ON THE NOTION OF AFFINITY OF SEVERAL DISTRIBUTIONS AND SOME OF ITS APPLICATIONS

KAMEO MATUSITA

(Received Feb. 20, 1967)

1. Introduction

The notion of distance between distributions often plays an important and useful role in statistics, and thus far, the author treated decision rules based on the distance in many occasions (see the references of [1]). Among many definitions of distance between distributions the author especially dealt with the following.

Let F_1 and F_2 be two distributions defined on the same space R, and let the probabilities of any (measurable) set E according to F_1 and F_2 be represented as

$$\int_E p_1(x)\,dm \qquad \text{and} \qquad \int_E p_2(x)\,dm\ , \qquad p_1(x),\ p_2(x) \geqq 0\ ,$$

respectively, where m denotes a measure (Lebesgue or counting or mixed) defined in R. Then, the distance between F_1 and F_2 is:

$$d_2(F_1, F_2) = \left[\int_R (\sqrt{p_1(x)} - \sqrt{p_2(x)}\,)^2\,dm\right]^{1/2}.$$

Employing this distance, we can treat problems of decision, estimation, and hypothesis testing. (As to the properties of this distance, see [2], for instance.) With this distance is closely associated the quantity

$$\rho_2(F_1, F_2) = \int_R \sqrt{p_1(x)p_2(x)}\,dm\ .$$

In fact, we have

$$d_2^2(F_1, F_2) = 2(1 - \rho_2(F_1, F_2))\ .$$

As is easily seen, $\rho_2(F_1, F_2)$ has the following properties:

(i) $0 \leqq \rho_2(F_1, F_2) \leqq 1$,

(ii) $\rho_2(F_1, F_2)=1$ when and only when $F_1=F_2$,

(iii) For a sequence of distributions $\{F_n\}$, $\rho(F_n, F_0)\to 1$ means that for any measurable set E, $F_n(E)\to F_0(E)$ uniformly in E, where $F_n(E)$, $F_0(E)$ respectively denote the probabilities of E according to F_n, F_0, and vice versa.

In this way, $\rho(F_1, F_2)$ represents the likeness of F_1 and F_2, and we have called it the *affinity* of F_1 and F_2. This notion of affinity can be used in place of the distance $d_2(\cdot, \cdot)$ (see [2], [3]).

The purpose of this paper is to extend the notion of affinity concerning two distributions to the case of several distributions, and to show some of its applications. In section 2 the notion of affinity of several distributions is introduced, and some of its properties are stated. In section 3 decision rules based on the affinity are given, and in section 4 its applications in multivariate analysis are mentioned. Some familiar expressions in multivariate analysis, for instance, the likelihood ratio for testing several covariance matrices being equal, appear as special cases.

2. Affinity of several distributions

Let $F_1, F_2, \cdots, F_r$ be distributions defined on the same space R with measure m, 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 define the affinity of $F_1, F_2, \cdots, F_r$ as

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

THEOREM 1. *We have*

$$0\leqq \rho_r^r(F_1, F_2, \cdots, F_r)\leqq \rho_{r-1}^{r-1}(F_{i_1}, F_{i_2}, \cdots, F_{i_{r-1}})\leqq \cdots \leqq \rho_2^2(F_{i_s}, F_{i_t})\leqq 1\ ,$$

where $(i_1, \cdots, i_{r-1})$ *is any set of* $r-1$ *integers out of* $(1, 2, \cdots, r)$ *and*

$$(i_s, i_t)\subset \cdots \subset (i_1, i_2, \cdots, i_{r-1})\subset (1, 2, \cdots, r)\ .$$

PROOF.

