_n\}$ be a sequence of Bayes solutions corresponding to a sequence of a priori distributions $\{\mu_n\}$, each δ_n being Bayes solution relative to an a priori distribution μ_n. Let, further, $\{\delta_n\}$ converge to a distribution δ_0. Then δ_0 is also a Bayes solution relative to some a priori distribution. That is, the limit distribution of a sequence of Bayes solutions is also a Bayes solution.

Proof As **M** is compact, we can select a convergent subsequence $\{\mu_{n'}\}$ from $\{\mu_n\}$. We denote its limit by μ. Now, let δ_μ be a Bayes solution relative to μ and assume that

$$r(\mu, \delta_0) - r(\mu, \delta_\mu) = \eta > 0.$$

Then, as $r(\mu_{n'}, \delta_{n'}) \to r(\mu, \delta_0)$ $(n' \to \infty)$, we have for an arbitrary positive number ε

$$r(\mu_{n'}, \delta_{n'}) > r(\mu, \delta_0) - \varepsilon = r(\mu, \delta_\mu) + \eta - \varepsilon$$

taking n' sufficiently large.

On the other hand, for a sufficiently large n' and an appropriate δ it holds

$$r(\mu, \delta_\mu) > r(\mu_{n'}, \delta) - \varepsilon$$

Thus we have

$$r(\mu_{n'}, \delta_{n'}) > r(\mu_{n'}, \delta) + \eta - 2\varepsilon.$$

When we take ε such as $\eta > 2\varepsilon$, then we have

$$r(\mu_{n'}, \delta_{n'}) > r(\mu_{n'}, \delta),$$

which contradicts the fact that $\delta_{n'}$ is a Bayes solution relative to $\mu_{n'}$. Hence

$$r(\mu, \delta_0) - r(\mu, \delta_\mu) = \eta = 0$$

and δ_0 is also a Bayes solution relative to μ.

When $\operatorname*{Max}_{\mu}\operatorname*{Min}_{\delta} r(\mu, \delta) = r(\lambda, \delta_\lambda)$, we call λ a *least favorable distribution*. Then we have clearly

THEOREM VI There exists a least favorable distribution.

THEOREM VII There exists a Bayes solution relative to every least favorable distribution, satisfying

$$r_0 \geqq r(F, \delta_\lambda) = \int_{\mathfrak{D}^*} r(F, \varphi) d\delta_\lambda$$

for all F in Ω, where r_0 denotes the value of $\operatorname*{Max}_{\mu}\operatorname*{Min}_{\delta} r(\mu, \delta)$.

Proof Let ε be an arbitrary positive number. Let $\Omega_1, \ldots\ldots, \Omega_m$ be subsets of Ω with

$$\Omega = \Omega_1 + \cdots + \Omega_m, \qquad \Omega_i \frown \Omega_j = 0 \qquad (i \neq j)$$

such that for each i it holds

$$|r(F, \varphi) - r(F', \varphi)| < \varepsilon \qquad \text{for } F, F' \epsilon \Omega_i$$

uniformly in φ. Further, let $\mathfrak{D}_1, \ldots\ldots, \mathfrak{D}_n$ be subsets of $\mathfrak{D}^*$ with

$$\mathfrak{D}^* = \mathfrak{D}_1 + \cdots + \mathfrak{D}_n, \qquad \mathfrak{D}_i \frown \mathfrak{D}_j = 0 \qquad (i \neq j)$$

such that for each i it holds

$$|r(F, \varphi) - r(F, \varphi')| < \varepsilon \qquad \text{for } \varphi, \varphi' \epsilon \mathfrak{D}_i$$

uniformly in F.

Take arbitrary elements $F_1, \ldots\ldots, F_m$ and $\varphi_1, \ldots\ldots, \varphi_n$ out of $\Omega_1, \ldots\ldots, \Omega_m$ and $\mathfrak{D}_1, \ldots\ldots, \mathfrak{D}_n$ respectively, and put

$$r_1 = \operatorname*{Max}_{(\xi)} \operatorname*{Min}_{\varphi_j} \sum_{i=1}^{m} \xi_i r(F_i, \varphi_j)$$

where $(\xi) = (\xi_1, \ldots\ldots, \xi_m)$ is restricted to satisfy $\xi_1 + \cdots + \xi_m = 1$ and $\xi_i \geqq 0$ $(i = 1, \ldots\ldots, m)$. Then the matrix

$$\begin{pmatrix} r_1 - r(F_1, \varphi_1) & r_1 - r(F_1, \varphi_2) \ldots\ldots & r_1 - r(F_1, \varphi_m) \\ r_1 - r(F_2, \varphi_1) & r_1 - r(F_2, \varphi_2) \ldots\ldots & r_1 - r(F_2, \varphi_n) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ r_1 - r(F_m, \varphi_1) & r_1 - r(F_m, \varphi_2) \ldots\ldots & r_1 - r(F_m, \varphi_m) \end{pmatrix}$$

satisfies the hypothesis of Lemma 9, consequently, there exists a $(\eta) = (\eta_1, \ldots\ldots, \eta_n)$ satisfying the above condition, such that

$$r_1 - \sum_{j=1}^{n} r(F_i, \varphi_j)\eta_j \geqq 0 \qquad i = 1, \ldots\ldots, m$$

As it is clear that $r_0 + \varepsilon \geqq r_1$, we have

$$r_0 + \varepsilon \geqq \sum_{j=1}^{n} r(F_i, \varphi_j)\eta_j \qquad i = 1, \ldots\ldots, m$$

therefore, when δ_ε denotes distribution on $\mathfrak{D}^*$ which gives to the points $\varphi_1, \ldots\ldots, \varphi_n$ the masses $\eta_1, \ldots\ldots, \eta_n$, respectively, we have

$$r_0 + \varepsilon \geqq r(F_i, \delta_\varepsilon) \qquad i = 1, \ldots, m$$

From this inequalities we obtain

$$r_0 + 2\varepsilon \geqq r(F, \delta_\varepsilon)$$

for all F in Ω. Now, let δ be a limit point of $\{\delta_\varepsilon\}$ when $\varepsilon \to 0$, then

$$r_0 \geqq r(F, \delta)$$

and consequently, for any least favorable distribution λ

$$r_0 \geqq r(\lambda, \delta) \geqq r(\lambda, \delta_\lambda) = r_0$$

where δ_λ denotes a Bayes solution relative to λ.

