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

Part I

By SADAO IKEDA

(Received March 23, 1961)

1. Introduction

Consider the class of all probability distributions which are absolutely continuous with respect to a σ-finite measure m on a σ-finite measure space (R, m), and denote by $f(z), g(z), \cdots$ their generalized probability density functions. Let H be a certain given class of probability density functions, and denote by $(H, g)_\alpha$ the problem of testing the hypotheses H that the distribution function under consideration has a member of H as its density function, against a simple alternative g at the significance level α.

Further let $\Lambda(H)$ be the class of all a priori probability distributions over a σ-field of the subsets of class H including all one point subsets and $H^{\Lambda(H)}$ the class of all probability density functions of the form

$$f_\lambda(z) = \int_H f(z) d\lambda(f) , \tag{1.1}$$

for f in H and λ in $\Lambda(H)$.

By the Neyman-Pearson fundamental lemma, there exists a most powerful test φ_λ of exact size α for the testing problem $(f_\lambda, g)_\alpha$ for each λ belonging to $\Lambda(H)$. If there exists an a priori probability distribution λ_0 in $\Lambda(H)$ such that the power of the most powerful test φ_{λ_0} of the testing problem $(f_{\lambda_0}, g)_\alpha$ is the smallest, i.e.,

$$E_g[\varphi_{\lambda_0}(Z)] \leqq E_g[\varphi_\lambda(Z)] , \quad \text{for all } \lambda \text{ in } \Lambda(H) , \tag{1.2}$$

then λ_0 is called a *least favorable distribution* over H, and the cor responding probability distribution with density function $f_{\lambda_0}(z)$ is called a *closest distribution* for the testing problem $(H, g)_\alpha$.

For most of the problems of testing hypotheses, we can easily obtain some idea about the least favorable distributions. For the problems of testing hypotheses where it is difficult to find out the least favorable distributions, E. L. Lehmann [1] suggested the following procedures: (i) Try to find a step function λ such that $E_f[\varphi_\lambda(Z)] = \alpha$ at all steps, or (ii) try to find an a priori distribution λ of the continuous type such that $E_f[\varphi_\lambda(Z)]=\alpha$ whenever the density of λ is positive. However, in problems of testing hypotheses, parametric or non-parametric, the closest distribution is often obtained as the nearest distribution defined by S. Kullback [6, Chap. 3, Sec. 2], which attains the minimum of a directed distance called the discrimination information measure, between two sets of probability density functions, the hypothesis and alternative. We, therefore, propose the following steps: first as certain the existence of the nearest distribution of a distance problem which is suitably constructed in connection with the testing problem under investigation, and if it exists, then examine whether the distribution is closest or not. To discuss this process is the purpose of the present paper.

In Section 2, the concept of the nearest distribution and its existence condition will be given. The basic idea and procedure for deriving a most powerful test will be stated in Section 3, and some examples of applications to the typical problems of testing hypotheses concerning the normal, binomial and Poisson populations are given in Section 4.

2. Nearest distribution in a distance problem

Let Z be a random variable defined on a σ-finite measure space (R, m), whose distribution is absolutely continuous with respect to the measure m, and let $T(Z)$ be a vector of s real statistics, $T_1(Z)$, $T_2(Z), \cdots, T_s(Z)$. For any given vector θ of s real numbers, $\theta_1, \theta_2, \cdots, \theta_s$, let $K(T, \theta)$ designate the class of all generalized probability density functions $f(z)$'s such that $E_f[T(Z)]=\theta$. Further in the present section, it is assumed that the carriers of the members of $K(T, \theta)$ and of $g(z)$ to be considered below are all identical. Thus we can assume, without any loss of generality, that they coincide with the whole space R.

The class of all probability density functions of the form

$$f_\tau(z)=g(z)e^{\tau T(z)}/M(\tau)\ , \tag{2.1}$$

were $\tau=(\tau_1, \tau_2, \cdots, \tau_s)$ is an s-dimensional real vector, $\tau T(z)=\sum_{i=1}^s \tau_i T_i(z)$ and $M(\tau)=E_g[e^{\tau T(Z)}]$, provided that the latter exists, will be called the *exponential family of probability density functions generated by g and T.* It is evident that the carriers of the members of this family are all identical with R(a.e. m), and that the identity $\tau T(z)=$const. (a.e. m)

implies $f_\tau(z)=g(z)$ (a.e. m). This class, $C(g, T)$, forms a homogeneous set of probability measures, that is, the members of $C(g, T)$ are absolutely continuous with respect to each other, and it is seen that the statistic $T(Z)$ is a sufficient statistic for this class.

For any pair $(f(z), g(z))$ of two probability density functions with the same carrier, the Kullback-Leibler mean information for discrimination defined by

$$I(f: g)=\int_R f(z)\log\frac{f(z)}{g(z)}dm(z)\ , \tag{2.2}$$

satisfies the inequality $I(f: g)\geqq 0$ with equality when and only when $f(z)=g(z)$ (a.e. m), and it may be regarded as a directed distance between f and g. Let H be a certain class of probability density functions whose carriers are all identical with R. A natural definition of the directed distance between H and a probability density function $g(z)$ may be given by $I(H: g)=\inf\{I(f: g); f\in H\}$. If there exists a member $f_0(z)$ of the class H such that $I(f_0: g)=I(H: g)$, then $I(f_0: g)$ is called a *minimum discrimination information*, and we shall say that $f_0(z)$ is a *nearest distribution* for the *distance problem* $(H: g)$.

As for the existence of the nearest distribution and its uniqueness for the distance problem $(K(T, \theta): g)$, the following theorem is a multi-statistics extension of the result by S. Kullback [6, Chap. 3, Theorem 2.1].

THEOREM 2.1. *Let D be the set of τ's corresponding to $C(g, T)$, i.e., $D=\{\tau; f_\tau(z)\in C(g, T)\}$, and suppose that* (i) *D is an open domain (including the origin) in the s-dimensional euclidean space,* (ii) $\tau T(z)=0$ (a.e. m) *if and only if $\tau=0$ for τ's belonging to D, and* (iii) $M(\tau)$ *is differentiable partially up to the second order with respect to the components of τ under the sign of the integration.*

Then, a necessary and sufficient condition in order that the nearest distribution exists for the distance problem $(K(T, \theta): g)$, is that the system of equations

$$\frac{\partial}{\partial\tau_i}\log M(\tau)=\theta_i\ ,\quad (i=1, 2, \cdots, s)\ , \tag{2.3}$$

is solvable in τ.

