DISTANCE AND DECISION RULES

KAMEO MATUSITA

1. Introduction

Distance between distribution functions often becomes a useful and convenient concept in statistics. In a class of distributions, distance can be defined in various ways. We, therefore, have to choose an adequate distance for a particular problem. Further, to treat the problem efficiently, we are required, or at least it is very desirable, to control all errors that may be committed in making decision or inference. For this purpose we have to formulate, or reformulate if necessary, the problem suitably. A proper formulation of the problem is very important for getting command over possible errors.

For instance, suppose that we want to know whether or not the random variable under consideration can be considered to have mean zero, when it is given that the random variable has the Gaussian distribution $N(\theta, 1)$. In this case, if we take the problem as the one inquiring just whether or not the mean of the random variable is zero, we can not control possible errors. For we can find a distribution which has mean not equal to 0, but which is as near to the distribution $N(0, 1)$ as desired, and this makes it impossible to control all possible errors in inference or decision making based on a finite number of observations. In terms of hypothesis testing, we can not make the first and second kinds of errors simultaneously as small as desired (i.e., while we can make the first kind of error smaller than any given (positive) value, this is not the case with the second kind of error). One way to avoid such inconvenience is to formulate the problem as follows. That is, introducing an adequate distance $d(\cdot\ ,\ \cdot)$ in the space of distributions concerned, we set the problem as making decision whether the random variable under consideration has $F_0 \equiv N(0, 1)$ or some F with $d(F, F_0) > \delta$ (>0), where δ is a constant which is to be predetermined from the actual situation of the problem. For the problem thus formulated we can control the errors (see Matusita [1], Matusita, Akaike [2]).

So far, the author has treated various problems with the same idea, the idea of controlling possible errors (see Matusita [1], [3], [5], [7], [8], Matusita, Akaike [2], [6], Matusita, Suzuki, Hudimoto [4]). There, employing a special distance, the author gave decision rules based on the distance, and he often dealt with finite discrete cases (because the finite discrete is the case with most actual problems, as the measurement has always a unit). Of course, as mentioned earlier (Matusita [5]), for some cases the distance can be replaced by other distances, and the space of definition of distributions can be made more general. The purpose of this paper is to treat a generalized version of decision rules based on the distance. At the same time, properties of distance which are used in giving reason to decision rules will be taken up and the concept of distance in a wider sense will be given. As a result the class of decision rules based on distance will be made more inclusive. For instance, decision rules by probability ratio and the maximum likelihood method can be counted within this class.

In section 2, a general formulation of the problem is given based on the distance, and it is seen that various classical problems are considered under this formulation. In section 3, a general decision rule based on distance and properties of distance between distributions are treated, and in section 4, the properties of the distance which are required for the decision rule are further investigated. Throughout these discussions, what is aimed at is to get command over possible errors. In section 5, the problem of estimation is dealt with. The minimum distance method is naturally considered here.

2. Formulation of the problem

Let X be a random variable, and let Ω be the set of distributions to which the distribution of X is known to belong. Further, let $\{\omega_\nu\}$ be a class of subsets of Ω. The problem we treat is to decide which ω_ν can be considered to contain the distribution of F. As is immediately seen, most problems in statistics can be reduced to this form. For efficient decision, we require here that for an adequate distance between distributions, $d(\cdot\,,\,\cdot)$, it holds that $d(\omega_\nu, \omega_\mu)>\alpha$ for any indices ν, μ with $\nu\neq\mu$, where α is a positive constant, and $d(\omega_\nu, \omega_\mu)=\inf_{\substack{F_\nu\in\omega_\nu\\ F_\mu\in\omega_\mu}} d(F_\nu, F_\mu)$. In some cases when $d(\omega_\nu, \omega_\mu)=0$, we can represent ω_ν by a single distribution F_ν such that $d(F_\nu, \omega_\mu)>0$. For such F_ν, we can consider the averaged distribution of ω_ν by an adequate distribution over ω_ν (see also Lehmann [9]). In the following we give some examples of the problem.

