DECISION RULE, BASED ON THE DISTANCE, FOR THE CLASSIFICATION PROBLEM

By KAMEO MATUSITA

(Received Oct. 20, 1956)

1. Introduction

The classification problem arises very often in our daily life, and thus far various methods have been presented for it (see [1], [2], [3], [4], [5]). It, however, seems that most of them have laid some assumptions, for instance, normality, on the functional form of distributions concerned and that the success rate has not been given except in a few cases such as when the distributions are completely specified (see [2], [5], [6]). In this paper we shall give a decision rule for this problem and a lower bound for its success rate when the distributions concerned are not specified. The rule we give here is very simple and the success rate tends to 1 as sizes of samples become large. The treatment by a linear combination in multivariate case will be also referred to.

Throughout the present paper, we shall consider only discrete distributions with a finite number of possible outcomes as in the previous papers [7], [8], [9], and our treatment is based on the following inequality or equality :

for any positive number η

$$(1) \qquad P(\|F-S_n\|>\eta)\leqq\frac{k-1}{n\eta^2} \quad \text{or} \quad \frac{k^2+k-1}{(n\eta^2)^2} \quad (k\geqq 2\,,\ n>k)$$

or

$$(2) \qquad P(\|F-S_n\|>\eta)\doteqdot P(\chi^2_{(k-1)}\geqq 4n\eta^2)$$

where F denotes the distribution of the discrete random variable under consideration, S_n the empirical distribution for n observations on the variable, k the number of values the variable takes on, and $\chi^2_{(k-1)}$ a random variable which has χ^2-distribution with $k-1$ degrees of freedom. The idea is just the same with which we have treated the two sample problem in [7], [8].

Our rule is always applicable with any distance which has a property like (1) or (2), though in this paper use is made only of distance $\|\ \ \|$.

2. Properties of distance $\| \quad \|$ and affinity

In connection with distance $\| \quad \|$, we have previously defined the affinity between two distributions, that is, for two distributions $F=\{p_i\}$, $G=\{q_i\}$, $(i=1, 2, \cdots, k)$ defined on the same space, R, we have defined the affinity as follows:

$$\rho(F, G)=\sum_{i=1}^{k}\sqrt{p_i}\sqrt{q_i}$$

while

$$\|F-G\|^2=\sum_{i=1}^{k}(\sqrt{p_i}-\sqrt{q_i})^2$$

Between $\|F-G\|$ and ρ it holds that

$$\|F-G\|^2=2(1-\rho) \tag{3}$$

and

$$\|F-G\|^2\leqq\sum_{i=1}^{k}|p_i-q_i|\leqq2\sqrt{1-\rho^2}\leqq2\|F-G\| \tag{4}$$

According to (3), ρ serves for the calculation of the distance and the computation of ρ is very simple. Further, from (4) it follows that for any subset of R, E,

$$|F(E)-G(E)|\leqq2\sqrt{1-\rho^2}\leqq2\|F-G\| \tag{5}$$

Therefore, when $\|F-G\|$ is small, consequently, ρ is large, the difference of probabilities defined by F and G is very small. For example, when $\rho=0.9998$, that is, $\|F-G\|=0.02$, we have always

$$|F(E)-G(E)|<0.03999.$$

Now, let X and Y be random variables with F and G, and $S_n=\left\{\frac{n_i}{n}\right\}$ and $S'_m=\left\{\frac{m_i}{m}\right\}$ the empirical distributions based on the observations on X and Y, respectively. Then we have:

when $F=G$,

$$P(\|S_n-S'_m\|<\eta)=P\left(\rho(S_n, S'_m)>1-\frac{\eta^2}{2}\right)$$

$$\geqq1-\frac{k-1}{\eta^2}\left(\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{m}}\right)^2$$

or

$$\geqq 1-\frac{4(k^2+k-1)}{\eta^4}\left(\frac{1}{n^2}+\frac{1}{m^2}\right)+\frac{16(k^2+k-1)^2}{\eta^8 n^2 m^2}$$

