DISCRETE DECISION PROBLEMS

By YUKIO SUZUKI

(Received Sept. 15, 1957)

1. Introduction.

In this paper we treat decision problems where probability distributions are all discrete ones which give positive probabilities over only finite points.

Similar situations have been also treated by K. Matusita and H. Akaike [1] by using of Matusita's affinity concept. They have derived a few theorems with respect to inequalities which give the bound of probability that the distance of true distribution and emprical one is smaller than a given positive number. By these inequalities, they have found the decision errors.

Our main purpose of this paper is to treat multiple decision problem from the viewpoint of minimum risk and its application to testing hypotheses. Further, the comparison between our procedure and procedure of testing hypotheses is given.

Especially it seems to be interesting that the optimal decision procedure obtained from our viewpoint is non-randomized ones.
A. Dvoretzky, A. Wald and J. Wolfowitz [2], [3] have proved that there exists optimal non-randomized decision function under the assumption that all probability measures are atomless.
Therefore, our result insures that their result is also relevant to discrete decision problems in spite of the complete breakness of their assumption.

Furthermore, it is to be noted that our results are analogous to the Neymann-Pearson's fundamental lemma and its three generalized lemmas obtained by A. Wald [4], H. Scheffé [5] and the author [10], respectively.

We have derived our results by using the concept of linear programming.

2. Multiple decision problem.

In this paper, we are concerned only with terminal decisions ; i. e. we assume that the decisions on experimentation have been made and the only decisions that remain to be made are terminal ones. Furthermore, we assume that we treat a finite class of discrete probability measures over a finite point set. Hence it follows that we are concerned only with a finite set of decisions.

Let $P_1, P_2, \cdots, P_m$ be a class of probability measures over a finite point set S, where $S=\{1, 2, \cdots, N\}$. We denote this class by $\mathfrak{P}$.

Our broblem is to find an optimal decision procedure to decide which probability measure of $\mathfrak{P}$ is the true probability measure of a random variable X on the basis of observations on X.

In general, a decision problem is said statistical when the choice of decisions is depend on unknown probability measures.

Now, let the class of decisions be $D=\{d_1, d_2, \cdots d_m\}$. Before we proceed to statistical one, we can solve non-statistical decision problem, i. e. to any $P_j(\in \mathfrak{P})$ a unique $d_j(\in D)$ can be corresponded which is optimum in the sense that, if we knew the probability measure of X to be P_j, then we would make the decision d_j and the remaining members of D will be, for P_j, incorrect decisions.

If n is the size of sample by obsevation on X, the number of the possible sample points is N^n. Now, denote these points by O_j $(j=1, 2, \cdots, N^n)$, respectively, where $O_j=(j_1, j_2, \cdots, j_n)$, $\begin{pmatrix} j_i \in S \\ i=1, 2, \cdots, n \end{pmatrix}$. Therefore, the sample space S_n is $\{O_j;\ j=1, 2, \cdots, N^n\}$.

According as Wald's definition, a non-randomized decision function δ is defined as follows, " δ is a univalued function defined over S_n whose values are in D".

Before we proceed to the definition of randomized decision functions, we must define randomized decisions. A randomized decision is determined uniquely by a probability measure defined over D according to which the choice of decision is made.

Since, in our case, D is a finite set, there is one-to-one correspondence between the class of all randomized decisions and the class, $\mathfrak{P}_D$, of all discrete probability measures over D.

Then, a randomized decision function δ can be defined as follows; " δ is a univalued function defined over S_n whose values are in $\mathfrak{P}_D$".

Let us now denote by $\delta(d_i | O_j)=x_{ji}$, $\begin{pmatrix} i=1, 2, \cdots, m \\ j=1, 2, \cdots, N^n \end{pmatrix}$, the probability that, when the sample point O_j has been obtained by observation on

X, the decision d_i is made.

Then, obviously, we get the following relations corresponding to a decision function δ

$$(2.1)\qquad \sum_{i=1}^{m} \delta(d_i \mid O_j) = \sum_{i=1}^{m} x_{ji} = 1 , \qquad x_{ji} \geqq 0 , \qquad \begin{pmatrix} i=1, 2, \cdots, m \\ j=1, 2, \cdots, N^n \end{pmatrix}.$$

These equations mean that, when any sample point has been obtained, it is not permitted to make no decision.

Let p_{ij} be the probability that sample point O_j is obtained when probability measure P_i is true;

$$(2.2)\qquad p_{ij} = \prod_{k=1}^{n} P_i(j_k) ,$$

where $O_j = (j_1, j_2, \cdots, j_n)$, $j_i \in S$, and $P_i(j_k)$ is the probability of occurrence of j_k under the probability measure P_i.

For our purpose the sample space can be reduced, without loss of generality, to the set of points

$$\{O_j \mid \exists_i : \ p_{ij} > 0 \quad (i=1, 2, \cdots m)\} .$$

By changing the sufices of O, we can describe the reduced sample space by $\{O_j ;\ (j=1, 2, \cdots, J)\}$. Hereafter, we will treat only the reduced sample space.

For any decision function δ, we define

$$(2.3)\qquad \begin{aligned} &\delta(d_k \mid O_j ;\ P_i) = p_{ij}\delta(d_k \mid O_j) = p_{ij}x_{jk} , \\ &\delta(d_k \mid P_i) = \sum_j \delta(d_k \mid O_j ;\ P_i) = \sum_{j=1}^{J} p_{ij}x_{jk} , \\ &(i, k=1, 2, \cdots, m) . \end{aligned}$$

Then, $\delta(d_k/P_i)$ represents the probability with which the decision d_k is made according as decision function δ when P_i is the true probability measure. As mentioned previously, in non-statistical decision problem, we can assume unique correct decision is d_j if it is known P_j to be true. Therefore, if $i \neq k$, $\delta(d_k \mid P_i)$ is regarded as the error committed by making d_k (incorrect decision) while P_i is true. On the contrary, if $i=k$, $\delta(d_i \mid P_i)$ is the probability of making the correct decision when P_i is true.

