AN APPLICATION OF THE DISCRIMINATION INFORMATION MEASURE TO THE THEORY OF TESTING HYPOTHESES PART II

BY SADAO IKEDA

(Received May 26, 1961)

The concept of discrimination information measure in the Kullback information theory was applied, in the preceding paper, Part I, to the theory of testing hypotheses in the case of parametric inference.

In the present paper, Part II, further application will be made to the problem of testing many-sided hypotheses (Section 5). Section 6 and the subsequent two sections will be devoted to applications of our method, given in Section 3 of Part I, to the problems of testing hypotheses in the non-parametric case. In Section 5, the general procedure will be reformulated for the purpose of application to the problems of testing non-parametric hypotheses. A unified method of deriving the most powerful tests for a special type of testing problems will be introduced in Section 7, and some of its examples are given in the final section.

Notation and terminology are the same as those of Part I.

5. Many-sided hypotheses

In the present section, an attempt will be made to apply the nearest distribution consideration to the problems of testing hypotheses which are so-called "many-sided". Actual procedures of derivations and realizations of the optimum tests for the many-sided hypotheses, in general, have many difficulties, and as far as the author knows, no respectable theory has been found in these directions, except for the discussion given by E. L. Lehmann [2, Chap. 3, Sec. 7] about the two-sided hypotheses concerning the one-parameter exponential families.

Now, consider a testing problem $(H, g)_\alpha$, and assume that the class of hypotheses H is covered by k (not necessarily disjoint) subclasses, $H \subset \bigcup_{i=1}^{k} H_i$, say. Suppose, further, that there exists a set of k statistics, $T_1(Z), T_2(Z), \cdots, T_k(Z)$ satisfying the conditions

$$E_g[T_i(Z)] > \sup_{f_i \in H_i} E_{f_i}[T_i(Z)], \quad i=1, 2, \cdots, k, \tag{5.1}$$

and

$$C(g, T_i)\cap L_0(T_i)^{\Lambda(L_0(T_i))}\neq 0\ ,\quad i=1,2,\cdots,k\ . \tag{5.2}$$

Then, there exists the nearest distribution for each distance problem $(H_i^{\Lambda(H_i)}: g)$ with the form

$$f_i^0(z)=g(z)\exp[\tau_i^0 T_i(z)]/M_i(\tau_i^0)\ , \tag{5.3}$$

where $M_i(\tau_i)=E_g[e^{\tau_i T_i(Z)}]$, and $\tau_i^0(<0)$ is the solution of the equation $\theta_i^0=(d/d\tau_i)\log M_i(\tau_i)$ with definition $\theta_i^0=E_{L_0}[T_i(Z)]$, for $i=1,2,\cdots,k$.

The probability density function $f_i^0(z)$ may be regarded as the nearest distribution in the i^{th} direction for the distance problem $(H{:}g)$, and the set of these f_i^0's for $i=1,2,\cdots,k$, is denoted by N_0. Let $\Lambda(N_0)$ be the set of all probability k-vectors, i.e.,

$$\Lambda(N_0)=\left\{\lambda=(\lambda_1,\lambda_2,\cdots,\lambda_k)\ ;\ \sum_{i=1}^{k}\lambda_i=1,\lambda_i\geqq 0,(i=1,\ 2,\cdots,k)\right\}. \tag{5.4}$$

Then it is a convex, closed and bounded subset of a hyperplane in the k-dimensional euclidean space. Define, as before,

$$N_0^{\Lambda(N_0)}=\left\{f_\lambda^0(z)=\sum_{i=1}^{k}\lambda_i f_i^0(z)\ ;\quad \lambda\in\Lambda(N_0)\right\}. \tag{5.5}$$

Then it is clear that $N_0^{\Lambda(N_0)}\subset H^{\Lambda(H)}$.

Consider the mean informations

$$I(f_\lambda^0: g)=\int_R f_\lambda^0(z)\log\frac{f_\lambda^0(z)}{g(z)}dm(z)\ ,\quad \lambda\in\Lambda(N_0)\ . \tag{5.6}$$

Then these form a family of continuous functions of λ, which are defined on the set $\Lambda(N_0)$. Therefore, it is obvious, from the convergence theorem of the Kullback-Leibler mean information given by the present author [9, Theorem 1.1], that there exists at least one member $\lambda^0=(\lambda_1^0,\lambda_2^0,\cdots,\lambda_k^0)$ of $\Lambda(N_0)$ which minimizes the mean information (5. 6), that is,

$$I(f_{\lambda^0}^0: g)=\min_{\lambda\in\Lambda(N_0)} I(f_\lambda^0: g)\ . \tag{5.7}$$

If we can find such a minimizing probability k-vector λ^0, then the nearest distribution, $f_{\lambda^0}^0(z)$, of the distance problem $(N_0^{\Lambda(N_0)}: g)$ may have the possibility to become the closest one, which is usable to derive an optimum test for the testing problem $(H, g)_\alpha$.

By virtue of the convexity of the function $x\log x$, it will easily be seen that, for any λ in $\Lambda(N_0)$,

$$I(f_\lambda^0: g) \leqq \sum_{i=1}^{k} \lambda_i I(f_i^0: g) , \tag{5.8}$$

with equality when and only when all the $f_i^0(z)$'s are identical with each other almost everywhere with respect to the measure m. In our present situation, the equality in (5.8) cannot hold, therefore the above inequality, with λ^0 in place of λ, shows that the minimum distance (5.7) is strictly smaller than the λ^0-weighted mean of $I(f_i^0:g)$'s.

In order to find a minimizing k-vector λ^0, it will be sufficient to minimize the following expression

$$I(f_\lambda^0: g)-a \sum_{i=1}^{k} \lambda_i = \int_R \left(\sum_{i=1}^{k} \lambda_i \psi_i \log \sum_{i=1}^{k} \lambda_i \psi_i\right) g dm - a \sum_{i=1}^{k} \lambda_i , \tag{5.9}$$

where a is Lagrange's multiplier, and

$$\psi_i(z) = f_i^0(z)/g(z) , \quad i=1, 2, \cdots, k . \tag{5.10}$$

It can be seen that λ^0 is a solution of a system of equations in λ such as

$$\int_R \left(\psi_j \log \sum_{i=1}^{k} \lambda_i \psi_i\right) g dm = \int_R \left(\sum_{i=1}^{k} \lambda_i \psi_i \log \sum_{i=1}^{k} \lambda_i \psi_i\right) g dm , \quad j=1, 2, \cdots, k . \tag{5.11}$$

In general, it is not easy to solve the above equation. We shall try to solve it in a special case for which some strong assumptions are imposed. Put, for each pair (i, j), $i, j=1, 2, \cdots, k$,

$$D_{ij} = \{z; \psi_i(z) > \psi_j(z)\} , \tag{5.12}$$

and, for simplicity, put $d\nu(z) = g(z) dm(z)$. Then we obtain the following

LEMMA 5.1. *Assume that, for each pair* (i,j), $i,j=1, 2, \cdots, k$, *there exists a one-to-one measurable transformation* u_{ij} *from* D_{ij} *onto* D_{ji} *such that* (i) *for any measurable subset* S *of* D_{ij}, $\nu(S) = \nu(u_{ij}(S))$, *and* (ii) $\psi_i(z) = \psi_j(u_{ij}(z))$ *and* $\psi_j(z) = \psi_i(u_{ij}(z))$. *Then, the vector* $\lambda^0 = (1/k, 1/k, \cdots, 1/k)$ *is a probability k-vector which minimizes the value of* $I(f_\lambda^0: g)$.

PROOF. From (5.11) we have for each pair (i,j), $i,j=1, 2, \cdots, k$, the equation,

$$\int_R (\psi_i - \psi_j) \log \left(\sum_{n=1}^{k} \lambda_n \psi_n\right) d\nu = 0 , \tag{5.13}$$

from which it follows, by definition (5.12), that

$$\int_{D_{ij}} (\psi_i - \psi_j) \log \left(\sum_{n=1}^{k} \lambda_n \psi_n\right) d\nu = \int_{D_{ji}} (\psi_j - \psi_i) \log \left(\sum_{n=1}^{k} \lambda_n \psi_n\right) d\nu . \tag{5.14}$$

By virtue of the conditions (i) and (ii) of the lemma, the left-hand member of the above equation becomes

$$(5.15)\quad \int_{D_{ij}} (\psi_i-\psi_j)\log\Big(\sum_{n=1}^{k}\lambda_n\psi_n\Big)d\nu=\int_{D_{ji}} (\psi_j-\psi_i)\log\Big(\sum_{n=1}^{k}\lambda_{n(i,j)}\psi_n\Big)d\nu\ ,$$

where $\{n(i,j);\ n=1,2,\cdots,k\}$ is the sequence obtained from $\{1,2,\cdots,k\}$ by interchanging i and j. Hence the equation (5.14) becomes

$$(5.16)\qquad \int_{D_{ji}} (\psi_j-\psi_i)\log\frac{\sum_{n=1}^{k}\lambda_{n(i,j)}\psi_n}{\sum_{n=1}^{k}\lambda_n\psi_n}\,d\nu=0\ .$$

Therefore, if the identities

$$(5.17)\qquad \sum_{n=1}^{k}\lambda_n\psi_n(z)=\sum_{n=1}^{k}\lambda_{n(i,j)}\psi_n(z)\ ,\quad (a.e.\ \nu)\ ,$$

$i,j=1,2,\cdots,k$, hold for a certain k-vector λ, then λ is a solution of the equation (5.11). In fact, the probability k-vector $\lambda^0=(1/k,1/k,\cdots,1/k)$ is a solution of (5.17), and moreover, it is unique unless $\psi_i(z)$'s are linearly dependent (ν), which completes the proof of our lemma.