THEOREM VIII We have

$$\operatorname*{Min}_{\delta} \operatorname*{Max}_{\mu} r(\mu, \delta) = \operatorname*{Max}_{\mu} \operatorname*{Min}_{\delta} r(\mu, \delta).$$

Proof It holds clearly

(*) $$\operatorname*{Min}_{\delta} \operatorname*{Max}_{\mu} r(\mu, \delta) \geqq \operatorname*{Max}_{\mu} \operatorname*{Min}_{\delta} r(\mu, \delta).$$

Now, let δ_0 be a Bayes solution such that

$$r(F, \delta_0) \leqq r_0 \qquad \text{for all } F \text{ in } \Omega.$$

Then, for an aribitrary distributibution μ on Ω we have

$$r(\mu, \delta_0) \leqq r_0$$

therefore

$$\operatorname*{Max}_{\mu} r(\mu, \delta_0) \leqq r_0$$

which means

(⁑) $$\operatorname*{Min}_{\delta} \operatorname*{Max}_{\mu} r(\mu, \delta) \leqq r_0$$

Combining (*) with (⁑) we have

$$\operatorname*{Min}_{\delta} \operatorname*{Max}_{\mu} r(\mu, \delta) = \operatorname*{Max}_{\mu} \operatorname*{Min}_{\delta} r(\mu, \delta).$$

THEOREM IX Let λ be an arbitrary least favorable distribution and δ_0 an arbitrary minimax solution. Then (λ, δ_0) is a saddle point of $r(\mu, \delta)$, i. e., $\operatorname*{Max}_{\mu} r(\mu, \delta_0) = r(\lambda, \delta_0)$, $\operatorname*{Min}_{\delta} r(\lambda, \delta) = r(\lambda, \delta_0)$.

Proof. As

$$\operatorname*{Min}_{\delta} r(\lambda, \delta) = r_0 = \operatorname*{Max}_{\mu} r(\mu, \delta_0)$$

we have

$$r(\lambda, \delta_0) \leqq \operatorname*{Max}_{\mu} r(\mu, \delta_0) = \operatorname*{Min}_{\delta} r(\lambda, \delta) \leqq r(\lambda, \delta)$$

and

$$r(\lambda, \delta_0) \geqq \operatorname*{Min}_{\delta} r(\lambda, \delta) = \operatorname*{Max}_{\mu} r(\mu, \delta_0) \geqq r(\mu, \delta_0).$$

This means (λ, δ_0) is a saddle point of $r(\mu, \delta)$.

From this proof follows obviously

$$r(\lambda, \delta_0) = r_0 = \underset{\delta}{\mathrm{Min}}\, r(\lambda, \delta)$$

therefore, we have

THEOREM X Any minimax solution is a Bayes solution relative to every least favorable distribution.

From this theorem and THEOREM IV we obtain the

THEOREM XI For any least favorable distribution λ and any minimax solution δ_0 we have

$$\delta_0\{\varphi_0; r(\lambda, \varphi_0) = \underset{\varphi}{\mathrm{Min}}\, r(\lambda, \varphi)\} = 1.$$

THEOREM XII For any minimax solution δ_0 we have

$$r(F, \delta_0) \leqq r_0 \qquad \text{for all } F \text{ in } \Omega$$

and conversely, if δ_0 is a distribution on $\mathfrak{D}^*$ such that $r(F, \delta_0) \leqq r_0$ for all F, then δ_0 is a minimax solution.

Proof Let δ_0 be a minimax solution. If $r(F_0, \delta_0) > r_0$ for a point F_0, then there exists an open set ω, at each point of which $r(F, \delta_0) > r_0$ holds. Let μ be an a priori distribution such that $\mu(\omega) = 1$. Then we have $r(\mu, \delta_0) > r_0 = r(\lambda, \delta_0)$, where λ is a least favorable distribution. This, however, contradicts the fact, that (λ, δ_0) is a saddle point. Therefore we have

$$r(F, \delta_0) \leqq r_0 \qquad \text{for all } F \text{ in } \Omega.$$

Conversely, assume that $r(F, \delta_0) \leqq r_0$ holds. Then it holds

$$\underset{\mu}{\mathrm{Max}}\, r(\mu, \delta_0) \leqq r_0$$

As $r_0 = \underset{\delta}{\mathrm{Min}}\, \underset{\mu}{\mathrm{Max}}\, r(\mu, \delta)$, we have $\underset{\mu}{\mathrm{Max}}\, r(\mu, \delta_0) = r_0$, which means that δ_0 is a minimax solution.

As an immediate consequence of this theorem we have

THEOREM XIII A necessary and sufficient condition that a decision function δ_0 is a minimax solution is that for any decision function δ we have

$$r(F, \delta_0) \leqq \underset{F}{\mathrm{Max}}\, r(F, \delta)$$

for all F.

From THEOREM X and the proof of THEOREM VI follows the

THEOREM XIV For an arbitrary positive number ε let $\mathfrak{D}_1, \ldots\ldots, \mathfrak{D}_n$ be subsets of $\mathfrak{D}^*$ with

$$\mathfrak{D}^* = \mathfrak{D}_1 + \cdots + \mathfrak{D}_n, \quad \mathfrak{D}_i \frown \mathfrak{D}_j = 0 \qquad (i \neq j)$$

such that for each i it holds

$$|r(F, \varphi) - r(F, \varphi')| < \varepsilon \qquad \text{for } \varphi, \varphi' \in \mathfrak{D}_i$$

uniformly in F, and let $\varphi_1, \ldots, \varphi_n$ be arbitrary elements out of $\mathfrak{D}_1, \ldots, \mathfrak{D}_n$ respectively. Then there exists a mixed decision function δ with $\delta(\varphi_1, \ldots, \varphi_n) = 1$, such that it holds

$$|r(F, \delta) - r(F, \delta_0)| < \varepsilon \qquad \text{uniformly in } F$$

where δ_0 denotes a minimax solution.

The way to determinate the masses $\eta_1, \ldots, \eta_n$, which are to be given by δ to the points $\varphi_1, \ldots, \varphi_n$ respectively, can also be seen in that proof of THEOREM VI. Generally, for any δ there exists $\eta = (\eta_1, \ldots, \eta_n)$ with $\sum_i \eta_i = 1$ and $\eta_i \geqq 0$ $(i = 1, \ldots, n)$ such that it holds

$$|r(F, \delta) - \sum_i r(F, \varphi_i)\eta_i| < \varepsilon$$

uniformly in F. Thus, up to the magnitude of ε we have only to consider a finite number of decision functions.