Thus, for any decision function δ, there corresponds a square matrix $M^{(\delta)}$, where $M^{(\delta)}$ is

$$M^{(\delta)}=(\delta(d_k\,|\,P_i)) \tag{2.4}$$
$$=\left(\sum_{j=1}^{J} p_{ij}x_{jk}\right) \qquad (k,\ i=1,\ 2,\ \cdots,\ m).$$

Let us denote by $\mathfrak{M}$ the class of $M^{(\delta)}$ corresponding to all decision functions.

If we regard each $M^{(\delta)}$ as the point of m^2-dimensional Euclidean space, the following lemma holds.

LEMMA 1. *$\mathfrak{M}$ is bounded, convex and closed.*

PROOF. Any element M of $\mathfrak{M}$ is written as

$$\left(\sum_{j=1}^{J} p_{ij}x_{jk}\right) \qquad (i,\ k=1,\ 2,\ \cdots,\ m)\ .$$

From (2.1), the boundedness of $\mathfrak{M}$ is easily obtained.

Let M_1, M_2 be different elements of $\mathfrak{M}$ and α, β be non-negative numbers such that $\alpha+\beta=1$.
Suppose

$$M_1=\left(\sum_j p_{ij}x_{jk}^{(1)}\right),$$
$$M_2=\left(\sum_j p_{ij}x_{jk}^{(2)}\right) \qquad (i,\ k=1,\ 2,\ \cdots,\ m).$$

Then

$$\alpha M_1+\beta M_2=(\sum p_{ij}(\alpha x_{jk}^{(1)}+\beta x_{jk}^{(2)}))\ .$$

Let

$$\alpha x_{jk}^{(1)}+\beta x_{jk}^{(2)}=y_{jk} \qquad \begin{pmatrix} j=1,\ 2,\ \cdots,\ J \\ k=1,\ 2,\ \cdots,\ m \end{pmatrix}.$$

Obviously $\{y_{jk}\}$ satisfies the condition (2.1), which means that $\{y_{jk}\}$ defines a decision function δ. Therefore, $\alpha M_1+\beta M_2=(\sum p_{ij}y_{jk})\in\mathfrak{M}$. Thus, the convexity of $\mathfrak{M}$ has been proved.

Since the class $\mathfrak{M}$ is obtained by applying the continuous transformation $\left(\sum_j p_{ij}x_{jk}\right)$ $(i,\ k=1,\ 2,\ \cdots,\ m)$ upon the closed domain defined by (2.1), it follows that $\mathfrak{M}$ is closed.

Henceforth, we will regard each M as the point of $m(m-1)$-dimensional Euclidean space, after elimination of diagonal elements, and use the same notation $\mathfrak{M}$ to represent the class of these M. Clearly, in this case, lemma 1 also holds.

It seems reasonable to judge the merit of any decision function δ entirely on the basis of $M^{(\delta)}$ associated with it.

Therefore, on the basis of $M^{(\delta)}$, the choice of decision function will be made. For this purpose, it is required to introduce a partial ordering into $\mathfrak{M}$, consequently into the space D of decision functions.

Let us define a partial ordering ($<$) as follows;

DEFINITION. *For any $M^{(\delta_1)}$, $M^{(\delta_2)}$ in $\mathfrak{M}$, $M^{(\delta_1)} < M^{(\delta_2)}$ if every off-diagonal element of $M^{(\delta_1)}$ is not smaller than the corresponping one of $M^{(\delta_2)}$ and at least one inequality holds.*

Of course other partial orderings may be possible, but our partial ordering seems to be natural, simple and general. By this partial ordering in $\mathfrak{M}$, "uniformly better" relation is induced into the space $\mathfrak{D}$ of decision functions.

DEFINITION. *The decision function δ_1 is uniformly better than δ_2 if the following inequality holds*

$$M^{(\delta_1)} < M^{(\delta_2)} .$$

Thus, in the same way as Wald, we can define the concepts of admissible decision functions, complete class and minimal complete class of decision functions.

DEFINITION. *A decision function δ^* is admissible if there exists no other decision function δ which is uniformly better than δ^*.*

DEFINITION. *A class C of decision functions is complete if for any δ not in C we can find such an element δ^* in C that is uniformly better than δ.*

DEFINITION. *A complete class C is a minimal complete class if no proper subclass of C is complete.*

Since $\mathfrak{M}$ is bounded, closed and convex, there exists supporting hyperplane such that

$$\text{(2.5)} \qquad \sum_{k \neq l}\sum a_{kl} \sum_j p_{kj} x_{jl} \geqq \sum_{k \neq l}\sum a_{kl} \sum_j p_{kj} x_{jl}^{(0)}$$

for every

$$M = \left(\sum_j p_{kj} x_{jl}\right) \in \mathfrak{M} ,$$

where

$$M^{(0)} = \left(\sum p_{kj} x_{jl}^{(0)}\right) \in \mathfrak{M} ,$$

$$a_{kl} > 0 \quad (k, l = 1, \cdots, m) . \atop k \neq l$$

Then, decision functions $\left\{x^0_{jl}\,;\,\begin{matrix} j=1, 2, \cdots, J \\ l=1, 2, \cdots, m\end{matrix}\right\}$ are clearly admissible.