Only the very special problems of testing hypotheses satisfy the conditions of the above lemma. Some of them will be shown later.

Now, the procedure of derivation of the most powerful test will be given in the following

LEMMA 5.2. *Under the situations mentioned above, if there exist a set* N_0 *of nearest distributions and a probability k-vector* $\lambda^0=(\lambda_1^0,\lambda_2^0,\cdots,\lambda_k^0)$ *which minimizes* $I(f_\lambda^0: g)$ *for* $\lambda\in\Lambda(N_0)$, *and moreover, if the test*

$$(5.18)\qquad \varphi_0(z)=\begin{cases}1, & \text{if } \sum_{i=1}^{k}\lambda_i^0\exp[\tau_i^0T_i(z)]/M_i(\tau_i^0)<c\ ,\\ a, & \text{if } \sum_{i=1}^{k}\lambda_i^0\exp[\tau_i^0T_i(z)]/M_i(\tau_i^0)=c\ ,\\ 0, & \text{otherwise,}\end{cases}$$

satisfies the size condition for the testing problem $(H,g)_\alpha$, *then it is a most powerful test for this problem.*

The proof of this lemma is omitted.

If the test φ_0 given above satisfies the size condition for $(H,g)_\alpha$, then it will easily be seen that

$$(5.19)\qquad E_{f_i^0}[\varphi_0(z)]=\alpha\ ,\quad i=1,2,\cdots,k\ .$$

In other words, (5.19) is a necessary condition in order that the test φ_0 is a most powerful test for the testing problem $(H,g)_\alpha$.

In general, the test (5.18) depends upon the alternative g through all the constant factors, $M_i(\tau_i^0)$'s, τ_i^0's, λ_i^0's, c and a. As was seen in Corollary 3.1, in our present procedure the possibility of enlargement of the alternative g to some composite alternatives A containing g, would be certified only when all the testing problems $(H,g_1)_\alpha$'s for $g_1 \in A$ possesses the same test as that of $(H,g)_\alpha$, as the most powerful ones, therefore, in the general cases, we cannot enlarge the alternative g in Lemma 5.2 to any class A of alternatives. Even if the test (5.18) is reduced to a test which is independent of the first three constant factors described above, i.e., M_i's, τ_i^0's and λ_i^0's, the above remark still remains true, because the possibility of such a reduction may be guaranteed by the special form of the generalized probability density function $g(z)$.

We shall show below some simple examples.

Example 5.1. First we shall consider a well-known two-sided hypotheses concerning a normal population. Let $X_1, X_2, \cdots, X_n$, be a random sample from a normal population $N(\mu,1)$ with unknown mean and variance unity. The testing problem concerned is $(H,g)_\alpha$, where

$$(5.20)\qquad \begin{cases} H : |\mu| \geqq \rho\ , & (\rho \text{ is positive and given})\ , \\ g : \mu = 0\ . \end{cases}$$

Since the sample mean $\bar{X} = (1/n) \sum_{i=1}^{n} X_i$ is a sufficient statistic for our present problem, we can assume, without any loss of generality, that

$$(5.21)\qquad \begin{cases} H = \left\{ f(\bar{x}) = \sqrt{\dfrac{n}{2\pi}} \exp\left[-\dfrac{n}{2}(\bar{x}-\mu)^2 \right];\ |\mu| \geqq \rho \right\}, \\ g = g(\bar{x}) = \sqrt{\dfrac{n}{2\pi}} \exp\left(-\dfrac{1}{2}\bar{x}^2 \right). \end{cases}$$

Subdividing the class H into two subclasses, put

$$(5.22)\qquad \begin{cases} H_1 : \mu \geqq \rho\ , \\ H_2 : \mu \leqq -\rho\ , \end{cases}$$

and examine the two statistics

$$(5.23)\qquad T_1(\bar{X}) = -\bar{X}\ , \quad T_2(\bar{X}) = \bar{X}\ ,$$

corresponding to the above subclasses, respectively. Clearly, these statistics satisfy the conditions (5.1) and (5.2), with $N(\rho, 1)$ and $N(-\rho, 1)$ as $L_0(T_1)$ and $L_0(T_2)$, respectively. After some calculations, by Lemma 2.2 ($s=1$), we can easily obtain

$$\begin{cases} \tau_1^0=\tau_2^0=-n\rho\ , \\ M_1(\tau_1^0)=M_2(\tau_2^0)=\exp(n\rho^2/2)\ , \end{cases} \tag{5.24}$$

from which we can find the nearest distributions such that

$$\begin{cases} f_1^0(\bar{x})=\sqrt{\dfrac{n}{2\pi}}\exp\left[-\dfrac{n}{2}(\bar{x}-\rho)^2\right], \\ f_2^0(\bar{x})=\sqrt{\dfrac{n}{2\pi}}\exp\left[-\dfrac{n}{2}(\bar{x}+\rho)^2\right], \end{cases} \tag{5.25}$$

and $N_0=\{f_1^0, f_2^0\}$.

On the other hand, it can be verified that the conditions (i) and (ii) of Lemma 5.1 are satisfied by the transformation $u_{12}(\bar{x})=-\bar{x}$ defined over the whole real line, since D_{12} and D_{21} are symmetric to each other with respect to the origin, and

$$\psi_i(\bar{x})=\exp\left[-n\rho(T_i(\bar{x})+\rho/2)\right]\ ,\quad i=1, 2\ . \tag{5.26}$$

Consequently, by Lemma 5.1, a minimizing vector is given by $\lambda^0=(1/2, 1/2)$. From Lemma 5.2 it follows that the test

$$\varphi_0(\bar{x})=\begin{cases} 1, & \text{if } \exp(n\rho\bar{x})+\exp(-n\rho\bar{x})\leqq c\ , \\ 0, & \text{otherwise,} \end{cases} \tag{5.27}$$

is a most powerful test for the testing problem $(H, g)_\alpha$, because this test satisfies the size condition for $(H, g)_\alpha$. This test may be reduced to

$$\varphi_0(\bar{x})=\begin{cases} 1, & \text{if } |\bar{x}|\leqq c\ , \\ 0, & \text{otherwise,} \end{cases} \tag{5.28}$$

by virtue of the symmetricity and convexity of the function $e^{bt}+e^{-bt}$, of t.

A natural extension of the above example to the two-sample case will be the following

Example 5.2. Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ be two random samples from the normal populations $N(\mu, 1)$ and $N(\nu, 1)$ respectively, where X_i's and Y_j's are assumed to be mutually independent.

Let $(H, g)_\alpha$ be a testing problem such that

$$\text{(5.29)} \qquad \begin{cases} H : |\mu| \geqq \rho \quad \text{and} \quad |\nu| \geqq \rho\ , \\ g : \mu = 0 \quad \text{and} \quad \nu = 0\ . \end{cases}$$

For this problem, we shall consider the four statistics such as

$$\text{(5.30)} \qquad \begin{cases} T_1(Z) = -(\bar{X} + \bar{Y})\ , \quad T_2(Z) = \bar{X} + \bar{Y}\ , \\ T_3(Z) = -(\bar{X} - \bar{Y})\ , \quad T_4(Z) = \bar{X} - \bar{Y}\ , \end{cases}$$

corresponding respectively to the four subclasses of H, given by

$$\text{(5.31)} \qquad \begin{cases} H_1 : \mu \geqq \rho, \nu \geqq \rho\ , \quad H_2 : \mu \leqq -\rho, \nu \leqq -\rho\ , \\ H_3 : \mu \geqq \rho, \nu \leqq -\rho, \quad H_4 : \mu \leqq -\rho, \nu \geqq \rho\ . \end{cases}$$

Instead of these subclasses, we shall first consider the following four subsets of the parameter space;

$$\text{(5.32)} \qquad \begin{cases} H'_1 : \mu + \nu \geqq 2\rho, \quad H'_2 : \mu + \nu \leqq -2\rho\ , \\ H'_3 : \mu - \nu \geqq 2\rho, \quad H'_4 : \mu - \nu \leqq -2\rho\ . \end{cases}$$

Then, obviously $H_i \subset H'_i$, $i = 1, 2, 3, 4$. Since the statistics $\bar{X}$ and $\bar{Y}$ are jointly sufficient for our testing problem (5.29), it will be sufficient to consider the 2-dimensional euclidean space of $z = (\bar{x}, \bar{y})$ as the sample space R.