1. The *classification problem*. This is just the above mentioned problem itself.

2. The case where $\{\omega_\nu\}$ consists of only two sets ω_1, ω_2 is of interest. Some familiar cases can be brought to this set-up.

(i) The *problem of fit*. This is the case where ω_1 consists of a single F_1, and $d(F_1, \omega_2)>0$. In the literature the problem of fit is often treated as the one which is concerned with a single F_1 and $\omega_2=\Omega-F_1$. In this case, however, it can happen that $d(F_1, \omega_2)=0$, which causes inconvenience for controlling errors.

(ii) The *two sample problem*. This is the problem concerned with whether two random variables X, Y have the same distribution. For this problem we consider as follows. Let G be the set of distributions which can be considered as the distributions of X, Y simultaneously. Then, the paired random variable (X, Y) has the direct product $F_1\times F_2$ as its distribution, where $F_1, F_2 \in G$. Now, set

$$\omega_1=\{F_1\times F_2 \mid F_1\equiv F_2,\ F_1, F_2 \in G\},$$

$$\omega_2=\{F_1\times F_2 \mid d(F_1\times F_2, \omega_1)>\alpha(>0),\ F_1, F_2 \in G\},$$

where $d(\cdot\ ,\ \cdot)$ denotes a distance in $G\times G$, and α is a constant. Then, $d(\omega_1, \omega_2)\geqq\alpha$, and we have the problem of the above mentioned type.

(iii) The *problem of independence*. For two random variables X, Y consider the pair (X, Y) and the set of the joint distributions of the pair, say Ω. Let F_1 and F_2 generally denote possible distributions of X, Y, respectively, and put $\omega_1=\{F_1\times F_2\}$ (the set of direct products) and $\omega_2=\{\Phi \mid d(\Phi, \omega_1)>\alpha(>0), \Phi \in \Omega\}$, where $d(\cdot\ ,\ \cdot)$ denotes a distance in Ω and α is a constant. Then, to decide whether the distribution of (X, Y) belongs to ω_1 or ω_2 is the problem of independence.

Besides these problems we can mention the *problem of invariance* (*symmetry*), a *linear regression model*, etc. Any problem whether or not a certain property is satisfied can be formulated as above.

3. The *problem of estimation*. In this problem each ω_ν consists of a single distribution or a set of distributions. In this problem, we, normally, do not require $d(\omega_i, \omega_j)>0$. We just take up one ω_ν based on some principles of reasoning.

3. Decision rules and desirable properties of the distance

When an adequate distance is introduced into the space of distributions, it is quite natural that, as the set which contains the distribution of the random variable under consideration, we take for a sample an ω_ν that minimizes $d(S, \omega_\nu)$, where S denotes the empirical distribution based on that sample, and $d(\cdot\ ,\ \cdot)$ the distance. The problem is to introduce a distance so that this decision rule can control the possible error, or get as high efficiency as desired. In the following we shall discuss properties of the distance which are required or at least desirable for such purposes.

First, under the word of distance between distributions in a space R, we understand the non-negative valued quantity which is uniquely defined for the distributions. Denote it generally by $d(F_i, F_j)$, F_i, F_j being distributions in space R. It is, of course, desirable that the distance is defined for any two distributions in R. However, as will be seen below, in some cases it is sufficient that the distance is defined in a restricted class of distributions. Now the first property we require is:

(i) $$d(F_i, F_j)=0 \qquad \text{when} \qquad F_i\equiv F_j\,.$$

As the second the following is considered.

(ii) $$d(F_i, F_j)=d(F_j, F_i), \quad \text{for any two distributions } F_i, F_j.$$

We do not always require this property, although it is desirable. The third is:

(iii) $$d(F_i, F_k)+d(F_j, F_k)\geqq d(F_i, F_j).$$

Concerning this inequality, it happens that in some cases we need only the special case where F_k is the empirical distribution (see below). These are the characteristic properties of the distance. However, we count the quantity $d(\cdot\,,\,\cdot)$ as a distance even when (ii) or (iii) is not always satisfied. By doing so, we can consider a wider class of quantities under the name of distance. The following property of the distance is important for our decision rule.