If the nearest distribution exists, then it is unique amd is a member of the family $C(g, T)$ with the form

(2.4) $$f_0(z)=g(z)e^{\tau^0 T(z)}/M(\tau^0)\ ,$$

where $\tau^0=(\tau_1^0, \tau_2^0, \cdots, \tau_s^0)$ *is the solution of the equation* (2.3). *The minimum discrimination information is given by*

(2.5) $$I(f_0\colon g)=\tau^0\theta-\log M(\tau^0)\ ,$$

where, of course, $\tau^0\theta=\sum_{i=1}^{s}\tau_i^0\theta_i$.

PROOF. Let $\tau=(\tau_1, \tau_2, \cdots, \tau_s)$ and λ (scalar) be the Lagrange multipliers, and consider the expression

(2.6) $$I(f\colon g)-E_f[\tau T(Z)]-\lambda E_f[1]=\int_R L(f(z), \tau, \lambda)dm\ ,$$

where the integrand of the right-hand side is of the form

(2.7) $$L(f(z), \tau, \lambda)=f(z)(\log\frac{f(z)}{g(z)}-\tau T(z)-\lambda)\ .$$

In order to minimize the mean information $I(f\colon g)$ subject to the condition that f belongs to the class $K(T, \theta)$, the expression (2.6) must be minimized under the condition that the function $f(z)$ is positive (a.e. m) on R. Here, we can assume, without any loss of generality, that the density function $g(z)$ is positive everywhere on R. In the general case where the σ-finite measure space (R, m) is arbitrary, however, it is not clear whether the usual method of minimizing the expression (2.6), regarding it as a functional of f, by calculating the variations is applicable.

Fortunately, it is possible, in our present case, to minimize the integrand (2.7) for each point z of R, as will be seen in the following. Since the real-valued function $h(x)=x(\log x/u-v)$ defined for $0<x<\infty$, where $u(>0)$ and v are any given constants, is minimized when and only when $x=ue^{v-1}$, the integrand (2.7) is minimized, for each point z of R, when and only when $f(z)$ is a function with the form

(2.8) $$f(z)=g(z)e^{\tau T(z)+\lambda-1}\ .$$

Clearly, this function is positive everywhere on R, integrable with respect to the measure m on R when τ belongs to the domain D, and from the condition (iii) of the theorem, $E_f[T_i(Z)]$ exists for every $i(=1, 2, \cdots, s)$ provided that τ belongs to D. Since m is a non-negative measure on R, it is obvious that the function $f(z)$ given by (2.8) minimizes the expression (2.6) for any τ and λ subject to the restriction mentioned above.

The condition that the function $f(z)$ given by (2.8) is a member of the class $K(T, \theta)$ implies that the function $f(z)$ is of the form

$$f_0(z)=g(z)e^{\tau^0 T(z)}/M(\tau^0) , \tag{2.9}$$

where, $\tau^0=(\tau_1^0, \tau_2^0, \cdots, \tau_s^0)$ is a real vector satisfying the system of equations (2.3). This proves the necessity of the solvability of the equation (2.3) for the existence of a nearest distribution for our distance problem $(K(T, \theta)\colon g)$.

Conversely, if the equation (2.3) has a solution τ^0 which belongs to the domain D, then the generalized probability density function $f_0(z)$ given by (2.9) is a member of the class $K(T, \theta)$. Then, as was seen above, $f_0(z)$ minimizes the mean information $I(f\colon g)$, and hence, it is the nearest distribution. This proves the sufficiency.

Suppose that a nearest distribution exists for the distance problem $(K(T, \theta)\colon g)$. Then, it is clear, from (2.9), that the nearest distribution belongs to the exponential family $C(g, T)$, and the minimum discrimination information is given by (2.5).

In order to show the uniqueness of the nearest distribution when it exists, it is sufficient to show the uniqueness of the solution of the system of equations (2.3) provided that it is solvable. For this, consider the quadratic form

$$U(x)=\sum_{i,j=1}^{s}\left(\frac{\partial^2}{\partial\tau_i\partial\tau_j}\log M(\tau)\right)x_i x_j , \tag{2.10}$$

where $x_1, x_2, \cdots, x_s$ are the real variables. Defining $M_i(\tau)=(\partial/\partial\tau_i)M(\tau)$ and $M_{ij}(\tau)=(\partial^2/\partial\tau_i\partial\tau_j)M(\tau)$, for the notational simplicity, one can easily find that

$$\begin{aligned} U(x)&=\frac{1}{M^2}\sum_{i,j=1}^{s}(M_{ij}M-M_iM_j)x_ix_j \\ &=\frac{1}{M^2}(E_g[(xT)^2e^{\tau T}]E_g[e^{\tau T}]-E_g^2[xTe^{\tau T}]) , \end{aligned} \tag{2.11}$$

where $xT=\sum_{i=1}^{s}x_iT_i$. Hence, by the Schwarz inequality, it follows that the quadratic form $U(x)$ is strictly positive for non-zero $x=(x_1, x_2, \cdots, x_s)$ unless $xT(z)=\text{const.}$ (a.e. m) on R, which means that the solution of the equation (2.3) is unique. The proof of the theorem is now complete.

In connection with this theorem, for the case when $s=1$, the properties of $M(\tau)$ and the solution of the equation (2.3) were discussed by S. Kullback in his book [6, Chap. 3, Sec. 4], some of which are listed without proof in the following lemma for the later use.

LEMMA 2.1. *Under the conditions of Theorem* 2.1 *for the case when* $s=1$, *let* $\tau(\theta)$ *be the solution of the equation* $(d/d\tau)\log M(\tau)=\theta$. *Then*

(i) $(d/d\tau)\log M(\tau)$ *is a monotone increasing function of* τ,

(ii) $\tau(\theta)$ *is a monotone increasing function of* θ, *and*

(iii) $\tau(\theta)<0, =0$ *and* >0 *corresponding to the cases when* $E_g[T(Z)] >\theta, =\theta$ *and* $<\theta$, *respectively.*

Now, under the situation of Theorem 2.1, suppose that the distance problem $(K(T, \theta): g)$ is solvable on a certain open and convex domain U of θ in the s-dimensional euclidean space. Let $K(T, \theta^0)$ be the set-theoretical sum of $K(T, \theta)$ for all θ such that $\theta_i \leqq \theta_i^0$, $i=1, 2, \cdots, s$, and $\theta \in U$, where θ^0, a vector whose components are $\theta_1^0, \theta_2^0, \cdots, \theta_s^0$, is fixed in the domain U. Then, as for the distance problem $(K(T, \theta^0): g)$, the following result can be obtained.

