SOME PROPERTIES OF AFFINITY AND APPLICATIONS

KAMEO MATUSITA

(Received Sept. 10, 1970)

1. Introduction

The "affinity" of distributions represents the likeness of the distributions ([1]). To speak this in other words, this means that the affinity represents the discrepancy among distributions, so the affinity can serve as a measure of discrimination among distributions. Actually, it has similar properties to those of information numbers thus far considered, for example, Kullback-Leibler's information for discrimination [5]. The purpose of this paper is to show some properties of the affinity of several distributions as a quantity which gives information for discrimination among distributions. An optimal decision rule for discriminating distributions, and bounds for the error of this rule will also be given. Employing this rule we can treat the coding problem in information transmission. However, this problem will not be dealt with in this paper. Further, by means of properties of the affinity we shall show that the limiting statistic of a sequence of sufficient statistics for a set of distributions is also sufficient for the set.

The affinity between two distributions is related to a distance between distributions, from which we can see that the distance has corresponding properties to those of the affinity. Actually, Kirmani [2], [3] show this fact. On the other hand, Kullback-Leibler's information is not symmetric in distributions. But the affinity is symmetric in distributions and has direct relationship with error probability when classification or discrimination is concerned.

Now, the definition of "affinity" is as follows. Let $F_1, F_2, \cdots, F_r$ be distributions defined in the same space R with measure m (Lebesgue or counting or mixed), and let $p_1(x), p_2(x), \cdots, p_r(x)$ be respectively their density functions with respect to m ($p_i(x) \geqq 0$). Then, we call the quantity

$$\rho_r(F_1, F_2, \cdots, F_r) = \int_R (p_1(x)p_2(x)\cdots p_r(x))^{1/r} dm$$

the affinity of $F_1, F_2, \cdots, F_r$.

2. Some theorems

In this section we shall prove some theorems concernig the affinity.

THEOREM 1. *Let* $\{F_{i\nu}\}$ $(i=1, 2, \cdots, r;\ \nu=1, 2, \cdots)$ *be sequences of distributions over* R *with density functions* $p_{i\nu}(x)$ *with respect to measure* m. *Further, let* $\lim_{\nu\to\infty} p_{i\nu}(x)=p_{i0}(x)$ *a.e. w.r.t.* m *in* R, $p_{i0}(x)$ *being a density function, and let* F_{i0} *denote the distribution defined by* $p_{i0}(x)$. *Then we have*

$$\rho_r(F_{1\nu}, F_{2\nu}, \cdots, F_{r\nu})\to\rho_r(F_{10}, F_{20}, \cdots, F_{r0}) \qquad (\nu\to\infty)\ .$$

The proof is straightforward.

THEOREM 2. *Let* $E=\sum_i E_i$, $E_i\cap E_j=\phi$, E_i, E_j *being measurable sets with respect to measure* m. *Then we get*

$$\int_E (p_1(x)p_2(x)\cdots p_r(x))^{1/r}dm \\ \leqq\sum_i\left(\int_{E_i} p_1(x)dm\int_{E_i} p_2(x)dm\cdots\int_{E_i} p_r(x)dm\right)^{1/r}\ .$$

The equality holds when and only when

$$\frac{p_1(x)}{P(E_i\,|\,p_1)}=\cdots=\frac{p_r(x)}{P(E_i\,|\,p_r)} \\ \text{a.e. in } E_i \text{ with } P(E_i\,|\,p_1)\cdots P(E_i\,|\,p_r)>0\ .$$

PROOF. We prove the theorem by mathematical induction.

(i) The case $r=2$.

By Schwarz' inequality we get

$$\int_E (p_1(x)p_2(x))^{1/2}dm=\sum_i\int_{E_i}(p_1(x)p_2(x))^{1/2}dm \\ \leqq\sum_i\left(\int_{E_i} p_1(x)dm\int_{E_i} p_2(x)dm\right)^{1/2}\ .$$

The equality holds when and only when for each i

$$p_1(x)=c_ip_2(x)\ , \qquad x\in E_i,\ c_i \text{ being a constant}\ .$$

The constant c_i can be determined from

$$P(E_i\,|\,p_1)=\int_{E_i} p_1(x)\,dm=c_i\int_{E_i} p_2(x)\,dm=c_iP(E_i\,|\,p_2)\ ,$$

that is, when $P(E_i\,|\,p_2)\neq 0$,

$$c_i=\frac{P(E_i\,|\,p_1)}{P(E_i\,|\,p_2)}\ .$$

When $P(E_i \mid p_2)=0$, we get also $P(E_i \mid p_1)=0$, consequently, $p_1(x)=0$, $p_2(x)=0$ a.e. in E_i. In this case, of course,

$$\int_{E_i} (p_1(x)p_2(x))^{1/2}dm = P(E_i \mid p_1)P(E_i \mid p_2) \qquad (=0) .$$

Therefore, the above equality holds when and only when

$$\frac{p_1(x)}{P(E_i \mid p_1)} = \frac{p_2(x)}{P(E_i \mid p_2)}$$
$$\text{a.e. in } E_i \text{ with } P(E_i \mid p_1)P(E_i \mid p_2)>0 .$$

(ii) Now, assuming that the theorem holds for $r=k-1$, we prove the theorem for $r=k$. By Hölder's inequality we get

$$\begin{aligned}
&\int_E (p_1(x)p_2(x)\cdots p_k(x))^{1/k}dm \\
&= \sum_i \int_{E_i} (p_1(x)p_2(x)\cdots p_k(x))^{1/k}dm \\
&\leqq \sum_i \left(\int_{E_i} p_1(x)dm\right)^{1/k}\left(\int_{E_i} (p_2(x)\cdots p_k(x))^{(1/k)(k/(k-1))}dm\right)^{(k-1)/k} .
\end{aligned}$$

The equality holds when and only when

$$(*) \qquad p_1(x)=c_i(p_2(x)\cdots p_k(x))^{1/(k-1)} , \qquad \text{a.e. in } E_i ,$$

where c_i are constants. According to the assumption we have

$$\begin{aligned}
&\int_{E_i} (p_2(x)\cdots p_k(x))^{1/(k-1)}dm \\
&\leqq \left(\int_{E_i} p_2(x)dm \int_{E_i} p_3(x)dm \cdots \int_{E_i} p_k(x)dm\right)^{1/(k-1)}
\end{aligned}$$

where the equality holds when and only when

$$(**) \qquad \frac{p_2(x)}{P(E_i \mid p_2)} = \cdots = \frac{p_k(x)}{P(E_i \mid p_k)}$$
$$\text{a.e. in } E_i \text{ with } P(E_i \mid p_2)\cdots P(E_i \mid p_k)>0 .$$

Thus we obtain

$$\begin{aligned}
&\int_E (p_1(x)\cdots p_k(x))^{1/k}dm \\
&\leqq \sum_i (P(E_i \mid p_1))^{1/k}[(P(E_i \mid p_2)\cdots P(E_i \mid p_k))^{1/(k-1)}]^{(k-1)/k} \\
&= \sum_i (P(E_i \mid p_1)P(E_i \mid p_2)\cdots P(E_i \mid p_k))^{1/k} .
\end{aligned}$$