(A) When a random variable X has the distribution F, it holds for an arbitrarily chosen, positive number δ, that

$$P(d(F, S_n)\geqq\delta\,)\rightarrow 0 \qquad (n\rightarrow\infty),$$

where S_n denotes the empirical distribution based on n observations on X.

With this property (A) the decision rule can control possible errors by increasing the number of observations in the case of a finite number of distributions. Namely, let $F_1, \cdots, F_s$ be the distributions under consideration. Then, for each i $(1\leqq i\leqq s)$ and any positive numbers ε and δ there exists an integer $N^{(i)}$ such that for $n\geqq N^{(i)}$

(*) $$P(d(F_i, S_n)\geqq\delta|F_i)<\varepsilon,$$

where $P(d(F_i, S_n)\geqq\delta|F_i)$ means the probability that $d(F_i, S_n)\geqq\delta$ when random variable X has distribution F_i. Therefore, when we put $N=\max_i N^{(i)}$, the relation (*) holds for $i=1, \cdots, s$ when $n\geqq N$. Assume here that we have taken

$$\delta=\frac{1}{2}\min_{(i,j)} d(F_i, F_j)$$

in advance, where the minimization is taken over the set of all pairs (i, j). When the symmetric property of the distance is not satisfied, we have to consider both combinations (i, j) and (j, i). Then, the decision rule is:

Make n $(\geqq N)$ observations on X. When

$$d(F_i, S_n) < d(F_j, S_n) \qquad \text{for any } j \neq i,$$

take F_i as the distribution of X. When

$$d(F_i, S_n) \leqq d(F_j, S_n) \qquad \text{for} \qquad j \neq i,$$

$$d(F_i, S_n) = d(F_k, S_n) \qquad \text{for some } k,$$

take F_i or some F_k as the distribution of X so that the probability according to any F_j that F_i is taken as the distribution of X is well defined.

With this decision rule is associated a subdivision of the sample space R^n. That is, subsets of R^n, $D_1, \cdots, D_s$, are associated with $F_1, \cdots, F_s$ so that all D_i are measurable with respect to all $F_1, \cdots, F_s$ and for a point $x = (x_1, \cdots, x_n) \in D_i$

$$d(F_i, S_n(x_1, \cdots, x_n)) \leqq d(F_j, S_n(x_1, \cdots, x_n)) \quad j \neq i, \quad (i = 1, \cdots, s)$$

and $D_i \frown D_j = 0$ $(i \neq j)$.

On the other hand, when the triangle inequality of the distance holds, we get

$$d(F_j, S_n) \geqq d(F_i, F_j) - d(F_i, S_n)$$

and, when $d(F_i, S_n) < \delta$,

$$d(F_j, S_n) > 2\delta - \delta = \delta.$$

Hence $(x_1, \cdots, x_n)$ belongs to D_i when $d(F_i, S_n(x_1, \cdots, x_n)) < \delta$. Consequently we have

$$P(D_i | F_i) \geqq P(d(F_i, S_n) < \delta | F_i) > 1 - \varepsilon \qquad (i = 1, \cdots, s)$$

$$P(D_i | F_j) \leqq P(R^n - D_j | F_j) < \varepsilon \qquad \text{for } i \neq j \qquad (i, j = 1, \cdots, s).$$

This means that the decision rule has error rate less than ε, or success rate greater than $1 - \varepsilon$. Notice that in this problem the distance need not be defined between empirical distributions, and that only such triangle inequalities as $d(F_i, S_n) + d(F_j, S_n) \geqq d(F_i, F_j)$ are required.