or

$$\doteqdot P(\chi^2_{(k-1)}<n\eta^2)\,P(\chi^2_{(k-1)}<m\eta^2)$$

where $\chi^2_{(k-1)}$ is a random variable having χ^2-distribution with $k-1$ degrees of freedom,

and when $\|F-G\|\geqq\delta_0\,(>\eta)$,

$$P(\|S_n-S'_m\|>\eta)=P\left(\rho(S_n\,,\,S'_m)<1-\frac{\eta^2}{2}\right)$$

$$\geqq 1-\frac{k-1}{(\delta_0-\eta)^2}\left(\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{m}}\right)^2$$

or

$$\geqq 1-\frac{4(k^2+k-1)}{(\delta_0-\eta)^4}\left(\frac{1}{n^2}+\frac{1}{m^2}\right)+\frac{16(k^2+k-1)}{(\delta_0-\eta)^8 n^2 m^2}$$

or

$$\doteqdot P(\chi^2_{(k-1)}<\eta(\delta_0-\eta)^2)P(\chi^2_{(k-1)}<m(\delta_0-\eta)^2)$$

(see [7], [9]). These relations will serve for study of the probabilistic behaviour of affinity.

3. Decision rule

Given two independent discrete random variables, X, Y, we want to decide whether the third variable Z has the same distribution as that of X or that of Y, provided that either of them holds. Let F and G be the distributions of X and Y, and S_n, S'_m, S''_l the empirical distributions based on observations X, Y and Z, respectively, where n, m, l mean the numbers of observations. Then our rule is :

When $\|S_n-S''_l\|\leqq\|S'_m-S''_l\|$, *decide that* Z *has* F,

and when $\|S_n-S''\|>\|S'_m-S''_l\|$, *decide that* Z *has* G,

which is also expressed as:

When $\rho(S_n\,,\,S''_l)\geqq\rho(S'_m\,,\,S''_l)$, *decide that* Z *has* F,

and when $\rho(S_n\,,\,S''_l)<\rho(S'_m\,,\,S''_l)$, *decide that* Z *has* G.

This rule can easily be extended to the case of several random variables. That is, when there are several random variables under consideration, say, X, X_1, X_2, $\cdots$, X_s, and when we want to decide which is the one among X_1, X_2, $\cdots$, X_s, whose distribution X is considered to have, make decision according to the distances $\|S_n-S^{(1)}_{n_1}\|$, $\|S_n-S^{(2)}_{n_2}\|$, $\cdots$, $\|S_n-S^{(s)}_{n_s}\|$; namely, when $\|S_n-S^{(i)}_{n_i}\|$ is the minimum among $\|S_n-S^{(1)}_{n_1}\|$, $\|S_n-S^{(2)}_{n_2}\|$, $\cdots$, $\|S_n-S^{(s)}_{n_s}\|$, then decide that X has the same distribution as that of X_i, where S_n, $S^{(1)}_{n_1}$, $\cdots$, $S^{(s)}_{n_s}$ mean the empirical distributions of X, X_1, $\cdots$, X_s, respectively. Of course, we assume here that the distribution of X is the same as one of the distributions of X_1, X_2, $\cdots$, X_s.

As is seen, our rule is very simple. The problem is how to evaluate the success rate. In the following we shall deal with this problem.

4. Evaluation of success rate I (Case of many observations)

In order to discuss the success rate of our rule, it is sufficient to treat the simplest case where only three variables, X, Y, Z, are concerned. Further, in this section we shall treat the case where a lot of observations on Z as well as on X and Y are available.