THEOREM XV For any least favorable distribution and any minimax solution δ_0 it holds

$$(*) \qquad r(F, \delta_0) \equiv r_0 \qquad \text{for all } F \text{ in } \overline{\Omega}_\lambda,$$

where Ω_λ represents the set of all points F in Ω such that any open set containing F has a positive λ-measure, and conversely if (*) holds for a least favorable distribution λ, then δ_0 is a minimax solution.

Proof For a minimax solution δ_0, we have

$$r(F, \delta_0) \leqq r_0 \qquad \text{for all } F \text{ in } \Omega.$$

Now, assume $r(F_0, \delta_0) < r_0$ for a point F_0 in $\overline{\Omega}_\lambda$. Then there exists an open set ω contained in $\overline{\Omega}_\lambda$, such that at each point F of ω we have $r(F, \delta_0) < r_0$ and

$$\int_\omega r(F, \delta_0)\,d\lambda < r_0 \int_\omega d\lambda$$

On the other hand

$$\int_{\Omega-\omega} r(F, \delta_0)\,d\lambda \leqq r_0 \int_{\Omega-\omega} d\lambda$$

Hence we have

$$r_0 = r(\lambda, \delta_0) < r_0$$

which is a contradiction. Therefore

$$r(F, \delta_0) \equiv r_0 \qquad \text{in } \overline{\Omega}_\lambda.$$

The second part of the theorem is obvious.

A decision function δ^* is called uniformly better than a decision function δ, when we have

$$r(F, \delta^*) \leqq r(F, \delta) \qquad \text{for all } F \text{ in } \Omega$$

and at least for one point F_0

$$r(F_0, \delta^*) < r(F, \delta)$$

and a decision function δ^* admissible, when there is no decision function, which is uniformly better than δ^*. A class of decision functions C is called complete relative to $\mathfrak{D}^*$, when for any δ of $\mathfrak{D}^*$, which is not in C, there exists a δ^* in C, which is uniformly better than δ. Then we have

THEOREM XVI The class of all the Bayes solutions is complete relative to $\mathfrak{D}^*$.

Proof Let δ_1 be an arbitrary decision function which is not a Bayes solution, and set

$$r^*(F, \varphi) = r(F, \varphi) - (F, \delta_1).$$

Then $r^*(F, \varphi)$ is a continuous function of (F, φ). Therefore, we have a minimax solution δ_0 with respect to $r^*(F, \varphi)$, i. e., a decision function δ_0 such that

$$\operatorname*{Max}_F r^*(F, \delta_0) \leqq \operatorname*{Max}_F r^*(F, \delta)$$

for any δ. We have clearly $r^*(F, \delta_1) \equiv 0$, consequently

$$r^*(F, \delta_0) \leqq 0 \qquad \text{for all } F \text{ in } \Omega$$

i. e.,

$$r(F, \delta_0) \leqq r(F, \delta_1) \qquad \text{for all } F \text{ in } \Omega.$$

Now, there is at least one point F_0, such that

$$r(F_0, \delta_0) < r(F, \delta_1)$$

For, if $r(F, \delta_0) \equiv r(F, \delta_1)$, then for any least favorable distribution λ we would have

$$r(\lambda, \delta_0) = r(\lambda, \delta_1)$$

against our hypothesis that δ_1 is not a Bayes solution.

5 Consideration of the cost

Now, we take into account of the cost of the experiment when we adopt a decision function φ. This cost depends also upon the sample x obtained. Therefore, the cost is a function of φ and x, which we denote by $c(x, \varphi)$. The upper bound of the admissible cost is usually preassigned before beginning the experiment. There are two ways of preassigning the cost. The one limits $c(x, \varphi)$ for each x, and the other limits the average of $c(x, \varphi)$. In both cases we make the following assumption

$(A. C. I)$ The cost function $c(x, \varphi)$ is a continuous function of φ for each x.

In the first case, a constant K is preassigned such that for each x

$c(x, \varphi) \leqq K$. Then the class $\mathfrak{D}^{**}$ of decision functions satisfying $c(x, \varphi) \leqq K$ for all x, is a closed subset of $\mathfrak{D}^*$, consequently $\mathfrak{D}^{**}$ is a compact, separable (metric) space. Therefore, considering $\mathfrak{D}^{**}$ instead of $\mathfrak{D}^*$, the previous theory is applicable, and we see the existence of an optimum solution also in this restricted case.

In the second case we make further two assumptions.

(*A. C. II*) The cost function $c(x, \varphi)$ as a function of φ is uniformly continuous with respect to x.

(*A. C. III*) The cost function $c(x, \varphi)$ is a measurable function of x.

According to these assumptions the average of cost with respect to the distribution F, $\int_R c(x, \varphi)dF$, is defined and is a continuous function of φ.

In this case the limit, K, of this average is supposed as pressigned. Then the class of decision functions, satisfying $\int_R c(x, \varphi)dF \leqq K$ for all F, forms a compact, separable (metric) space, and again our theory may be applied.

6 The case where Ω consists of a finite number of distributions

In this section we shall show under reasonable conditions that when we are concerned with a finite number of distributions we can make the risk as small as desired by making the sample size sufficiently large, and at the same time give a formula to determinate the sample-size and a method to obtain the decision function explicitly, so as to make the risk smaller than the preassigned value. The case, when Ω consist of two distributions, will be treated in more detail, and we shall obtain a further results.

Definition For any pair of distributions F_1, F_2 we call the number

$$\rho = \rho(F_1, F_2) = \int_R \sqrt{p_1(x)}\,\sqrt{p_2(x)}\,dm$$

the *affinity* between F_1 and F_2, where $p_1(x)$ and $p_2(x)$ denote the density functions of F_1 and F_2 respectively.