The equality holds when and only when (*) and (**) hold. Assume that (*) and (**) hold, and set

$$\frac{p_2(x)}{P(E_i \mid p_2)} = \cdots = \frac{p_k(x)}{P(E_i \mid p_k)} = d_i(x)$$
$$\text{a.e. in } E_i \text{ with } P(E_i \mid p_2)\cdots P(E_i \mid p_k) > 0 \,.$$

Then we have

$$p_j(x) = P(E_i \mid p_j) d_i(x) \,, \qquad \text{a.e. in } E_i,\ j=2,\cdots,k \,,$$

and

$$p_1(x) = c_i(P(E_i \mid p_2)\cdots P(E_i \mid p_k))^{1/(k-1)} d_i(x) \,, \qquad \text{a.e. in } E_i \,.$$

Consequently, we get

$$P(E_i \mid p_j) = P(E_i \mid p_j) \int_{E_i} d_i(x) dm \,,$$

hence

$$\int_{E_i} d_i(x) dm = 1 \,,$$

and

$$\int_{E_i} p_1(x) dm = c_i (P(E_i \mid p_2)\cdots P(E_i \mid p_k))^{1/(k-1)} \int_{E_i} d_i(x) dm \,,$$

that is,

$$P(E_i \mid p_1) = c_i(P(E_i \mid p_2)\cdots P(E_i \mid p_k))^{1/(k-1)} \,.$$

Therefore, we obtain,

$$p_1(x) = P(E_i \mid p_1) d_i(x) \qquad \text{a.e. in } E_i \,.$$

Thus, when $P(E_i \mid p_1) > 0$, we get

$$\frac{p_1(x)}{P(E_i \mid p_1)} = \cdots = \frac{p_k(x)}{P(E_i \mid p_k)} = d_i(x) \qquad \text{a.e. in } E_i \,.$$

Conversely, assume that this relation holds for each E_i with $P(E_i \mid p_1) \cdots P(E_i \mid p_k) > 0$. Then we have for such an E_i

$$\int_{E_i} d_i(x) dm = 1 \,,$$

and when we notice that $P(E_i \mid p_j) = 0$ implies $p_j(x) = 0$ a.e. in E_i, we obtain

$$\int_E (p_1(x)\cdots p_k(x))^{1/k} dm$$
$$= \sum_i \int_{E_i} (p_1(x)\cdots p_k(x))^{1/k} dm$$

$$=\sum_i \int_{E_i} (P(E_i \mid p_1)\cdots P(E_i \mid p_k))^{1/k} d_i(x)dm$$
$$=\sum_i (P(E_i \mid p_1)\cdots P(E_i \mid p_k))^{1/k} .$$

Thus the proof was completed.

THEOREM 3. *It holds that*

$$\int_E (p_1(x)\cdots p_r(x))^{1/r}dm=\inf \sum_i (P(E_i \mid p_1)\cdots P(E_i \mid p_r))^{1/r}$$

where inf *is taken over all partitions* $\{E_i\}$ *of* E *such that* $E=E_1+\cdots+E_r$, $E_i \cap E_j=\phi$ $(i\neq j)$.

Especially, when $E=R$, *we have*

$$\rho_r(F_1, F_2,\cdots, F_r)=\inf \sum_i (P(E_i \mid p_1)\cdots P(E_i \mid p_r))^{1/r}$$

where inf *is taken over all partitions* $\{E_i\}$ *of* R *such that* $R=E_1+\cdots+E_r$, $E_i \cap E_j=\phi$ $(i\neq j)$.

PROOF. When there are zero-points of $p_1(x)$, let $E_0=\{x; p_1(x)=0, x\in E\}$. Then we have only to prove the theorem with respect to $E-E_0$. Therefore, in the following, we assume that $p_1(x)>0$ for any $x \in E$.

Since $\int_R (p_1(x)p_j(x))^{1/2}dm \leqq 1$, $j=1, 2,\cdots, r$, there exists a number L (>1) such that for $E_L^{(j)}=\{x; p_j(x)\geqq Lp_1(x)\}$, we have $\int_{E_L^{(j)}} (p_1(x)p_j(x))^{1/2}dm<\varepsilon$, where ε is a positive number fixed arbitrarily in advance. For $\int_{E_L^{(j)}} (p_i(x)p_j(x))^{1/2}dm\leqq \frac{1}{L^{1/2}} \int_{E_L^{(j)}} p_j(x)dm\leqq \frac{1}{L^{1/2}}$. The latter tends to zero as $L\rightarrow\infty$.

Now, for an arbitrarily fixed positive number δ, let

$$0=l_0<l_1<\cdots<l_n=L , \qquad l_{i+1}-l_i<\delta$$

and

$$E_i^{(j)}=\left\{x;\ l_i\leqq\frac{p_j(x)}{p_1(x)}<l_{i+1},\ x\in E\right\} \qquad i=0, 1,\cdots, n-1$$

and

$$E_n^{(j)}=\left\{x;\ l_n\leqq\frac{p_j(x)}{p_1(x)},\ x\in E\right\}=E\cap E_L^{(j)} .$$

Then we have

$$E=E_0^{(j)}+\cdots+E_n^{(j)}, \qquad E_i^{(j)}\cap E_k^{(j)}=\phi \qquad (i\neq k) .$$

Set

$$E_\infty=E_n^{(2)}\cup\cdots\cup E_n^{(r)}$$

and let $E_1, E_2,\cdots, E_N$ be the sets that are defined by taking intersection of $E_0^{(j)},\cdots, E_{n-1}^{(j)}$, $j=2,\cdots, r$, i.e., $E_i=E_{\nu(2,i)}^{(2)}\cap\cdots\cap E_{\nu(n-1,i)}^{(n-1)}$ with $0\leqq \nu(2, i),\cdots, \nu(n-1, i)\leqq n-1$, so that

$$E_1+E_2+\cdots+E_N=E-E_\infty\ , \qquad E_i\cap E_j=\phi \qquad (i\neq j)$$

Then we have

$$E_i\cap E_\infty=\phi\ , \qquad i=1, 2,\cdots, N$$

and

$$E=E_1+E_2+\cdots+E_N+E_\infty\ .$$

Hence, we have

$$\begin{aligned}
&\sum_{i=1}^N \inf_{x\in E_i}\left(\frac{p_2(x)}{p_1(x)}\cdots\frac{p_r(x)}{p_1(x)}\right)^{1/r}P(E_i\mid p_1)\\
&\qquad+\inf_{x\in E_\infty}\left(\frac{p_2(x)}{p_1(x)}\cdots\frac{p_r(x)}{p_1(x)}\right)^{1/r}P(E_\infty\mid p_1)\\
&\leqq\int_E\left(\frac{p_2(x)}{p_1(x)}\cdots\frac{p_r(x)}{p_1(x)}\right)^{1/r}p_1(x)dm\\
&=\int_E (p_1(x)\cdots p_r(x))^{1/r}dm\\
&=\sum_{i=1}^N\int_{E_i}(p_1(x)\cdots p_r(x))^{1/r}dm\\
&\qquad+\int_{E_\infty}(p_1(x)\cdots p_r(x))^{1/r}dm\\
&\leqq\sum_{i=1}^N (P(E_i\mid p_1)\cdots P(E_i\mid p_r))^{1/r}\\
&\qquad+\int_{E_\infty}(p_1(x)\cdots p_r(x))^{1/r}dm\ .
\end{aligned}$$