As $p_i(x)\geqq 0$, it is obvious that

$$0\leqq \int_R (p_1(x) \cdots p_r(x))^{1/r}\, dm\ .$$

Now, by Hölder's inequality, we have

$$\int_R (p_1(x) \cdots p_r(x))^{1/r} dm \leqq \left(\int_R p_1(x) dm\right)^{1/r} \left(\int_R (p_2(x) \cdots p_r(x))^{[1/r]\cdot[r/(r-1)]} dm\right)^{(r-1)/r}$$
$$= \left(\int_R (p_2(x) \cdots p_r(x))^{1/(r-1)} dm\right)^{(r-1)/r},$$

that is,

$$\left(\int_R (p_1(x) \cdots p_r(x))^{1/r} dm\right)^r \leqq \left(\int_R (p_2(x) \cdots p_r(x))^{1/(r-1)} dm\right)^{r-1}.$$

Thus, we have

$$\left(\int_R (p_1(x) \cdots p_r(x))^{1/r} dm\right)^r \leqq \left(\int_R (p_2(x) \cdots p_r(x))^{1/(r-1)} dm\right)^{r-1} \leqq \cdots$$
$$\leqq \left(\int_R (p_{r-1}(x) p_r(x))^{1/2} dm\right)^2 \leqq 1.$$

COROLLARY 1. *It holds that*

$$(\rho_r(F_1, F_2, \cdots, F_r))^r = (\rho_{r-1}(F_2, F_3, \cdots, F_r))^{r-1}$$

when and only when

$$p_1(x) = (p_2(x) \cdots p_r(x))^{1/(r-1)} \frac{1}{\int_R (p_2(x) \cdots p_r(x))^{1/(r-1)} dm}$$

except for a set of m-measure 0.

This can be shown from the condition that the equality sign hold in Hölder's inequality.

COROLLARY 2.

$$\rho_r(F_1, F_2, \cdots, F_r) \leqq \min_{(i,j)} (\rho_2(F_i, F_j))^{2/r}.$$

This is an immediate consequence of Theorem 1.

COROLLARY 3.

$$\rho_r(F_1, F_2, \cdots, F_r) = 1$$

when and only when $F_1 = F_2 = \cdots = F_r$.

When $F_1 = \cdots = F_r$, it is obvious that $\rho_r(F_1, F_2, \cdots, F_r) = 1$. Conversely, when $\rho_r(F_1, F_2, \cdots, F_r) = 1$, we have $\rho_2(F_i, F_j) = 1$ for any pair (i, j) $(i, j = 1, 2, \cdots, r)$, from which it follows that $F_1 = F_2 = \cdots = F_r$.

As mentioned above, the affinity of two distributions F_1, F_2 is connected to the distance $d_2(F_1, F_2) = \left(\int_R (\sqrt{p_1(x)} - \sqrt{p_2(x)})^2 dm\right)^{1/2}$ by an equality. The affinity of r distributions has connection to the distance $d_r(F_1, F_2)=\left|\int_R (p_1^{1/r}(x)-p^{1/r}(x))^r\, dm\right|^{1/r}$. However, the relation is not expressed in terms of equality as the case of two distributions, but in terms of inequality.

THEOREM 2. *It holds that*

$$\rho_r(F_1, F_2, \cdots, F_r)\geqq 1-(r-1)\delta\ ,$$

when, for any pair (i, j) $(i, j=1, 2, \cdots, r)$, *we have* $d_r(F_i, F_j)\leqq\delta$.

PROOF.