As A. Wald proved, the complete class of admissible decision functions is minimal complete. Hereafter, our attention is centred on the minimal complete class C of decision functions. Therefore, it is our main problem to obtain the procedure to get the decision functions which are optimal in the sense that it belongs to C, for given $\{a_{kl}\}$ $\left(\begin{matrix} k, l=1, \cdots, m \\ k\neq l\end{matrix}\right)$, where a_{kl} may be interpreted as the relative weight put on the errors which are committed if decision d_l is made erraneously while the true probability measure is P_k. From (2.5), the optimal decision function can be determined by the solution of the following problem which is a special type of the so-called linear programming problem; minimize

$$\text{(A)} \qquad \sum_{k\neq l}\sum a_{kl} \sum_j p_{kj} x_{jl}$$

subject to the following conditions

$$\text{(B)} \qquad \sum_{l=1}^{m} x_{jl}=1\ ,$$

$$x_{jl}\geqq 0\ , \qquad (j=1, 2, \cdots, J).$$

However, (A) is rewritten:

$$\text{(C)} \qquad \sum_j \left\{\sum_{l=1}^{m}\left(\sum_{k\neq l} a_{kl}p_{kj}\right)x_{jl}\right\}\ .$$

Therefore, the first linear programming problem is solved by solving the following linear programming problems separately for each j ($j=1, \cdots, J$):

Minimize

$$\text{(D)} \qquad \sum_{l=1}^{m}\left(\sum_{k\neq l} a_{kl}p_{kj}\right)x_{jl}$$

subject to

$$\text{(E)} \qquad \sum_{l=1}^{m} x_{jl}=1\ ,$$

$$x_{jl}\geqq 0\ , \qquad (l=1, \cdots, m).$$

Obviously, a solution of the last problem is

$$\text{(F)} \qquad x_{jl}=\begin{cases}1, & \text{if } i=l_0(j) \\ 0 & \text{otherwise} \quad (j=1, \cdots, J)\end{cases}$$

where $l_0(j)$ is defined by

$$(G) \qquad \min_{1\leq l\leq m}\Big(\sum_{k\neq l} a_{kl}p_{kj}\Big)=\sum_{k\neq l_0(j)} a_{kl_0(j)}p_{kj}$$

$$(j=1, 2, \cdots, J)\ .$$

The optimal decision function δ_0 determined by (F) is a non-randomized one. Of course, when $l_0(j)$ is not univalued, there exist infinitely many randomized decision functions. However, these are all equivalent to the decision function δ_0.

Thus we obtain the following theorem.

THEOREM 1. *There exists always the non-randomized decision function which is equivalent to each optimal decision function.*

REMARK. Under the assumption of non-atomicity of the probability measures proposed, other authors have proved the possibility of elimination of randomization in optimal decision function. However, though the assumption of non-atomicity is completely broken in our case, the existence of non-randomized decision function equivalent to each optimal one has been proved by the above theorem.

3. Testing hypothesis I.—simple case

In this section we treat the classical type of testing hypothesis, where probability measures considered are all discrete ones defined over a finite point set $S=\{1, 2, \cdots, N\}$. The same notations as in § 2 are used. Let P_1 and P_2 be such probability measures. By a hypothesis H_1 we mean a statement that the unkown probability measure of a random variable X is P_1 and a by a alternative H_2 we mean a statement the unkown probability measure of X is P_2. We want to test the hypothesis H_1 against the alternative H_2 on the basis of a random sample of size n. In this case, the decision space is $D=\{d_1, d_2\}$, where d_i is the decision that the unknown prob. measure is $H_i(i=1, 2)$. Let the decision function δ be defined over the reduced sample space S_n, which takes the value in D.

Now, denote as in § 2.

$$\delta(d_i|O_j)=x_{ji}=x_{ji} \quad \begin{pmatrix} i=1, 2, \cdots, m \\ j=1, 2, \cdots, J \end{pmatrix},$$

where $O_j \in S_n$.

Then we obtain

$$(3.1) \qquad x_{j1}+x_{j2}=1 \quad (j=1, 2, \cdots, J),$$
$$x_{j1}, x_{j2} \geqq 0 .$$

If the significance level is α ($0<\alpha<1$), the following equation concerning with the first kind of error holds;

$$(3.2) \qquad \delta(d_2 | P_1) = \sum_{j=1}^{N^n} p_{ir} x_{j2} = \alpha .$$

On the other hand, the second kind of error is

$$(3.3) \qquad \partial(d_1 | P_2) = \sum_{j=1}^{N^n} p_{2j} x_{j1} .$$

Then, the most powerful test is obtained by minimizing (3.3) subject to the conditions (3.1) and (3.2).

This minimizing problem is nothing but a linear programming problem. As well-known, the domain specified by the condition is a convex polyhedron in $2J$-dimensional Euclidean space and the form (3.3) is minimized at least at a extreme point of it.