From this relation we get

$$\begin{aligned}
0&\leqq\sum_{i=1}^N (P(E_i\mid p_1)\cdots P(E_i\mid p_r))^{1/r}+\int_{E_\infty}(p_1(x)\cdots p_r(x))^{1/r}dm\\
&\quad-\int_E (p_1(x)\cdots p_r(x))^{1/r}dm\\
&\leqq\sum_{i=1}^N (P(E_i\mid p_1)\cdots P(E_i\mid p_r))^{1/r}+\int_{E_\infty}(p_1(x)\cdots p_r(x))^{1/r}dm\\
&\quad-\sum_{i=1}^N \inf_{x\in E_i}\left(\frac{p_2(x)}{p_1(x)}\cdots\frac{p_r(x)}{p_1(x)}\right)^{1/r}P(E_i\mid p_1)\\
&\quad-\inf_{x\in E_\infty}\left(\frac{p_2(x)}{p_1(x)}\cdots\frac{p_r(x)}{p_1(x)}\right)^{1/r}P(E_\infty\mid p_1)
\end{aligned}$$

$$\leqq \sum_{i=1}^{N} (P(E_i \mid p_1)\cdots P(E_i \mid p_r))^{1/r} + \int_{E_\infty} (p_1(x)\cdots p_r(x))^{1/r} dm$$

$$-\sum_{i=1}^{N} \left(\inf_{x\in E_i} \frac{p_2(x)}{p_1(x)} \cdot \inf_{x\in E_i} \frac{p_3(x)}{p_1(x)} \cdots \inf_{x\in E_i} \frac{p_r(x)}{p_1(x)} \right)^{1/r} P(E_i \mid p_1)$$

$$\leqq \sum_{i=1}^{N} [P(E_i \mid p_1)\cdot l_{\nu(2,i)+1} P(E_i \mid p_1)\cdot l_{\nu(3,i)+1} P(E_i \mid p_1)\cdots l_{\nu(r,i)+1} P(E_i \mid p_1)]^{1/r}$$

$$+\int_{E_n^{(2)}} (p_1(x)\cdots p_r(x))^{1/r} dm + \int_{E_n^{(3)}-E_n^{(2)}} (p_1(x)\cdots p_r(x))^{1/r} dm$$

$$+\int_{E^{(4)}-E^{(2)}\cup E^{(3)}} (p_1(x)\cdots p_r(x))^{1/r} dm + \cdots$$

$$+\int_{E_n^{(r)}-E_n^{(2)}\cup E_n^{(3)}\cup\cdots E_n^{(r-1)}} (p_1(x)\cdots p_r(x))^{1/r} dm$$

$$-\sum_{i=1}^{N} (l_{\nu(2,i)}\cdots l_{\nu(r,i)})^{1/r} P(E_i \mid p_1)$$

$$\leqq \sum_{i=1}^{N} [(l_{\nu(2,i)+1}\cdots l_{\nu(r,i)+1})^{1/r} - (l_{\nu(2,i)}\cdots l_{\nu(r,i)})^{1/r}] P(E_i \mid p_1)$$

$$+\left(\int_{E_n^{(2)}} (p_1(x)p_2(x))^{1/2} dm\right)^{2/r} + \left(\int_{E_n^{(3)}} (p_1(x)p_3(x))^{1/2} dm\right)^{2/r} + \cdots$$

$$+\left(\int_{E_n^{(r)}} (p_1(x)p_r(x))^{1/2} dm\right)^{2/r} *$$

$$\leqq \sum_{i=1}^{N} [(l_{\nu(2,i)+1}\cdots l_{\nu(r,i)+1})^{1/r} - (l_{\nu(2,i)}\cdots l_{\nu(r,i)})^{1/r}] P(E_i \mid p_1)$$

$$+ r\varepsilon^{2/r} .$$

In general, when $0 \leqq x < L-\delta$ $(\delta > 0)$

$$y = (x+\delta)^{1/r} - x^{1/r}$$

is a decreasing function, because $y' = [(x+\delta)^{(1/r)-1} - x^{(1/r)-1}]/r < 0$. Hence $(x+\delta)^{1/r} - x^{1/r} \leqq \delta^{1/r}$.

Accordingly, we get

$$(l_{\nu(2,i)+1} l_{\nu(3,i)+1}\cdots l_{\nu(r,i)+1})^{1/r} - (l_{\nu(2,i)} l_{\nu(3,i)}\cdots l_{\nu(r,i)})^{1/r}$$

$$\leqq ((l_{\nu(2,i)}+\delta)(l_{\nu(3,i)}+\delta)\cdots(l_{\nu(r,i)}+\delta))^{1/r} - (l_{\nu(2,i)} l_{\nu(3,i)}\cdots l_{\nu(r,i)})^{1/r}$$

$$= (l_{\nu(2,i)}+\delta)^{1/r}(l_{\nu(3,i)}+\delta)^{1/r}\cdots(l_{\nu(r,i)}+\delta)^{1/r}$$

$$- l_{\nu(2,i)}^{1/r}(l_{\nu(3,i)}+\delta)^{1/r}\cdots(l_{\nu(r,i)}+\delta)^{1/r}$$

$$+ l_{\nu(2,i)}^{1/r}(l_{\nu(3,i)}+\delta)^{1/r}\cdots(l_{\nu(r,i)}+\delta)^{1/r}$$

$$- l_{\nu(2,i)}^{1/r} l_{\nu(3,i)}^{1/r}(l_{\nu(4,i)}+\delta)^{1/r}\cdots(l_{\nu(r,i)}+\delta)^{1/r}$$

$$+ \cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots$$

* For any measurable set E, we have

$$\left(\int_E (p_1(x)p_2(x))^{1/2} dm\right)^2 \geqq \left(\int_E (p_1(x)p_2(x)p_3(x))^{1/3} dm\right)^3 \geqq \cdots \geqq \left(\int_E (p_1(x)\cdots p_r(x))^{1/r} dm\right)^r.$$