LEMMA 2.2. *If the statistics* $T_1(Z), T_2(Z), \cdots, T_s(Z)$, *which are the components of the vector of statistics* $T(Z)$, *are mutually independent under the distribution of* Z *whose probability density function is* $g(z)$, *and moreover, if* $E_g[T_i(Z)]>\theta_i^0$, $i=1, 2, \cdots, s$, *then the nearest distribution for the distance problem* $(K(T, \theta^0): g)$ *belongs to the class* $K(T, \theta^0)$.

PROOF. From Theorem 2.1, the minimum discrimination information for the distance problem $(K(T, \theta): g)$ is given by

$$I(f_{\tau(\theta)}: g)=\tau(\theta)\theta-\log M(\tau(\theta)), \tag{2.12}$$

where $\tau(\theta)=(\tau_1(\theta), \tau_2(\theta), \cdots, \tau_s(\theta))$ is the solution of the system of equations (2.3). Since the independence of the statistics $T_i(Z)$'s under g implies that

$$M(\tau)=\prod_{i=1}^{s} M_i(\tau_i) \tag{2.13}$$

with definitions $M_i(\tau_i)=E_g[e^{\tau_i T_i(Z)}]$, $i=1, 2, \cdots, s$, the equation (2.3) and the minimum discrimination information (2.12) are reduced, respectively, to

$$\frac{d}{d\tau_i}\log M_i(\tau_i)=\theta_i\ ,\quad i=1,2,\cdots,s\ , \tag{2.14}$$

and

$$I(f_{\tau(\theta)}:g)=\sum_{i=1}^{s}(\tau_i(\theta_i)\theta_i-\log M_i(\tau_i(\theta)))\ , \tag{2.15}$$

where $\tau(\theta)=(\tau_1(\theta_1),\tau_2(\theta_2),\cdots,\tau_s(\theta_s))$ is the solution of the equation (2.14).

Let θ' and θ'' be any two points fixed in the domain U such that $\theta'_i\leqq\theta''_i$, $i=1,2,\cdots,s$, (not all equal), and put $\theta(p)=(1-p)\theta'+p\theta''$, where p is a real number such that $0\leqq p\leqq 1$. From the convexity of the domain U, it is obvious that $\theta(p)$ is again a point of U for any p $(0\leqq p\leqq 1)$, and hence the distance problem $(K(T,\theta(p)):g)$ is solvable for $0\leqq p\leqq 1$. Then, from (2.15) we have, after some calculations,

$$\frac{d}{dp}I(f_{\tau(\theta(p))}:g)=\sum_{i=1}^{s}\tau_i(\theta_i(p))(\theta''_i-\theta'_i)\ . \tag{2.16}$$

Since $\tau_i(\theta_i(p))$'s are the solutions of the equations of (2.14) respectively, taking $\theta_i(p)$'s instead of θ_i's on the righthand sides of the equations in)2.14), it follows, from (iii) of Lemma 2.1, that $\tau_i(\theta_i(p))<0$, and hence we have $(d/dp)\ I(f_{\tau(\theta(p))}:g)<0$. This means that the minimum discrimination information $I(f_{\tau(\theta)}:\mathrm{g})$, which is a function of θ, is minimized when $\theta=\theta^0$ under the restriction that $\theta_i\leqq\theta^0_i$, $i=1,2,\cdots,s$. Thus the proof of our lemma is complete.

It is not clear whether the result of the above lemma remains true or not, when the assumption of the independence for the statistics is removed.

The condition (iii) concerning the regularity of $M(\tau)$ in Theorem 2.1 requires the existence of $E_g[T_iT_je^{\tau T}]$ for all $i,j=1,2,\cdots,s$, and for all τ in D. A sufficient condition in order that the regularity assumption (iii) in the theorem is fulfilled together with other two assumptions, (i) and (ii), will be given in the following lemma. The proof is easy and is omitted.

LEMMA 2.3. *If the statistics* $T_1(Z), T_2(Z),\cdots,T_s(Z)$, *the components of the vector of statistics* $T(Z)$, *are all bounded with probability one under the distribution of* Z *whose probability density function is* $g(z)$, *then* $M(\tau)$ *satisfies the required assumptions* (i) *to* (iii) *in Theorem* 2.1, *where the domain* D *may be taken to be the s-*

dimensional euclidean whole space.

In the applications of these results to the statistical inference, it will be convenient to specialize the structure of the space R. Suppose that the measure space (R, m) is the n-product of a σ-finite measure space (R_0, m_0), that is, R is the cartesian n-product of a certain (abstract) component space R_0, on which a σ-field is defined as usual from those of component spaces, and m is a σ-finite measure on R, defined over the above σ-field such that $m(A_1\times A_2\times\cdots\times A_n)=m_0(A_1)\times m_0(A_2)\times\cdots\times m_0(A_n)$ whenever A_i's are the measurable subsets of R_0. Under this situation, the class of all generalized probability density functions of the distributions of the random vector $Z=(X_1, X_2, \cdots, X_n)$ defined on the measure space (R, m) is denoted by Q_0, whereas Q_1 and Q_2 designate the classes of all generalized probability density functions of the forms $f(z)=\prod_{i=1}^n f_i(x_i)$ and $f(z)=\prod_{i=1}^n f(x_i)$ respectively, where $f(x)$'s are the generalized probability density functions of the component variables X's of the random vector Z, which are defined on the component spaces. Clearly, $Q_2\subset Q_1\subset Q_0$.

The following corollaries are the direct consequences of Theorem 2.1, and the proofs are omitted.

COROLLARY 2.1. *Let* $T(Z)=(T_1(X_1), T_2(X_2), \cdots, T_n(X_n))$, *and let* $g(z)=\prod_{i=1}^n g_i(x_i)$ *be a member of the class* Q_1 *defined above. Assume that* $M_i(\tau_i)=E_{g_i}[e^{\tau_i T_i(X_i)}]$ *is differentiable twice in* τ_i *under the sign of integration, for each* i, *and moreover, the equation in* τ_i, $(d/d\tau_i)\log M_i(\tau_i)=\theta_i$, *has a solution* τ_i^0, *for each* i.