Let E_1, E_2, $\cdots$, E_k $(k\geqq 2)$ be the values the variables take on. Then F, G, S'_m, S''_l are respectively represented by, say, $\{p_i\}$, $\{q_i\}$, $\left\{\frac{n_i}{n}\right\}$, $\left\{\frac{m_i}{m}\right\}$, $\left\{\frac{l_i}{l}\right\}$ $(i=1, 2, \cdots, k)$. We assume here that a lower bound for $\|F-G\|$, d, is known. Then it holds:

When Z has distribution F, we have

$$P(\|S_n-S''_l\|\leqq\|S'_m-S''_l\|)\geqq\left(1-\frac{16(k-1)}{nd^2}\right)\left(1-\frac{16(k-1)}{md^2}\right)\left(1-\frac{16(k-1)}{ld^2}\right)$$

for n, m, l greater than $\dfrac{16(k-1)}{d^2}$, *and*

$$P(\|S_n-S''_l\|\leqq\|S'_m-S''_l\|)$$
$$\geqq\left(1-\frac{16(k^2+k-1)}{(nd^2)^2}\right)\left(1-\frac{16(k^2+k-1)}{(md^2)^2}\right)\left(1-\frac{16(k^2+k-1)}{(ld^2)^2}\right)$$

for n, m, l greater than $\dfrac{4\sqrt{k^2+k-1}}{d^2}$ *and k, and*

$$P(\|S_n-S''_l\|\leqq\|S'_m-S''_l\|)$$
$$\doteqdot P\left(\chi^2_{(k-1)}<4n\frac{d^2}{16}\right)P\left(\chi^2_{(k-1)}<4m\frac{d^2}{16}\right)P\left(\chi^2_{(k-1)}<4l\frac{d^2}{16}\right)$$

for sufficiently large n, m, l, where $\chi^2_{(k-1)}$ *denotes a random variable having* χ^2*-distribution with* $k-1$ *degrees of freedom.*

PROOF: First we have:

$$\|S_n - S_l''\| \leqq \|S_n - F\| + \|S_l'' - F\|,$$

$$\|S_m' - S_l''\| \geqq \|F - G\| - \|S_m' - G\| - \|S_l'' - F\|.$$

From this it follows that

$$\|S_m' - S_l''\| - \|S_n - S_l''\| \geqq \|F - G\| - \|S_n - F\| - 2\|S_l'' - F\| - \|S_m' - G\|$$
$$\geqq d - \|S_n - F\| - 2\|S_l'' - F\| - \|S_m' - G\|.$$

Therefore, when $\|S_n - F\| + 2\|S_l'' - F\| + \|S_m' - G\|$ is less than d, we have

$$\|S_m' - S_l''\| - \|S_n - S_l''\| \geqq d - d = 0,$$

that is,

$$\|S_m' - S_l''\| \geqq \|S_n - S_l''\|.$$

Thus we have

$$P(\|S_n - S_l''\| \leqq \|S_m' - S_l''\|) \geqq P(S_n - F\| + 2\|S_l'' - F\| + \|S_m' - G\| < d)$$

$$\geqq P\left(\|S_n - F\| < \frac{d}{4}, \|S_l'' - F\| < \frac{d}{4}, \|S_m' - G\| < \frac{d}{4}\right)$$

$$= P\left(\|S_n - F\| < \frac{d}{4}\right) P\left(\|S_l'' - F\| < \frac{d}{4}\right) P\left(\|S_m' - G\| < \frac{d}{4}\right)$$

$$\geqq \left(1 - \frac{16(k-1)}{nd^2}\right)\left(1 - \frac{16(k-1)}{ld^2}\right)\left(1 - \frac{16(k-1)}{md^2}\right)$$

for n, m, l greater than $\dfrac{16(k-1)}{d^2}$, and

$$\geqq \left(1 - \frac{16(k^2+k-1)}{(nd^2)^2}\right)\left(1 - \frac{16(k^2+k-1)}{(ld^2)}\right)\left(1 - \frac{16(k^2+k-1)}{(md^2)^2}\right)$$

for n, l, m greater than $\dfrac{4\sqrt{k^2+k-1}}{d^2}$ and k, and

$$\doteqdot P_r\left(\chi^2_{(k-1)} < 4n\frac{d^2}{16}\right) P_r\left(\chi^2_{(k-1)} < 4l\frac{d^2}{16}\right) P_r\left(\chi^2_{(k-1)} < 4m\frac{d^2}{16}\right)$$

for sufficiently large n, l, m.