Corollary 6

$$\rho(F_1, F_2) \begin{cases} = 1 & \text{when } \|F_1 - F_2\| = 0 \\ < 1 & \text{when } \|F_1 - F_2\| > 0 \end{cases}$$

and

$$\rho(F_1, F_2) < \rho(F_1, F_3) \qquad \text{when } \|F_1 - F_2\| > \|F_1 - F_3\|$$

This Corollary follows immediately from

$$\|F_1 - F_2\|^2 = 2 - 2\rho(F_1, F_2).$$

For example, when $p_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\alpha)^2}{2\sigma^2}}$, $p_2(x) = \frac{1}{\sqrt{2\pi}\,\sigma'} e^{-\frac{(x-\alpha')^2}{2\sigma'^2}}$ we have $\rho = \frac{\sqrt{2\sigma\sigma'}}{\sqrt{\sigma^2+\sigma'^2}} e^{-\frac{(\alpha-\alpha')^2}{4(\sigma^2+\sigma'^2)}}$, and when $p_1(m) = \frac{\alpha^m}{m!}e^{-\alpha}$, $p_2(m) = \frac{\alpha'^m}{m!}e^{-\alpha}$ $(m = 0, 1, 2, \ldots\ldots)$, $\rho = e^{-\left(\frac{\alpha+\alpha'}{2}-\sqrt{\alpha\alpha'}\right)}$.

THEOREM XVII) Let ρ be the affinity between two distinct distributions F_1 and F_2. Let ε be an arbitrary postive number and k an integer such that $\rho^k < \varepsilon$. When we put*

$$E_1 = \left\{(x_1, \ldots, x_k); p_1(x_1)\ldots p_1(x_k) \geqq p_2(x_1)\ldots p_2(x_k)\right\}$$

in the direct product space $R^{(k)}$ and $E_2 = E_1^c$, then we have

$$F_1^{(k)}(E_1) > 1 - \varepsilon,$$
$$F_2^{(k)}(E_2) > 1 - \varepsilon$$

where $F_1^{(k)}$ and $F_2^{(k)}$ represent the extended distributions on $R^{(k)}$ of F_1 and F_2, respectively.

Proof We have clearly

$$\begin{aligned} F_1^{(k)}(E_1) &= \int_{E_1} p_1(x_1)\ldots p_1(x_k)dm \ldots dm = 1 - \int_{E_2} p_1(x_1)\ldots p_1(x_k)dm \ldots dm \\ &\geqq 1 - \int_{E_2} \sqrt{p_1(x_1)\ldots p_1(x_k)p_2(x_2)\ldots p_2(x_k)}\, dm \ldots dm \\ &\geqq 1 - \int_{R^{(k)}} \sqrt{p_1(x_1)\ldots p_1(x_k)p_2(x_2)\ldots p_2(x_k)}\, dm \ldots dm \\ &= 1 - \left(\int_R \sqrt{p_1(x)p_2(x)}\, dm\right)^k \\ &= 1 - \rho^k > 1 - \varepsilon \end{aligned}$$

and similarly

$$F_2^{(k)}(E_2) > 1 - \varepsilon.$$

Now, let Ω consist of s distinct distributions $F_1, \ldots, F_s$, and for each i $(i = 1, \ldots, s)$ let d_i denote the correct decision when F_i is the true distribution and assume $W(F_i, d_i) = 0$. Let ε be an arbitrary positive number and k an integer such that $(s-1)\rho^k < \varepsilon$, where $\rho = \underset{\substack{(i,j)\\(i \neq j)}}{\mathrm{Max}}\, \rho(F_i, F_j)$.

Further, put

$$E_{ij} = \left\{(x_1, \ldots, x_k); p_i(x_1)\ldots p_i(x_k) \geqq p_j(x_i)\ldots p_j(x_k)\right\}$$
$$(i \neq j,\ i, j = 1, \ldots, s)$$

*) As to this theorem cf. also S. Kakutani, *On Equivalence of Infinite Product Measure*, Ann. Math. vol. 49 (1948).

and

$$E_i = \bigcap_{j \neq i} E_{ij}.$$

Then for every i we have from the above theorem

$$F_i^{(k)}(E_{ij}) > 1 - \rho^k,$$

consequently

$$\begin{aligned} F_i^{(k)}(E_i) &= 1 - F_i^{(k)}(E_i^c) \\ &= 1 - F_i^{(k)}(\bigcup_{j \neq i} E_{ij}^c) \\ &\geqq 1 - \sum_{i \neq j} F_i^{(k)}(E_{ij}^c) \\ &> 1 - (s-1)\rho^k > 1 - \varepsilon. \end{aligned}$$

All E_i are clearly disjoint with each other, and are measurable. Now, put

$$\varphi(x) = d_i \qquad \text{when} \qquad x \epsilon E_i \ (i = 1, \ \ , s)$$

and the value of $\varphi(x)$ is defined arbitarily for $x \epsilon \left(R^{(k)} - \sum_{i=1}^{s} E_i\right)$ so far as the thus obtained $\varphi(x)$ is measurable. Then this $\varphi(x)$ makes the risk smaller than ε. In fact, we have

$$\begin{aligned} r(F_i, \varphi) &= \int_{R^{(k)}} W(F_i, \varphi(x)) dF_i^{(k)} \\ &= \int_{E_1} + \cdots + \int_{E_s} + \int_{E_0} \\ &\leqq \int_{E_i^c} dF_i^{(k)} < \varepsilon. \end{aligned}$$

where $E_0 = R - \sum_{i=1}^{s} E_i$. Thus we have the

THEOREM XVIII)* *When Ω consists of a finite number of distinct distributions we can make the risk smaller than the aribitarily preassigned value by taking the sample-size and a decision function as mentioned above.*

Thus a minimax solution makes clearly the risk smaller than the same preassigued value. Of course, we assume here that a priori distributions on $R^{(k)}$, $\{F_1^{(k)}\}$, are the same as on Ω.

Now, for the decision function φ, the property

$$r(F_1, \varphi) = \cdots = r(F_s, \varphi)$$

is desirable. For the fulfillment of this condition the decision function

---

*) As to this theorem cf. also R. von Mises, *On the Problem of Testing Hypotheses*, Ann. Math. Stat. vol. 14 (1943) and H. Kudo, *On the "Power" Functions*, Research Memoirs of the Institute of Statistical Mathematics vol. 4 (1948) (in Japanese)

obtained above must be modified. In the following we shall give a decision function which satisfies this condition in the case where Ω consists of two distinct continuous distributions F_1 and F_2.