Simultaneous linear equations in (3.1) and (3.2) is in vector notation

$$(3.4) \qquad \begin{pmatrix} 1 & 1 & 0 & 0 \cdots\cdots & 0 \\ 0 & 0 & 1 & 1 & \\ \vdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 0 & 1 \quad 1 \\ \hline 0 & p_{11} & 0 & p_{12} \cdots 0 & p_{1J} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \\ \vdots \\ x_{J,1} \\ x_{J,2} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \alpha \end{pmatrix} .$$

Now, let us use the following notations;

$$(35) \qquad \left.\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right\} s = E_s , \qquad \left.\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ p_{1s} \end{pmatrix}\right\} s = E'_s , \qquad \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \alpha \end{pmatrix} = E_0 ,$$
$$(s=1, 2, \cdots, J) .$$

As well-known in the theory of linear programming, there is one to one correspondence between extreme points of the polyhedron and non-degenerate feasible bases. Here the non-degenerate feasible basis means the class of $J+1$ independent vectors, say $E_{s_i}, E'_{s_j}\begin{pmatrix} i=1, 2, \cdots, k \\ j=k+1, \cdots, J+1 \end{pmatrix}$, such that for positive $\lambda_{s_i}, \lambda_{s_j}$,

$$\sum_{i=1}^{k} \lambda_{s_i} E_{s_i} + \sum_{k+1}^{J+1} \lambda_{s_j} E'_{s_j} = E_0 . \tag{3.6}$$

As easily be seen, from (3.4), in every non-degenerate feasible basis there must appear only one pair of vectors, $E_{s_{i_0}}, E'_{s_{j_0}}$, such that $s_{i_0}=s_{j_0}$.

Above all, a minimizing solution of (3.3) subject to (3.4) can be represented as the solution of the following equations; if the optimal basis is non-degenerate,

$$\sum_{i=1}^{k} E_{t_i} x_{t_i1} + \sum_{i=k+1}^{J-1} E'_{t_i} x_{t_i2} + E_{t_{j_0}} x_{t_{j_0}1} + E'_{t_{j0}} x_{t_{j_0}2} = E_0 , \tag{3.7}$$

where

$$\{E_{t_1}, E_{t_2}, \cdots, E_{t_k}, E'_{t_{k+1}}, \cdots E'_{t_{J-1}}, E_{t_{j_0}}, E'_{t_{j_0}}\}$$

is a feasible basis for optimal solution.

Therefore, the minimizing solution is

$$\begin{aligned} x_{t_i1}&=1, \quad & i&=1, \cdots, k, \\ x_{t_i2}&=1, \quad & i&=k+1, \cdots, J-1, \\ x_{t_{j_0}1}+x_{t_{j_0}2}&=1, \quad & 0&<x_{t_{j_0}1}, x_{t_{j_0}2}<1. \end{aligned}$$

This means that there exists an optimal decision function by which randomized decision is made only on one sample point and non-randomized decisions are made on all of the remaining sample points.

In degenerate case, the minimizing solution can be represented as the solution of (3.7) or the following equations

$$\sum_{i=1}^{k} E_{t_i} x_{t_i1} + \sum_{i=k+1}^{J} E'_{t_i} x_{t_i2} = E_0 .$$

In the this case,

$$x_{t_i1}=1 \ (i=1, 2, \cdots, k), \ x_{t_i2}=1 \ (i=k+1, \cdots, J) .$$

Therefore, our optimal decision function is a non-randomized one. Thus we obtain the following theorem.

THEOREM 2. *In testing the hypothesis $H_1(P_1)$ against the alternative*

$H_2(P_2)$, *for a given significance level* α, *there exists an optimal decision function* (*i.e. most powerful test*) *by which randomized decision is necessary at most for one sample point.*

Now, we treat the similar problem by the procedure described in § 2. Then our error matrix is

$$\begin{pmatrix}\delta(d_1 \mid P_1), & \delta(d_2 \mid P_1)\\ \delta(d_1 \mid P_2), & \delta(d_2 \mid P_2)\end{pmatrix}=\begin{pmatrix}\sum_j p_{1j}x_{j1}, & \sum_j P_{1j}x_{j2}\\ \sum_j p_{2j}x_{j1}, & \sum_j p_{2j}x_{j2}\end{pmatrix}.$$

Therefore our problem is to solve the following minimizing problem for given coefficients a_{12}, a_{21} ;

$$\text{(3.8)} \qquad \min. \quad a_{12}\sum_{j=1}^{N^n} p_{1j}x_{j2}+a_{21}\sum_{j=1}^{N^n} p_{2j}x_{j1}$$

subject to

$$\text{(3.9)} \qquad \begin{cases} x_{j1}+x_{j2}=1 & (j=1, 2, \cdots, J)\ , \\ x_{ji}\geqq 0 & (i=1, 2)\ . \end{cases}$$

The minimizing solution is obtained by the following principle:

$$\begin{aligned} x_{j2}&=1 \quad \text{if} \quad a_{21}p_{2j}\geqq a_{12}p_{1j}\ , \\ x_{j1}&=1 \quad \text{if} \quad a_{21}p_{2j}< a_{12}p_{1j}\ . \end{aligned}$$

This is the same as given by the Neyman-Pearson's Fundamental Lemma, except its consideration of the significance level.

4. Testing hypothesis II. (Simple hypothesis and composite alternative).

Let $P_0, P_1, P_2, \cdots, P_m$ be discrete probability measures defined over $S=\{1, 2, \cdots, N\}$. We treat the problem of testing the hypothesis H_1 against the alternative H_2 on the basis of a sample of size n, where H_1 is the statement that the unknown prob. measure is P_0 and H_2 is the one that the unknown prob. measure is one of $P_1, P_2, \cdots, P_m$.

Let O_j be sample points $(j=1, 2, \cdots, J^n)$, and define $p_{ij}=P_i(O_j)$ $\begin{pmatrix} i=0, 1, 2, \cdots, m \\ j=1, 2, \cdots, J \end{pmatrix}$.