According to this fact, it can be seen that *our rule has success rate not less than*

$$\left(1-\frac{16(k-1)}{nd^2}\right)\left(1-\frac{16(k-1)}{md^2}\right)\left(1-\frac{16(k-1)}{ld^2}\right)$$

or

$$\left(1-\frac{16(k^2+k-1)}{(nd^2)^2}\right)\left(1-\frac{16(k^2+k-1)}{(md^2)^2}\right)\left(1-\frac{16(k^2+k-1)}{(ld^2)^2}\right)$$

or

$$P\left(\chi^2_{(k-1)}<4n\frac{d^2}{16}\right)P\left(\chi^2_{(k-1)}<4m\frac{d^2}{16}\right)P\left(\chi^2_{(k-1)}<4l\frac{d^2}{16}\right).$$

These values shows that *the success rate of our rule tends to 1 as the numbers of observations become larger and larger.*

5. Evaluation of success rate II (Case of a single observation)

Now consider the case where we have to make decision based on only one observation on Z. In this case *we can consider the success rate of our rule as not less than min* $\left(\Sigma_1\frac{n_i}{n}, \Sigma_2\frac{m_i}{m}\right)-\eta$ *with probability equal to or greater than* $\left(1-\frac{k-1}{n\eta^2}\right)\left(1-\frac{k-1}{m\eta^2}\right)$ *or* $\left(1-\frac{(k^2+k-1)}{n^2\eta^4}\right)\left(1-\frac{(k^2+k-1)}{m^2\eta^4}\right)$ *or* $P(\chi^2_{(k-1)}<4n\eta^2)\,P(\chi^2_{(k-1)}<4m\eta^2)$, *where the summation* Σ_1 *runs over such i's that* $\frac{n_i}{n}\geqq\frac{m_i}{m}$, *and* Σ_2 *runs over such j's that* $\frac{n_j}{n}<\frac{m_j}{m}$, *and* η *is a positive number less than* $\frac{d}{2}$.

For

$$P(\|S_n-F\|<\eta\,,\|S'_m-G\|<\eta)=P(\|S_n-F\|<\eta)P(\|S'_m-G\|<\eta)$$
$$\geqq\left(1-\frac{k-1}{n\eta^2}\right)\left(1-\frac{k-1}{m\eta^2}\right)\quad\text{or}\quad\left(1-\frac{k^2+k-1}{n^2\eta^4}\right)\left(1-\frac{k^2+k-1}{m^2\eta^4}\right)$$

or

$$\fallingdotseq P(\chi^2_{(k-1)}<4n\eta^2)P(\chi^2_{(k-1)}<4m\eta^3)\,,$$

that is, it holds with probability equal to or greater than $\left(1-\frac{k-1}{n\eta^2}\right)\left(1-\frac{k-1}{m\eta^2}\right)$ or $\left(1-\frac{k^2+k-1}{n^2\eta^4}\right)\left(1-\frac{k^2+k-1}{m^2\eta^4}\right)$ or $P(\chi^2_{(k-1)}<4n\eta^2)P(\chi^2_{(k-1)}<4m\eta^2)$ that $\|S_n-F\|<\eta$ and $\|S'_m-G\|<\eta$.

Let $P(E_i:F)$ and $P(E_i:G)$ denote the probabilities of E_i by F and G, respectively. In [5], [6], we presented the decision rule which decides that Z has F when the observation on Z falls into the set $\mathfrak{S}=\{E_i; P(E_i:F)\geqq P(E_i:G)\}$, and that Z has G otherwise. This rule has the success rate not less than $\min(\sum_{E_i\in\mathfrak{S}} P(E_i:F), \sum_{E_j\in\mathfrak{S}} P(E_j:G))$. Our present rule is a modification of this one. We use here the empirical distribution instead of the true distribution there.