When $F_1, \cdots, F_s$ are unknown, we use the empirical distributions in place of them. Let $S^{(i)} = S^{(i)}(x_1, \cdots, x_{n_i})$ denote the empirical distribution based on a sample of size n_i from F_i. Assume that distances between F_i and an empirical distribution, between F_i and F_j, and between empirical distributions are defined and that for those distances the triangle inequality holds. Further, for arbitrarily chosen positive numbers ε, δ, let

$$P(d(F_i, S_n) \geqq \delta | F_i) < \varepsilon \qquad (i=1, \cdots, s)$$

hold for $n \geqq N$, where $S_n = S(x_1, \cdots, x_n)$ is the empirical distribution based on $(x_1, \cdots, x_n)$, and assume that $n_i \geqq N$. Now, set

$$\overline{D}_i = \{(x_1, \cdots, x_n) | \, d(S^{(i)}, S_n) \leqq d(S^{(j)}, S_n), \ j \neq i\} \qquad (i=1, \cdots, s)$$

and remove some boundary points from $\overline{D}_i$ so that the resulting set D_i is measurable with respect to all F_k, and has no common points with other D_j $(j \neq i)$. This D_i clearly contains

$$D_i^0 = \{(x_1, \cdots, x_n) | \, d(S^{(i)}, S_n) < d(S^{(j)}, S_n), \ j \neq i\}.$$

Then, when $\delta = \dfrac{1}{4} \min\limits_{(i,j)} d(F_i, F_j)$, we have

$$D_i \supseteq \{(x_1, \cdots, x_n) | \, d(F_i, S_n) < \delta\}.$$

In fact, from $d(F_i, S_n) < \delta$ it follows that

$$\begin{aligned} d(F_j, S_n) &\geqq d(F_i, F_j) - d(F_i, S_n) \\ &> 4\delta - \delta = 3\delta ; \end{aligned}$$

from $d(F_i, S^{(i)}) < \delta$, $d(F_i, S_n) < \delta$ it follows that $d(S^{(i)}, S_n) < 2\delta$; and from $d(F_j, S^{(j)}) < \delta$, $d(F_i, S_n) < \delta$ it follows that

$$\begin{aligned} d(S^{(j)}, S_n) &\geqq d(F_i, F_j) - d(F_j, S^{(j)}) - d(F_i, S_n) \\ &> 4\delta - \delta - \delta = 2\delta. \end{aligned}$$

Since

$$P(d(F_i, S^{(i)}) < \delta) > 1 - \varepsilon \qquad (i=1, \cdots, s)$$

for $n \geqq N$, we have

$$P(d(F_i, S^{(i)}) < \delta, \ i=1, \cdots, s) > (1-\varepsilon)^s.$$

It is assumed here that $S_n, S^{(1)}, \cdots, S^{(s)}$ are independent. Then, we obtain

$$\begin{aligned} &P(d(S^{(i)}, S_n) < d(S^{(j)}, S_n), \ \ j \neq i \mid F_i) \\ &\quad \geqq P(d(F_i, S^{(i)}) < \delta, \ d(F_j, S^{(j)}) < \delta \ (j \neq i) \text{ and } d(F_i, S_n) < \delta \mid F_i) \\ &\quad = P(d(F_i, S^{(i)}) < \delta, \ i=1, \cdots, s) P(d(F_i, S_n) < \delta \mid F_i) \end{aligned}$$

$$>(1-\varepsilon)^s(1-\varepsilon)$$
$$=(1-\varepsilon)^{s+1}\equiv 1-\varepsilon' \quad \text{(say)}.$$

From this we also have

P(there exists an integer j_0 such that $d(S^{(i)}, S_n)>d(S^{(j_0)}, S_n)$, $j_0\neq i|F_i)\leqq\varepsilon'$.

This holds for $i=1, \cdots, s$. Consequently we get

$$P(d(S^{(i)}, S_n)<d(S^{(j)}, S_n)|F_j)\leqq\varepsilon', \qquad j\neq i.$$

Thus, the decision rule concerned can control the possible errors.