Clearly the conditions (5.1) and (5.2) are satisfied by the statistics (5.30) and the subclasses (5.32), and, it can easily be seen that

$$\text{(5.33)} \qquad \begin{cases} \tau_i^0 = -\dfrac{2nm}{n+m}\rho\ (= \tau^0)\ , \\ M_i(\tau_i^0) = \exp\left\{\dfrac{2nm}{n+m}\rho^2\right\}\ (= M(\tau^0))\ , \quad i = 1, 2, 3, 4\ . \end{cases}$$

Hence, the set N_0 of the nearest distributions consists of the following four members:

$$\text{(5.34)} \qquad \begin{cases} f_1^0(z) = g(z) \exp\left[-\tau^0(\bar{x} + \bar{y})\right] / M(\tau^0)\ , \\ f_2^0(z) = g(z) \exp\left[\tau^0(\bar{x} + \bar{y})\right] / M(\tau^0)\ , \\ f_3^0(z) = g(z) \exp\left[-\tau^0(\bar{x} - \bar{y})\right] / M(\tau^0)\ , \\ f_4^0(z) = g(z) \exp\left[\tau^0(\bar{x} - \bar{y})\right] / M(\tau^0)\ , \end{cases}$$

where, of course,

$$\text{(5.35)} \qquad g(z) = \frac{\sqrt{nm}}{2\pi} \exp\left[-\frac{n}{2}\bar{x}^2 - \frac{m}{2}\bar{y}^2\right].$$

From (5.34) we have

$$(5.36)\qquad \begin{cases} \psi_1(z)=\exp[-\tau^0(\bar{x}+\bar{y})]/M(\tau^0), & \psi_2(z)=\exp[\tau^0(\bar{x}+\bar{y})]/M(\tau^0)\ , \\ \psi_3(z)=\exp[-\tau^0(\bar{x}-\bar{y})]/M(\tau^0), & \psi_4(z)=\exp[\tau^0(\bar{x}-\bar{y})]/M(\tau^0)\ , \end{cases}$$

and

$$(5.37)\qquad \begin{cases} D_{12}=\{(\bar{x},\bar{y});\ \bar{x}+\bar{y}<0\}\ , & D_{21}=\{(\bar{x},\bar{y});\ \bar{x}+\bar{y}>0\}\ , \\ D_{13}=\{(\bar{x},\bar{y});\ \bar{y}<0\}\ , & D_{31}=\{(\bar{x},\bar{y});\ \bar{y}>0\}\ , \\ D_{14}=\{(\bar{x},\bar{y});\ \bar{x}<0\}\ , & D_{41}=\{(\bar{x},\bar{y});\ \bar{x}>0\}\ , \\ D_{23}=\{(\bar{x},\bar{y});\ \bar{x}>0\}\ , & D_{32}=\{(\bar{x},\bar{y});\ \bar{x}<0\}\ , \\ D_{24}=\{(\bar{x},\bar{y});\ \bar{y}>0\}\ , & D_{42}=\{(\bar{x},\bar{y});\ \bar{y}<0\}\ , \\ D_{34}=\{(\bar{x},\bar{y});\ \bar{x}-\bar{y}<0\}\ , & D_{43}=\{(\bar{x},\bar{y});\ \bar{x}-\bar{y}>0\}\ . \end{cases}$$

For each pair (D_{ij}, D_{ji}), there exists a one-to-one transformation u_{ij} from D_{ij} to D_{ji}, defined by

$$(5.38)\qquad \begin{cases} u_{12}(\bar{x},\bar{y})=(-\bar{x},-\bar{y}), & u_{13}(\bar{x},\bar{y})=(\bar{x},-\bar{y}), & u_{14}(\bar{x},\bar{y})=(-\bar{x},\bar{y})\ , \\ u_{23}(\bar{x},\bar{y})=(-\bar{x},\bar{y}), & u_{24}(\bar{x},\bar{y})=(\bar{x},-\bar{y}), & u_{34}(\bar{x},\bar{y})=(-\bar{x},-\bar{y})\ , \end{cases}$$

which satisfies the conditions of Lemma 5.1, therefore, a minimizing vector is given by $\lambda^0=(1/4, 1/4, 1/4, 1/4)$.

Consequently, we have a most powerful test for the testing problem $(f^0_{\lambda^0}, g)_\alpha$

$$(5.39)\qquad \varphi_0(z)=\begin{cases} 1, & \text{if } T(z)\leqq c\ , \\ 0, & \text{otherwise,} \end{cases}$$

where the statistic $T(Z)$ is defined by

$$(5.40)\qquad \begin{aligned} T(Z)=\exp[-\tau^0(\bar{X}+\bar{Y})]&+\exp[\tau^0(\bar{X}+\bar{Y})] \\ +\exp[-\tau^0(\bar{X}-\bar{Y})]&+\exp[\tau^0(\bar{X}-\bar{Y})]\ . \end{aligned}$$

Now, it is readily seen that the members of N_0, given by (5.34), correspond to the normal populations of the random variable $Z=(\bar{X}, \bar{Y})$ with common variance $(n+m)/nm$ and respective means (ρ, ρ), $(-\rho, -\rho)$, $(\rho, -\rho)$ and $(-\rho, \rho)$, therefore, they are the members of H_1, H_2, H_3 and H_4 defined by (5.31), respectively.

The rejection region of test (5.39) is a convex domain of the sample space R, symmetric with respect to the origin, to both $\bar{x}$- and $\bar{y}$-axis, and to the straight lines $\bar{x}+\bar{y}=0$ and $\bar{x}-\bar{y}=0$, which spreads with increasing c. It is also shown that, if the rejection region of test (5.39) is included by the square domain determined by four points $(\pm\rho, \pm\rho)$, for short, then the test (5.39) satisfies the size condition for the testing problem $(H, g)_\alpha$ given by (5.29).

For such a case, in order to realize the test (5.39), which is now a most powerful test for $(H, g)_\alpha$, it is necessary to solve the size condition

$$P_{f^0_{\lambda c}}\{T(Z) \leqq c\} = \alpha \,. \tag{5.41}$$

Unfortunately the author has not been successful in it yet.

In the above examples, closest distributions are easily guessed and we can derive the most powerful tests also by applying the Neyman-Pearson fundamental lemma (extended), on account of (5.19).

Example 5.3. We shall return again to the situation of the first example of this section, and consider the problem of testing hypotheses $(H, g)_\alpha$ where

$$\begin{cases} H : \mu \geqq \rho_1 \text{ or } \mu \leqq -\rho_2 \,, & (\rho_1, \rho_2 > 0, \text{ given}) \,, \\ g : \mu = 0 \,. \end{cases} \tag{5.42}$$

The class H of hypotheses may be divided into two subclasses, $H = H_1 \cup H_2$ (say), corresponding respectively to the expressions in the definition of H given above. For these two subclasses, taking the statistics

$$T_1(Z) = -\bar{X} \quad \text{and} \quad T_2(Z) = \bar{X} \,, \tag{5.43}$$

we obtain, analogously to Example 5.1,

$$\begin{cases} \tau_1^0 = -n\rho_1 \,, \quad \tau_2^0 = -n\rho_2 \,, \\ M_1(\tau_1^0) = \exp\left[(n\rho_1^2)/2\right] \,, \quad M_2(\tau_2^0) = \exp\left[(n\rho_2^2)/2\right] \,, \end{cases} \tag{5.44}$$

hence, the set N_0 of the nearest distributions consists of the following two probability density functions:

$$\begin{cases} f_1^0(\bar{x}) = g(\bar{x}) \exp\left[n\rho_1\bar{x} - (n\rho_1^2)/2\right] \,, \\ f_2^0(\bar{x}) = g(\bar{x}) \exp\left[-n\rho_2\bar{x} - (n\rho_2^2)/2\right] \,, \end{cases} \tag{5.45}$$

where $g(\bar{x})$ is the same as that given by (5.21).

For the distance problem $(N_0^{A(N_0)} : g)$, the nearest distribution is found by obtaining the minimizing probability vector $\lambda^0 = (\lambda_0, 1-\lambda_0)$, which is the solution of the equation

$$\int_R (\psi_1(\bar{x}) - \psi_2(\bar{x})) \log\left(\lambda\psi_1(\bar{x}) + (1-\lambda)\psi_2(\bar{x})\right) g(\bar{x})d\bar{x} = 0 \,, \tag{5.46}$$

where $\psi_1(\bar{x})$ and $\psi_2(\bar{x})$ are the second factors of the right-hand expressions of (5.45), respectively, and R is the real line.

It is difficult to solve the above equation explicitly, but an approximate solution would be found, if necessary, by applying Newton's method and using a high-speed computer.

Now, the most powerful test for the testing problem $(f^0_{\lambda_0}, g)_\alpha$ is given by

$$\varphi_0(\bar{x})=\begin{cases}1, & \text{if } T(\bar{x})\leqq c\ ,\\ 0, & \text{otherwise,}\end{cases} \tag{5.47}$$

where the statistic $T(\bar{X})$ is defined by

$$T(\bar{X})=\lambda_0 \exp[n\rho_1\bar{X}-n\rho_1^2/2]+(1-\lambda_0)\exp[-n\rho_2\bar{X}-n\rho_2^2/2]\ . \tag{5.48}$$

The rejection region of this test is a closed interval which is contained in the interval $(-\rho_2, \rho_1)$, and moreover, from (5.19) the rejection probabilities under f^0_1 and f^0_2 must be equal to α, that is,

$$P_{f^0_1}(W_0)=P_{f^0_2}(W_0)=\alpha\ , \tag{5.49}$$

where the set W_0 is the rejection region, defined by $W_0=\{\bar{x};\ T(\bar{x})\leqq c\}$.

In general, the condition (5.49) determines the rejection region completely, but the procedure of determination is not easy. It is an outstanding question whether the test φ_0 defined by (5.47) satisfies the condition (5.49), and if it does so our procedure will provide us with a useful method for the derivation and the realization of an optimum test for the testing problem under consideration.

6. Most powerful tests for non-parametric problems

In the present section, the method of derivation of the most powerful tests developped in Section 3 will be specialized for some problems of testing hypotheses in the non-parametric case.

For most non-parametric problems, the second condition of Theorem 3.1 (or Theorem 3.2) will be readily fulfilled by a certain statistic satisfying the first condition, because the classes of hypotheses are usually very wide, but, on the contrary, the validity of the third condition for the size of test is likely to be interrupted, due to just the same reason as above. Hence, if the testing problem, is one such that the class of hypotheses specifies directly or indirectly the tail probability of the probability distribution of a usable statistic, then our method mentioned in Section 3 will be applicable and a most powerful test will be obtained. The problem of testing hypotheses which is solvable by a sign test is a typical example of such problems.