As to the distance $\|S_n-S_1''\|$, when the observation on Z falls on E_i, we have

$$\|S_n-S_1''\|^2=2\left(1-\sqrt{\frac{n_i}{n}}\right),$$

$$\|S_m'-S_1''\|^2=2\left(1-\sqrt{\frac{m_i}{m}}\right)$$

where n_i and m_i mean the numbers of E_i in the previous observations on X and Y, respectively. Therefore,

$$\|S_n-S_1''\|\leqq S_m'-S_1''\|$$

means

$$\frac{n_i}{n}\geqq\frac{m_i}{m}.$$

Thus our rule is equivalent to the rule:

Decide that Z has F when the observation value on Z falls in the set $\mathfrak{S}'=\left\{E_i;\ \frac{n_i}{n}\geqq\frac{m_i}{m}\right\}$ and that Z has G otherwise.

Hence, the rule has the success rate $\min\left(P(\mathfrak{S}':F), P((\mathfrak{S}')^c:G)\right)$, where $(\mathfrak{S}')^c$ means the complement of $\mathfrak{S}'$.

Now, from $\|S_n-F\|<\eta$ and $\|S_m'-G\|<\eta$, it follows that

$$|P(\mathfrak{S}':F)-P(\mathfrak{S}':S_n)|\leqq 2\eta,$$

$$|P(\mathfrak{S}':G)-P(\mathfrak{S}':S_m')|\leqq 2\eta,$$

$$|P((\mathfrak{S}')^c:G)-P((\mathfrak{S}')^c:S_m')|\leqq 2\eta.$$

Consequently,

$$P(\mathfrak{S}':F)\geqq P(\mathfrak{S}':G)-4\eta,$$

$$P(\mathfrak{S}':F)\geqq P(\mathfrak{S}':S_n)-2\eta,$$

$$P((\mathfrak{S}')^c:G)\geqq P((\mathfrak{S}')^c:S_m')-2\eta.$$

Therefore, our rule has the success rate not less than $\min\Big(P(\mathfrak{S}'; S_n), P((\mathfrak{S}')^c; S_n)\Big)-2\eta$, when $\|S_n-F\|<\eta$ and $\|S'_m-G\|<\eta$, and we can say as mentioned in the beginning of this section.

6. Treatment by linear combination

In some cases where the random variables under consideration take vector values a linear combination of the components of the vector value is considered rather than the vector value itself. However, this can not always be recommended. In this section, we shall give some explanation about these circumstances.

Let

$$\begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ x_{1t} \end{pmatrix}, \begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2t} \end{pmatrix}, \cdots, \begin{pmatrix} a_{k1} \\ a_{k2} \\ \vdots \\ a_{kt} \end{pmatrix}$$

be the vector values the random variables $\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_t \end{pmatrix}, \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \end{pmatrix}$ take on with probabilities $(p_1, p_2, \cdots, p_k)$ and $(q_1, q_2, \cdots, q_k)$, and represent generally those vector values by $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_t \end{pmatrix}$. For a set of numbers $(\alpha_1, \alpha_2, \cdots, \alpha_t)$ we consider linear combinations

$$X=\alpha_1 X_1+\alpha_2 X_2+\cdots+\alpha_t X_t ,$$
$$Y=\alpha_1 Y_1+\alpha_2 Y_2+\cdots+\alpha_t Y_t .$$

These quantities take the values $\alpha_1 a_1+\alpha_2 a_2+\cdots+\alpha_t a_t$. When we want to treat the classification problem employing these X, Y, we evidently should choose $(\alpha_1, \alpha_2, \cdots, \alpha_t)$ such that the distributions of X and Y lies apart from each other as much as possible. For this, when $p_1, p_2, \cdots, p_k$, $q_1, q_2, \cdots, q_k$ are known, we classify $I=\{1, 2, \cdots, k\}$ into subgroups $I_1=\{i_{11}, i_{12}, \cdots, i_{1k_1}\}$, $\cdots$, $I_g=\{i_{g1}, i_{g2}, \cdots, i_{gk_g}\}$ such that the affinity