Let the decision space be $D=\{d_0, d_1, \cdots, d_m\}$, where d_i is the decision that the unknown prob. measure is $P_i(i=0, 1, 2, \cdots, m)$, and $D^*=\{d_0^*, d_1^*\}$, where $d_0^*=d_0$, $d_1^*=\{d_1, \cdots, d_m\}$. As in § 2, consider the decision function δ^* for this decision problem and let

$$\delta^*(d_i^*|O_j)=x_{ji} \qquad \begin{pmatrix} i=0, 1, \\ j=1, 2, \cdots, J \end{pmatrix}.$$

Then

$$\sum_{i=0}^{1} x_{ji}=1\,, \qquad (j=1, 2, \cdots, J). \tag{4.1}$$

The first kind of error is

$$\delta^*(d_1^*|P_0)=\sum_{j=1}^{J} p_{0j}x_{j1}=\alpha\ . \tag{4.2}$$

In this case, there are m errors of judgement;

$$\sum_{j=1}^{J} p_{ij}x_{j0}\,, \tag{4.3}$$

if d_0 is taken when the true prob. measure is $P_i(i=1, 2, \cdots, m)$.
If we define the optimal test by the principle that a test is optimal if it minimizes the maximum error of (4.3), it is obtained by solving the following linear programming problem: minimize

$$\beta \tag{4.4}$$

subject to

$$\begin{cases} x_{j0}+x_{j1}=1\,, & (j=1, 2, \cdots, J) \\ \sum_{j=1}^{J} p_{0j}x_{j1}=\alpha\,, & \\ \sum_{j=1}^{J} p_{ij}x_{j0}\leqq\beta\,, & (i=1, 2, \cdots, m)\,. \end{cases} \tag{4.5}$$

By considering new non-negative variables $y_i(i=1, 2, \cdots, m)$, i. e. slack variables, (4.5) is rewritten as follows:

$$\begin{cases} x_{j0}+x_{j1}=1 \quad (i=1, 2, \cdots, J)\,, \\ \sum_{j} p_{0j}x_{j1}=\alpha\,, \\ \sum_{j} p_{ij}x_{j0}-\beta+y_i=0 \quad (i=1, 2, \cdots, m)\,. \end{cases} \tag{4.6}$$

The coefficient matrix of these equations is

$$
(4.7)\qquad
\underbrace{\begin{array}{l}
\left.\begin{array}{r} j=1 \\ 2 \\ \vdots \\ J \end{array}\right. \\
J+1+m \quad 1 \\
\left.\begin{array}{r} i=1 \\ 2 \\ \vdots \\ m \end{array}\right.
\end{array}}_{}
\overbrace{\left(\begin{array}{cccccccc|c|ccc}
1 & 1 & 0 & \cdots & & & & & 0 & & & \\
0 & 0 & 1 & 1 & & & & & \vdots & & 0 & \\
\vdots & \vdots & \vdots & \vdots & \ddots & & & & \vdots & & & \\
0 & 0 & \vdots & \vdots & & & 1 & 1 & 0 & & & \\
\hline
0 & p_{01} & 0 & p_{02} & & & 0 & p_{0J} & & & & \\
\hline
p_{11} & 0 & p_{12} & 0 & & & p_{1J} & 0 & -1 & 1 & & 0 \\
\vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots & & \ddots & \\
p_{m1} & 0 & p_{m2} & 0 & & & p_{mJ} & 0 & -1 & 0 & & 1
\end{array}\right)}^{2J+1+m}
\begin{pmatrix} x_{10} \\ x_{11} \\ x_{20} \\ x_{21} \\ \vdots \\ x_{J0} \\ x_{J0} \\ \beta \\ y_1 \\ \vdots \\ y_m \end{pmatrix}
=
\begin{pmatrix} 1 \\ \vdots \\ 1 \\ \hline \alpha \\ \hline 0 \\ \vdots \\ 0 \end{pmatrix}
$$

As well-known in the theory of linear programming, mimimum value of β is attained at an extreme point of the polyhedron defined by (4.6). i.e. (4.7). The basis corresponding to this extreme point has the following properties :

Let

$$
(4.8)\qquad
\begin{array}{r} J+1 \\ \\ m \end{array}
\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ \hline 0 \\ \hline p_{1j} \\ \vdots \\ p_{mj} \end{pmatrix}
\begin{array}{l} \}\, j \\ \\ \\ \end{array}
= V_j\,, \quad
\begin{array}{r} N^n \\ 1 \\ m \end{array}
\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ \hline p_{0j} \\ 0 \\ \vdots \\ 0 \end{pmatrix}
\begin{array}{l} \}\, j \\ \\ \\ \end{array}
= U_j\,, \quad
\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \hline -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}
= B\,,
$$

$$
\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \hline 0 \\ \hline 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}
\begin{array}{l} \}\, J \\ \}\, 1 \\ \}\, i \\ \\ \end{array}
= E_i\,, \qquad
\begin{pmatrix} 1 \\ \vdots \\ 1 \\ \hline \alpha \\ \hline 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}
\begin{array}{l} \}\, J \\ \\ \\ \end{array}
= W\,.
$$

In case of $J>m$,

0. at least one of V_j, U_j is in it.
1. it contains B.

   This fact easily follows from (4.7).
2. if $J+k$ of V or U vectors are chosen, $m-k$ E_i-vectors must be chosen, where $k \leqq m$. Thus the optimal basis contains just k pairs of $\{U_j, V_j\}$. This means the optimal decision function requires randomized decisions on just k sample points, where $k \leqq m$.

In case of $J \leqq m$, of course the number of sample points which need the randomized decision for the optimal decision function cannot be over J. Hence we obtain the following theorem.

THEOREM 3. *The optimal test for testing $H_0(P_0)$ against $H_1(P_1, \cdots, P_m)$, for a given significance level $\alpha(0<\alpha<1)$, requires randomized decisions at most for* $\min(J, m)$ *sample points.*

This theorem is clearly a generalization of theorem 1.

The similar problem can be treated from the view-point of § 2. Here, the decision space D^* consists of two decisions d_0^*, d_1^* as mentioned above.