Let, as before, $T(Z)$ be a vector of s real statistics of one dimension, $T_1(Z)$, $T_2(Z)$, $\cdots$, $T_s(Z)$, where Z is a random variable whose distribution is absolutely continuous with respect to a measure m on a σ-finite measure space (R, m).

In order to widen the scope of application of our procedure, it will be convenient to extend the definition of the mean information to the case where the carriers of two generalized probability density functions are different from each other (with respect to the measure m). Let D_f and D_g be the carriers of the density functions $f(z)$ and $g(z)$ respectively. Then, the following two cases may occur; $a)$ $m(D_f-D_g)>0$ and $b)$ $m(D_f-D_g)=0$. For both of these cases, we shall define the mean information for discrimination in favor of f against g, as before,

$$I(f:g)=\int_R f(z)\log\frac{f(z)}{g(z)}dm\ . \tag{6.1}$$

In this definition, the "power" of discrimination of each sample point z, which is obtained by an observation on the random variable Z whose density function is $f(z)$, is measured by the value of $\log\,(f(z)/g(z))$.

Therefore, the sample points belonging to the difference D_f-D_g contribute an infinite amount of mean information if $m(D_f-D_g)>0$, while those belonging to D_g-D_f do no amount. From these, as is shown also mathematically, it follows that $I(f:g)=\infty$ in the case $a)$, and $I(f:g)=\int_{D_f} f(z)\log\,(f(z)/g(z))\,dm$ in the case $b)$. It is easily shown that $I(f:g)\geqq 0$ in the latter case.

For the above definition of the mean information, Theorem 2.1 still holds true, where, of course, the class $K(T,\theta)$ is wider than that given in Section 2 for the distance problem $(K(T,\theta):g)$, or more precisely, the class $K(T,\theta)$ considered in Section 2 is a subclass of $K(T,\theta)$ in the present case so that the carriers of the members of the former are all the same as that of $g(z)$.

In the present and the subsequent two sections, the concept of directed distance is based upon the mean information defined by (6.1), and this extended definition brings the wider applicability of our method.

First, we shall consider the testing problem $(K(T,\theta), g)_\alpha$, where $K(T,\theta)$ is defined as

$$K(T,\theta)=\{f(z)\,;\,E_f[T(Z)]=\theta\}\ . \tag{6.2}$$

From Theorem 2.1 and the proof of Theorem 3.1, we easily obtain the following

THEOREM 6.1. *For some statistic $T(Z)$, if a system of equations*

$$\frac{\partial}{\partial\tau_i}\log M(\tau)=\theta_i\ , \quad (i=1, 2, \cdots, s)\ , \tag{6.3}$$

has a solution $\tau^0=(\tau_1^0, \tau_2^0, \cdots, \tau_s^0)$, then the nearest distribution for the distance problem $(K(T,\theta):g)$ exists in $K(T,\theta)$ and is of the form

$$f_0(z)=g(z)\exp[\tau^0 T(z)]/M(\tau^0)\ . \tag{6.4}$$

If the most powerful test for the testing problem $(f_0, g)_\alpha$, given by

$$\varphi_0(z)=\begin{cases}1, & \text{if } \tau^0 T(z)<c\ ,\\ a, & \text{if } \tau^0 T(z)=c\ ,\\ 0, & \textit{otherwise,}\end{cases} \tag{6.5}$$

satisfies the size condition for the testing problem $(K(T,\theta), g)_\alpha$, then it is a most powerful test for the problem $(K(T,\theta), g)_\alpha$, and hence $f_0(z)$ given by (6.4) *becomes the closest distribution.*

The proof of this theorem is omitted.

Now, when we want to test a class H of hypotheses, which is contained in $K(T,\theta)$ for a certain statistic $T(Z)$, against an alternative g, it is necessary to see whether the nearest distribution (6.4) is a member of H, and if it is so, the test given by (6.5) will be most powerful for the testing problem $(H, g)_\alpha$, provided that it satisfies the size condition for this problem.

For a certain type of problems of testing non-parametric hypotheses, it is often the case that we can find a statistic such that the class of hypotheses and the alternative are specified almost completely by the expected value of that statistic, or more precisely, the hypothetical class H coincides with or is contained in $K(T,\theta)$. In such a case, the "specifying" statistic can be taken as a test statistic, as will be seen in the following section.

Very few of the non-parametric problems of testing hypotheses would possess the most powerful tests, and it seems, to the author, that the methods which have been introduced in order to derive the most powerful tests in the literatures so far are diverse and divided. In the present paper, we shall try to give a unified method, applying the result in Section 3, to some of these problems.

7. Non-parametric hypotheses which are specified by the characteristic variables

In the present section, we shall be concerned with a certain type of non-parametric problems of testing hypotheses, for which the class of hypotheses is specified by some characteristic variables associated with a partition of the sample space which suitably corresponds to the testing problem under investigation.

Let (R, m) be a σ-finite measure space which is the n-product of a component space (R_0, m_0), that is, $(R, m)=(R_0, m_0)\times(R_0, m_0)\times\cdots\times(R_0, m_0)$, where (R_0, m_0) is a certain σ-finite measure space. As were defined in Section 2 (the corollaries 2.1 and 2.2), Q_0, Q_1 and Q_2 respectively denote the classes of the generalized probability density functions of all probability distributions, with the forms $f(z)=\prod_{i=1}^{n} f_i(x_i)$ and $f(z)=\prod_{i=1}^{n} f(x_i)$ of a vector of random variables $Z=(X_1, X_2, \cdots, X_n)$ where X_i is a random variable defined on the i^{th} component space (R_0, m_0) with a generalized probability density function $f_i(x_i)$ in general. In other words, the class Q_1 stands for the case in which the n component variables are mutually independent, and Q_2 stands for the case in which they are independently and identically distributed, while for the class Q_0 such restrictions are not imposed. Clearly it holds that $Q_0 \supset Q_1 \supset Q_2$.

Let $\{W_1, W_2, \cdots, W_s\}$ be an (m)-partition of the space R, i.e., $W_1, W_2, \cdots, W_s$ are the measurable and mutually disjoint (m) subsets of R such that $R=\sum_{i=1}^{n} W_i$. Let $T_1(z), T_2(z), \cdots, T_s(z)$ be the defining functions of $W_1, W_2, \cdots, W_s$, respectively. Then the statistics $T_1(Z)$, $T_2(Z), \cdots, T_s(Z)$ are called the characteristic variables associated with the partition $\{W_1, W_2, \cdots, W_s\}$ of the sample space R. Put $T(Z)=(T_1(Z), T_2(Z), \cdots, T_s(Z))$. Furthermore, let θ and θ' be two probability s-vectors such that the components are all positive, and $\theta \neq \theta'$. For a probability density function $g(z)$ belonging to the class Q_0 suppose that $E_g[T(Z)]=\theta'$. Clearly the necessary assumptions of regularity and existence of $M(\tau)=E_g[e^{\tau T(Z)}]$, as a function of $\tau=(\tau_1, \tau_2, \cdots, \tau_s)$, are fulfilled in our present case.

Under these situations, consider the distance problem $(K(T, \theta) : g)$. Then, as an extension of the result given by S. Kullback [6, Chap. 3, Example 2.3], we can prove the following

LEMMA 7.1. (i) *The equation* (6.3) *has always a solution* $\tau^0=(\tau_1^0, \tau_2^0, \cdots, \tau_s^0)$ *such that*

$$\tau_i^0=\log \rho\gamma_i\ , \quad (i=1, 2, \cdots, s) \tag{7.1}$$

where ρ *is an arbitrary positive constant and* $\gamma_i=\theta_i/\theta_i'$, $i=1, 2, \cdots, s$.

(ii) *If* $\gamma_1\leqq\gamma_2\leqq\cdots\leqq\gamma_s$, *then it holds that*

$$\tau_1^0\leqq\tau_2^0\leqq\cdots\leqq\tau_s^0\ . \tag{7.2}$$

PROOF. Since $M(\tau)=\sum_{i=1}^{s}\theta_i'\exp(\tau_i)$, changing the variables from τ_i's to $y_i=\theta_i'\exp(\tau_i)$'s, we can obtain from (6.3) a system of equations

$$\frac{y_i}{\sum_{k=1}^{s} y_k}=\theta_i\ , \quad (i=1, 2, \cdots, s)\ , \tag{7.3}$$

which has infinitely many solutions such that $y_i=\rho\theta_i$, $(i=1, 2, \cdots, s)$ where ρ is any positive constant. From this, the two statements of the lemma follow.

It will easily be seen that the nearest distribution becomes independent of ρ, i.e.,

$$f_0(z)=g(z)\exp\left\{\sum_{i=1}^{s}(\log\gamma_i)\ T_i(z)\right\}/\text{const.}\ , \tag{7.4}$$

by virtue of the relation $\sum_{i=1}^{s}T_i(z)=1\ (m)$ on R. Therefore, we can, without any loss of generality, take the constant to be unity, and

$$\tau_i^0=\log\gamma_i\ , \quad (i=1, 2, \cdots, s)\ . \tag{7.5}$$

Now, an (m)-partition of the sample space R, $\{W_1, W_2, \cdots, W_s\}$, is said, for short, to be "n-decomposable", if there exists a (m_0)-partition $\{V_i^1, V_i^2, \cdots, V_i^{s_i}\}$ of the i^{th} component space (R_0, m_0) for each i, $(i=1, 2, \cdots, n)$, such that the members W_j's are all direct products of members of the partitions of n component spaces, that is,

$$W_j=V_1^{k_1}\times V_2^{k_2}\times\cdots\times V_n^{k_n}\ , \tag{7.6}$$

and consequently $s=\prod_{i=1}^{n}s_i$.