$$\rho_1=\sum_{j=1}^{g}\sqrt{\sum_{i\in I_j} p_i}\sqrt{\sum_{i\in I_j} q_i}$$

attains its minimum value. Then, from a system of linear equations

$$\alpha_1 a_{i_{j1}1}+\cdots+\alpha_t a_{i_{j1}t}=\alpha_1 a_{i_{j2}1}+\cdots+\alpha_t a_{i_{j2}t}$$
$$=\cdots=\alpha_1 a_{i_{jk_j}1}+\cdots+\alpha_t a_{i_{jk_j}t}$$
$$(j=1, 2, \cdots, g)$$

we can find a system $(\alpha_1, \alpha_2, \cdots, \alpha_t)$ suitable for the above purpose.

With each value $a=\alpha_1 a_1+\cdots+\alpha_t a_t$ for $(\alpha_1, \alpha_2, \cdots, \alpha_t)$ thus obtained are associated probabilities defined by $\{p_i\}$ and $\{q_i\}$, that is, to $a=\alpha_1 a_1+\cdots+\alpha_t a_t$ with $a_1=a_{i_{j1}1}, \cdots, a_t=a_{i_{j1}t}$ are given $\sum_{i\in I_j} p_i$ and $\sum_{i\in I_j} q_i$. We put

$$E_1=\{a; P(a:\{p_i\})\geqq P(a:\{q_i\})\},$$
$$E_2=\{a; P(a:\{p_i\})<P(a:\{q_i\})\}$$

where $P(a:\{p_i\})$ and $P(a:\{q_i\})$ denote probabitities of a defined by $\{p_i\}$ and $\{q_i\}$, respectively. Then, for an observed system of values $\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_t \end{pmatrix}$ on the third variable $Z=\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_t \end{pmatrix}$, which has $(p_1, p_2, \cdots, p_k)$ or $(q_1, q_2, \cdots q_k)$ as its distribution, we calculate $z=\alpha_1 z_1+\cdots+\alpha_t z_t$ and decide that Z has the distribution $(p_1, p_2, \cdots, p_k)$ when z falls in E_1 and that Z has $(q_1, q_2, \cdots, q_k)$ otherwise. The success rate of this procedure is not less than $1-\rho_1$.

The above-mentioned method to determine $(\alpha_1, \alpha_2, \cdots \alpha_t)$ depends on the values of $(p_1, p_2, \cdots, p_k)$ and $(q_1, q_2, \cdots, q_k)$, but when $(p_1, p_2, \cdots, p_k)$ and $(q_1, q_2, \cdots, q_k)$ are not known, we make use of the empirical distributions on $\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_t \end{pmatrix}$ and $\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \end{pmatrix}$ instead of them and proceed as mentioned before (sections 4, 5).

Another method to determine $(\alpha_1, \alpha_2, \cdots, \alpha_t)$ is to make $|E(X-Y)| =|\sum_{i=1}^{t}\alpha_i(E(X_i)-E(Y_i))|$ as large as possible and at the same time $D^2(X-Y)=\sum_{i=1}^{t}\alpha_i^2(D^2(X_i)+D^2(Y_i))$ as small as possible, for example, to maximize $\frac{|E(X-Y)|}{D^2(X-Y)}$ (see [4]). This means in a sense setting apart the distributions of X and Y as much as possible.

Now, for X and Y with any system $(\alpha_1, \alpha_2, \cdots, \alpha_t)$, we generally have

$$\rho(X, Y) \geqq \rho = \sum_{i=1}^{k} \sqrt{p_i}\sqrt{q_i}$$

as is easily seen. This implies that it can not always be recommended to consider the problem by taking linear combinations of the components of the vector variables concerned.*) Of course, the situation may be different, when the precision of estimation of the distributions is much raised by taking the linear combinations, as in the case when the distribution of a linear combination approaches a Gaussian distribution. However, in the discrete case, this seldom takes place for the number of the components which is not so large.

7. Parametric case

Thus far we have considered the case where the distributions under consideration are completely unknown. In this section we shall treat the case where some parameters of the distributions are unkown, while their functional forms are known.