Then, the nearest distribution exists for the distance problem $(K(T, \theta): g)$ *as a member of the intersection* $K(T,\theta)\cap Q_1$ *with the form*

$$f_0(z)=\prod_{i=1}^n[g_i(x_i)\exp(\tau_i^0 T_i(x_i))/M_i(\tau_i^0)]\,. \tag{2.17}$$

The minimum discrimination information is given by

$$I(f_0: g)=\sum_{i=1}^n[\tau_i^0\theta_i-\log M_i(\tau_i^0)]\,. \tag{2.18}$$

COROLLARY 2.2. *Let* $T(Z)=(T_0(X_1), T_0(X_2), \cdots, T_0(X_n))$, *and let* $g(z)=\prod_{i=1}^n g_0(x_i)$ *be a member of the class* Q_2. *Assume that* $M_0(\tau)=E_{g_0}[e^{\tau T_0(X)}]$ *is differentiable twice in* τ *(scalar) under the sign of integration,* *and that the equation in* τ, $(d/d\tau)\log M_0(\tau)=\theta_0$ *has the solution* τ_0 *(scalar).*

Then, for the distance problem $(K(T, \theta): g)$ *with* $\theta=(\theta_0, \theta_0, \cdots, \theta_0)$ *there exists the nearest distribution belonging to the class* Q_2 *with the form*

$$f_0(z)=\prod_{i=1}^{n}[g_0(x_i)e^{\tau_0 T_0(x_i)}/M_0(\tau_0)] , \tag{2.19}$$

and the minimum discrimination information is given by

$$I(f_0: g)=n[\tau_0\theta_0-\log M_0(\tau_0)] . \tag{2.20}$$

In general, in order to apply the concept of the information measure for discrimination introduced above, to the theory of testing hypotheses, two different ways may be considered.

The one is, in short, the "nearest distribution consideration", that is to say, the nearest distribution of a distance problem $(H: g)$ appears frequently to be least favorable for the corresponding problem of the testing hypotheses, $(H, g)_\alpha$, and hence, if one wants to find a least favorable distribution for the testing problem $(H, g)_\alpha$, it would be useful to inquire, in the first step, whether the corresponding distance problem $(H: g)$ is solvable or not, examining some exponential families $C(g, T)$'s with different statistics T's.

The other way is the "information statistic consideration", which has been summarized in the book of S. Kullback [6]. An estimate of the minimum discrimination information from a random sample, which is called a minimum discrimination information statistic, may be considered as a measure of directed distance, which is obtainable by a statistician, and he can construct a test procedure basing upon that estimate.

The purpose of the present paper is to work out the former idea in order to derive an optimum (most powerful) test, and the procedure will be described in the following section.

3. Nearest distribution and the derivation of most powerful test

In order to make the situations clear, under which our discussions will be set forth, we shall give, in the first place, the set-theoretical interpretations of the basic result by E. L. Lehmann and C. Stein [4] together with our present method.

Denote by $\mathscr{T}_\alpha(H)$ the class of all test functions for the testing problem $(H, g)_\alpha$, i.e.,

$$(3.1)\quad \mathscr{T}_\alpha(H)=\{\psi;\ 0\leqq\psi(z)\leqq 1(a.\,e.\,m)\ ,\quad E_f[\psi(Z)]\leqq\alpha \text{ for all } f \text{ in } H\}\ .$$

If the assumption concerning the joint measurability of $f(z)$'s for the class H is satisfied, then Fubini's theorem provides us with the identity $\mathscr{T}_\alpha(H)=\mathscr{T}_\alpha(H^{\Lambda(H)})$, and the set $\mathscr{M}_\alpha(H)$ of all the most powerful tests for the testing problem $(H, g)_\alpha$ is identical with the set $\mathscr{M}_\alpha(H^{\Lambda(H)})$ for the testing problem $(H^{\Lambda(H)}, g)_\alpha$. Therefore, we shall consider the testing problem $(H^{\Lambda(H)}, g)_\alpha$ instead of the original problem $(H, g)_\alpha$ itself.

For each f_λ in $H^{\Lambda(H)}$, there exists a most powerful test φ_λ of the testing problem $(f_\lambda, g)_\alpha$, and we designate the set of all these φ_λ's corresponding to the class $H^{\Lambda(H)}$ by $\mathscr{F}_\alpha(H^{\Lambda(H)})$. Let, further, H^0 be the class of all closest distributions for the testing problem $(H^{\Lambda(H)}, g)_\alpha$, and denote by $\mathscr{F}_\alpha(H^0)$ the class of most powerful tests φ_λ's of the testing problems $(f_\lambda, g)_\alpha$'s for all f_λ belonging to H^0.

The result by Lehmann-Stein [4] states the inclusion relation

$$(3.2)\qquad \mathscr{F}_\alpha(H^{\Lambda(H)})\cap\mathscr{T}_\alpha(H^{\Lambda(H)})\subset\mathscr{F}_\alpha(H^0)\cap\mathscr{M}_\alpha(H)\ .$$

Therefore, in order to derive the most powerful test, it will be sufficient to find a member f_λ of $H^{\Lambda(H)}$ such that the most powerful test for the testing problem $(f_\lambda, g)_\alpha$ satisfies the size condition for the original problem $(H, g)_\alpha$ or equivalently for $(H^{\Lambda(H)}, g)_\alpha$. In other words, the first class of the left-hand member of (3.2) would be a class of reference, in which a most powerful test of $(H, g)_\alpha$ will be found.

On the contrary, in our present method as will be seen below, a class $\mathscr{F}_\alpha(H^{\Lambda(H)}\cap C(g, T)\cap K(T, \theta^0))$ consisting of at most one member, considering a suitable statistic T, will be examined whether its element is eligible for a member of $\mathscr{T}_\alpha(H)$.

Now, our procedure proceeds as follows: Let $T(Z)=(T_1(Z), T_2(Z), \cdots, T_s(Z))$ be a vector of s real statistics, and let $\{L(T)\}$ be a partition of the class H such that the statistic $T(Z)$ is distributed identically under all members of each subclass $L(T)$. Then, it is clear that the distributions of $T(Z)$ under the members of the class $L(T)^{\Lambda(L(T))}$ are identical with each other, and the family of all these classes, $\{L(T)^{\Lambda(L(T))}\}$, constitutes a disjoint family of subclasses of the class $H^{\Lambda(H)}$. Furthermore, let $E_{L(T)}[T(Z)]=\theta_L$.