$$\begin{aligned}
1-&\int(p_1\cdots p_r)^{1/r}\,dm\\
&=\left[\int(\underbrace{p_1\cdots p_1}_{r})^{1/r}\,dm-\int(\underbrace{p_1\cdots p_1}_{r-1}p_r)^{1/r}\,dm\right]\\
&\quad+\left[\int(\underbrace{p_1\cdots p_1}_{r-1}p_r)^{1/r}\,dm-\int(\underbrace{p_1\cdots p_1}_{r-2}p_{r-1}p_r)^{1/r}\,dm\right]\\
&\quad+\cdots\cdots\cdots\cdots\cdots\\
&\quad+\left[\int(p_1p_1p_3\cdots p_r)^{1/r}\,dm-\int(p_1p_2\cdots p_r)^{1/r}\,dm\right]\\
&=\int(p_1^{1/r}-p_r^{1/r})(p_1\cdots p_1)^{1/r}\,dm+\int(p_1^{1/r}-p_{r-1}^{1/r})(p_1\cdots p_1p_r)^{1/r}\,dm\\
&\quad+\cdots+\int(p_1^{1/r}-p_2^{1/r})(p_1p_3\cdots p_r)^{1/r}\,dm\\
&\leqq\left(\int|p_1^{1/r}-p_r^{1/r}|^r\,dm\right)^{1/r}\left(\int(\underbrace{p_1\cdots p_1}_{r-1})^{[1/r]\cdot[r/(r-1)]}dm\right)^{(r-1)/r}\\
&\quad+\left(\int|p_1^{1/r}-p_{r-1}^{1/r}|^r\,dm\right)^{1/r}\left((\underbrace{p_1\cdots p_1}_{r-2}p_r)^{[1/r]\cdot[r/(r-1)]}dm\right)^{(r-1)/r}\\
&\quad+\cdots+\left(\int|p_1^{1/r}-p_2^{1/r}|^r\,dm\right)^{1/r}\left(\int(p_1p_3\cdots p_r)^{[1/r]\cdot[r/(r-1)]}dm\right)^{(r-1)/r}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{(by Hölder's inequality)}\\
&\leqq\left(\int|p_1^{1/r}-p_r^{1/r}|^r\,dm\right)^{1/r}+\left(\int|p_1^{1/r}-p_{r-1}^{1/r}|^r\,dm\right)^{1/r}\\
&\quad+\cdots+\left(\int|p_1^{1/r}-p_2^{1/r}|^r\,dm\right)^{1/r}\\
&\leqq(r-1)\delta\ .
\end{aligned}$$

Now, the distance $d_1(F_1, F_2)$ has a direct meaning for the probabilities of any set E according to F_1 and F_2, and as stated concerning

$\rho_2(F_1, F_2)$ in the introduction, $d_2(F_1, F_2)$ also has a probabilistic meaning. In the following we shall show that the same can be said for $d_r(F_1, F_2)$ $(r>1)$, by establishing some relations between $d_r(F_1, F_2)$ $(r>1)$ and $d_1(F_1, F_2)$.

THEOREM 3. *It holds that*

$$d_1(F_1, F_2) \leqq r \cdot d_r(F_1, F_2) .$$

PROOF. Let f_1, f_2 be non-negative, measurable functions such that $f_1^{r/(r-1)}$, $f_2^{r/(r-1)}$ are also measurable. Then, we have

$$\int_R |f_1^r - f_2^r|\, dm = \int_R |f_1 - f_2| (f_1^{r-1} + f_1^{r-2} f_2 + \cdots f_2^{r-1})\, dm$$
$$\leqq \left(\int_R |f_1 - f_2|^r\, dm\right)^{1/r} \cdot \left(\int_R (f_1^{r-1} + f_1^{r-2} f_2 + \cdots + f_2^{r-1})^{r/(r-1)}\, dm\right)^{(r-1)/r}$$

(by Hölder's inequality).

Since $\left(\int_R |f|^{r/(r-1)}\, dm\right)^{(r-1)/r}$ makes a norm in the $L_{r/(r-1)}$ space, we have

$$\left(\int_R (f_1^{r-1} + f_1^{r-2} f_2 + \cdots + f_2^{r-1})^{r/(r-1)}\, dm\right)^{(r-1)/r}$$
$$\leqq \sum_{1=0}^{r-1} \left(\int_R (f_1^{r-1-i} f_2^i)^{r/(r-1)}\, dm\right)^{(r-1)/r} .$$

Further, we have

$$\int_R (f_1^{r-1-i} f_2^i)^{r/(r-1)}\, dm = \int_R f_1^{r(r-1-i)/(r-1)} f_2^{ri/(r-1)}\, dm$$
$$\leqq \left(\int f_1^{[r(r-1-i)/(r-1)]\cdot[(r-1)/(r-1-i)]}\, dm\right)^{(r-1-i)/(r-1)} \left(\int_R f_2^{[ri/(r-1)]\cdot[(r-1)/i]}\, dm\right)^{i/(r-1)}$$
$$= \left(\int_R f_1^r\, dm\right)^{(r-1-i)/(r-1)} \left(\int_R f_2^r\, dm\right)^{i/(r-1)} ,$$

hence,

$$\left(\int_R (f_1^{r-1-i} f_2^i)^{r/(r-1)} dm\right)^{(r-1)/r} \leqq \left(\int_R f_1^r\, dm\right)^{(r-1-i)/r} \left(\int_R f_2^r\, dm\right)^{i/r} .$$