For any non-negative number γ put

$$E_\gamma = \left\{(x_1, \ldots, x_k); p_1(x_1)\ldots p_1(x_k) < \gamma p_2(x)\ldots p_2(x_k)\right\}$$

Then we have clearly

$$E_\gamma \subset E_{\gamma'} \quad \text{when} \quad \gamma < \gamma'$$

and when we put $g(\gamma) = W(F_1, d_2)\, F_1(E_\gamma)$, $g(\gamma)$ is a monotone-increasing continuous function of γ such that $g(0) = 0$, $\lim_{\gamma\to\infty} g(\gamma) = W(F_1, d_2)$. On the other hand, when we put $h(\gamma) = W(F_2, d_1)\, F_2(E_\gamma^c)$, $h(\gamma)$ is a monotone-decreasing continuous function such that $h(0) = W(F_2, d_1)$, $\lim_{\gamma\to\infty} h(\gamma) = 0$. Therefore, there exists a positive number γ_0 such that $g(\gamma_0) = h(\gamma_0)$, provided that both $W(F_1, d_2)$ and $W(F_2, d_1)$ are positive. (The case where one of $W(F_1, d_2)$ and $W(F_2, d_1)$ is zero is trivial.) Then put

$$\varphi_0(x) = d_1 \quad \text{when} \quad x \in E_{\gamma_0}^c$$
$$\varphi_0(x) = d_2 \quad \text{when} \quad x \in E_{\gamma_0}$$

and we have

$$r(F_1, \varphi_0) = r(F_2, \varphi_0) \qquad (= g(\gamma_0) = h(\gamma_0) < \rho^k)$$

This φ_0 is a minimax solution in the *strict* sense, i. e., $\operatorname*{Min}_{\varphi}\operatorname*{Max}_{F} r(F, \varphi) = r(F_1, \varphi_0) = r(F_2, \varphi_0)$. ($\varphi_0$ is, further, a minimax solution in the *general* sense, i. e., $\operatorname*{Min}_{\delta}\operatorname*{Max}_{\mu} r(\mu, \delta) = \operatorname*{Max}_{\mu} r(\mu, \varphi_0)$.*) To show that, for an arbitrary decision function φ put

$$A = \left\{x = (x_1, \ldots, x_k); \varphi(x) = d_1\right\},$$

$$B = \left\{x = (x_1, \ldots, x_k); \varphi(x) = d_2\right\}$$

Here we, of course, assume $d_1 \neq d_2$. If $r(F_1, \varphi_0) = r(F_2, \varphi_0) > r(F_1, \varphi)$, $r(F_2, \varphi)$, then we would have

$$\int_{E_{\gamma_0}} p_1(x_1)\ldots p_1(x_k)dm \ldots dm > \int_B p_1(x_1)\ldots p_1(x_k)dm \ldots dm$$

and

$$\int_{E_{\gamma_0}^c} p_2(x_1)\ldots p_2(x_k)dm \ldots dm > \int_A p_2(x_1)\ldots p_2(x_k)dm \ldots dm$$

---

*) cf. the Addendum below.

therefore,

$$\int_{E_{\gamma_0}-B} p_1(x_1)\dots p_1(x_k)dm\dots dm > \int_{B-E_{\gamma_0}} p_1(x_1)\dots p_1(x_k)dm\dots dm$$

and

$$\int_{E^c_{\gamma_0}-A} p_2(x_1)\dots p_2(x_k)dm\dots dm > \int_{A-E^c_{\gamma_0}} p_2(x_1)\dots p_2(x_k)dm\dots dm$$

As $E_{\gamma_0} - B \supset A - E^c_{\gamma_0}$, $E^c_{\gamma_0} - A \supset B - E_{\gamma_0}$, we have

$$\int_{E^c_{\gamma_0}-A} \gamma_0 p_2(x)\dots p_2(x_k)dm\dots dm > \int_{A-E^c_{\gamma_0}} \gamma_0 p_2(x_1)\dots p_2(x_k)dm\dots dm$$

$$> \int_{E_{\gamma_0}-B} p_1(x_1)\dots p_1(x_k)dm\dots dm > \int_{B-E^c_{\gamma_0}} p_1(x_1)\dots p_1(x_k)dm\dots dm$$

which contradicts

$$\int_{E^c_{\gamma_0}-A} \gamma_0 p_2(x_2)\dots p_2(x_k)dm\dots dm \leqq \int_{E^c_{\gamma_0}-A} p_1(x_1)\dots p_1(x_k)dm\dots dm.$$

Therefore, at least one of the two values $r(F_1, \varphi)$ and $r(F_2, \varphi)$ is equal or greater than $r(F_1, \varphi_0) = r(F_2, \varphi_0)$, which means that $\varphi_0(x)$ is a minimax solution in the strict sense.

Addendum

In the following we shall show that under certain conditions mixed decision functions may be eliminated from statistical decision rules.

THEOREM XIX When each F of Ω is atomless, and when for any finite number of distributions $F_1, \dots, F_l$ from Ω there exists only a finite number of decisions, which may be made concerning $F_1, \dots, F_l$, then for any mixed decision function δ there exists a (pure) decision function $\varphi(x)$ such that

$$r(F, \varphi) = r(F, \delta) \qquad \textit{for all } F \textit{ in } \Omega.$$

Proof By THEOREM XIV there exist for any positive ϵ a finite number of decision functions $\varphi_1(x), \dots, \varphi_n(x)$ and a real vector $(\eta) = (\eta_1, \dots, \eta_n)$ with $\sum_{k=1}^{n} \eta_k = 1$ and $\eta_k \geqq 0$ $(k = 1, \dots, n)$, such that it holds

$$|r(F, \delta) - \sum_{k=1}^{n} r(F, \varphi_k)\eta_k| < \epsilon$$

*) As to this Addendum cf. A. Dvoretzky, A. Wald and J. Wolfowitz, *Elimination in Certain Statistical Decision Procedures and Zero-Sum Two-Person Games*, Ann. Math. Stat. Vol. 22 (1951).

uniformly in F. Let $\{F_1, \ldots, F_l\}$ be a finite subset of Ω such that for any F there exists F_i with

$$|r(F, \varphi) - r(F_i, \varphi)| < \epsilon \qquad \text{uniformly in } \varphi$$

and we have clearly

$$|r(F, \delta) - \sum_{k=1}^{n} r(F_i, \varphi_k)\eta_k| < 2\,\epsilon$$

Now, for any measurable set E in R put

$$\mu_{ik}(E) = \int_E W(F_i, \varphi_k(x))dF_i$$
$$(i = 1, \ldots, l;\ j,\ k = 1, \ldots, n)$$