Hereafter throughout the present paper, we shall assume that the regularity for $M(\tau)$ described in Theorem 2.1 is always satisfied. The nearest distribution for the distance problem $(K(T, \theta_L): g)$ will be given, if it exists at all, by

$$f_{\tau_L}(z)=g(z)\exp(\tau_L T(z))/M(\tau_L)\ , \tag{3.3}$$

where τ_L is a real vector, as was seen in Theorem 2.1. Let

$$W(L, T)=\{z; \tau_L T(z)<c_L\}\ , \tag{3.4}$$

$$V(L, T)=\{z; \tau_L T(z)=c_L\}\ , \tag{3.5}$$

where the constant c_L is so chosen that, with some $a_L(0<a_L<1)$

$$P_L(W(L, T))+a_L P_L(V(L, T))=\alpha\ . \tag{3.6}$$

Then, the most powerful test of the testing problem $(f_{\tau L}, g)_\alpha$ is given by

$$\varphi_L(z)=\begin{cases}1, & \text{if } z\in W(L, T)\ ,\\ a_L, & \text{if } z\in V(L, T)\ ,\\ 0, & \text{otherwise.}\end{cases} \tag{3.7}$$

For some statistic $T(Z)$, if there exists a subclass $L_0(T)^{A(L_0(T))}$ which contains the nearest distribution $f\tau_{L_0}(z)$ for the distance problem $(H^{A(H)}: g)$, and if the test $\varphi_{L_0}(z)$ satisfies the size condition for the testing problem $(H, g)_\alpha$, then, the test φ_{L_0} is the most powerful test for $(H, g)_\alpha$ and consequently, $f_{\tau_{L_0}}(z)$ becomes to be closest.

These considerations lead us to the following

THEOREM 3.1. *If there exist a statistic* $T(Z)=(T_1(Z), T_2(Z), \cdots, T_s(Z))$ *and a subclass* $L_0(T)$ *satisfying the following three conditions;*

(i) $E_g[T_i(Z)]>\sup_{f\in H} E_f[T_i(Z)]\ ,\quad i=1, 2, \cdots, s\ ,$

(ii) $C(g, T)\cap L_0(T)^{A(L_0(T))}\neq 0\ ,$

(iii) $P_{L_0}(W(L_0, T))+a_{L_0}P_{L_0}(V(L_0, T))\geqq P_L(W(L_0, T))+a_{L_0}P_L(V(L_0, T))$ for all $L(T)\ ,$

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

$$\varphi_{L_0}(z)=\begin{cases}1, & \textit{if } \tau_{L_0}T(z)<c_{L_0}\ ,\\ a_{L_0}, & \textit{if } \tau_{L_0}T(z)=c_{L_0}\ ,\\ 0, & \textit{otherwise}\ .\end{cases} \tag{3.8}$$

The nearest distribution for the distance problem $(H^{A(H)}: g)$ *lies in the class* $L_0(T)^{A(L_0(T))}$, *and it is the closest distribution for the testing problem* $(H, g)_\alpha$.

Moreover, if the distribution of Z *has density function* $g(z)$ *and if the statistics* $T_1(Z), T_2(Z), \cdots, T_s(Z)$, *are mutually independent, then the subclass* $L_0(T)$ *will be determined as a class satisfying the conditions*

$$E_{L_0}[T_i(Z)] = \max_L E_L[T_i(Z)] \ , \quad i=1, 2, \cdots, s \ . \tag{3.9}$$

PROOF. Put $E_{L_0}[T(Z)] = \theta^0$. Then $L_0(T)^{A(L_0(T))} \subset K(T, \theta^0)$. From Theorem 2.1 and the condition (ii) of the present theorem, there exists only one member $f_0(z)$, the nearest distribution for the distance problem $(K(T, \theta^0): g)$, which belongs to the subclass $L_0(T)^{A(L_0(T))}$, with the form

$$f_0(z) = g(z)e^{\tau^0 T(z)}/M(\tau^0) \ , \tag{3.10}$$

where τ^0 is the unique solution of the equation (2.3) for the distance problem $(K(T, \theta^0): g)$.

The most powerful test of the testing problem $(f_0, g)_\alpha$ is given by (3.8), and by the condition (iii) of the present theorem, this test satisfies the size condition for the testing problem $(H, g)_\alpha$. Consequently, the test given by (3.8) becomes most powerful, for the testing problem $(H, g)_\alpha$, which completes the proof of the first statement of the theorem.

The second statement is obvious from Lemma 2.2, and the proof of our theorem is complete.

For practical applications, it will be convenient to specialize the above theorem to the case when $s=1$. Let

$$t_L(\alpha) = \max \left\{ t_L ; \int_{t > t_L} dP_L^T(t) > \alpha \right\} \ , \tag{3.11}$$

where $T(Z)$ is a single real statistic and $P_L^T(t)$ denotes the distribution function of the statistic $T(Z)$ induced by a subclass $L(T)$. If $P_L^T(t)$ is continuous, then $t_L(\alpha)$ becomes the 100α percent-point on the right tail.

Now, the above theorem can be restated in the following

THEOREM 3.2. *If there exist a statistic* $T(Z)$ *and a subclass* $L_0(T)$ *satisfying the following three conditions;*

(i) $E_g[T(Z)] > \sup_{f \in H} E_f[T(Z)]$,

(ii) $C(g, T) \cap L_0(T)^{A(L_0(T))} \neq 0$,

(iii) $t_{L_0}(\alpha) \geqq t_L(\alpha)$, *and* $P_{L_0}\{T(Z)=t_{L_0}(\alpha)\} \geqq P_L\{T(Z)=t_{L_0}(\alpha)\}$ *for all* $L(T)$,

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

$$\varphi_0(z)=\begin{cases}1, & \textit{if } T(z)>t_{L_0}(\alpha) ,\\ a_0, & \textit{if } T(z)=t_{L_0}(\alpha) ,\\ 0, & \textit{otherwise.}\end{cases} \tag{3.12}$$

The nearest distribution for the distance problem $(H^{A(H)}: g)$ *coincides with the closest distribution for the testing problem* $(H, g)_\alpha$, *and lies in the subclass* $L_0(T)^{A(L_0(T))}$ *which satisfies the condition*

$$E_{L_0}[T(Z)]=\max_L E_L[T(Z)] . \tag{3.13}$$

PROOF. From Lemma 2.1, (iii), the test defined by (3.8) in the preceding theorem ($s=1$) is now reduced to that in (3.12). (3.13) follows from Lemma 2.2. The proof of the theorem is thus complete.

In the case when $s=1$, as was seen in the above theorem, the most powerful test of the testing problem $(H, g)_\alpha$ depends only on the sign of the solution of the equation (2.3) for the distance problem $(H^{A(H)}: g)$ (for $s=1$), and hence the following corollary is an immediate consequence.