Therefore, we obtain

$$\int_R |f_1^r - f_2^r|\, dm$$
$$\leqq \left(\int_R |f_1 - f_2|^r\, dm\right)^{1/r} \sum_{i=0}^{r-1} \left(\int_R f_1^r\, dm\right)^{(r-1-i)/r} \left(\int_R f_2^r\, dm\right)^{i/r} .$$

When we put $f_1 = p_1^{1/r}$, $f_2 = p_2^{1/r}$, we have

$$\int_R |p_1-p_2|\,dm \leqq \left(\int_R |p_1^{1/r}-p_2^{1/r}|^r\,dm\right)^{1/r}\cdot\sum_{i=0}^{r-1}1=r\left(\int_R |p_1^{1/r}-p_2^{1/r}|^r\,dm\right)^{1/r},$$

which completes the proof.

THEOREM 4. *It holds that*

$$d_1(F_1, F_2)\geqq(d_r(F_1, F_2))^r .$$

PROOF. Generally, for any two real numbers a, b with $a\geqq b\geqq 0$, we have $a^r-b^r\geqq(a-b)^r$. Therefore, when we set

$$E_1=\{x\mid p_1(x)\geqq p_2(x)\}\ ,\qquad E_2=\{x\mid p_1(x)<p_2(x)\}\ ,$$

we have

$$p_1(x)-p_2(x)\geqq(p_1^{1/r}(x)-p_2^{1/r}(x))^r\qquad \text{on}\quad E_1 ,$$

and

$$-p_1(x)+p_2(x)\geqq(-p_1^{1/r}(x)+p_2^{1/r}(x))^r\qquad \text{on}\quad E_2 ,$$

from which it follows that

$$\int_R |p_1(x)-p_2(x)|\,dm\geqq\int_R |p_1^{1/r}(x)-p_2^{1/r}(x)|^r\,dm .$$

Through Theorems 3 and 4 we can know about the relation of $d_r(F_1, F_2)$ to the difference of F_1 and F_2.

As a generalization of Theorem 4, we have

THEOREM 5. *It holds that*

$$(d_{r-1}(F_1, F_2))^{r-1}\geqq(d_r(F_1, F_2))^r .$$

Consequently, we have

$$d_1(F_1, F_2)\geqq(d_2(F_1, F_2))^2\geqq\cdots\geqq(d_r(F_1, F_2))^r .$$

PROOF. Let $y=(x^r-1)^{1/r}$, $x>1$. Then, y is an increasing function of r. Hence,

$$(x^r-1)^{1/r}>(x^{r-1}-1)^{1/(r-1)}\qquad (r>1) .$$

Let $x=u^{1/r(r-1)}$, $u>1$. Then we have $(u^{1/(r-1)}-1)^{1/r}>(u^{1/r}-1)^{1/(r-1)}$, that is, $(u^{1/(r-1)}-1)^{r-1}>(u^{1/r}-1)^r$. Therefore, for x with $p_1(x)>p_2(x)>0$, we get

$$\left(\left(\frac{p_1}{p_2}\right)^{1/(r-1)}-1\right)^{r-1}>\left(\left(\frac{p_1}{p_2}\right)^{1/r}-1\right)^r ,$$

consequently,

$$(p_1^{1/(r-1)}-p_2^{1/(r-1)})^{r-1}>(p_1^{1/r}-p_2^{1/r})^r .$$

For x with $p_2(x)=0$ or $p_1(x)=p_2(x)$, we have

$$(p_1^{1/(r-1)}-p_2^{1/(r-1)})^{r-1}=(p_1^{1/r}-p_2^{1/r})^r .$$

Therefore, we obtain

$$\int_{E_1}(p_1^{1/(r-1)}-p_2^{1/(r-1)})^{r-1}\,dm \geqq \int_{E_1}(p_1^{1/r}-p_2^{1/r})^r\,dm$$

where $E_1=\{x \mid p_1(x)\geqq p_2(x)\}$. Similarly, we obtain

$$\int_{E_2}(p_2^{1/(r-1)}-p_1^{1/(r-1)})^{r-1}\,dm \geqq \int_{E_2}(p_2^{1/r}-p_1^{1/r})^r\,dm ,$$

where $E_2=\{x \mid p_1(x)<p_2(x)\}$. Hence we get

$$\int_R |p_1^{1/(r-1)}-p_2^{1/(r-1)}|^{r-1}\,dm \geqq \int_R |p_1^{1/r}-p_2^{1/r}|^r\,dm .$$