$$
\begin{aligned}
&+l_{\nu(2,i)}^{1/r}\cdots l_{\nu(r-1,i)}^{1/r}(l_{\nu(r,i)}+\delta)^{1/r}-l_{\nu(2,i)}^{1/r}\cdots l_{\nu(r-1,i)}^{1/r}l_{\nu(r,i)}^{1/r} \\
\leqq\ &\delta^{1/r}(l_{\nu(3,i)}+\delta)^{1/r}\cdots(l_{\nu(r,i)}+\delta)^{1/r} \\
&+l_{\nu(2,i)}^{1/r}\delta^{1/r}(l_{\nu(4,i)}+\delta)^{1/r}\cdots(l_{\nu(r,i)}+\delta)^{1/r} \\
&+\cdots\cdots\cdots\cdots \\
&+l_{\nu(2,i)}^{1/r}\cdots l_{\nu(r-1,i)}^{1/r}\delta^{1/r} \\
\leqq\ &(r-1)L^{(r-2)/r}\delta^{1/r}\,.
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
0\leqq\ &\sum_{i=1}^{N}(P(E_i\,|\,p_1)\cdots P(E_i\,|\,p_r))^{1/r}+(P(E_\infty\,|\,p_1)\cdots P(E_\infty\,|\,p_r))^{1/r} \\
&-\int_E (p_1(x)\cdots p_r(x))^{1/r}dm \\
\leqq\ &(r-1)L^{(r-2)/r}\delta^{1/r}\sum_{i=1}^{N}P(E_i\,|\,p_1)+r\varepsilon^{2/r} \\
\leqq\ &(r-1)L^{(r-2)/r}\delta^{1/r}+r\varepsilon^{2/r}\,.
\end{aligned}
$$

Since ε and δ can arbitrarily be chosen, this relation shows that the theorem holds true.

4. Transformation and affinity

Let $T(*)$ be a measurable transformation from the sample space $(\mathcal{X}, \mathcal{A}, m)$ to the space $(\mathcal{Y}, \mathcal{B}, \mu)$, where for any set $G\in\mathcal{B}$, we have $T^{-1}(G)\in\mathcal{A}$, and $\mu(G)=m(T^{-1}(G))$. Let p_i be probability densities in $\mathcal{X}$ with respect to m. Then, for $G\in\mathcal{B}$, the probability given by p_i is defined as

$$P(G)=P(T^{-1}(G)\,|\,p_i)=\int_{T^{-1}(G)}p_i(x)dm\,.$$

Since $\mu(G)=m(T^{-1}(G))$, $\mu(G)=0$ means $m(T^{-1}(G))=0$ and $P(T^{-1}(G)\,|\,p_i)=0$. Therefore, the above $P(G)$ is absolutely continuous with respect to μ, hence there exists a μ-measurable non-negative function $q_i(y)$ defined in $\mathcal{Y}$ such that

$$P(G)=\int_G q_i(y)d\mu\,.$$

Hereafter, we denote the above probability $P(G)$ by $P(G\,|\,q_i)$. Then we have

$$\int_G q_i(y)d\mu=\int_{T^{-1}(G)}q_i(T(x))dm=\int_{T^{-1}(G)}p_i(x)dm\,.$$

THEOREM 4. *We have*

$$\int_{\mathcal{Y}} (q_1(y)\cdots q_r(y))^{1/r} d\mu \geqq \int_{\mathcal{X}} (p_1(x)\cdots p_r(x))^{1/r} dm .$$

The equality holds when and only when

$$\frac{q_1(T(x))}{p_1(x)} = \cdots = \frac{q_r(T(x))}{p_r(x)} \qquad a.e.\ w.r.t.\ m\ in\ \mathcal{X} - E_0$$

where

$$E_0 = \{x;\ p_i(x) = 0\ for\ some\ i,\ x \in \mathcal{X}\} .$$

PROOF. Let G be a μ-measurable set in $\mathcal{Y}$ and let $G = G_1 + \cdots + G_n$, $G_i \cap G_j = \phi$ $(i \neq j)$, G_i: μ-measurable sets, be any partition of G. Then we have

$$\int_G (q_1(y)\cdots q_r(y))^{1/r} d\mu \leqq \sum_{i=1}^{n} (P(G_i \mid q_1)\cdots P(G_i \mid q_r))^{1/r} .$$

Let

$$E = T^{-1}(G) \quad \text{and} \quad E_i = T^{-1}(G_i) , \qquad i = 1, 2, \cdots, n .$$

Then

$$E = E_1 + \cdots + E_n, \qquad E_i \cap E_j = \phi \qquad (i \neq j) ,$$
$$E,\ E_i \text{ are } m\text{-measurable sets,}$$

and

$$P(G_j \mid q_i) = \int_{G_j} q_i(y) d\mu = \int_{T^{-1}(G_j)} p_i(x) dm = P(E_j \mid p_i) ,$$
$$i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, n .$$

Therefore

$$\begin{aligned}\int_E (p_1(x)\cdots p_r(x))^{1/r} dm &\leqq \sum_{j=1}^{n} (P(E_j \mid p_1)\cdots P(E_j \mid p_r))^{1/r} \\ &= \sum_{j=1}^{n} (P(G_j \mid q_i)\cdots P(G_j \mid q_r))^{1/r} .\end{aligned}$$

Hence

$$\int_E (p_1(x)\cdots p_r(x))^{1/r} dm \leqq \inf \sum_{j=1}^{n} (P(G_j \mid q_1)\cdots P(G_j \mid q_r))^{1/r} ,$$

where inf is taken over all finite partitions of G. This implies, according to Theorem 3, that

$$\int_E (p_1(x)\cdots p_r(x))^{1/r} dm \leqq \int_G (q_1(y)\cdots q_r(y))^{1/r} d\mu .$$

When $G=\mathcal{Y}$, we have $E=\mathcal{X}$ and

$$\int_{\mathcal{X}} (p_1(x)\cdots p_r(x))^{1/r}dm \leqq \int_{\mathcal{Y}} (q_1(y)\cdots q_r(y))^{1/r}d\mu .$$

Thus the first part of the theorem was proved.

Now, when

$$(*) \qquad \frac{q_1(T(x))}{p_1(x)}=\cdots=\frac{q_r(T(x))}{p_r(x)} \qquad \text{a.e. w.r.t. } m \text{ in } \mathcal{X}-E_0 ,$$

let

$$G_0=\{y\,;\,q_i(y)=0 \text{ for some } i,\ y\in\mathcal{Y}\} .$$

Then $T^{-1}(G_0)\subset E_0$ a.s. and we have

$$\begin{aligned}
&\int_{\mathcal{Y}} (q_1(y)\cdots q_r(y))^{1/r}d\mu \\
&=\int_{\mathcal{Y}-G_0} (q_1(y)\cdots q_r(y))^{1/r}d\mu \\
&=\int_{\mathcal{X}-T^{-1}(G_0)} \left(\frac{q_2(T(x))}{q_1(T(x))}\cdots\frac{q_r(T(x))}{q_1(T(x))}\right)^{1/r} p_1(x)dm \\
&=\int_{\mathcal{X}-E_0} \left(\frac{p_2(x)}{p_1(x)}\cdots\frac{p_r(x)}{p_1(x)}\right)^{1/r} p_1(x)dm \\
&=\int_{\mathcal{X}-E_0} (p_1(x)\cdots p_r(x))^{1/r}dm \\
&=\int_{\mathcal{X}} (p_1(x)\cdots p_r(x))^{1/r}dm .
\end{aligned}$$