COROLLARY 3.1. *Let* A *be a class of the generalized probability density functions containing the density function* $g(z)$. *If it holds that*

$$\inf_{g_1 \in A} E_{g_1}[T(Z)] > \sup_{f \in H} E_f[T(Z)] , \tag{3.14}$$

and

$$C(g_1, T) \cap L_0(T)^{A(L_0(T))} \neq 0 , \quad \text{for all } g_1 \in A , \tag{3.15}$$

for the statistic $T(Z)$ *in Theorem* 3.2, *then the test given by* (3.12) *is uniformly most powerful for the testing problem* $(H, A)_\alpha$.

The proof of this corollary is omitted.

In order that the test (3.8) in Theorem 3.1 is realizable, it will, in general, be necessary to know the induced distributions of the statistic $T(Z)$ under this problem, while in the case when $s=1$, the test (3.12) will be realizable if it is known that the statistic $T(Z)$ is a monotone function of a certain statistic whose distributions are completely specified under the problem.

Here, we shall remark that we can restrict ourselves to the statistics which satisfy the first condition of Theorem 3.1 (or, of Theorem 3.2), in order to derive the most powerful test for the problem of testing hypotheses $(H, g)_\alpha$. For this, suppose that there exists a statistic $T'(Z)$ which satisfies the condition

$$\text{(i)}' \qquad E_g[T'_i(Z)] < \inf_{f\in H} E_f[T'_i(Z)]\ , \quad i=1, 2, \cdots, s\ ,$$

together with other conditions (ii) and (iii) in Theorem 3.1. Let $T(Z) = -T'(Z)$. Then it will easily be seen that

$$\{L(T')^{A(L(T'))}\} = \{L(T)^{A(L(T))}\} \text{ and } M'(\tau) = E_g[e^{\tau T'(Z)}] = M(-\tau)\ .$$

The most powerful test based on the statistic $T'(Z)$ may be written, corresponding to (3.8), as

$$\text{(3.16)} \qquad \varphi'_{L_0}(z) = \begin{cases} 1, & \text{if } \tau'_{L_0} T'(z) < c'_{L_0}\ , \\ a'_{L_0}, & \text{if } \tau'_{L_0} T'(z) = c'_{L_0}\ , \\ 0, & \text{otherwise.} \end{cases}$$

But, since $L_0(T') = L_0(T)$, $\tau'_{L_0} = -\tau_{L_0}$ and $c'_{L_0} = c_{L_0}$, the statistic $T(Z)$ satisfies the conditions (i) to (iii) of Theorem 3.1 and the test given by (3.16) becomes the same with that given by (3.8), which justifies our assertion.

Possibilities of the applications of our method developed in the present section would not cover the whole class of problems of testing hypotheses, which posess the most powerful tests derived by means of the least favorable distributions, but most of the problems hitherto considered in literatures, seem to have the simple structures from the viewpoint of the present considerations.

4. Normal, binomial and Poisson populations

In the present section, we shall consider several problems of testing hypotheses concerning some populations having the distributions of exponential type, i.e., the distributions admitting the sufficient statistics. These families of distributions seem to be susceptible to contain the members of the class $C(g, T)$ when the alternative g is also of exponential type.

EXAMPLE 4.1. Let $X_1, X_2, \cdots, X_n$ be independently and identically distributed according to a normal distribution $N(\xi, \sigma^2)$ with unknown mean and variance. Let us consider the problem $(H, g)_\alpha$ of testing hypotheses, where

$$(4.1)\qquad \begin{cases} H:\ \sigma^2=\sigma_0^2\ , \\ g:\ \ \sigma^2=\sigma_1^2\ ,\quad \xi=\xi_1\ ,\quad (\sigma_1^2<\sigma_0^2)\ . \end{cases}$$

Instead of the most powerful similar test based on the sample variance given by J. Neyman and E. S. Pearson [7], E. L. Lehmann [1] introduced the most powerful test based on the statistic

$$(4.2)\qquad T(Z)=-\sum_{i=1}^{n}(X_i-\xi_1)^2\ .$$

From the viewpoint of the present consideration, it will easily be seen that the statistic (4.2) satisfies all the conditions (i) to (iii) of Theorem 3.2 when we take $N(\xi_1, \sigma_0^2)$ as $L_0(T)$.

EXAMPLE 4.2. Under the same situation as the above example, let us examine the problem $(H, g)_\alpha$ where

$$(4.3)\qquad \begin{cases} H:\ \sigma^2\leqq\sigma_0^2 \\ g:\ \ \sigma^2=\sigma_1^2\ ,\quad \xi=\xi_1\ ,\quad (\sigma_1^2>\sigma_0^2)\ . \end{cases}$$

We take the unbiased estimate of the variance as the statistic $T(Z)$, i.e.,

$$(4.4)\qquad T(Z)=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\ .$$

Then it is evident that $T(Z)$ satisfies the conditions (i) and (iii) in Theorem 3.2, when we put $L_0(T)=\{N(\xi, \sigma_0^2):\ -\infty<\xi<\infty\}$.

Condition (ii) in the theorem will be examined as follows; since

$$(4.5)\qquad M(\tau)=\left(1-\frac{2\sigma_1^2}{n-1}\tau\right)^{-\frac{n-1}{2}}\ ,$$

solving the equation

$$(4.6)\qquad \frac{d}{d\tau}\log M(\tau)=\sigma_0^2\ ,$$

we obtain

$$(4.7)\qquad \tau_0=\frac{n-1}{2}\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)\ ,$$

and

$$(4.8)\qquad M(\tau_0)=\left(\frac{\sigma_0}{\sigma_1}\right)^{n-1}\ .$$

The condition (ii) in Theorem 3.2 is now reduced to the solvability of the integral equation in $\lambda(\xi)$

$$(4.9)\qquad \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n\int_{-\infty}^{\infty}\exp\left[-\frac{n}{2\sigma_0^2}(\bar{x}-\xi)^2-\frac{n-1}{2\sigma_0^2}T\right]\lambda(\xi)d\xi$$
$$=\left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^n\exp\left[-\frac{n}{2\sigma_1^2}(\bar{x}-\xi_1)^2-\frac{n-1}{2\sigma_1^2}T+\tau_0T\right]\Big/M(\tau_0)$$

or equivalently

$$(4.9)'\qquad \sqrt{\frac{n}{2\pi\sigma_0^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{n}{2\sigma_0^2}(\bar{x}-\xi)^2\right]\lambda(\xi)d\xi=\sqrt{\frac{n}{2\pi\sigma_1^2}}\exp\left[-\frac{n}{2\sigma_1^2}(\bar{x}-\xi_1)^2\right],$$

subject to the restriction that $\lambda(\xi)$ is a probability density function on the real line. Passing to the moment generating functions of both sides of the above equation, we have

$$(4.10)\qquad \int_{-\infty}^{\infty}\exp\left[\xi t+\frac{\sigma_0^2}{2n}t^2\right]\lambda(\xi)d\xi=\exp\left[\xi_1t+\frac{\sigma_1^2}{2n}t^2\right],$$

or equivalently

$$(4.11)\qquad \int_{-\infty}^{\infty}e^{\xi t}\lambda(\xi)d\xi=\exp\left[\xi_1t+\frac{1}{2n}(\sigma_1^2-\sigma_0^2)t^2\right],$$

for all real t.