An example of n-decomposable (m)-partition may be given as follows:

let $\{V_i^0, V_i^1\}$ be an (m_0)-partition of the i^{th} component space R_0 for each i. Then the 2^n subsets of R which are direct products $V_1^{k_1}\times V_2^{k_2}\times\cdots\times V_n^{k_n}$, $(k_i=0, 1;\ i=1, 2, \cdots, n)$, form an n-decomposable partition of R, $\{W_0, W_1, \cdots, W_s\}$ where $s+1=2^n$. If, moreover, for a probability density function $f(z)$ belonging to the class Q_1, $P_{f_i}(V_i^1)=p_i$, $(i=1, 2, \cdots, n)$, then we have

$$(7.7)\qquad P_f(W_j)=P_f(V_1^{k_1}\times V_2^{k_2}\times\cdots\times V_n^{k_n})=\prod_{i=1}^{n} p_i^{k_i}(1-p_i)^{1-k_i}\ ,$$

$k_i=0, 1\,;\ i=1, 2, \cdots, n$. Thus we have an n-decomposable partition $\{W_0, W_1, W_2, \cdots, W_s\}$ of R with a system of probabilities $\{\theta_0, \theta_1, \theta_2, \cdots, \theta_s\}$, where $\theta_j=\prod_{i=1}^{n} p_i^{k_i}(1-p_i)^{1-k_i}$.

In general, if there exist a partition $\{W_1, W_2, \cdots, W_s\}$ of the sample space R and a certain given probability s-vector $(\theta_1, \theta_2, \cdots, \theta_s)$, then they determine the class of generalized probability density functions such as $H=\{f\colon E_f[T_j(Z)]=\theta_j, j=1, 2, \cdots, s\}$, where $T_j(Z)$'s are the characteristic variables associated with W_j's respectively. We shall call the above partition with a system of probabilities a *probability scheme*. If a partition $\{W_1, W_2, \cdots, W_s\}$ is n-decomposable and a probability s-vector $(\theta_1, \theta_2, \cdots, \theta_s)$ associated with the partition is n-decomposable, i.e., $\theta_j= p_{1k_1}\cdot p_{2k_2}\cdots p_{nk_n}$ corresponding to the expression $W_j=V_1^{k_1}\times V_2^{k_2}\times\cdots\times V_n^{k_n}$, $k_i=1, 2, \cdots, s_i$; $i=1, 2, \cdots, n$; $j=1, 2, \cdots, s$, where $\sum_{k_i=1}^{s_i} p_{ik_i}=1$, $i= 1, 2, \cdots, n$, then the probability scheme is called an *n-decomposable probability scheme*. Given a problem of testing hypotheses, a probability scheme can be constructed for it, and our procedure of deriving a most powerful test will be performed, as will be seen later.

We shall show, as examples, an "equi-probability scheme" and a "binomial probability scheme" for convenience of the later use. A probability scheme $\{W_1, W_2, \cdots, W_s\}$ with a probability s-vector $(\theta_1, \theta_2, \cdots, \theta_s)$ such that $\theta_1=\theta_2=\cdots=\theta_s(=1/s)$ will be called an "equi-probability scheme". On the other hand, in an n-decomposable probability scheme $\{W_0, W_1, W_2, \cdots, W_s\}$ with probability $(s+1)$-vector $(\theta_0, \theta_1, \theta_2, \cdots, \theta_s)$, if each member of the partition is of the form $W_j=V_1^{k_1}\times V_2^{k_2}\times\cdots\times V_n^{k_n}$ where $k_i=0, 1$; $i=1, 2, \cdots, n$, and if the corresponding probability is of the form

$$\theta_j = p^{\sum_{i=1}^{n} k_i} (1-p)^{n-\sum_{i=1}^{n} k_i},$$

then it is called a "binomial probability scheme". In this case $s+1=2^n$, and the members of the partition may be divided into $n+1$ groups, $\mathscr{E}_0, \mathscr{E}_1, \cdots, \mathscr{E}_n$, where $\mathscr{E}_i$ contains $\binom{n}{i}$ W_j's with the probability $p^i(1-p)^{n-i}$, that is, for example,

$$\begin{aligned} &\mathscr{E}_0 = \{W_0 = V_1^0 \times V_2^0 \times \cdots \times V_n^0\}, \\ &\mathscr{E}_1 = \{W_1 = V_1^1 \times V_2^0 \times \cdots \times V_n^0,\ W_2 = V_1^0 \times V_2^1 \times \cdots \times V_n^0,\ \cdots,\ W_n = \\ &\qquad\qquad V_1^0 \times V_2^0 \times \cdots \times V_n^1\}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots, \\ &\mathscr{E}_n = \{W_s = V_1^1 \times V_2^1 \times \cdots \times V_n^1\}, \end{aligned} \tag{7.8}$$

with sum of probabilities $\binom{n}{i}p^i(1-p)^{n-i}$ in $\mathscr{E}_i$ for $i=0, 1, \cdots, n$.

There would be other schemes, but in the present paper, we do not enter into further details. In connection with a binomial scheme, it will be convenient, for the later use, to show the following

LEMMA 7.2. *Let $\theta_i = p^i(1-p)^{n-i}$ and $\theta_i' = p_1^i(1-p_1)^{n-i}$, $i=0, 1, 2, \cdots, n$. If $p > p_1$, then the ratio $\gamma_i = \theta_i/\theta_i'$ increases monotonically with increasing i.*

PROOF. The proof of this lemma follows immediately from the inequality

$$\log \gamma_i - \log \gamma_{i-1} = \log \frac{p(1-p_1)}{p_1(1-p)} > 0, \tag{7.9}$$

for each i $(=1, 2, \cdots, n)$.

Now, consider the problem of testing hypotheses $(H_0, g)_\alpha$. For the case when the hypothetical class H_0 is included in Q_0, in general, we have the following theorem.

THEOREM 7.1. *In a testing problem $(H_0, g)_\alpha$, if a class H_0 of hypotheses can be specified by a vector of characteristic variables, $T(Z) = (T_1(Z), T_2(Z), \cdots, T_s(Z))$, associated with a probability scheme $\{W_1, W_2, \cdots, W_s\}$ with a probability s-vector $\theta = (\theta_1, \theta_2, \cdots, \theta_s)$, (that is, the class H_0 is specified by a probability scheme), i.e.,*

$$H_0 = \{f(z)\,;\, E_f[T(Z)] = \theta\}, \tag{7.10}$$

then the test

$$\varphi_0(z)=\begin{cases}1, & \text{if } \sum_{i=1}^{s}\tau_i^0\,T_i(z)<c\ ,\\ a, & \text{if } \sum_{i=1}^{s}\tau_i^0\,T_i(z)=c\ ,\\ 0, & \text{otherwise,}\end{cases} \tag{7.11}$$

where $\tau^0=(\tau_1^0, \tau_2^0, \cdots, \tau_s^0)$ *is a solution given by* (7.5) *of the equation*(6.3) *with* $\theta'=E_g[T(Z)]$, *is a most powerful test for the testing problem* $(H_0, g)_\alpha$. *Moreover, it becomes uniformly most powerful for the testing problem* $(H_0, A_0)_\alpha$ *where*

$$A_0=\{g_1(z);\ E_{g_1}[T(Z)]=\theta'\}\ . \tag{7.12}$$

PROOF. From Theorem 6.1 and Lemma 7.1, it follows that the nearest distribution of the distance problem $(H_0 : g)$ exists and belongs to the class H_0, with the form

$$f_0(z)=g(z)\exp[\tau^0 T(z)]/M(\tau^0)\ . \tag{7.13}$$

The class H_0 is just the same as $K(T, \theta)$ given by (6.2), and the size of the test (7.10) now depends only on θ, that is, the test is similar for the hypothetical class H_0, therefore we have the first statement of the theorem.

It can be seen that, from Lemma 7.1, the solution τ^0 is independent of a special alternative in the class A_0 defined in the theorem, that is, τ_0 depends upon θ' only, therefore, the second statement of the theorem follows from Corollary 3.1. Thus, the proof of our theorem is complete.

If, in a testing problem $(H_1, g)_\alpha$, the class H_1 of hypotheses specifies a class of generalized probability density functions belonging to Q_1, and the alternative g also belongs to Q_1, then we have the following

THEOREM 7.2. *In a testing problem* $(H_1, g)_\alpha$, *if the class* H_1 *is specified by a vector of characteristic variables* $T(Z)=(T_1(Z), T_2(Z), \cdots, T_s(Z))$ *associated with a certain n-decomposable probability scheme* $\{W_1, W_2, \cdots, W_s\}$ *with a probability s-vector* $\theta=(\theta_1, \theta_2, \cdots, \theta_s)$, *in such a manner that*

$$H_1=\left\{f(z)=\prod_{i=1}^{n}f_i(x_i);\ E_f[T(Z)]=\theta\right\}\ , \tag{7.14}$$

then the test given by (7.11) *in the preceding theorem is a most powerful* *test for the problem* $(H_1, g)_\alpha$, *provided that the alternative* g *belongs to* Q_1. *The second statement of the preceding theorem also remains true in the present case, if we take the class* A_1 *defined by*

$$A_1=\left\{g_1(z)=\prod_{i=1}^{n} g_{1i}(x_i);\ E_{g_1}[T(Z)]=\theta'\right\}, \tag{7.15}$$

instead of A_0 *in that theorem.*

PROOF. Taking account of the proof of the preceding theorem, it will be sufficient to show that the nearest distribution of the distance problem $(H_1 : g)$ belongs to the class H_1.