Suppose that we are concerned with a finite number of sets of distributions, $\omega_1, \cdots, \omega_s$. The problem is to pick up one among $\omega_1, \cdots, \omega_s$ as the set to which the distribution of the random variable under consideration belongs. Assume that the distance is defined in the space of distributions concerned and empirical distributions. Defining

$$d(S, \omega_i)=\inf_{F\in\omega_i} d(S, F),$$

we can deal with the problem by comparing $d(S, \omega_1), \cdots, d(S, \omega_s)$. Here we assume

$$d(\omega_i, \omega_j)>0 \qquad (d(\omega_i, \omega_j)=\inf_{\substack{F_i\in\omega_i\\ F_j\in\omega_j}} d(F_i, F_j)).$$

When each ω_i contains only a finite number of distributions, the problem becomes exactly the above stated. However, when some or all of ω_i contain infinitely many distributions, we need to take up a distance which has a stronger property than the previous one. That is, we require the distance with the following property:

(B) For arbitrarily chosen positive numbers ε, δ, there exists an integer N such that when $n\geqq N$ we have

$$P(d(F, S)\geqq\delta|F)<\varepsilon$$

for any distribution F under consideration.

With the distance with this property we can treat the case of infinitely many distributions just as before, that is, when it is given that $d(\omega_\nu, \omega_\mu)>\alpha>0$ for $\nu\neq\mu$, our decision rule can control possible errors. As a matter of fact, when we previously dealt with decision rules in the finite discrete case, we took advantage of the property B (see Matusita [5], [7], Matusita, Akaike [6]. Concerning property B see also Hoeffding and Wolfowitz [10]).

Now, the set of distributions in a Euclidean space R is separable with respect to the distance

$$d_0(F_i, F_j)=\left\{\int_R(\sqrt{p_i(x)}-\sqrt{p_j(x)})^2\,dm\right\}^{1/2},$$

$p_i(x)$, $p_j(x)$ being probability densities with respect to measure m, of F_i and F_j, respectively. As a result, $\{\omega_\nu\}$, a class of sets of distributions in R, can not contain more than countably many ω_ν when $d(\omega_\nu, \omega_\mu)>\alpha>0$ for the distance $d(\cdot, \cdot)$ in R which is equivalent to, or cruder than, $d_0(\cdot, \cdot)$. Besides, in some actual cases, it comes out by a priori knowledge, or information obtained from samples about the a priori distribution, especially about its range, that we can do with a finite number of ω_i. Further, when, for any two ω_i, ω_j, there exist boundary distributions F_{i_0}, F_{j_0} such that

$$d(F_{i_0}, \omega_i)=0,\ d(F_{j_0}, \omega_j)=0,\ d(F_{i_0}, F_{j_0})>\alpha>0,\ (\alpha \text{ being a constant}),$$
$$d(F_{i_0}, \omega_j)\leqq d(\omega_i, \omega_j),\ d(F_{j_0}, \omega_i)\leqq d(\omega_i, \omega_j),$$

we can treat the problem with such boundary distributions $\{F_{i_0}\}$ (see Matusita, Suzuki, Hudimoto [4]). When the number of ω_i is finite, the number of such boundary distributions is also finite, and we come again to the former finite case.

4. More on properties of the distance

Let us investigate further the basic properties of the distance from our standpoint of making decision.

First, let us consider the case where the distributions concerned are known. We use the notion of distance to express that an empirical distribution is nearer to one distribution than to another, that is, we use that notion for comparison such as $d(F_i, S)<d(F_j, S)$. What is essentially needed, therefore, is that

(I) a linear order relation holds among $\{d(F_i, S)\}$.