The error matrix is

$$\left(\begin{array}{c|c} \delta(d_0^*|P_0) & \delta(d_1^*|P_0) \\ \hline \delta(d_0^*|P_1) & d(d_1^*|P_1) \\ \vdots & \vdots \\ \delta(d_0^*|P_m) & \delta(d_1^*|P_m) \end{array}\right) = \left(\begin{array}{c|c} \sum_j p_{0j}x_{j0} & \sum_j p_{0j}x_{j1} \\ \hline \sum_j p_{1j}x_{j0} & \sum_j p_{1j}x_{j1} \\ \vdots & \vdots \\ \sum_j p_{mj}x_{j0} & \sum_j p_{mj}x_{j1} \end{array}\right). \tag{4.9}$$

Therefore, the optimal decision function, for a given weight of errors $\{a_{01}, a_{10}, a_{20}, \cdot\cdot, a_{m,0}\}, a_{ij} \geqq 0$, is found by solving the problem below:

$$\min. \quad a_{01} \sum_j p_{0j}x_{j1} + \sum_i a_{i0} \sum_j p_{ij}x_{j0} \tag{4.10}$$

subject to

$$\begin{aligned} x_{j0}+x_{j1}&=1 \quad (j=1, 2, \cdots), \\ x_{ji}&\geqq 0 \quad (i=0, 1). \end{aligned} \tag{4.11}$$

By the same procedure in § 2, we obtain the following solution:

$$\begin{aligned} &x_{j0}=0,\ x_{j1}=1, \quad \text{if} \quad a_{01}p_{0j} < \sum_{i=1}^{m} a_{i0}p_{ij}, \\ &x_{j0}=1,\ x_{j1}=0, \quad \text{otherwise}. \end{aligned} \tag{4.12}$$

Thus we have obtained an optimal and non-randomized decision function.

5. Testing hypothesis III. (Composite hypothesis and simple alternative).

Suppose $P_0, P_1, P_2, \cdots, P_m$ be defined as in § 4. On the basis of a random sample of size n, we treat the problem of testing the hypothesis H_0 against the alternative H_1, where H_0 is the statement that the unknown prob. measure is one of $P_1, \cdots, P_m$ and H_1 is the one that the unknown prob. measure is P_0.

Let O_j be sample points, and define

$$p_{ij}=P_i(O_j)\ , \qquad \begin{pmatrix} i=0,1,2,\cdots,m \\ j=1,2,\cdots,J \end{pmatrix}.$$

In this case, the decision space is $D=\{d_0^*, d_1^*\}$, where $d_i^*(i=0, 1)$ is the decision that $H_i(i=0, 1)$ is true as in § 2, consider the decision function δ for this decision problem and let

$$\delta(d_i^*|O_j)=x_{ji}\ , \qquad \begin{pmatrix} i=0,1 \\ j=1,2,\cdots J \end{pmatrix}.$$

Then

$$x_{j0}+x_{j1}=1\ , \qquad (j=1, 2, \cdots, J). \tag{5.1}$$

The error matrix corresponding δ is

$$\left(\begin{array}{c|c} \delta(d_0^*|P_1) & \delta(d_1^*|P_1) \\ \delta(d_0^*|P_2) & \delta(d_1^*|P_2) \\ \vdots & \vdots \\ \delta(d_0^*|P_m) & \delta(d_1^*|P_m) \\ \hline \delta(d_0|P_0) & \delta(d_1|P_0) \end{array}\right) = \left(\begin{array}{c|c} \sum_j p_{1j}x_{j0} & \sum_j p_{1j}x_{j1} \\ \sum_j p_{2j}x_{j0} & \sum_j p_{2j}x_{j1} \\ \vdots & \vdots \\ \sum_0 p_{mj}x_{j0} & \sum_j x_{mj}x_{j1} \\ \hline \sum_j p_{0j}x_{j0} & \sum_j x_{0j}x_{j1} \end{array}\right).$$

For given levels of significance, $\alpha_1, \cdots, \alpha_m$, most powerful test can be obtained by solving the minimizing problem below.

$$\min.\ \sum_j p_{0j}x_{j0} \tag{5.2}$$

subject to

$$x_{j0}+x_{j1}=1\ , \qquad (j=1, 2, \cdots J), \tag{5.3}$$
$$x_{ji}\geqq 0\ , \qquad (i=0, 1),$$

$$\sum p_{kj}x_{j1}+\lambda_k\leqq\alpha_k\ , \qquad (k=1, 2, \cdots, m), \tag{5.4}$$

where $\lambda_k(\geqq 0)(k=1, 2, \cdots, m)$ are slack variables. By matrix notation restrictions (5.3) and (5.4) can be represented as follows;

$$\left(\begin{array}{cccccc|ccc} 1 & 1 & 0 & 0 & 0 & 0 & & & \\ 0 & 0 & 1 & 1 & \vdots & \vdots & & & \\ \vdots & \vdots & 0 & 0 & \vdots & \vdots & & 0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots\; 0 & 0 & & & \\ 0 & 0 & 0 & 0 & 1 & 1 & & & \\ \hline 0 & p_{11} & 0 & p_{12} & 0 & p_{1J} & 1 & & \\ \vdots & p_{21} & \vdots & \vdots & \vdots & \vdots & & 1 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & & \ddots \\ 0 & p_{m2} & 0 & p_{m2} & 0 & p_{mJ} & & & 1 \end{array}\right) \begin{pmatrix} x_{10} \\ x_{11} \\ x_{20} \\ x_{21} \\ \vdots \\ x_{J0} \\ x_{J1} \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} .$$

By the analogous reasoning obout the optimal basis of linear programing problem in § 4, we obtain the following theorem.