For this, let us examine the nearest distribution given by (7.13). Since $g(z)$ is a member of the class Q_1, let $g(z)=\prod_{i=1}^{n} g_i(x_i)$.

By the assumption that the (m)-partition $\{W_1, W_2, \cdots, W_s\}$ is n-decomposable, there exists an (m_0)-partition $\{V_i^1, V_i^2 \cdots, V_i^{s_i}\}$ of the i^{th} component space R_0, on which the i^{th} component variable X_i of a random vector Z is distributed according to a probability distribution with density function $f_i(x_i)$ (under f belonging to H_1) or $g_i(x_i)$ (under g), for each $i(=1, 2, \cdots, n)$. Let $T_i^{k_i}(x_i)$ be the defining function of $V_i^{k_i}$, and let $p_{ik_i}=E_{f_i}[T_i^{k_i}(X_i)]$ and $p'_{ik_i}=E_{g_i}[T_i^{k_i}(X_i)]$, $k_i=1, 2, \cdots, s_i$; $i=1, 2, \cdots, n$. Then, from the n-decomposability of the probability scheme described in the theorem, which specifies the class H_1, it follows that, if $W_j= V_1^{k_1}\times V_2^{k_2}\times \cdots \times V_n^{k_n}$ then $\theta_j=p_{1k_1}\cdot p_{2k_2}\cdot \cdots \cdot p_{nk_n}$ and $E_g[T_j(Z)]=\theta'_j=p'_{1k_1}\cdot p'_{2k_2} \cdot\cdots\cdot p'_{nk_n}$.

It will easily be seen that

$$\begin{aligned} \tau^0 T(z) &=\log\left(\prod_{i=1}^{n}\prod_{k_i=0}^{s_i} p_{ik_i}^{T_i^{k_i}(x_i)}\Big/\prod_{i=1}^{n}\prod_{k_i=1}^{s_i} p'^{\,T_i^{k_i}(x_i)}_{ik_i}\right) \\ &=\sum_{i=1}^{n}\sum_{k_i=1}^{s_i} T_i^{k_i}(x_i)\log\left(p_{ik_i}/p'_{ik_i}\right), \end{aligned} \tag{7.16}$$

from which it follows that the nearest distribution (7.13) becomes

$$f_0(z)=\prod_{i=1}^{n}\left(g_i(x_i)\exp\left\{\sum_{k_i=1}^{s_i} T_i^{k_i}(x_i)\log\gamma_{ik_i}\right\}\right)/\text{const}., \tag{7.17}$$

where $\gamma_{ik_i}=p_{ik_i}/p'_{ik_i}$, which is a generalized probability density function belonging to the class Q_1. From this our theorem follows.

Analogously we can show the following theorem for the testing problem $(H_2, g)_\alpha$ in the case where both H_2 and g are included in the class Q_2.

THEOREM 7.3. *Let* $T(Z)=(T_1(Z), T_2(Z), \cdots, T_s(Z))$ *be a vector of characteristic variables associated with an n-decomposable probability scheme* $\{W_1, W_2, \cdots, W_s\}$ *with probability s-vector* $\theta=(\theta_1, \theta_2, \cdots, \theta_s)$, *which is composed of the same partition* $\{V_1, V_2, \cdots, V_{s_0}\}$ *(say) of* R_0 *for every component. Assume that a class* H_2 *of probability density functions and* g *are contained in the class* Q_2.

If the class H_2 *is specified by the statistic* $T(Z)$ *in such a manner that*

(7.18)
$$H_2=\left\{f(z)=\prod_{i=1}^{n} f(x_i);\ E_f[T(Z)]=\theta\right\},$$

then the problem of testing hypotheses $(H_2, g)_\alpha$ *possesses a most powerful test with the same form as that given by* (7.11). *The second assertion in Theorem* 7.1 *holds true, if we take the class*

(7.19)
$$A_2=\left\{g_1(z)=\prod_{i=1}^{n} g_1(x_i);\ E_{g_1}[T(Z)]=\theta'\right\},$$

instead of A_0 *in that theorem, where* $\theta'=E_g[T(Z)]$.

PROOF. From the assumption we can put $g(z)=\prod_{i=1}^{n} g(x_i)$. For f in H_2 and for g, put $p_i=E_f[T_i(X)]$ and $p'_i=E_g[T_i(X)]$, $i=1, 2, \cdots, s_0$, where $T_i(x)$'s are the defining functions of V_i's respectively.

Parallel to the proof of the preceding theorem, we can obtain the result that the nearest distribution (7.13) of the distance problem $(K(T, \theta) : g)$ is given by

(7.20)
$$f_0(z)=\prod_{i=1}^{n}\left(g(x_i)\exp\left\{\sum_{k=1}^{s_0} T_k(x_i)\log\frac{p_k}{p'_k}\right\}\right)/\text{const.},$$

which is a member of the class Q_2.

Now, the assertions of our present theorem will be confirmed in a manner analogous to the proof of Theorem 7.1.

To realize the test given by (7.11) which is a most powerful test for the testing problem $(H_0, g)_\alpha$, $(H_1, g)_\alpha$ or $(H_2, g)_\alpha$ corresponding to each case in the above three theorems, we must determine the constants c and a such that the test satisfies the required size condition. The procedure of the determination will be shown in the following

Lemma 7.3. *Throughout the three theorems stated above, if the members of the (m)-partition of the probability scheme are arranged in such a way that* $\gamma_1 \leqq \gamma_2 \leqq \cdots \leqq \gamma_s$ *where* $\gamma_j = \theta_j / \theta'_j$, $j = 1, 2, \cdots, s$, *then the test* (7.11) *becomes equivalent to*

$$\varphi_0(z) = \begin{cases} 1, & \text{if } \sum_{j=1}^{k_0} T_j(z) = 1 \,, \\ a_0, & \text{if } T_{k_0+1}(z) = 1 \,, \\ 0, & \text{otherwise,} \end{cases} \tag{7.21}$$

where k_0 *and* a_0 *are the constants given by*

$$\begin{cases} k_0 = \max\left\{k \,;\, \sum_{j=1}^{k} \theta_j < \alpha\right\} \,, \\ a_0 = \left(\alpha - \sum_{j=1}^{k_0} \theta_j\right) \Big/ \theta_{k_0+1} \,. \end{cases} \tag{7.22}$$

Proof. By the assumption of this lemma, and Lemma 7.1, it holds that the size of the test φ_0 given by (7.11) becomes

$$E_f[\varphi_0(Z)] = \sum_{j=1}^{k} \theta_j + a\theta_{k+1} \,, \tag{7.23}$$

where the constant k represents the number of τ_j^0's which are smaller than c. Then, the size condition implies that $k = k_0$ and $a = a_0$ given by (7.22), and the proof of our lemma is complete.

It will easily be noticed that, if all the θ_j's are identical with each other, then

$$k_0 = [s\alpha] \,, \quad a_0 = s\alpha - [s\alpha] \,, \tag{7.24}$$

where the bracket [ ] designates the Gauss symbol. This is the case of an equi-probability scheme.

It is also remarked that the function $\sum_{j=1}^{k_0} T_j(z)$ in the expression (7.21) is the defining function of the subset of R, $W^{(k_0)} = \sum_{j=1}^{k_0} W_j$. In general, we define $W^{(k)} = \sum_{j=1}^{k} W_j$, $k = 1, 2, \cdots, s$.

When we consider the problem of testing hypotheses $(H, g)_\alpha$, where the class of hypotheses, H, is a proper subset of H_0, H_1 or H_2 in the above theorems, the following two cases may occur: a) the class H contains the nearest distribution of the distance problem $(H_0 : g)$, $(H_1 : g)$ or $(H_2 : g)$, and b) H does not contain the nearest distribution. In the first case a), the test given by (7.21) is most powerful for the testing problem $(H, g)_\alpha$, but in the second case b), it is not necessarily true. For such a case, a sufficient condition that the test (7.21) becomes most powerful for the testing problem $(H, g)_\alpha$, will be given by the following lemma.

LEMMA 7.4. *Under each situation of the above three theorems, if there exist a generalized probability density function $h(z)$ in the class $H^{A(H)}$ and a positive constant d such that*

$$\begin{cases} g(z) \geqq dh(z) & \text{on } W^{(k_0+1)}, \\ g(z) \leqq dh(z) & \text{on } R - W^{(k_0+1)}, \end{cases} \tag{7.25}$$

then the test given by (7.21) *is most powerful for the testing problem* $(H, g)_\alpha$.

The proof of this lemma follows from Theorem 1 given by E. L. Lehmann and C. Stein [4, p. 497]. In fact, since $W^{(k_0+1)}$ is a rejection region of the test (7.21) for the testing problem $(H, g)_\alpha$, the existence of such a density function $h(z)$ and a constant d in the present lemma implies that the test φ_0 in (7.21) has a form of probability ratio test for the testing problem $(H^{A(H)}, g)_\alpha$, and hence it is most powerful for the testing problem $(H, g)_\alpha$.

8. Applications to the sign-test and others

First, we shall examine a single sample location parameter problem treated by D. A. S. Fraser [8, 10]. This problem possesses a most powerful test, so called "sign-test" or more precisely "binomial-test", derived by means of a least favorable distribution. As was suggested by D. A. S. Fraser [8, Chap. 5, Sec. 2] or by E. L. Lehmann [2, Chap. 3, Sec. 8, Example 8], in the usual procedure hietherto introduced to derive the sign-test as an optimum one, the concept of the "generalized sufficient statistic" is considered, and in that procedure the explicit form of the closest distribution is required.