In fact, we have

$$q_1(y)\cdots q_r(y)=0 \qquad \text{for } y\in G_0$$

and

$$\int_{G_0} (q_1(y)\cdots q_r(y))^{1/r}d\mu=0 ,$$

consequently,

$$\int_{T^{-1}(G_0)} (p_1(x)\cdots p_r(x))^{1/r}dm=0$$

as

$$\int_{G_0} (q_1(y)\cdots q_r(y))^{1/r}d\mu \geqq \int_{T^{-1}(G_0)} (p_1(x)\cdots p_r(x))^{1/r}dm ,$$

hence

$$p_1(x)\cdots p_r(x)=0 \qquad \text{a.e. in } T^{-1}(G_0) .$$

Now, we shall prove the converse. That is, when

$$(**) \qquad \int_{\mathcal{Y}} (q_1(y)\cdots q_r(y))^{1/r} d\mu=\int_{\mathcal{X}} (p_1(x)\cdots p_r(x))^{1/r} dm ,$$

we shall show that the relation (*) holds.

Let E be an m-measurable set in $\mathcal{X}$ w.r.t. m. Set

$$p_i^{(1)}(x)=\begin{cases} p_i(x) & \text{for } x\in E \\ 0 & \text{for } x\in E^c , \end{cases}$$

and

$$p_i^{(2)}(x)=\begin{cases} 0 & \text{for } x\in E \\ p_i(x) & \text{for } x\in E^c . \end{cases}$$

Then we have

$$p_i(x)=p_i^{(1)}(x)+p_i^{(2)}(x) .$$

Further, as $\int_{T^{-1}(G)} p_i^{(1)}(x)dm$, $\int_{T^{-1}(G)} p_i^{(2)}(x)dm$, defined for any μ-measurable set G in $\mathcal{Y}$, are absolutely continuous set functions w.r.t. μ (although they are not necessarily probability measures), there exist μ-measurable non-negative functions $q_i^{(1)}(y)$, $q_i^{(2)}(y)$ such that

$$\int_G q_i^{(1)}(y)d\mu=\int_{T^{-1}(G)} p_i^{(1)}(x)dm ,$$

and

$$\int_G q_i^{(2)}(y)d\mu=\int_{T^{-1}(G)} p_i^{(2)}(x)dm .$$

For these $q_i^{(1)}(y)$, $q_i^{(2)}(y)$, we can prove, quite similarly to the above case of probability measures,

$$\int_G (q_1^{(1)}(y)\cdots q_r^{(1)}(y))^{1/r}d\mu \geqq \int_{T^{-1}(G)} (p_1^{(1)}(x)\cdots p_r^{(1)}(x))^{1/r}dm$$

and

$$\int_G (q_1^{(2)}(y)\cdots q_r^{(2)}(y))^{1/r}d\mu \geqq \int_{T^{-1}(G)} (p_1^{(2)}(x)\cdots p_r^{(2)}(x))^{1/r}dm$$

for any μ-measurable set G in $\mathcal{Y}$.

On the other hand, since

$$\int_{T^{-1}(G)} p_i^{(1)}(x)dm=\int_{T^{-1}(G)\cap E} p_i(x)dm ,$$

$$\int_{T^{-1}(G)} p_i^{(2)}(x)dm = \int_{T^{-1}(G)\cap E^c} p_i(x)dm ,$$

we get

$$\int_G (q_i^{(1)}(y)+q_i^{(2)}(y))d\mu = \int_{T^{-1}(G)} p_i(x)dm .$$

This holds for any μ-measurable set G. Hence we have

$$q_i(y)=q_i^{(1)}(y)+q_i^{(2)}(y) \qquad \text{a.e. w.r.t. } \mu .$$

Consequently, we get, for any μ-measurable G,

$$\begin{aligned}
(1) \qquad & \int_G (q_1(y)\cdots q_r(y))^{1/r}d\mu \\
&= \int_G ((q_1^{(1)}(y)+q_1^{(2)}(y))\cdots(q_r^{(1)}(y)+q_r^{(2)}(y))^{1/r}d\mu \\
&\geqq \int_G [(q_1^{(1)}(y)\cdots q_r^{(1)}(y))^{1/r}+(q_1^{(2)}(y)\cdots q_r^{(2)}(y))^{1/r}]d\mu \\
&\geqq \int_{T^{-1}(G)} (p_1^{(1)}(x)\cdots p_r^{(1)}(x))^{1/r}dm + \int_{T^{-1}(G)} (p_1^{(2)}(x)\cdots p_r^{(2)}(x))^{1/r}dm \\
&= \int_{T^{-1}(G)\cap E} (p_1(x)\cdots p_r(x))^{1/r}dm + \int_{T^{-1}(G)\cap E^c} (p_1(x)\cdots p_r(x))^{1/r}dm \\
&= \int_{T^{-1}(G)} (p_1(x)\cdots p_r(x))^{1/r}dm .
\end{aligned}$$

Now, in order that

$$\int_{\mathcal{Y}} (q_1(y)\cdots q_r(y))^{1/r}d\mu = \int_{\mathcal{X}} (p_1(x)\cdots p_r(x))^{1/r}dm ,$$

it is necessary and sufficient that for any μ-measurable G in $\mathcal{Y}$ we have

$$\int_G (q_1(y)\cdots q_r(y))^{1/r}d\mu = \int_{T^{-1}(G)} (p_1(x)\cdots p_r(x))^{1/r}dm .$$

From (1) we can see that this, in turn, reduces to:

$$\begin{aligned}
(2) \qquad & ((q_1^{(1)}(y)+q_1^{(2)}(y))\cdots(q_r^{(2)}(y)+q_r^{(2)}(y))^{1/r} \\
&= (q_1^{(1)}(y)\cdots q_r^{(1)}(y))^{1/r}+(q_1^{(2)}(y)\cdots q_r^{(2)}(y))^{1/r} \qquad \text{a.e. w.r.t. } \mu ,
\end{aligned}$$

and

$$(3) \qquad \int_{\mathcal{Y}} (q_1^{(1)}(y)\cdots q_r^{(1)}(y))^{1/r}d\mu = \int_{\mathcal{X}} (p_1^{(1)}(x)\cdots p_r^{(1)}(x))^{1/r}dm ,$$

and

$$(4) \qquad \int_{\mathcal{Y}} (q_1^{(2)}(y)\cdots q_r^{(2)}(y))^{1/r}d\mu = \int_{\mathcal{X}} (p_1^{(2)}(x)\cdots p_r^{(2)}(x))^{1/r}dm .$$

When, for any particular value of y,

$$\begin{pmatrix} q_1^{(1)}(y) \\ q_1^{(2)}(y) \end{pmatrix}, \cdots, \begin{pmatrix} q_r^{(1)}(y) \\ q_r^{(2)}(y) \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

a necessary and sufficient condition that (2) hold, is that

$$(5) \qquad \begin{pmatrix} q_1^{(1)}(y) \\ q_1^{(2)}(y) \end{pmatrix}, \cdots, \begin{pmatrix} q_r^{(1)}(y) \\ q_r^{(2)}(y) \end{pmatrix} \quad \text{are all proportional*.}$$