THEOREM 4. (Composite hypothesis-simple alternative) *For given significance levels, there exists the optimal decision function such that the number of the sample points, for which randomized decision is needed, is at most* min (J, m).

This theorem is also a generalization of the theorem in § 3.

By our multiple decision procedure, the above decision problem can be reduced to the following linear programing problem for a given weight $\{a_{00}, a_{11}, \cdots, a_{m1}\}$:

$$\min\left\{\sum_{h=1}^{m} a_{h1} \sum_{j=1}^{J} p_{hj} x_{j1} + a_{00} \sum_{j=1}^{J} p_{0j} x_{j0}\right\} \tag{5.5}$$

subject to

$$x_{j1}+x_{j0}=1\ , \quad x_{ji}\geqq 0\ , \qquad \begin{pmatrix} i=0,1 \\ j=1,2,\cdots,J \end{pmatrix}. \tag{5.6}$$

This problem can be reduced to J simple problems :

$$\min\left(\sum_{h=1}^{m} a_{h1} p_{hj}\right) x_{j1} + a_{00} p_{0j} x_{j0} \tag{5.7}$$

subject to

$$x_{j1}+x_{j0}=1\ , \quad x_{ji}\geqq 0\ , \qquad (j=1, 2, \cdots, J). \tag{5.8}$$

Therefore, we obtain the optimal, non-randomized decision function by the following procedure :

$$\begin{aligned} &\sum_{h=1}^{m} a_{h1} p_{hj} \geqq a_{00} p_{0j} \rightarrow x_{j0}=1\ , \\ &\sum_{h=1}^{m} a_{h1} p_{hj} < a_{00} p_{0j} \rightarrow x_{j1}=1\ , \end{aligned} \tag{5.9}$$

6. Testing hypotheses IV. (Composite hypothesis and Composite alternative).

Suppose $P_1, P_2, \cdots, P_m, Q_1, Q_2, \cdots, Q_{m'}$ be discrete prob. measures definded as preceding sections. We treat the problem of testing hypothesis H_0 against the alternative H_1, where H_0 is the statement that the unknown prob. measure is one of $P_1, \cdots, P_m$ and H_1 is the one that the unknown prob. measure is one of $Q_1, Q_2, \cdots, Q_{m'}$.

Let O_j be the sample point of the reduced sample space, and define

$$p_{jk}=P_k(O_j)\ , \qquad (k=1, 2, \cdots, m),$$
$$q_{hj}=Q_h(O_j)\ , \qquad (h=1, 2, \cdots, m')$$
$$(j=1, 2, \cdots, J).$$

In this case, the decision space is $D=\{d_0^*, d_1^*\}$, where d_i^* is the decision that $H_i(i=0, 1)$ is true.

For any decision function δ we define

$$\delta(d_i^*|O_j)=x_{ji}\ , \qquad \begin{pmatrix} i=0, 1 \\ j=1, 2, \cdots, J \end{pmatrix}.$$

The loss matrix corresponding to δ is

$$\left(\begin{array}{c|c} \delta(d_0^*|P_1) & \delta(d_1^*|P_1) \\ \vdots & \vdots \\ \delta(d_0^*|P_m) & \delta(d_1^*|P_m) \\ \hline \delta(d_0^*|Q_1) & \delta(d_1^*|Q_1) \\ \vdots & \vdots \\ \delta(d_0^*|Q_{m'}) & \delta(d_1^*|Q_{m'}) \end{array}\right) = \left(\begin{array}{c|c} \sum_j p_{1j}x_{j0} & \sum_j p_{1j}x_{j1} \\ \vdots & \vdots \\ \sum_j p_{mj}x_{j0} & \sum_j p_{mj}x_{j1} \\ \hline \sum_j q_{1j}x_{j0} & \sum_j q_{1j}x_{j1} \\ \vdots & \vdots \\ \sum_j q_{m'j}x_{j0} & \sum_j q_{m'j}x_{j1} \end{array}\right).$$

For given levels of significance, $\alpha_1, \alpha_2, \cdots, \alpha_m$, decision function δ must satisfy the following equations:

$$\text{(6.1)} \qquad \begin{cases} \sum_{j=1}^{J} p_{1j}x_{j2} \leqq \alpha_1 \\ \quad \vdots \\ \sum_{j=1}^{J} p_{mj}x_{j2} \leqq \alpha_m\ . \end{cases}$$

The 2nd errors comitted by the decision function δ are

$$\text{(6.2)} \qquad \sum_{j=1}^{J} q_{hj}x_{j1}\ , \qquad (h=1, 2, \cdots, m').$$

As optimal decision function let us choose such a decision function that attains the minimum value of the maximum of (6.2).

From this viewpoint, we obtain following linear programing problem;

(6.3) $$\min \beta$$

subject to

$$x_{j0}+x_{j1}=1\ ,\quad x_{ji}\geqq 0\ ,\qquad (i=0,1),$$

(6.4)
$$\begin{cases} \sum_{j=1}^{J} p_{k1}x_{j1}\leqq \alpha_{1k}\ , & (k=1,2,\cdots,m),\\ \sum_{j=1}^{J} q_{hj}x_{j0}\leqq \beta_1\ , & (k=1,2,\cdots,m'). \end{cases}$$