On the contrary, in our method developed in the preceding section, the procedure becomes very simple. In that, of course, the explicit form of the closest distribution is not needed.

Example 8.1. Let $X_1, X_2, \cdots, X_n$ be identically and independently distributed according to a probability distribution of the continuous type on the real line, with density function $f(x)$. Let $\xi_p(f)$ be the p-quantile of this distribution and consider the testing problem $(H, A)_\alpha$, where, for the given constants $p(0<p<1)$ and ξ_0,

$$\begin{cases} H: & \xi_p(f)\leqq\xi_0 , \\ A: & \xi_p(f)>\xi_0 . \end{cases} \tag{8.1}$$

Here, of course, the classes H and A should be regarded as the classes of the probability density functions of the random vector $Z=(X_1, X_2, \cdots, X_n)$ defined on the n-dimensional euclidean space R with the usual Euclid-Lebesgue measure m over it, which satisfy the conditions given by (8.1) for every marginal distribution, therefore, they are both subclasses of Q_2.

As usual, in the first step, we shall examine the testing problem $(H_0, g)_\alpha$ such that

$$H_0 : \xi_p(f)=\xi_0 , \tag{8.2}$$

and the alternative g is a certain fixed member of the class A, for which we shall assume that $\xi_p(g)=\xi_1(>\xi_0)$.

Denote the interval $(-\infty, \xi_0]$ by V_i^1 for the i^{th} axis, $i=1, 2, \cdots, n$. Then, $P_f(V_i^1)=p$ under the hypotheses and $P_g(V_i^1)=p_1(<p)$ under the alternative, for each component. Let V_i^0 be the complementary set of V_i^1, for $i=1, 2, \cdots, n$. As was considered in the preceding section, we can compose an n-decomposable binomial probability scheme, $\{W_0, W_1, W_2, \cdots, W_s\}$ with a probability $(s+1)$-vector $\theta=(\theta_0, \theta_1, \theta_2, \cdots, \theta_s)$, where we assume that the members of the partition, W_j's, are arranged in such a manner as (7.8). Thus, if W_j belongs to the group $\mathscr{E}_i$ then the corresponding θ_j is equal to $p^i(1-p)^{n-i}$, $j=0, 1, 2, \cdots, s$; $i=0, 1, 2, \cdots, n$.

The class of hypotheses, H_0, is specified by a vector of characteristic variables, $T(Z)=(T_0(Z), T_1(Z), T_2(Z), \cdots, T_s(Z))$ associated with the above partition. In fact, we have

$$H_0=\{f ; E_f[T(Z)]=\theta\}\cap Q_2 . \tag{8.3}$$

It follows from the lemmas 7.1, 7.2, 7.3 and Theorem 7.3 that the test

$$\varphi_0(z)=\begin{cases} 1, & \text{if } \sum_{j=0}^{k_0} T_j(z)=1 , \\ a_0, & \text{if } T_{k_0+1}(z)=1 , \\ 0, & \text{otherwise,} \end{cases} \tag{8.4}$$

is most powerful for the testing problem $(H_0, g)_\alpha$. From the construction of the binomial probability scheme used above, it can easily be seen that the test (8.4) is equivalent to

$$\varphi_0(z)=\begin{cases}1, & \text{if } \sum_{i=1}^{n} T(x_i)>c\,,\\ a, & \text{if } \sum_{i=1}^{n} T(x_i)=c\,,\\ 0, & \text{otherwise,}\end{cases} \tag{8.5}$$

where $T(x)$ is the defining function of the interval (ξ_0, ∞) on the real line.

The test (8.4) or (8.5) satisfies the size condition for the testing problem $(H, g)_\alpha$, hence it becomes most powerful for $(H, g)_\alpha$. It is also concluded from Corollary 3.1 that the test given above is uniformly most powerful for the original testing problem $(H, A)_\alpha$ given by (8.1).

Scrutinizing this example, it can easily be noticed that the problem of testing hypotheses $(H, A)_\alpha$ such that, for a certain fixed sequence $V_1, V_2, \cdots, V_n$ of the subsets of respective component spaces (V_i is a fixed subset of the i^{th} component space),

$$\begin{cases}H=\left\{f(z)=\prod_{i=1}^{n} f_i(x_i);\ P_{f_i}(V_i)=p\right\},\\ A=\left\{g(z)=\prod_{i=1}^{n} g_i(x_i);\ P_{g_i}(V_i)=p_1\right\},\end{cases} \tag{8.6}$$

where p and p_1 are fixed constants which are positive and less than unity, will be solvable by an optimum test, which will be derived by the similar procedure as the above example.

Example 8.2. As an example of the construction of a multinomial probability scheme, let us consider the testing problem $(H, A)_\alpha$ such that, under the situation of the preceding example,

$$\begin{cases}H: & \xi_p(f)=\xi_0 \text{ and } \eta_p(f)=\eta_0\,,\\ A: & \xi_{p'}(f)=\xi_0 \text{ and } \eta_{p'}(f)=\eta_0\,,\end{cases} \tag{8.7}$$

where $\xi_p(f)=\inf\{\xi; P_f((-\infty, \xi])=p\}$ is the p-quantile as before, while we define $\eta_p(f)=\sup\{\eta; P_f([\eta, \infty))=p$, and, ξ_0 and η_0 are the fixed constants such that $\xi_0<\eta_0$. Here, of course, we shall assume that p and p' are also fixed and $0<p'<p<1/2$.

Each component space R_0 (the real line) is partitioned into three subsets, $V_1=(-\infty, \xi_0)$, $V_2=(\eta_0, \infty)$ and $V_0=(\xi_0, \eta_0)$. We can construct an n-decomposable multinomial probability scheme, $\{W_1, W_2, \cdots, W_s\}$ with probabilities $\{\theta_1, \theta_2, \cdots, \theta_s\}$ such that, if W_j is composed of u V_1's, v V_2's and $(n\text{-}u\text{-}v)$ V_0's, then $\theta_j = p^u(1-2p)^{n-u-v}p^v = p^{u+v}(1-2p)^{n-u-v}, j=1, 2, \cdots, s$. In this case $s=3^n$.

It will easily be noticed that this scheme is essentially equivalent to a binomial probability scheme, hence, analogously to the preceding example, it can be seen that a uniformly most powerful test for our testing problem (8.7) is given by

$$\varphi_0(z) = \begin{cases} 1, & \text{if } T(z) < c\ , \\ a, & \text{if } T(z) = c\ , \\ 0, & \text{otherwise}, \end{cases} \tag{8.8}$$

where the statistic $T(Z)$ is defined by $T(Z) = \sum_{i=1}^{n} T_1(X_i) + \sum_{i=1}^{n} T_2(X_i)$ for the defining functions $T_1(x)$ and $T_2(x)$ of the sets V_1 and V_2 respectively.

For the testing problems of this type, for which the tail probabilities, specified by hypotheses or by the alternative, are not equal, the determination of an optimum rejection regions is not easy, because the order relation of γ_j's (ratios $P_H(W_j)/P_A(W_j)$'s of a probability scheme to be used) is not clear.

Example 8.3. We shall be concerned, in the present example, with a problem of testing hypotheses of invariance under all permutations of the components of sample point.

Let R be an n-dimensional euclidean space with element $z = (x_1, x_2, \cdots, x_n)$ and let m be the Euclid-Lebesgue measure on it. Let, further, $Z=(X_1, X_2, \cdots, X_n)$ be a random variable of dimensions n, being distributed according to an n-dimensional probability distribution absolutely continuous with respect to the measure m, with density function $f(z)$. Hence, we can assume, without any loss of generality, that the sample points are all such that $x_i \neq x_j\ (i \neq j)$ for all $i, j=1,2, \cdots, n$. For each sample point $z=(x_1, x_2, \cdots, x_n)$, let $z^i=(x_{i_1}, x_{i_2}, \cdots, x_{i_n})$, $i=1, 2, \cdots, N\ (N=n!)$, be the sample points obtained by the N permutations of the components $x_1, x_2, \cdots, x_n$ of z.

Consider the testing problem of the hypotheses

$$H_0 : f(z^1) = f(z^2) = \cdots = f(z^N), \text{ for all } z\ , \tag{8.9}$$

against a simple alternative g_0, basing upon a single observation on the random vector $Z=(X_1, X_2, \cdots, X_n)$. Later, some additional assumptions will be imposed on the alternative g_0.

Rearranging z^i's in such a way that

$$g_0(z^1) \geqq g_0(z^2) \geqq \cdots \geqq g_0(z^N) , \tag{8.10}$$

for each sample point z, we define the following N mutually disjoint subsets of the sample space R.

$$W_i = \{z^i ; z \in R\} , \quad (i=1, 2, \cdots, N) . \tag{8.11}$$

We assume that these subsets are all measurable. Then they constitute an (m)-partition of the sample space R. Let $T_i(z)$ be the defining function of W_i, $i=1, 2, \cdots, N$. Then, putting $E_{H_0}[T_i(Z)]=\theta$ and $E_{g_0}[T_i(Z)]=\theta_i$, $i=1, 2, \cdots, N$, we have

$$\theta = 1/N , \quad \theta_1 \geqq \theta_2 \geqq \theta_3 \geqq \cdots \geqq \theta_N , \tag{8.12}$$

where we assume that $\theta_N > 0$. As will be noted later, this assumption is not necessarily needed. It is also assumed that θ_i's are not all equal.