In this case, we get

$$(6) \qquad \frac{q_i^{(1)}(y)}{q_1^{(1)}(y)} = \frac{q_i^{(2)}(y)}{q_1^{(2)}(y)} = \frac{q_i(y)}{q_1(y)}$$

provided that $q_1^{(1)}(y)$, $q_1^{(2)}(y)$ (hence $q_i^{(1)}(y)$, $q_i^{(2)}(y)$) are not equal to zero. When $q_1^{(1)}(y)=0$, we have $q_i^{(1)}(y)=0$, $i=2, 3, \cdots, r$, and $q_i^{(2)}(y)=q_i(y)$, $i=1, 2, \cdots, r$. Similarly, when $q_1^{(2)}(y)=0$, we have $q_i^{(2)}(y)=0$, $i=2, 3, \cdots, r$, and $q_i^{(1)}(y)=q_i(y)$, $i=1, 2, \cdots, r$.

Now, assume that (**) holds.

Let $G_{01}=\{y;\ q_1^{(1)}(y)=0,\ y \in \mathcal{Y}\}$. Then we get

$$\begin{aligned}
&\int_{\mathcal{Y}} (q_1^{(1)}(y)\cdots q_r^{(1)}(y))^{1/r} d\mu \\
&= \int_{\mathcal{Y}-G_{01}} (q_1^{(1)}(y)\cdots q_r^{(1)}(y))^{1/r} d\mu \\
&= \int_{\mathcal{Y}-G_{01}} \left(\frac{q_2^{(1)}(y)}{q_1^{(1)}(y)} \cdots \frac{q_r^{(1)}(y)}{q_1^{(1)}(y)} \right)^{1/r} q_1^{(1)}(y) d\mu \\
&= \int_{\mathcal{Y}-G_{01}} \left(\frac{q_2(y)}{q_1(y)} \cdots \frac{q_r(y)}{q_1(y)} \right)^{1/r} q_1^{(1)}(y) d\mu \\
&= \int_{\mathcal{X}-T^{-1}(G_{01})} \left(\frac{q_2(T(x))}{q_1(T(x))} \cdots \frac{q_r(T(x))}{q_1(T(x))} \right)^{1/r} p_1^{(1)}(x) dm \\
&= \int_{E-E\cap T^{-1}(G_{01})} \left(\frac{q_2(T(x))}{q_1(T(x))} \cdots \frac{q_r(T(x))}{q_1(T(x))} \right)^{1/r} p_1(x) dm\ .
\end{aligned}$$

Let

$$E_i = \left\{ x;\ \frac{q_j(T(x))}{q_i(T(x))} \geqq \frac{p_j(x)}{p_i(x)},\quad \begin{matrix} j=1, 2, \cdots, r, \text{ and at least for one } j \\ \text{strict inequality holds};\ x \in \mathcal{X}-E_0 \end{matrix} \right\}$$

and put, for instance, E_1 in place of the above E. Then, if $m(E_1)>0$, we obtain

$$\int_{\mathcal{Y}} (q_1^{(1)}(y)\cdots q_r^{(1)}(y))^{1/r} d\mu$$

* See, for example, Hardy, Littlewood, Polyà, *Inequalities*, p. 22.

$$=\int_{E_1}\left(\frac{q_2(T(x))}{q_1(T(x))}\cdots\frac{q_r(T(x))}{q_1(T(x))}\right)^{1/r}p_1(x)dm$$
$$>\int_{E_1}\left(\frac{p_2(x)}{p_1(x)}\cdots\frac{p_r(x)}{p_1(x)}\right)^{1/r}p_1(x)dm$$
$$\text{(notice: } T^{-1}(G_{01})\subset E_0,\ E_i\cap E_0=\phi)$$
$$=\int_{E_1}(p_1(x)\cdots p_r(x))^{1/r}dm$$
$$=\int_{\mathcal{X}}(p_1(x)\cdots p_r(x))^{1/r}dm\ .$$

Since we have (3), this presents a contradiction. Hence, when (3) holds, $m(E_1)$ must be zero. Similarly, when (3) holds, all $m(E_i)$ must be zero. This means

$$\frac{q_1(T(x))}{p_1(x)}=\cdots=\frac{q_r(T(x))}{p_r(x)} \qquad \text{a.e. w.r.t. } m \text{ in } \mathcal{X}-E_0\ .$$

COROLLARY 1. *When the set E_0 defined above has measure zero,*

$$\int_{\mathcal{Y}}(q_1(y)\cdots q_r(y))^{1/r}d\mu=\int_{\mathcal{X}}(p_1(x)\cdots p_r(x))^{1/r}dm$$

is a necessary and sufficient condition that $T(*)$ be a sufficient statistic for discriminating among $p_1(x),\cdots,p_r(x)$.

COROLLARY 2. *In order that $T(*)$ be a sufficient statistic for $p_1(x),\cdots,p_r(x)$, it is necessary and sufficient that we have*

$$\int_{\mathcal{Y}}(q_{\nu_1}(y)\cdots q_{\nu_s}(y))^{1/s}d\mu=\int_{\mathcal{X}}(p_{\nu_1}(x)\cdots p_{\nu_s}(x))^{1/s}dm$$

for any s-tuple $(\nu_1,\cdots,\nu_s)$ from $\{1,2,\cdots,r\}$ with $2\leqq s\leqq r$.

This corollary has a meaning in the necessity assertion.

COROLLARY 3. *In order that $T(*)$ be a sufficient statistic for $p_1(x),\cdots,p_r(x)$, it is necessary and sufficient that we have*

$$\int_{\mathcal{Y}}(q_{\nu_1}(y)q_{\nu_2}(y))^{1/2}d\mu=\int_{\mathcal{X}}(p_{\nu_1}(x)p_{\nu_2}(x))^{1/2}dm$$

for any pair (ν_1,ν_2) from $\{1,2,\cdots,r\}$.

5. Limit of a sequence of sufficient statistics

In this section we shall treat the problem of whether the limit of a sequence of sufficient statistics is also sufficient, by means of properties of affinity (concerning this problem, see [4]).