Now, the property (A) or (B) has been used to show that

(II) for given ε (>0), there exists an integer N so that for $n\geqq N$

$$P(d(F_i, S)<d(F_j, S)|F_i, i\neq j)>1-\varepsilon.$$

Further, the triangle relation $d(F_i, S)+d(F_j, S)\geqq d(F_i, F_j)$ has been used with (A) or (B) to show that

(III) for $n\geqq N$

$$P(d(F_i, S)<d(F_j, S)|F_j, i\neq j)<\varepsilon.$$

Hence, when it is known by other reasoning that (I), (II), (III) hold, we need not pay attention to whether (A) or (B), or the triangle relation is satisfied. Our decision rule applies without regard to such properties.

For example, consider the case where two different distributions F_1, F_2 with probability densities $p_1(x)$, $p_2(x)$ with respect to the same measure in R are concerned. Define the distance between F_i and the empirical distribution S_n based on a sample $(x_1, \cdots, x_n)$, as

$$d(F_i, S_n)=c_i\{p_i(x_1)\cdots p_i(x_n)\}^{-1} \qquad (i=1, 2),$$

where c_i are positive constants. As the distance between F_1 and F_2 we can take any definition. Then, (I) is obvious. (II) and (III) are also satisfied (see Matusita [1], [3]). Thus, the decision rule based on this definition of distance can control the possible errors. On the other hand, $d(F_1, S_n)<d(F_2, S_n)$ implies

$$\frac{1}{c_1}p_1(x_1)\cdots p_1(x_n)>\frac{1}{c_2}p_2(x_1)\cdots p_2(x_n), \text{ and vice versa.}$$

This means that the decision rule by probability ratio can be considered within the category of decision rules based on the distance.

When the distributions concerned are unknown, we have to consider the empirical distribution for each of them. Let $S_n^{(\nu)}$ denote the empirical distribution based on a sample from F_ν, and S_n the empirical distribution based on n observations. Then, when the following properties (I)′, (II)′, (III)′ are satisfied by a distance $d(\cdot\,,\cdot)$, the decision rule based on this distance satisfies our requirement.

(I)′ A linear order relation holds among $\{d(S_n^{(\nu)}, S_n)\}_\nu$.

(II)′ For given ε (>0) there exists an integer N so that for $n\geqq N$

$$P(d(S_n^{(\nu)}, S_n)<d(S_n^{(\mu)}, S_n)\,|\,F_\nu,\ \nu\neq\mu)>1-\varepsilon.$$

(III)′ For given ε (>0) there exists an integer N so that for $n\geqq N$

$$P(d(S_n^{(\nu)}, S_n)<d(S_n^{(\mu)}, S_n)\,|\,F_\mu,\ \nu\neq\mu)<\varepsilon.$$

(These (I)′, (II)′, (III)′ can, of course, be derived from (A) or (B) and the triangle inequality of the distance.)

Thus far, we have aimed at controlling the possible errors in decision making. When nothing is known about the a priori distribution, it is desirable to make the sizes of the possible errors equal to each other. In some cases, this can be done by modifying the definition of distance, or replacing the fundamental relation, say, $d(F_i, S_n)<d(F_j, S_n)$, by $c_id(F_i, S_n)<c_jd(F_j, S_n)$, c_i, c_j being appropriate constants. Further, when two decision rules have the same size of the possible errors, the one with smaller sample size is clearly better. The optimality in this sense depends on the choice of the distance.

5. Estimation

Let X be the random variable under consideration, S the empirical distribution based on observations on X, and Ω the set of distributions each of which may possibly be the distribution of X. Then, the problem of estimation is to choose one among Ω for an observed S as the distribution of X. When the distance $d_2(\cdot\ ,\ \cdot)$ is defined between any F in Ω and S, the decision rule which is immediately considered is to take the one which minimizes $d_2(F, S)$. This is the minimum distance method. Let S_n be the empirical distribution based on a sample of size n as before, let

$$d_2(F_{e,n}, S_n)=\min_{F\in\Omega} d_2(F, S_n),$$

and let F_0 be the distribution of X. Then

$$d_2(F_{e,n}, S_n)\leqq d_2(F_0, S_n).$$

Let $d_1(\cdot\ ,\ \cdot)$ be the distance defined in Ω with the following properties:

1) $d_1(F_1, F_1)=0$,

2) when $d_1(F_1, F_2)=0$, F_1 and F_2 have the same characteristic with respect to the point under consideration,

3) for a sequence of distributions $\{F_n\}$ and a distribution F_0, $d_1(F_n, F_0)\to 0$ means that the characteristic under consideration of F_n approaches that of F_0. When the distance functions $d_1(\cdot\ ,\ \cdot)$, $d_2(\cdot\ ,\ \cdot)$ are such that the triangle inequality

$$d_1(F_{e,n}, F_0)\leqq d_2(F_{e,n}, S_n)+d_2(F_0, S_n)$$

holds, we have

$$d_1(F_{e,n}, F_0)\leqq 2d_2(F_0, S_n).$$

Hence, when (A) holds, for any positive ε, δ, there exists an integer N such that

$$P(d_1(F_{e,n}, F_0)<\delta|F_0)>1-\varepsilon$$

for $n>N$. This means

$$d_1(F_{e,n}, F_0)\to 0 \qquad \text{with probability 1.}$$

That is, the minimum distance method provides a strongly consistent estimate.

Further, even when the above triangle relation does not always hold, the minimum distance method can provide a consistent estimate. For instance, let $p_\nu(x)$ denote the probability density of distribution F_ν with respect to a Lebesgue or counting measure m defined in space R. Distance in Ω can be defined in any way so long as it makes sense for the problem, for example, $d_1(F_1, F_2)=\left\{\int_R(\sqrt{p_1(x)}-\sqrt{p_2(x)})^2\,dm\right\}^{1/2}$ or $\int_R|p_1(x)-p_2(x)|dm$. As $d_2(F, S_n)$ $(S_n=S(x_1, \cdots, x_n))$ we define

$$d_2(F_\nu, S_n)=\{p_\nu(x_1)\cdots p_\nu(x_n)\}^{-1}.$$

Then, the minimum d_2-distance method is the maximum likelihood method, so the minimum d_2-distance method provides a consistent estimate under some conditions. From this example it can be seen that the maximum likelihood method can be taken as a decision rule based on the distance (in a wide sense).

We can also treat the problem to estimate the true distribution (parameter) by a set of distributions (parameters), based on the notion of distance (see Matusita [5]).

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] K. Matusita, "On the theory of statistical decision functions," *Ann. Inst. Stat. Math.*, 3 (1951), 17-35.
[2] K. Matusita and H. Akaike, "Note on the decision problem," *Ann. Inst. Stat. Math.*, 4 (1952), 11-14.
[3] K. Matusita, "Decision rule by probability ratio," *Ann. Inst. Stat. Math.*, 6 (1954), 143-151.
[4] K. Matusita, Y. Suzuki and H. Hudimoto, "On testing statistical hypotheses," *Ann. Inst. Stat. Math.*, 6 (1954), 133-141.
[5] K. Matusita, "Decision rules based on the distance, for problems of fit, two samples, and estimation," *Ann. Math. Statist.*, 26 (1955), 613-640.
[6] K. Matusita and H. Akaike, "Decision rules, based on the distance, for the problems of independence, invariance and two samples," *Ann. Inst. Stat. Math.*, 7 (1956), 67-80.
[7] K. Matusita, "Decision rule, based on the distance for the classification problem," *Ann. Inst. Stat. Math.*, 8 (1956), 67-77.
[8] K. Matusita, "Interval estimation based on the notion of affinity," *Bull. Inter. Stat. Inst.*, 38 (1961), Part 4, 241-244.
[9] E. L. Lehmann, *Testing Statistical Hypothesis*, John Wiley, New York, 1959, Chap. 8.
[10] W. Hoeffding and J. Wolfowitz, "Distinguishability of sets of distributions," *Ann. Math. Statist.*, 29 (1958), 700-718.