Thus, we have an equi-probability scheme $\{W_1, W_2, \cdots, W_N\}$. In the first place, we shall consider the problem of testing hypotheses $(H, A)_\alpha$ where

$$\begin{cases} H = \{f(z);\ E_f[T_i(Z)] = \theta , & i=1, 2, \cdots, N\} , \\ A = \{g(z);\ E_g[T_i(Z)] = \theta_i , & i=1, 2, \cdots, N\} . \end{cases} \tag{8.13}$$

Clearly, $H_0 \subset H$ and $g_0 \in A$.

From Lemma 7.1, Theorem 7.1 and Lemma 7.3, it will be seen that, for any member $g(z)$ in the class A, the testing problem $(H, g)_\alpha$ possesses a most powerful test such as

$$\varphi_0(z) = \begin{cases} 1, & \text{if } \sum_{i=1}^{[N\alpha]} T_i(z) = 1 , \\ a_0, & \text{if } T_{[N\alpha]+1}(z) = 1 , \\ 0, & \text{otherwise.} \end{cases} \tag{8.14}$$

By Corollary 3.1 it is concluded that the test (8.14) becomes uniformly most powerful for the testing problem $(H, A)_\alpha$.

Obviously the class H_0 defined by (8.9) is a proper subset of the class H given by (8.13), hence, in general, the nearest distribution for the distance problem $(H : g_0)$ does not necessarily belong to the class H_0. But, if we consider a function such as

$$h_0(z) = g_0(z^{[N\alpha]+1}), \text{ for all } z , \tag{8.15}$$

then this is non-negative and integrable(m) over R, hence, multiplying a suitable constant λ, we have a probability density function $h(z)=\lambda h_0(z)$. Clearly $h(z)$ belongs to the class H_0 and satisfies the condition of Lemma 7.4, from which it follows that the test φ_0 given by (8.14) is most powerful for the original problem $(H_0, g_0)_\alpha$.

It is noted that the condition that $\theta_N>0$ can be removed. Suppose that $\theta_s>0$ and $\theta_{s+1}=\theta_{s+2}=\cdots=\theta_N=0$ in (8.12). Examining our method precisely, we can see that a most powerful test for the testing problem $(H, g_0)_\alpha$ is given by the one defined by (8.14). Therefore, if $s\geqq[N\alpha]+1$, then, as was seen above, optimulity of the test is guaranteed by Lemma 7.4, and if $s<[N\alpha]+1$, then the power of this test is unity, hence it becomes most powerful for the testing problem $(H_0, g_0)_\alpha$.

Testing of an invariance hypotheses has been treated by E. L. Lehmann and C. Stein [11] under more general situations, and the problem considered in the present example is a special case of theirs. It will be noticed that our method is applicable to the testing problems of invariance hypotheses under fairly general situations. Lehmann and Stein have derived the most powerful test given by (8.14), basing upon investigation of a test of "structure $S(\alpha)$", and our method of derivation considered in the present example may be regarded as an actual procedure of realizing their idea.

Example 8.4. (D. A. S. Fraser [8, Chap. 5, Sec. 2, Example 2.1])

We shall consider the following two-sample problem. Let $X_1, X_2, \cdots, X_{n_1}$ be independently and identically distributed according to a common distribution on the real line with a probability density function $f_1(x)$ with respect to the Euclid-Lebesgue measure. Similarly, let $X_{n_1+1}, X_{n_1+2}, \cdots, X_{n_1+n_2}$ be independently and identically distributed according to a common distribution on the real line with density function $f_2(x)$. Assume that X_i's and X_{n_1+j}'s are mutually independent.

Consider the testing problem $(H_0, g_0)_\alpha$ such that

$$(8.16)\qquad \begin{cases} H_0: & f_1(x)=f_2(x)\ (=f^0(x),\ \text{unknown}),\\ g_0: & f_1(x)=g_1^0(x)\ \text{and}\ f_2(x)=g_2^0(x), \end{cases}$$

where $g_1^0(x)$ and $g_2^0(x)$ are certain given density functions.

Let $n_1+n_2=n$ and let R be the n-dimensional euclidean whole space. Let $f_0(z)$ and $g_0(z)$ be the probability density functions of the joint distributions of $X_1, X_2, \cdots, X_{n_1}, X_{n_1+1}, X_{n_1+2}, \cdots, X_n$, under the hypotheses and the alternative, respectively, that is, $f_0(z)=\prod_{i=1}^{n} f^0(x_i)$ and $g_0(z)=\prod_{i=1}^{n_1} g_1^0(x_i) \prod_{j=1}^{n_2} g_2^0(x_{n_1+j})$.

As was seen in the preceding example, we can construct an equi-probability scheme with the same definition as that given by (8.10) and (8.11), i.e., a partition $\{W_1, W_2, \cdots, W_N\}$, where $N=n!$, with equal probabilities $\{1/N, 1/N, \cdots, 1/N\}$. Let $T_k(Z)$ be the characteristic variable associated with W_k, $k=1, 2, \cdots, N$.

In the first place, we shall consider the following problem of testing hypotheses,

$$\begin{cases} H=\{f(z);\ \ E_f[T_k(Z)]=1/N,\ k=1, 2, \cdots, N\}\ , \\ A=\{g(z);\ \ E_g[T_k(Z)]=\theta_k,\ \ \ k=1, 2, \cdots, N\} \cap Q_1\ , \end{cases} \tag{8.17}$$

where $\theta_k=E_{g_0}[T_k(Z)]$, $k=1, 2, \cdots, N$, which are so arranged that $\theta_1 \geqq \theta_2 \geqq \cdots \geqq \theta_N$. It follows from Theorem 7.2 and Lemma 7.3 that the test given by

$$\varphi_0(z)=\begin{cases} 1, & \text{if } \sum_{k=1}^{[N\alpha]} T_k(z)=1\ , \\ a, & \text{if } T_{[N\alpha]+1}(z)=1\ , \\ 0, & \text{otherwise,} \end{cases} \tag{8.18}$$

is a most powerful test for the testing problem $(H, g)_\alpha$, where $g(z)$ is any member of the class A. From Corollary 3.1, it is also concluded that the test becomes uniformly most powerful for the problem $(H, A)_\alpha$.

In this case, however, the nearest distribution for the distance problem $(H:g_0)$ is not necessarily contained in the class H_0. (From (8.17) it is clear that $H_0 \subset H$ and $g_0 \in A$.) Hence, in general, the test (8.18) is not a most powerful test for the testing problem $(H_0, g_0)_\alpha$ with which we are concerned.

In particular, if we take a special alternative g_0 which will be defined below in (8.19), then, as was stated by D. A. S. Fraser [8], the test given by (8.18) becomes most powerful and similar for the testing problem $(H_0, g_0)_\alpha$, by virtue of the sufficiency and the bounded completeness of the order statistic for the class H_0.

We shall state the procedure of realization of the above test, taking a special alternative such as

(8.19)
$$g_1^0(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right],$$
$$g_2^0(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(x-\mu+\delta)^2\right],$$

where $\sigma(>0)$, μ and $\delta(>0)$ are certain given constants. The equi-probability scheme $\{W_1, W_2, \cdots, W_N\}$ was constructed in a following manner,

(8.20) $$W_k=\{z^k;\ g_0(z^1)\geqq g_0(z^2)\geqq\cdots\geqq g_0(z^k)\geqq\cdots\geqq g_0(z^N),\ z\in R\},$$

$k=1, 2, \cdots, N$, where z^k's are the points obtained by the N permutations of the components of z.

Let $z^k=(x_{k_1}, x_{k_2}, \cdots, x_{k_n})$ for each $z=(x_1, x_2, \cdots, x_n)$, $k=1, 2, \cdots, N$. Then

(8.21) $$g_0(z^k)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n$$
$$\cdot\exp\left\{-\frac{1}{2}\sigma^2\left[\sum_{i=1}^n x_i^2-2\mu\sum_{i=1}^n x_i+n_1\mu^2+n_2(\mu-\delta)^2+\frac{\delta}{\sigma^2}\sum_{i=n_1+1}^n x_{k_i}\right]\right\}.$$

Thus, corresponding to the partition $\{W_1, W_2, \cdots, W_N\}$, we have

(8.22) $$\sum_{i=1}^{n_1} x_{1_i}\geqq\sum_{i=1}^{n_1} x_{2_i}\geqq\cdots\cdots\geqq\sum_{i=1}^{n_1} x_{N_i}.$$

Therefore, if an actual sample point $z=(x_1, \cdots, x_{n_1}, x_{n_1}+1, \cdots, x_n)$ is obtained, then we must calculate the first $s\,(=[N\alpha]+1)$ members of the above inequality, together with $\sum_{i=1}^{n_1} x_i$, and if the inequality

(8.23) $$\sum_{i=1}^{n_1} x_i\geqq\sum_{i=1}^{n_1} x_{s_i},$$

holds, then we reject the hypotheses H_0 with probability one or a, according as the strict inequality or the equality holds.

It is interesting to examine the properties of our method, i.e., for example, whether the tests derived by the unified procedure, developed in the last three sections, for the problems of testing non-parametric hypotheses are the best ones obtainable.

The author expresses his hearty thanks to Prof. Junjiro Ogawa for his scrutiny of the present paper and his kind encouragement.

DEPT. MATH., COLLEGE OF SCI. AND ENG., NIHON UNIV.

References

[9] S. Ikeda, "A remark on the convergence of Kullback-Leibler's mean information," *Ann. Inst. Stat. Math.*, Vol. 12 (1960), pp. 81-88.

[10] D. A. S. Fraser, "Non-parametric theory, scale and location parameters," *Can. Jour. Math.*, Vol. 6 (1953), pp. 46–68.

[11] E. L. Lehmann and C. Stein, "On the theory of some non-parametric hypotheses," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 28–45.