Combining the above results, we have

THEOREM 6. *When $d_2(F_i, F_j)\leqq\delta_0$ for any pair (i, j) $(i\neq j,\ i, j=1, 2, \cdots, r)$ and $\rho_0=\min\rho_2(F_i, F_j)$, it holds that*

$$1-(r-1)2^{1/r}\delta_0^{1/r}\leqq\rho_r(F_1, F_2, \cdots, F_r)\leqq\rho_0^{2/r} .$$

In fact,

$$(d_r(F_i, F_j))^r\leqq d_1(F_i, F_j)\leqq 2d_2(F_i, F_j)\leqq 2\delta_0$$

and the inequality follows from Theorem 2. The second inequality is Corollary 2 to Theorem 1 itself.

3. Decision rules based on the affinity

To apply the notion of affinity to statistical problems, we have to consider the affinity of distributions which are respectively determined by samples from the distributions under consideration. Let F_i $(i=1, 2, \cdots, r)$ be the unknown distributions under consideration and $p_i(x)$ the density function of F_i with respect to a measure m on R, as above. Further, let S_i be the distribution determined by a sample from F_i, and $q_i(x)$ the density function of S_i with respect to m. The way S_i is determined by a sample depends on the functional form of F_i. When the functional form is unknown, we substitute some appropriate or reasonable form for it, considering the nature of the distribution. Then we can consider the affinity of $S_1, S_2, \cdots, S_r$,

$$\rho_r(S_1, S_2, \cdots, S_r)=\int_R (q_1(x)\, q_2(x) \cdots q_r(x))^{1/r}\, dm .$$

When the problem is to decide whether or not $F_1=F_2=\cdots=F_r$, taking some value ρ_0, we make a decision as follows.

We decide that $F_1=F_2=\cdots=F_r$ when $\rho_r(S_1, S_2, \cdots, S_r)\geqq\rho_0$,

We decide that $F_1, F_2, \cdots, F_r$ are not identical when $\rho_r(S_1, S_2, \cdots, S_r)<\rho_0$.

The decision errors are evaluated when the distribution of $\rho_r(S_1, S_2, \cdots, S_r)$ is known. Also in the case when, for arbitrary $\varepsilon, \delta>0$, we can choose the number of observations, n_0, so that we have $\mathrm{P}(d_2(F_i, S_i)<\delta, i=1, \cdots, r)\geqq 1-\varepsilon$, S_i being based on a sample of $n_i>n_0$, we can control the errors. To show that exactly, however, we have to modify the problem as follows: Decide whether $F_2=\cdots=F_r$ or the distance between some pair of $F_1, F_2, \cdots, F_r$ is greater than or equal to a given positive constant. In fact,

$$1-(r-1)2^{1/r}\delta_0^{1/r}\leqq\rho_r(S_1, S_2, \cdots, S_r)\leqq\rho_2^{2/r} ,$$

when $\delta_0=\max_{(i,j)} d_2(S_i, S_j)$ and $\rho_2=\min_{(i,j)} \rho(S_i, S_j)$. On the other hand, we have

$$d_2(S_i, S_j)\leqq d_2(F_i, S_i)+d_2(F_j, S_j)+d_2(F_i, F_j) ,$$

$$d_2(S_i, S_j)\geqq d_2(F_i, F_j)-d_2(F_i, S_i)-d_2(F_j, S_j) .$$

Suppose that the sample size n_i is chosen for $\delta_1>0$ so that

$$\mathrm{P}(d_2(F_i, S_i)<\delta_1,\ i=1, 2, \cdots, r)\geqq 1-\varepsilon .$$

Then, when $F_1=F_2=\cdots=F_r$, we have

$$\mathrm{P}(d_2(S_i, S_j)<2\delta_1,\ i, j=1, 2, \cdots, r)\geqq 1-\varepsilon ,$$

hence,

$$\mathrm{P}(\rho_r(S_1, S_2, \cdots, S_r)\geqq 1-(r-1)2^{2/r}\delta_1^{1/r})\geqq 1-\varepsilon .$$