The restriction (6.4) can be written by matrix notation as follows:

$$\begin{array}{r} J\left\{\begin{array}{c}\\ \\ \\ \\ \end{array}\right. \\ m\left\{\begin{array}{c}\\ \\ \\ \end{array}\right. \\ m'\left\{\begin{array}{c}\\ \\ \\ \\ \end{array}\right. \end{array}
\overbrace{}^{2J}\ \overbrace{}^{m}\ \overbrace{}^{1}\ \overbrace{}^{m'}
\left(\begin{array}{cccccc|ccc|c|ccc}
1 & 1 & 0\cdots & 0 & 0 & & & & & 0 & & & \\
0 & 0 & \ddots & & \vdots & \vdots & & 0 & & \vdots & & 0 & \\
\vdots & \vdots & & \ddots & 0 & 0 & & & & \vdots & & & \\
0 & 0\cdots & 0 & & 1 & 1 & & & & 0 & & & \\
\hline
0 & p_{11} & & & 0 & p_{1J} & 1 & & 0 & 0 & & & \\
\vdots & \vdots & & & \vdots & \vdots & & \ddots & & \vdots & & 0 & \\
0 & p_{m1} & & & 0 & p_m & & & 1 & 0 & & & \\
\hline
q_{11} & 0 & & & q_{1J} & 0 & & & & -1 & 1 & & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & 0 & & -1 & & \ddots & \\
\vdots & \vdots & & & \vdots & \vdots & & & & \vdots & & & \\
q_{m'1} & 0 & & & q_{m'J} & 0 & & & & -1 & 1 & & 1
\end{array}\right)
\begin{pmatrix} x_{10}\\ x_{11}\\ x_{20}\\ x_{22}\\ \vdots\\ x_{J0}\\ x_{J1}\\ \lambda_1\\ \vdots\\ \lambda_m\\ \beta\\ \mu_1\\ \vdots\\ \mu_{m'} \end{pmatrix}
=
\begin{pmatrix} 1\\ 1\\ \vdots\\ 1\\ \hline \alpha_1\\ \alpha_2\\ \vdots\\ \alpha_m\\ \hline 0\\ \vdots\\ 0 \end{pmatrix},$$

where λ, μ are slack variables. Again, by the similar reasoning as the previous sections, we can conclude that the number of sample points corresponding to randomized decisions needed by the optimal decision function is at most $\min(J, m+m'-1)$.

THEOREM 5. (Composite hypothesis-Composite alternative) *For given signifivance levels, there exists the optimal decision function such that the number of the sample points, for which randomized decision is needed, is at most* $\min(J, m+m'-1)$.

This result is clearly the generalized one of theorems of previous sections.

From the point of view in § 2, the above decision problem can be solved by solving the following linear programing problem for given weight $\{a_{11}, \cdots, a_{m1}, b_{10}, \cdots, b_{m'0}\}$:

$$\text{(6.5)} \qquad \min.\left\{\sum_{k=1}^{m} a_{k1} \sum_{j=1}^{J} p_{kj}x_{j1} + \sum_{h=1}^{m'} b_{h0} \sum_{j=1}^{J} q_{hj}x_{j0}\right\}$$

subject to

$$\text{(6.6)} \qquad x_{j0}+x_{j1}=1\ , \quad x_{ji}\geqq 0\ , \qquad \begin{pmatrix} i=0, 1 \\ j=1, 2, \cdots J \end{pmatrix}.$$

Again, this problem can be reduced to J simple problems:

$$\min\left(\sum_{k=1}^{m} a_{k1}p_{kj}\right)x_{j1}+\left(\sum_{h=1}^{m'} b_{h0}q_{hj}\right)x_{j0}$$

subject to

$$x_{j1}+x_{j0}=1, \quad x_{ji}\geqq 0\ , \qquad (i=0, 1).$$

Thus, we obtain the optimal, non-randomized decision function by the following procedure:

$$\sum_{k=1}^{m} a_{k1}p_{kj} \geqq \sum_{h=1}^{m'} b_{h0}q_{hj} \rightarrow x_{j0}=1\ ,$$

$$\sum_{k=1}^{m} a_{k1}p_{kj} < \sum_{h=1}^{m'} b_{h0}q_{hj} \rightarrow x_{j1}=1\ .$$

This procedure is analogous to the one obtained by H. Scheffé in connection with the generalization of Neyman-Pearson's fundamental lemma.

REFERENCES

[1] K. Matusita and H. Akaike, Decision rules, based on the distance, for the broblems of independence, invariance and two samples, *Ann. Stat. Math.* Vol. (1955).

[2] A. Dvoretzky, A. Wald, and J. Wolfowitz, Elimination of randomization in certain problems of statistics and of the theory of games, *Proc. Nat. Acad. Sci.* U.S.A. Vol. 36 (1950).

[3] ———, Relations among certain ranges of vector measures. *Pacific Journal of Math.* Vol. 1 (1951).

[4] G. B. Dantzig and A. Wald, on the fundamental lemma of Neyman and Pearson. *Ann. Math. Stat.* 1951.

[5] H. Scheffe and Herman Chernoff, A generalization of the Neyman-Pearson's fundamental lemma *Ann. Mach. Stat.* 1952.

[6] R. Halmos, The range of a vector measure, *Bull. Amer. Math. Soc.* Vol. 54 (1948).

[7] D. Blackwell, The range of certain vector integrals, *Proc. Amer. Math. Soc.* Vol. 2 1951.

[8] A. Wald, Statistical decision functions, Wiley, New York (1950).

[9] D. V. Lindley, Statistical inference, *Jour. Royal Stat. Soc.* Vol. 15 (1953).

[10] Y. Suzuki, Note on Neymen-Pearson's fundamental lemma, *Ann. Stat. Math.* Vol. 6 (1954).

[11] A. Charnes, W. W. Cooper and A. Henderson, An introduction to linear programming, Wiley, New York, (1953).

[12] T. C. Koopmans (ed.). Activity analysis of production and allocation, Wiley, New York, (1951).