Let $F(x; \theta_1)$ and $G(x; \theta_2)$ be the distributions of X and Y, respectively, and assume θ_1 and θ_2 (generally vectors) are unknown. First, we estimate θ_1 and θ_2 by minimizing $\|F-S_n\|$ and $\|G-S'_m\|$. Denote the estimates thus obtained by $\hat{\theta}_1$ and $\hat{\theta}_2$, and F and G with $\hat{\theta}_1$, $\hat{\theta}_2$ replacing θ_1, θ_2 by $F_{e,n}$ and $G_{e,m}$, respectively. Then we have

$$P(\|F_{e,n}-S_n\|<\eta) \geqq 1-\frac{k-1}{n\eta^2} \text{ or } 1-\frac{k^2+k-1}{n^2\eta^4}$$

or

$$\doteqdot P(\chi^1_{(k-1)}<4n\eta^2) \quad \text{when } n \text{ is sufficiently large}$$

and

$$P(\|G_{e,m}-S'_m\|<\eta) \geqq 1-\frac{k-1}{m\eta^2} \text{ or } 1-\frac{k^2+k-1}{m^2\eta^4}$$

or

$$\doteqdot P(\chi^2_{(k-1)}<4m\eta^2) \quad \text{when } m \text{ is sufficiently large.}$$

*) Note that we consider here the classification problem, not the mere ordering of vector values.

(see [7], [9].) Now, put

$$\mathfrak{X} = \{E_i ; F_{e,n}(E_i) \geqq G_{e,m}(E_i)\}$$

and our rule is equivalent to:

Decide that Z has F when S_1'' lies in $\mathfrak{X}$, and that Z has G otherwise.

As for the success rate of this rule we can say with probability not less than $\left(1-\frac{k-1}{n\eta^2}\right)\left(1-\frac{k-1}{m\eta^2}\right)$ or $\left(1-\frac{k^2+k-1}{n^2\eta^4}\right)\left(1-\frac{k^2+k-1}{m^2\eta^4}\right)$, or asymptotically equal to $P(\chi^2_{(k-1)} < 4n\eta^2)P(\chi^2_{(k-1)} < 4m\eta^2)$, that it is not less than $\min(P(\mathfrak{X}: F_{e,n}),\ P(\mathfrak{X}^c : G_{e,m})) - 2\eta$.

Concerning the estimation of the parameters, we can employ the χ^2-minimum method instead of minimum $\| \ \|$ method, when $F(E_i:\theta_1)$ and $G(E_1: \theta_2)$ have a positive lower bound for possible values of θ_1 and θ_2.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Wald, A., "On a statistical problem arising in the classification of an individual into one of two groups," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 145–162.

[2] von Mises, R., "On the classification of observation data into distinct groups," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 68–73.

[3] Rao, C. Radhakrishna, *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc. (1952), Chap. 8, pp. 273–350. (Many references are given at the end of this chapter).

[4] Hayashi, C., "On the Prediction of Phenomena from Qualitative Data from the Mathematico-Statistical Point of View," *Ann. Inst. Stat. Math.*, Vol. III (1952), pp. 69–98.

[5] Matusita, K., "On the theory of statistical decision functions," *Ann. Inst. Stat. Math.*, Vol. III (1951), pp. 17–35.

[6] Matusita, K. and H. Akaike, "Note on the decision problem," *Ann. Inst. Stat. Math.*, Vol. IV (1952), pp. 11–14.

[7] Matusita, K., "Decision rules, based on the distance, for problems of fit, two samples, and estimation," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 631–640.

[8] Matusita, K. and H. Akaike, "Decision rules, based on the distance, for the problems of independence, invariance and two samples," *Ann. Inst. Stat. Math.*, Vol. VII (1956), pp. 67–80.

[9] Matusita, K., and M. Motoo, "On the fundamental theorem for the decision rule based on distance $\| \ \|$," *Ann. Inst. Stat. Math.*, Vol. VII (1956), pp. 137–142.