When $F_1, F_2, \cdots, F_r$ are not identical, say $d_2(F_\mu, F_\nu)=\delta'>0$, we have

$$d_2(S_\mu, S_\nu)\geqq\delta'-d_2(F_\mu, S_\mu)-d_2(F_\nu, S_\nu) .$$

Suppose that we have taken δ_1 so that $\delta'>2\delta_1$. Then

$$\mathrm{P}(d_2(S_\mu, S_\nu)\geqq\delta'-2\delta_1)\geqq 1-\varepsilon ,$$

hence,

$$P(\max d_2(S_i, S_j) \geqq \delta' - 2\delta_1) \geqq 1 - \varepsilon ,$$

consequently,

$$P\left(\rho_r(S_1, S_2, \cdots, S_r) \leqq \left(1 - \frac{(\delta' - 2\delta_1)^2}{2}\right)^{1/r}\right) \geqq 1 - \varepsilon .$$

Therefore, when δ_1 satisfies

$$\left(1 - \frac{(\delta' - 2\delta_1)^2}{2}\right)^{1/r} \leqq 1 - (r-1)2^{2/r}\delta_1^{1/r}$$

besides $\delta' > 2\delta_1$, $1 > (r-1)2^{2/r}\delta_1$, taking ρ_0 such that

$$\left(1 - \frac{(\delta' - 2\delta_1)^2}{2}\right)^{1/r} \leqq \rho_0 \leqq 1 - (r-1)2^{2/r}\delta_1^{1/r} ,$$

we obtain

$$P(\rho_r(S_1, S_2, \cdots, S_r) \geqq \rho_0 \,|\, F_1 = F_2 = \cdots = F_r) \geqq 1 - \varepsilon ,$$

and

$$P(\rho_r(S_1, S_2, \cdots, S_r) < \rho_0 \,|\, d_2(F_\mu, F_\nu) \geqq \delta' \text{ for some } (\mu, \nu)) \geqq 1 - \varepsilon .$$

In this way the decision errors are controlled to be less than ε.

When sample sizes n_i are fairly different, the order of approximation of S_i to F_i also differs from distribution to distribution, so, in this case, when each n_i is not so large as S_i can be considered to give a satisfactory approximate distribution to F_i, we put a weight on S_i in the affinity, which depends on n_i. For instance, we take the weighted affinity

$$\rho_{rw}(S_1, S_2, \cdots, S_r) = \rho_n(\underbrace{S_1, \cdots, S_1}_{n_1}, \underbrace{S_2, \cdots, S_2}_{n_2}, \cdots, \underbrace{S_r, \cdots, S_r}_{n_r}) ,$$

where $n = n_1 + n_2 + \cdots + n_r$.

4. Applications in multivariate analysis*)

Let $F_1, F_2, \cdots, F_r$ be respectively k-dimensional Gaussian distributions $N(a_1, A_1^{-1})$, $N(a_2, A_2^{-1})$, $\cdots$, $N(a_r, A_r^{-1})$, where $a_1, a_2, \cdots, a_r$ are k-dimensional vectors and $A_1, A_2, \cdots, A_r$ are positive definite (symmetric) $(k \times k)$-matrices. Then we have

$$\rho_r(F_1, F_2, \cdots, F_r) = \frac{\prod_i |A_i|^{1/2r}}{\left|\frac{1}{r}\sum_i A_i\right|^{1/2}} \exp\left[\frac{1}{2r}\left\{(\sum_i A_i a_i, (\sum_i A_i)^{-1} \sum_i A_i a_i) - \sum_i (A_i a_i, a_i)\right\}\right] .$$

*) Concerning this section see [4], [5].