Let $\{M_n\}$ $(n=1, 2, \cdots)$ be a sequence of families of probability densities in the space $\mathcal{X}$ with respect to measure m, and let M_0 be a family of probability densities in $\mathcal{X}$ with respect to m such that every member $p_0(x)$ of M_0 is the limit of a sequence $\{p_n(x)\}$, where $p_n(x) \in M_n$, i.e., there exists a sequence $\{p_n(x\}$, $p_n(x) \in M_n$, such that $\lim p_n(x)=p_0(x)$ a.e. w.r.t. m. In the following, the term "w.r.t. m" will be omitted for simplicity. For a sequence $\{p_n(x)\}$, $p_n(x) \in M_n$, which converges to $p_0(x) \in M_0$ a.e., let X_n be a random variable with $p_n(x)$ $(p=1, 2, \cdots)$ and X_0 a random variable with $p_0(x)$. Then X_n tends to X_0 in law. Generally, when X_n tends to X_0 a.s., X_n tends to X_0 in law. Let $T^{(n)}(*)$ be a transformation from $\{\mathcal{X}, \mathcal{A}, m\}$ into the space $\{\mathcal{Y}, \mathcal{B}, \mu\}$ $(n=1, 2, \cdots)$. Further, assume that when $\{X_n\}$, X_n having probability density of M_n, converges to X_0, X_0 having a probability density of M_0, in law, $T^{(n)}(X_n)$ tends to $T^{(0)}(X_0)$ in law. When $T^{(n)}(X_n) \to T^{(0)}(X_0)$ a.s. for X_n tending to X_0 a.s., this assumption is satisfied. Then we have:

THEOREM 5. *When $T^{(n)}(*)$ gives a sufficient statistic for M_n $(n=1, 2, \cdots)$, so does also $T^{(0)}(*)$ for M_0.*

PROOF. Let $p_0(x)$, $q_0(x)$ be any two members of M_0, and let $\lim p_n(x) =p_0(x)$, $p_n(x) \in M_n$, and $\lim_{n\to\infty} q_n(x)=q_0(x)$, $q_n(x) \in M_n$. Further, let X_n and Y_n be random variables with probability densities $p_n(x)$ and $q_n(x)$, respectively, and denote the distributions of X_n, Y_n, and $T^{(n)}(X_n)$, $T^{(n)}(Y_n)$ by F_{X_n}, F_{Y_n}, $F_{T(X_n)}$, $F_{T(Y_n)}$, respectively. Then $F_{X_n} \to F_{X_0}$, $F_{Y_n} \to F_{Y_0}$, $F_{T^{(n)}(X_n)} \to F_{T^{(0)}(X_0)}$, $F_{T^{(n)}(Y_n)} \to F_{T^{(0)}(Y_0)}$. Since $T^{(n)}(*)$ gives a sufficient statistic for M_n, we get

$$\rho_2(F_{T^{(n)}(X_n)}, F_{T^{(n)}(Y_n)})=\rho_2(F_{X_n}, F_{Y_n}) , \qquad n=1, 2, \cdots .$$

Letting $n \to \infty$, we obtain

$$\rho_2(F_{T^{(0)}(X_0)}, F_{T^{(0)}(Y_0)})=\rho_2(F_{X_0}, F_{Y_0})$$

due to the continuity of the affinity with respect to the distributions concerned. This equality means that $T^{(0)}(*)$ gives a sufficient statistic for M_0.

6. Affinity as a measure of discrimination

As is mentioned in the Introduction, and as can be seen from the above properties, the affinity can serve as a measure of discrimination among distributions. When we are concerned with r distributions, F_1, $F_2, \cdots, F_r$, we can consider as such measures

$$\rho_2(F_1, F_2), \rho_2(F_1, F_3), \cdots, \rho_2(F_{r-1}, F_r) ;$$

$$\rho_3(F_1, F_2, F_3), \rho_3(F_1, F_2, F_4), \cdots, \rho_3(F_{r-2}, F_{r-1}, F_r):$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots ;$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots ;$$

$$\rho_r(F_1, F_2, \cdots, F_r);$$

and their weighted sums. For instance, we can consider

$$w_{12}\rho_2(F_1, F_2)+w_{13}\rho_2(F_1, F_3)+\cdots+w_{r-1r}\rho_2(F_{r-1}, F_r)$$

with $\sum w_{ij}=1$, $w_{ij}\geqq 0$, or simply

$$\rho_2(F_1, F_2)+\rho_2(F_1, F_3)+\cdots+\rho_2(F_{r-1}, F_r)\Big/\binom{r}{2}^{*}.$$

In the case when $\rho_r(F_1, F_2, \cdots, F_r)\neq 0$, $\rho_r(F_1, F_2, \cdots, F_r)$ alone will be able to serve as a quantity representing the discrepancy of the distributions. However, for two sets of distributions, $\{F_1, \cdots, F_r\}$, $\{G_1, \cdots, G_r\}$, with $\rho_r(F_1, \cdots, F_r)=\rho_r(G_1, \cdots, G_r)$, it will be necessary to consider

$$[\rho_{r-1}(F_2, \cdots, F_r)+\cdots+\rho_{r-1}(F_1, \cdots, F_{r-1})]/r$$

and

$$[\rho_{r-1}(G_2, \cdots, G_r)+\cdots+\rho_{r-1}(G_1, \cdots, G_{r-1})]/r$$

for comparison of the discrepancies of $\{F_1, \cdots, F_r\}$ and $\{G_1, \cdots, G_r\}$. When these two are equal, we will have to proceed further.

7. Decision rule

Given distributions $F_1, F_2, \cdots, F_r$ over $(\mathcal{X}, \mathcal{A}, m)$, assume that a sample x comes from one of the $\{F_i\}$, which, however, is not known to the experimenter. Further, assume that the probability that a sample comes from F_i is known to be w_i $(i=1, 2, \cdots, r)$ $(\geqq 0;\ \sum_{i=1}^{r} w_i=1)$. Then the problem is to decide from which F_i sample x comes. For that, we partition the space $\mathcal{X}$ into a set of subsets $\{E_i\}$ $(i=1, 2, \cdots, r)$ so that when $x \in E_i$ we decide on F_i, and that the error rate will be minimum. The success rate of this case is expressed as

$$\int_{E_1} w_1 p_1(x)dm+\cdots+\int_{E_r} w_r p_r(x)dm,$$

where $p_i(x)$ denotes the probability density of F_i $(i=1, 2, \cdots, r)$. There-

---

* Added in proof-reading.
The idea of weighted distance is also found in [6].

fore, the error rate becomes

$$\begin{aligned}1-&\left(\int_{E_1} w_1p_1(x)dm+\cdots+\int_{E_r} w_rp_r(x)dm\right)\\&=\int_{E_1^c} w_1p_1(x)dm+\cdots+\int_{E_r^c} w_rp_r(x)dm\\&=\int_{E_1}(w_2p_2(x)+\cdots+w_rp_r(x))dm+\cdots\\&\quad+\int_{E_r}(w_1p_1(x)+\cdots+w_{r-1}p_{r-1}(x))dm\ .\end{aligned}$$

Now, let $\{A_1,\cdots,A_r\}$ be a partition of $\mathcal{X}$ such that

(1) for $x\in A_i$, it holds that $w_ip_i(x)\geqq w_jp_j(x)$, $\quad j\neq i$,

(2) $A_i\cap A_j=\phi$ $(i\neq j)$; $\quad A_i$ are m-measurable,

and

(3) $A_1+A_2+\cdots+A_r=\mathcal{X}$.