Then all μ_{ik} are atomless, therefore, by a result of Dvoretzky, Wald and Wolfowitz*) there exist subsets $S_1, \ldots, S_n$ of R such that

$$\mu_{ik}(S_j) = \eta_j\mu_{ik}(R) \qquad i = 1, \ldots, l, j = 1, \ldots, n,\ k = 1, \ldots, n.$$

and $$S_i \frown S_j = 0 \qquad (i \neq j).$$

Consequently we have

$$\sum_k \eta_k\mu_{ik}(R) = \sum_k \mu_{ik}(S_k) = \sum_{k=1}^{n} \int_{S_k} W(F_i,\ \varphi_k(x))dF_i$$

Let $D(F_1, \ldots, F_l) = \{d_1, \ldots d_m\}$ be all decisions concerning $\{F_1, \ldots, F_l\}$ and $M_{jk} = \{x\,;\, \varphi_k(x) = d_j\}$ $(j = 1, \ldots, m\ k = 1, \ldots, n)$. Then

$$\sum_{k=1}^{n} r(F_i, \varphi_k)\eta_k = \sum_{k=1}^{n} \eta_k \int_R W(F_i, \varphi_k(x))dF_i$$
$$= \sum_{k=1}^{n}\sum_{j=1}^{m} W(F_i, d_j)\ \ F_i(S_k \frown M_{jk})$$
$$= \sum_{j=1}^{m} W(F_i, d_j) \sum_{k=1}^{n} F_i(S_k \frown M_{jk})$$
$$(i = 1, \ldots, l)$$

Putting $\sum_k (S_k \frown M_{jk}) = N_j$, we have

$$\sum_{k=1}^{n} r(F_i, \varphi_k)\eta_k = \sum_{j=1}^{m} W(F_i, d_j)\, F_i(N_j)$$

Therefore, when we put

$$\varphi_0(x) = d_j \qquad \text{for} \quad x \epsilon N_j \qquad (j = 1, \ldots, m)$$

we have

$$\sum_{k=1}^{n} r(F_i, \varphi_k)\eta_k = r(F_i, \varphi_0) \qquad (i = 1, \ldots, l)$$

---

*) A. Dvoretzky, A. Wald and J. Wolfowitz, *Relations among certain Ranges of Vector Measures, Lemma 2*, Pacific J. Math. Vol. I.

Consequently, for any F in Ω there exists an F_i in $\{F_1, \dots F_l\}$ such that

$$|r(F, \delta) - r(F_i, \varphi_0)| < 2\epsilon$$

Hence we have

$$|r(F, \delta) - r(F, \varphi_0)| < 3\epsilon \qquad \text{for all } F \text{ in } \Omega.$$

As ϵ may be taken arbitrarily small, Ω is separable and $\mathfrak{D}^*$ is compact, there exists a decision function φ such that

$$r(F, \delta) = r(F, \varphi) \qquad \text{for all } F \text{ in } \Omega.$$

Added in proof. After sending this paper to the printer I read the interesting paper of A. Wald and J. Wolfowitz, *Two Methods of Randomization in Statistics and the Theory of Games*, Annals of Mathematics, May, 1951, which deals the equivalence between Wald's randomized decision function and our mixed one.

Institute of Statistical Mathematics

CORRECTIONS TO

" ON THE THEORY OF STATISTICAL DECISION FUNCTIONS "

KAMEO MATUSITA

The following correction should be made to the above-mentioned paper (this Annals, Vol. 3, No. 1 (1951), 17-35).

i) The proof of Corollary 4 should be replaced by the following:

For an arbitrary open set e in D, let $e=\cup e_i$, where e_i are the neighborhoods with their closures $\bar{e}_i$ contained in e. Then we have

$$\{x;\ \varphi(x)\in e\}=\bigcup_i \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{x;\ \varphi_n(x)\in \bar{e}_i\}\in \mathfrak{R}\,.$$

This proves the corollary.

ii) The passage from the line 3 to the line 6 on the page 20, "Then ...is a closed set" should be omitted. Though it is not wrong, it is liable to raise misunderstanding.

iii) The statement in Lemma 3, "$r(F, \varphi)$ is a continuous function on $\Omega\times\mathfrak{D}$ and satisfies", should be replaced by "When $F_i\to F$, $\varphi_i(x)\to\varphi(x)$ $(i\to\infty)$, then we have $r(F_i, \varphi_i)\to r(F, \varphi)$ $(i\to\infty)$ and". The proof remains available.

iv) The passages from the line 9 to the line 29 on the page 21, "Lemma 4...is a compact, metric and separable space", should be replaced by the following:

LEMMA 4′. *As a function of F in Ω $r(F, \varphi)$ is uniformly continuous with respect to φ.*

PROOF. As $W(F, d)$ is continuous on the compact space $\Omega\times D$, for any positive number ε there exists a positive number δ such that

$$||F_1-F_2||<\delta \rightarrow |W(F_1, d)-W(F_2, d)|<\varepsilon \qquad \text{for all } d.$$

Since $||F_1-F_2||<\delta$ implies

$$|F_1(E)-F_2(E)|\leqq 2||F_1-F_2||<2\delta$$

for any set E of $\mathfrak{R}$, we have

$$
\begin{aligned}
|r(F_1, \varphi)-r(F_2, \varphi)| &= \left| \int_R W(F_1, \varphi(x))\, dF_1 - \int_R W(F_2, \varphi(x))\, dF_2 \right| \\
&\leqq \left| \int_R \{W(F_1, \varphi(x)) - W(F_2, \varphi(x))\}\, dF_1 \right| \\
&\quad + \left| \int_R W(F_2, \varphi(x))\, dF_2 - \int_R W(F_2, \varphi(x))\, dF_1 \right| \\
&< \varepsilon + 2\delta .
\end{aligned}
$$

This proves the lemma.

Now we introduce a metric in $\mathfrak{D}$.

DEFINITION. Let φ_1, φ_2 be two elements of $\mathfrak{D}$. We define the distance $\rho_{\mathfrak{D}}$ between φ_1, φ_2 by

$$
\rho_{\mathfrak{D}}(\varphi_1, \varphi_2) = \max_{F \in \Omega} |r(F, \varphi_1) - r(F, \varphi_2)| .
$$

Clearly we have $\rho_{\mathfrak{D}}(\varphi_1, \varphi_2) \leqq 1$. Let $\mathfrak{D}^*$ denote the metric space obtained from $\mathfrak{D}$, in identifying the elements between which the $\rho_{\mathfrak{D}}$ distances are zero, and metrized by $\rho_{\mathfrak{D}}$.