Especially, when $A_1=A_2=\cdots=A_r=A$ (say),

$$\rho_r(F_1, F_2, \cdots F_r)=\exp\left[\frac{1}{2r}\left\{(A\sum_i a_i, r^{-1}\sum_i a_i)-\sum_i (Aa_i, a_i)\right\}\right],$$

and when $a_1=a_2=\cdots=a_r$,

$$\rho_r(F_1, F_2, \cdots, F_r)=\frac{\prod_i |A_i|^{1/2r}}{\left|\frac{1}{r}\sum_i A_i\right|^{1/2}},$$

Now, assume that $F_1, F_2, \cdots, F_r$ are unknown. The problem is to decide whether or not $F_1, F_2, \cdots, F_r$ are identical. Let $x_1^{(i)}, x_2^{(i)}, \cdots, x_{n_i}^{(i)}$ be a sample from $N(a_i, A_i^{-1})$, and

$$\bar{x}^{(i)}=\frac{1}{n_i}\sum_{\alpha=1}^{n_i} x_\alpha^{(i)}$$

$$V_i=\frac{1}{n_i-1}\sum_{\alpha=1}^{n_i}(x_\alpha^{(i)}-\bar{x}^{(i)})(x_\alpha^{(i)}-\bar{x}^{(i)})',$$

$$U_i=V_i^{-1},$$

$$U=\frac{1}{r}\sum_{i=1}^{r} U_i.$$

When $A_1=A_2=\cdots=A_r=A$ but A is unknown, we consider

$$\rho_{r1}=\exp\left[\frac{1}{2r}\left\{\left(U\sum_i \bar{x}^{(i)}, \frac{1}{r}\sum_i \bar{x}^{(i)}\right)-\sum_i (U\bar{x}^{(i)}, \bar{x}^{(i)})\right\}\right]$$

as a test statistic for deciding whether or not $a_1=a_2=\cdots=a_r$.

When $A_1=A_2=\cdots=A_r=A$, and A itself is known, we consider

$$\rho_{r2}=\exp\left[\frac{1}{2r}\left\{\left(A\sum_i \bar{x}^{(i)}, \frac{1}{r}\sum_i \bar{x}^{(i)}\right)-\sum_i (A\bar{x}^{(i)}, \bar{x}^{(i)})\right\}\right].$$

As a statistic for deciding whether or not $A_1=A_2=\cdots=A_r$, we consider

$$\rho_{r3}=\frac{\prod_i |U_i|^{1/2r}}{\left|\frac{1}{r}\sum_i U_i\right|^{1/2}}.$$

As a statistic for deciding whether or not $a_1=a_2=\cdots=a_r$ and $A_1=A_2=\cdots=A_r$ at the same time, we consider $\rho_r=\rho_{r1}\rho_{r3}$.

When we take the logarithm of ρ_{r1} or ρ_{r2}, we obtain

$$-2r \log \rho_{r1} = \sum_i (U\bar{x}^{(i)}, \bar{x}^{(i)}) - \left(U\sum_i \bar{x}^{(i)}, \frac{1}{r}\sum_i \bar{x}^{(i)}\right)$$

$$= \sum (U(\bar{x}^{(i)} - \bar{\bar{x}}), (\bar{x}^{(i)} - \bar{\bar{x}}))$$

and

$$-2r \log \rho_{r2} = \sum_i (A\bar{x}^{(i)}, \bar{x}^{(i)}) - \left(A\sum_i \bar{x}^{(i)}, \frac{1}{r}\sum_i \bar{x}^{(i)}\right)$$

$$= \sum (A(\bar{x}^{(i)} - \bar{\bar{x}}), (\bar{x}^{(i)} - \bar{\bar{x}})) ,$$

where $\bar{\bar{x}} = \frac{1}{r}\sum \bar{x}^{(i)}$.

Now, since the distributions of $\rho_{2\alpha}(F_i, S_i)$ $(\alpha = 1, 2, 3)$ are obtained (see [4]), we can evaluate the decision errors when employing the above test statistics.