The solution of the equation (4.11) will be given by

$$(4.12)\qquad \lambda_0(\xi)=\frac{1}{\sqrt{2\pi}\sigma_\xi}\exp\left[-\frac{1}{2\sigma_\xi^2}(\xi-\xi_1)^2\right],$$

where $\sigma_\xi^2=(\sigma_1^2-\sigma_0^2)/n$, and this is the least favorable distribution. The closest distribution is obtained as the nearest distribution which is the right-hand member of the equation (4.9). We obtain from (4.7) and (4.8)

$$(4.13)\qquad f_{\lambda_0}(z)=\left(\frac{1}{\sqrt{2\pi}}\right)^n\frac{1}{\sigma_1\sigma_0^{n-1}}\exp\left[-\frac{n}{2\sigma_1^2}(\bar{x}-\xi_1)^2-\frac{1}{2\sigma_0^2}\sum_{i=1}^n(x_i-\bar{x})^2\right].$$

The most powerful test for $(H,g)_\alpha$ will be given by the following

$$(4.14)\qquad \varphi_0(z)=\begin{cases}1, & \text{if } \sum_{i=1}^n(x_i-\bar{x})^2\geqq\sigma_0^2\chi_{n-1}^2(\alpha)\,,\\ 0, & \text{otherwise,}\end{cases}$$

where $\chi_{n-1}^2(\alpha)$ is the (upper) 100α percent-point of the chi-square distribution of $n-1$ degrees of freedom. This test will be uniformly most powerful for the testing problem $(H,A)_\alpha$, if A is any subclass of the class of normal distributions $\{N(\xi, \sigma^2); \sigma^2 > \sigma_0^2\}$.

As for the alternative method of derivation of this test, see, for example, E. L. Lehmann [1, Chap. II, Sec. 2], or D. A. S. Fraser [8, Chap. 2, Sec. 3]. Their method requires some intuitive reasoning in order to find a least favorable distribution, but in our present method, as was seen above, it is found out by a formal calculation, if we examine a "existence condition" of the nearest distribution for the corresponding distance problem, i.e., the second condition of Theorem 3.2. Such an improvement of our present method will be recognized in the subsequent examples, too.

EXAMPLE 4.3. We shall consider a two sample problem. Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_{n'}$ be the random samples from the normal populations $N(\mu, 1)$ and $N(\nu, 1)$, respectively, where X_i's and Y_j's are mutually independent. Let us consider the testing problem $(H, A)_\alpha$, where

$$\text{(4.15)} \qquad \begin{cases} H: \ \mu=\nu \ , \\ A: \ \mu<\nu \ . \end{cases}$$

For any fixed μ_1 and ν_1 in the class A of alternatives, let

$$\text{(4.16)} \qquad g(z)=\left(\frac{1}{\sqrt{2\pi}}\right)^{n+n'} \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\mu_1)^2+\sum_{j=1}^{n'}(y_j-\nu_1)^2\right)\right] ,$$

while the class of hypotheses, H, will be represented as

$$\text{(4.17)} \quad H=\left\{f(z)=\left(\frac{1}{\sqrt{2\pi}}\right)^{n+n'} \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\mu)^2+\sum_{j=1}^{n'}(y_j-\mu)^2\right)\right] ; \right.$$
$$\left. -\infty<\mu<\infty\right\} .$$

In the first step, examine the testing problem $(H, g)_\alpha$. The statistic $T(Z)=\bar{Y}-\bar{X}$, where $\bar{X}$ and $\bar{Y}$ stand for the sample means of the respective samples, satisfies the conditions (i) and (iii) of Theorem 3.2, taking the normal distribution $N\left(0, \dfrac{1}{n}+\dfrac{1}{n'}\right)$ as the distribution of the statistic T induced from the class $L_0(T)$ in the theorem.

Since

$$\text{(4.18)} \qquad \begin{cases} M(\tau)=\exp\left[(\nu_1-\mu_1)\tau+\dfrac{n+n'}{2nn'}\tau^2\right] , \\ \tau_0=-\dfrac{n+n'}{2nn'}(\nu_1-\mu_1) \ , \end{cases}$$

the nearest distribution for the distance problem $(H: g)$ becomes

$$(4.19) \qquad f_0(z)=\left(\frac{1}{\sqrt{2\pi}}\right)^{n+n'} \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2+\sum_{j=1}^{n'}(y_j-\mu_0)^2\right)\right],$$

where $\mu_0=(n\mu_1+n'\nu_1)/(n+n')$, which is a member of the class H. Hence, the condition (ii) in Theorem 3.2 is fulfilled, and the test

$$(4.20) \qquad \varphi_0(z)=\begin{cases}1, \text{ if } \bar{y}-\bar{x}\geqq\sqrt{\frac{n+n'}{nn'}}t_0(\alpha) , \\ 0, \text{ otherwise,}\end{cases}$$

where $t_0(\alpha)$ denotes the (upper) 100α percent-point of the standard normal distribution, will be most powerful for the testing problem $(H, g)_\alpha$. This test is also uniformly most powerful for the problem (4.15).

EXAMPLE 4.4. Let $X_1, X_2, \cdots, X_n$ be a sequence of mutually independent random variables, where each X_i is distributed according to a Poisson distribution with unknown mean λ_i, $i=1, 2, \cdots, n$. Let, for the sum of means, $\theta=\sum_{i=1}^n\lambda_i$,

$$(4.21) \qquad \begin{cases}H: \theta\leqq\theta_0 , \\ A: \theta>\theta_0 ,\end{cases}$$

and consider the testing problem $(H, A)_\alpha$. Here, θ_0 is a given constant. (See E. L. Lehmann [2, Chap. 3, Prob. 27].)