LEMMA 4″. *$\mathfrak{D}^*$ is a conditionally compact space.*

PROOF. Let ε be an arbitrarily chosen positive number. Then, according to Lemma 4′, every F has a neighborhood, $U(F)$, such that

$$
F' \in U(F) \to |r(F, \varphi) - r(F', \varphi')| < \varepsilon \qquad \text{for all } \varphi .
$$

As Ω is compact, Ω can be covered by a finite number of such neighborhoods, i.e.,

$$
\Omega = U(F_1) \smile U(F_2) \smile \dots \smile U(F_s) .
$$

Now put

$$
\rho'(\varphi_1, \varphi_2) = \max_{1 \leqq i \leqq s} |r(F_i, \varphi_1) - r(F_i, \varphi_2)| .
$$

This $\rho'(\varphi_1, \varphi_2)$ gives also a metric in $\mathfrak{D}$ and satisfies $0 \leqq \rho'(\varphi_1, \varphi_2) \leqq 1$. We denote by $\mathfrak{D}'$ the metric space obtained from $\mathfrak{D}$ by this ρ'. Then, an element of $\mathfrak{D}'$, say φ, is represented by a system of numbers $\{r(F_1, \varphi), r(F_2, \varphi), \cdots, r(F_s, \varphi)\}$. Taking into account of the fact $0 \leqq r(F, \varphi) \leqq 1$, we see immediately that $\mathfrak{D}'$ is conditionally compact. Now it holds

$$
0 \leqq \rho_{\mathfrak{D}}(\varphi_1, \varphi_2) - \rho'(\varphi_1, \varphi_2) < 2\varepsilon
$$

from which follows the lemma.

From Lemma 3 follows immediately

LEMMA 5. *When* $\varphi_n(x) \to \varphi(x)$ $(n \to \infty)$, $r(F, \varphi_n) \to r(F, \varphi)$ *uniformly in* F.

Now, let $\gamma(\mathfrak{D}^*)$ be the metric space of all fundamental sequences in $\mathfrak{D}^*$. The distance between two fundamental sequences $\varphi_1^* = \{\varphi_{1n}\}$, $\varphi_2^* = \{\varphi_{2n}\}$ is, as usual, given by

$$\rho(\varphi_1^*, \varphi_2^*) = \lim_{n\to\infty} \rho(\varphi_{1n}, \varphi_{2n}) .$$

Then, $\gamma(\mathfrak{D}^*)$ is complete and totally bounded, consequently compact. $\mathfrak{D}^*$ is clearly embedded in $\gamma(\mathfrak{D}^*)$, the closure of $\mathfrak{D}^*$, $\mathfrak{D}^*$ is $\gamma(\mathfrak{D}^*)$ and $r(F, \varphi)$ defined on $\Omega \times \mathfrak{D}^*$ is continuously extended on $\Omega \times \gamma(\mathfrak{D}^*)$. Therefore, we can extend the notion of decision function to the whole space $\gamma(\mathfrak{D}^*)$. The real meaning of an element of $\gamma(\mathfrak{D}^*)$ not belonging to $\mathfrak{D}^*$, say φ^*, consists in the fact, that for any positive number ε there exists a decision function φ in $\mathfrak{D}^*$ such that

$$|r(F, \varphi^*) - r(F, \varphi)| < \varepsilon \qquad \text{for all } F .$$

When $\mathfrak{D}^*$ is compact from the beginning, as in the case where the sample space R is a countable set, we have, of course, $\mathfrak{D}^* = \gamma(\mathfrak{D}^*)$. In the following we denote $\gamma(\mathfrak{D}^*)$ merely by $\mathfrak{D}^*$. *$\mathfrak{D}^*$ is, thus, a compact metric space.*

v) The passage from the line 6 from the bottom to the line 5 from the bottom on the page 22, "and are...from Lemma 7 the", should be replaced by the following:

"We define the distances in M and Δ, respectively, as follows:

$$\rho(\mu_1, \mu_2) = \max_{\delta} |r(\mu_1, \delta) - r(\mu_2, \delta)|$$

$$\rho(\delta_1, \delta_2) = \max_{\mu} |r(\mu, \delta_1) - r(\mu, \delta_2)| .$$

Then we have for $n \to \infty$

$$\mu_n \to \mu \text{ (in the ordinary sense)} \to \rho(\mu_n, \mu) \to 0$$

$$\delta_n \to \delta \text{ (in the ordinary sense)} \to \rho(\delta_n, \delta) \to 0 .$$

Thus, M and Δ are compact metric spaces. We obtain immediately the following".

vi) In Theorem IV $r(\mu, \varphi)$ means $\int_\Omega r(E, \varphi)\, d\mu$.

vii) In the line 2 from the bottom on the page 30, the sign "$\geqq$" should be replaced by "$>$".

Correction to the Paper "On the Theory of Statistical Decision Functions"

This Annals Vol. III

By Kameo MATUSITA

(Received June 30, 1952)

The following correction should be made to the above-mentioned paper.

1 The proof of Corollary 4 should be replaced by the following:

For an arbitrary open set e in D, let $e = \bigcup e_i$, where e_i are the neighborhoods with their closures $\bar{e}_i$ contained in e. Then we have

$$\{x; \varphi(x) \in e\} = \bigcup_i \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{x; \varphi_n(x) \in \bar{e}_i\} \in \mathfrak{R}$$

This proves the corollary.

2 The passage from the line 3 to the line 6 on the page 20, "Then ......is a closed set" should be omitted. Though it is not wrong, it is liable to raise misunderstanding.

3 The statement in Lemma 3, "$r(F, \varphi)$ is a continuous function on $\Omega \times \mathfrak{D}$ and satisfies", should be replaced by "When $F_i \to F$, $\varphi_i(x) \to \varphi(x)$ $(i \to \infty)$, then we have $r(F_i, \varphi_i) \to r(F, \varphi)$ $(i \to \infty)$ and". The proof remains available.

4 The passages from the line 9 to the line 29 on the page 21, "Lemma 4......is a compact, metric and separable space", should be replaced by the following:

Lemma 4′ As a function of F in Ω $r(F, \varphi)$ is uniformly continuous with respect to φ.