In the case where n_i are not so large and are fairly different from each other, we consider the weighted affinity

$$\rho_{rw} = \rho_{rw}(S_1, \cdots, S_r) = \rho_n(\underbrace{S_1, \cdots, S_1}_{n_1}, \underbrace{S_2, \cdots, S_2}_{n_2}, \cdots, \underbrace{S_r, \cdots, S_r}_{n_r})$$

$$= \frac{\prod_{i=1}^{r} |U_i|^{n_i/2n}}{\left|\frac{1}{n}\sum_{i=1}^{r} n_i U_i\right|} \exp\left[\frac{1}{2n}\left\{\left(\sum_i n_i U_i \bar{x}^{(i)}, (\sum_i n_i U_i)^{-1} \sum_i n_i U_i \bar{x}^{(i)}\right)\right.\right.$$

$$\left.\left. - \sum n_i (U_i \bar{x}^{(i)}, \bar{x}^{(i)})\right\}\right]$$

where $n = n_1 + n_2 + \cdots + n_r$. Hence, when $A_1 = A_2 = \cdots = A_r = A$, but A is unknown, we consider

$$\rho_{rw1} = \exp\left[\frac{1}{2n}\left\{\left(\hat{U}\sum_i n_i \bar{x}^{(i)}, \frac{1}{n}\sum_i n_i \bar{x}^{(i)}\right) - \sum_i n_i (\hat{U}\bar{x}^{(i)}, \bar{x}^{(i)})\right\}\right] ,$$

where $\hat{U} = \frac{1}{n}\sum_i n_i U_i$ and when $A_1 = A_2 = \cdots = A_r = A$ and A is known,

$$\rho_{rw2} = \exp\left[\frac{1}{2n}\left\{\left(A\sum_i n_i \bar{x}^{(i)}, \frac{1}{n}\sum_i n_i \bar{x}^{(i)}\right) - \sum_i n_i (A\bar{x}^{(i)}, \bar{x}^{(i)})\right\}\right] ,$$

Taking the logarithm of these ρ_{rw1}, ρ_{rw2}, we obtain

$$-2n \log \rho_{rw1} = \sum_{i,j} (\hat{U}(x_j^{(i)} - x_0), (x_j^{(i)} - x_0))$$

and

$$-2n \log \rho_{rw2} = \sum_{i,j} (A(x_j^{(i)} - x_0), (x_j^{(i)} - x_0)) ,$$

where $x_0 = \frac{1}{n}\sum_i x_j^{(i)}$. These statistics have known distributions (F- or χ^2-) under the null hypothesis. Further, corresponding to ρ_{r3}, we consider

$$\rho_{rw3} = \frac{\prod_{i=1}^{r} |U_i|^{n_i/2n}}{\left|\frac{1}{n}\sum_i n_i U_i\right|^{1/2}} .$$

When $\frac{n_i}{n_i - 1} \doteqdot 1$, we can take

$$\rho'_{rw3} = \frac{\prod_i |U_i|^{n_i/2n}}{\left|\frac{1}{n}\sum_i (n_i - 1) U_i\right|^{1/2}} ,$$

or

$$\rho''_{rw3} = \frac{\prod_i |U_i|^{(n_i-1)/2n}}{\left|\frac{1}{n}\sum_i (n_i - 1) U_i\right|^{1/2}} .$$

These two statistics are essentially the same as the likelihood ratio or a modified one for the problem under consideration. In other words, the likelihood ratio in this case can be interpreted as representing the affinity of distributions.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] K. Matusita, "Distance and decision rules," *Ann. Inst. Statist. Math.*, 16 (1964), 305-315.
[2] K. Matusita, "Decision rules based on the distance for problems of fit, two samples and estimation," *Ann. Math. Statist.*, 26 (1955), 613-640.
[3] K. Matusita and H. Akaike, "Decision rules based on the distance for the problems of independence, invariance and two samples," *Ann. Inst. Statist. Math.*, 7 (1956), 67-80.
[4] K. Matusita, "A distance and related statistics in multivariate analysis," *Multivariate Analysis* (ed. by P. R. Krishnaiah), Academic Press, New York, (1966), 187-200.
[5] K. Matusita, "Classification based on distance in multivariate Gaussian cases," *Fifth Berkeley Symposium on Math. Statist. and Prob.*, Univ. Calif. Press, 1967.