If we consider the statistic

$$(4.22) \qquad T(Z)=\sum_{i=1}^{n}X_i ,$$

then the distribution of this statistic is again the Poisson distribution with mean θ. Let $g(z)$ be a member of the class A of alternatives, that is,

$$(4.23) \qquad g(z)=\prod_{i=1}^{n}e^{-\lambda_{i1}}\frac{\lambda_{i1}^{x_i}}{x_i!} , \quad \left(\theta_1=\sum_{i=1}^{n}\lambda_{i1}>\theta_0\right) .$$

For the testing problem $(H, g)_\alpha$, the statistic (4.22) satisfies the conditions (i) and (iii) in Theorem 3.2, if we take the class of all Poisson distributions of Z such that $\sum_{i=1}^n\lambda_i=\theta_0$ as the class $L_0(T)$ in the theorem.

The condition (ii) in the teorem is examined as follows: Since

$$(4.24) \qquad M(\tau)=\exp\left[\theta_1(e^\tau-1)\right] ,$$

the equation in τ, $(d/d\tau)\log M(\tau)=\theta_0$ has the solution $\tau_0=\log(\theta_0/\theta_1)$, from which we obtain $M(\tau_0)=e^{\theta_0-\theta_1}$. Hence, the nearest distribution of the distance problem $(K(T,\theta_0)\colon g)$ becomes

$$(4.25)\qquad f_0(z)=\left(\prod_{i=1}^{n}\exp(-\lambda_{i1})\frac{\lambda_{i1}^{x_i}}{x_i!}\right)\cdot\exp(\tau_0^T/M(\tau_0)$$
$$=\prod_{i=1}^{n}\exp(-\lambda_{i0})\frac{\lambda_{i0}^{x_i}}{x_i!}\ ,$$

where $\lambda_{i0}=\theta_0\lambda_{i1}/\theta_1$, $i=1,2,\cdots,n$, which is a member of the hypothesical class H, or more precisely, of the class $L_0(T)$. Hereby, the test

$$(4.26)\qquad \varphi_0(z)=\begin{cases}1, & \text{if } T(z)>c\ ,\\ a, & \text{if } T(z)=c\ ,\\ 0, & \text{otherwise}\ ,\end{cases}$$

becomes most powerful for the problem $(H,g)_\alpha$, and it will be uniformly most powerful for the original problem (4.21).

EXAMPLE 4.5. (E. L. Lehmann [2, Chap. 3, Prob. 30]) Let X and Y be two random variables, which are distributed independently according to the binomial distributions $B(n,p_1)$ and $B(n,p_2)$ respectively.

Examine the testing problem $(H,A)_\alpha$ such that

$$(4.27)\qquad \begin{cases}H\colon\ p_1\geqq p_2\ ,\\ A\colon\ p_1<p_2 \text{ and } p_1+p_2=1\ .\end{cases}$$

Let, for fixed p_{11} and p_{21} in the class A,

$$(4.28)\qquad g(z)=\binom{n}{x}p_{11}^x(1-p_{11})^{n-x}\binom{n}{y}p_{21}^y(1-p_{21})^{n-y}\ .$$

Obviously, the statistic $T(z)=Y-X$ satisfies the first condition (i) in Theorem 3.2, for the testing problem $(H,g)_\alpha$.

The second condition (ii) in the theorem will be confirmed as follows; Since

$$(4.29)\qquad M(\tau)=(1-p_{11}+p_{11}e^{\tau})^n(1-p_{21}+p_{21}e^{\tau})^n\ ,$$

solving the equation in τ, $(d/d\tau)\log M(\tau)=0$ we have

$$(4.30)\qquad \begin{cases}\tau_0=\log\dfrac{p_{11}}{p_{21}}=\log\dfrac{p_{11}}{1-p_{11}}\ ,\\ M(\tau_0)=(4p_{11}p_{21})^n\ ,\end{cases}$$

and the nearest distribution, for the distance problem $(K(T,0)\colon g)$, now becomes

(4.31)
$$f_0(z)=g(z)e^{\tau_0 T(z)}/M(\tau_0)$$
$$=\binom{n}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{n-x}\binom{n}{y}\left(\frac{1}{2}\right)^y\left(1-\frac{1}{2}\right)^{n-y},$$

which is a member of the class H such that $p_1=p_2=1/2$.

Denote by $\beta(p_1, p_2)(p_1\geqq p_2)$ the probability of rejection of the test

(4.32)
$$\varphi_0(z)=\begin{cases}1, & \text{if } T(z)>c,\\ a, & \text{if } T(z)=c,\\ 0, & \text{otherwise},\end{cases}$$

that is, put

(4.33)
$$\beta(p_1, p_2)=P_{p_1p_2}\{Y-X>c\}+aP_{p_1p_2}\{Y-X=c\}.$$

When c is sufficiently close to n, i.e., the level of the testing problem, α is sufficiently small, it holds that

(4.34)
$$\beta(p_1, p_2)\leqq\beta(p, p)\leqq\beta\left(\frac{1}{2}, \frac{1}{2}\right),$$

as was remarked by E. L. Lehmann [2]. This means that the test (4.32) above satisfies the size condition for the problem $(H, g)_\alpha$, and then, this test is most powerful for the problem $(H, g)_\alpha$. Moreover, it will easily be seen that this test becomes uniformly most powerful for the original problem (4.27).

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

REFERENCES

[1] E. L. Lehmann, *Theory of testing hypotheses*, Lecture Notes, 1949, Associated Student Store, Univ. of Calif., Berkeley.

[2] E. L. Lehmann, *Testing Statistical Hypotheses*, 1959, J. Wiley & Sons.

[3] E. L. Lehmann, "On the existence of least favorable distributions," *Ann. Math. Stat.*, Vol. 23 (1952), 408-416.

[4] E. L. Lehmann and C. Stein, "Most powerful tests of composite hypotheses. I. Normal distributions," *Ann. Math. Stat.*, vol., 19 (1948), 495-516.

[5] E. L. Lehmann, "Some principles of the theory of testing hypotheses", *Ann. Math. Stat.*, Vol. 21 (1950), 1-26.

[6] S. Kullback, *Information Theory and Statistics*, 1959, J. Wiley & Sons.

[7] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc.*, London, A, 231 (1933), 289-338.

[8] D. A. S. Fraser, *Non-parametric Methods in Statistics*, 1957, J. Wiley & Sons.