Proof As $W(F, d)$ is continuous on the compact space $\Omega \times D$, for any positive number ε there exists a positive number δ such that

$$\| F_1 - F_2 \| < \delta \to | W(F_1, d) - W(F_2, d) | < \varepsilon \qquad \text{for all } d.$$

Since $\| F_1 - F_2 \| < \delta$ implies

$$| F_1(E) - F_2(E) | \leqq 2 \| F_1 - F_2 \| < 2\delta$$

for any set E of $\mathfrak{R}$, we have

$$| r(F_1, \varphi) - r(F_2, \varphi) | = \left| \int_R W(F_1, \varphi(x)) dF_1 - \int_R W(F_2, \varphi(x)) dF_2 \right|$$

$$\leq \left| \int_R \{ W(F_1, \varphi(x)) - W(F_2, \varphi(x)) \} dF_1 \right|$$
$$+ \left| \int_R W(F_2, \varphi(x)) dF_2 - \int_R W(F_2, \varphi(x)) dF_1 \right|$$
$$< \varepsilon + 2\delta$$

This proves the lemma.

Now we introduce a metric in $\mathfrak{D}$.

Definition Let φ_1, φ_2 be two elements of $\mathfrak{D}$. We define the distance $\rho_{\mathfrak{D}}$ between φ_1, φ_2 by

$$\rho_{\mathfrak{D}}(\varphi_1, \varphi_2) = \max_{F \in \Omega} | r(F, \varphi_1) - r(F, \varphi_2) |$$

Clearly we have $\rho_{\mathfrak{D}}(\varphi_1, \varphi_2) \leqq 1$. Let $\mathfrak{D}^*$ denote the metric space obtained from $\mathfrak{D}$, in identifying the elements between which the $\rho_{\mathfrak{D}}$ distances are zero, and metrized by $\rho_{\mathfrak{D}}$.

Lemma 4'' $\mathfrak{D}^*$ *is a conditionally compact space.*

Proof Let ε be an arbitrarily chosen positive number. Then, according to Lemma 4', every F has a neighborhood, $U(F)$, such that

$$F' \in U(F) \rightarrow | r(F, \varphi) - r(F', \varphi') | < \varepsilon \qquad \text{for all } \varphi.$$

As Ω is compact, Ω can be covered by a finite number of such neighborhoods, i.e.,

$$\Omega = U(F_1) \smile U(F_2) \smile \cdots \smile U(F_s)$$

Now put

$$\rho'(\varphi_1, \varphi_2) = \max_{1 \leqq i \leqq s} | r(F_i, \varphi_1) - r(F_i, \varphi_2) |$$

This $\rho'(\varphi_1, \varphi_2)$ gives also a metric in $\mathfrak{D}$ and satisfies $0 \leqq \rho'(\varphi_1, \varphi_2) \leqq 1$. We denote by $\mathfrak{D}'$ the metric space obtained from $\mathfrak{D}$ by this ρ'. Then, an element of $\mathfrak{D}'$, say φ, is represented by a system of numbers $\{ r(F_1, \varphi), r(F_2, \varphi), \ldots, r(F_s, \varphi) \}$. Taking into account of the fact $0 \leqq r(F, \varphi) \leqq 1$, we see immediately that $\mathfrak{D}'$ is conditionally compact. Now it holds

$$0 \leqq \rho_{\mathfrak{D}}(\varphi_1, \varphi_2) - \rho'(\varphi_1, \varphi_2) < 2\varepsilon$$

from which follows the lemma.

From Lemma 3 follows immediately

Lemma 5 *When* $\varphi_n(x) \rightarrow \varphi(x)$ $(n \rightarrow \infty)$, $r(F, \varphi_n) \rightarrow r(F, \varphi)$ *uniformly in* F.

Now, let $\gamma(\mathfrak{D}^*)$ be the metric space of all fundamental sequences in $\mathfrak{D}^*$. The distance between two fundamental sequences $\varphi_1^* = \{\varphi_{1n}\}$, $\varphi_2^* = \{\varphi_{2n}\}$ is, as usual, given by

$$\rho(\varphi_1^*, \varphi_2^*) = \lim_{n\to\infty} \rho(\varphi_{1n}, \varphi_{2n})$$

Then, $\gamma(\mathfrak{D}^*)$ is complete and totally bounded, consequently compact. $\mathfrak{D}^*$ is clearly embedded in $\gamma(\mathfrak{D}^*)$, the closure of $\mathfrak{D}^*$, $\mathfrak{D}^*$ is $\gamma(\mathfrak{D}^*)$ and $r(F, \varphi)$ defined on $\Omega \times \mathfrak{D}^*$ is continuously extended on $\Omega \times \gamma(\mathfrak{D}^*)$. Therefore, we can extend the notion of decision function to the whole space $\gamma(\mathfrak{D}^*)$. The real meaning of an element of $\gamma(\mathfrak{D}^*)$ not belonging to $\mathfrak{D}^*$, say φ^*, consists in the fact, that for any positive number ε there exists a decision function φ in $\mathfrak{D}^*$ such that

$$|r(F, \varphi^*) - r(F, \varphi)| < \varepsilon \qquad \text{for all } F$$

When $\mathfrak{D}^*$ is compact from the beginning, as in the case where the sample space R is a countable set, we have, of course, $\mathfrak{D}^* = \gamma(\mathfrak{D}^*)$. In the following we denote $\gamma(\mathfrak{D}^*)$ merely by $\mathfrak{D}^*$. *$\mathfrak{D}^*$ is, thus, a compact metric space.*

5 The passage from the line 6 from the bottom to the line 5 from the bottom on the page 22, "and are......from Lemma 7 the", should be replaced by the following:

"We define the distances in M and Δ, respectively, as follows:

$$\rho(\mu_1, \mu_2) = \max_{\delta} |r(\mu_1, \delta) - r(\mu_2, \delta)|$$

$$\rho(\delta_1, \delta_2) = \max_{\mu} |r(\mu, \delta_1) - r(\mu, \delta_2)|$$

Then we have for $n \to \infty$

$$\mu_n \to \mu \text{ (in the ordinary sense)} \longrightarrow \rho(\mu_n, \mu) \to 0$$

$$\delta_n \to \delta \text{ (in the ordinary sense)} \longrightarrow \rho(\delta_n, \delta) \to 0$$

Thus, M and Δ are compact metric spaces. We obtain immediately the following"

6 In Theorem IV $r(\mu, \varphi)$ means $\int_\Omega r(E, \varphi) d\mu$.

7 In the line 2 from the bottom on the page 30 the sign "$\geqq$" should be replaced by "$>$"

Institute of Statistical Mathematics