Then we have

THEOREM 6. *The decision rule, φ, defined by the above $\{A_i\}$ is optimum for discriminating $F_1,\cdots,F_r$ in the sense that it gives the smallest error rate. The error rate is bounded from above by*

$$\sqrt{w_1w_2}\,\rho_2(F_1,F_2)+\cdots+\sqrt{w_{r-1}w_r}\,\rho_2(F_{r-1},F_r)\ ,$$

and from below by

$$\frac{(r-1)(w_1w_2\cdots w_r)\rho_r^r(F_1,F_2,\cdots,F_r)}{r^{r-1}}\ .$$

PROOF. First notice that any decision rule for discriminating F_1, $\cdots,F_r$ is defined by a partition of $\mathcal{X}$. Therefore, we can consider only partitions of $\mathcal{X}$. Let $\{E_1,\cdots,E_r\}$ be any partition of $\mathcal{X}$, i.e. E_i are m-measurable, $E_i\cap E_j=\phi$ $(i\neq j)$, and $E_1+\cdots E_r=\mathcal{X}$. Denote the decision rule defined by this partition by ψ, that is, ψ assigns x to F_i when $x\in E_i$. Then the success rate of ψ is

$$\int_{E_1} w_1p_1(x)dm+\cdots+\int_{E_r} w_rp_r(x)dm\ ,$$

which is rewritten as

$$\begin{aligned}&\int_{E_1\cap A_1} w_1p_1(x)dm+\int_{E_1\cap A_2} w_1p_1(x)dm+\cdots+\int_{E_1\cap A_r} w_1p_1(x)dm\\&\qquad+\cdots\cdots\cdots\cdots\cdots\end{aligned}$$

$$+\cdots\cdots\cdots\cdots\cdots$$
$$+\int_{E_r\cap A_1} w_r p_r(x)dm+\int_{E_r\cap A_2} w_r p_r(x)dm+\cdots+\int_{E_r\cap A_r} w_r p_r(x)dm\ ,$$

hence

$$\leqq\int_{E_1\cap A_1} w_1 p_1(x)dm+\int_{E_1\cap A_2} w_2 p_2(x)dm+\cdots+\int_{E_1\cap A_r} w_r p_r(x)dm$$
$$+\cdots\cdots\cdots\cdots\cdots$$
$$+\cdots\cdots\cdots\cdots\cdots$$
$$+\int_{E_r\cap A_1} w_1 p_1(x)dm+\int_{E_r\cap A_2} w_2 p_2(x)dm+\cdots+\int_{E_r\cap A_r} w_r p_r(x)dm$$
$$=+\int_{A_1} w_1 p_1(x)dm+\int_{A_2} w_2 p_2(x)dm+\cdots+\int_{A_r} w_r p_r(x)dm\ ,$$

which is the success rate of φ. To speak this in another way, the error rate of φ is less than or equal to that of ϕ.

The error rate of φ, $e(\varphi)$, is:

$$e(\varphi)=\int_{A_1^c} w_1 p_1(x)dm+\cdots+\int_{A_r^c} w_r p_r(x)dm$$
$$=\int_{A_2} w_1 p_1(x)dm+\cdots+\int_{A_r} w_1 p_1(x)dm$$
$$+\cdots\cdots\cdots\cdots\cdots$$
$$+\cdots\cdots\cdots\cdots\cdots$$
$$+\int_{A_1} w_r p_r(x)dm+\cdots+\int_{A_{r-1}} w_r p_r(x)dm$$

which, in turn, is

$$\leqq\int_{A_2}\sqrt{w_1 w_2 p_1(x)p_2(x)}\,dm+\cdots+\int_{A_r}\sqrt{w_1 w_r p_1(x)p_r(x)}\,dm$$
$$+\cdots\cdots\cdots\cdots\cdots$$
$$+\cdots\cdots\cdots\cdots\cdots$$
$$+\int_{A_1}\sqrt{w_1 w_2 p_1(x)p_r(x)}\,dm+\cdots+\int_{A_{r-1}}\sqrt{w_{r-1} w_r p_{r-1}(x)p_r(x)}\,dm$$
$$=\int_{A_1+A_2}\sqrt{w_1 w_2 p_1(x)p_2(x)}\,dm+\cdots+\int_{A_{r-1}+A_r}\sqrt{w_{r-1} w_r p_{r-1}(x)p_r(x)}\,dm$$
$$\leqq\sqrt{w_1 w_2}\,\rho_2(F_1, F_2)+\cdots+\sqrt{w_{r-1} w_r}\,\rho_2(F_{r-1}, F_r)\ .$$

On the other hand, we have

$$e(\varphi)=\int_{A_1}(w_2 p_2(x)+\cdots+w_r p_r(x))dm+\cdots$$
$$+\int_{A_r}(w_1 p_1(x)+\cdots+w_{r-1} p_{r-1}(x))dm$$

$$
\begin{aligned}
&\geqq (r-1)\int_{A_1}(w_2\cdots w_r p_2(x)\cdots p_r(x))^{1/(r-1)}dm+\cdots\\
&\quad+(r-1)\int_{A_r}(w_1\cdots w_{r-1}p_1(x)\cdots p_{r-1}(x))^{1/(r-1)}dm\\
&\geqq (r-1)\int_{A_1}w_1p_1(x)dm\int_{A_1}(w_2\cdots w_r p_2(x)\cdots p_r(x))^{1/(r-1)}dm\\
&\quad+\cdots\cdots\cdots\cdots\cdots\\
&\quad+(r-1)\int_{A_r}w_rp_r(x)dm\int_{A_r}(w_1\cdots w_{r-1}p_1(x)\cdots p_{r-1}(x))^{1/(r-1)}dm\\
&\geqq (r-1)\left[\left\{\int_{A_1}(w_1w_2\cdots w_r p_1(x)p_2(x)\cdots p_r(x))^{1/r}dm\right\}^r+\cdots\right.\\
&\quad\left.+\left\{\int_{A_r}(w_1w_2\cdots w_r p_1(x)p_2(x)\cdots p_r(x))^{1/r}dm\right\}^r\right]\\
&\geqq (r-1)(w_1w_2\cdots w_r)\frac{\rho_r^r(F_1,F_2,\cdots,F_r)}{r^{r-1}}\ .
\end{aligned}
$$

Thanks are due to K. Takahasi for his useful comments.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications, *Ann. Inst. Statist. Math.*, **19**, 181-192.

[2] Kirmani, S. N. U. A. (1968). Some results on Matusita's measure of distance, *Jour. Indian Statist. Assoc.*, **6**, 89-98.

[3] Kirmani, S. N. U. A. (1971). Some limiting properties of Matusita's measure of distance, this issue of the *Ann. Inst. Statist. Math.*, **23**, 157-162.

[4] Kudo, H. (1970). On an approximation to a sufficient statistic including a concept of asymptotic sufficiency, *Journ. Faculty of Science, Univ. of Tokyo*, Section 1, **17**, 273-290.

[5] Kullback, S. (1959). *Information Theory and Statistics*, John Wiley, New York.

Added in proof-reading.

[6] Glick, Ned. Optimum correct classification probability as an indicator of distance between or separation among distributions, submitted to the *Ann. Inst. Statist. Math.*