SOME PROPERTIES OF MATUSITA'S MEASURE OF AFFINITY OF SEVERAL DISTRIBUTIONS

GODFRIED T. TOUSSAINT

(Received March 17, 1972)

1. Introduction

Let $F_1, F_2, \cdots, F_r$ be distributions defined on the same space $R$ with measure $m$ and let $p_i(x), p_2(x), \cdots, p_r(x)$ be respectively their density functions with respect to $m$ ($p_i(x) \geq 0$ and $m$ is Lebesgue, counting, or mixed). Then Matusita [1] has defined the affinity of $F_1, F_2, \cdots, F_r$ as

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

For the case of two distributions $F_1$, $F_2$ the affinity is given by

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

Matusita [1] has shown that

\[
\rho_r(F_1, F_2, \cdots, F_r) \leq \min_{i,j} [\rho_i(F_1, F_j)]^{1/r}.
\]

He has also shown that $\rho_2(F_1, F_2)$ is uniquely related to a distance measure between $F_1$ and $F_2$. The distance measure in question is given as

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

and the relation is given by

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

Matusita [1] has also shown that $\rho_r(F_1, F_2, \cdots, F_r)$ is related to a generalization of $d_2(F_1, F_2)$. The generalization is given by

\[
d_r(F_1, F_2) = \left( \int_R |p_1^{1/r}(x) - p_2^{1/r}(x)|^r dm \right)^{1/r},
\]

and the relation is given by

\[
\rho_r(F_1, F_2, \cdots, F_r) \geq 1 - (r-1)\delta \,.
\]
when, for any pair \((i, j) (i, j = 1, 2, \cdots, r)\), we have \(d_s(F_i, F_j) \leq \delta\).

Equation (5) suggests that other distance measures besides \(d_s(F_1, F_2)\) are good measures of affinity. One well known measure of distance between two distributions is Kullback-Leibler's information given by

\[
I(F_1, F_2) = \int_R p_i(x) \log \frac{p_2(x)}{p_i(x)} \, dm.
\]

Unfortunately this measure is not symmetric in distributions as is the affinity. However, the divergence, which is the sum of \(I(F_1, F_2)\) and \(I(F_2, F_1)\), is symmetric and hence is a suitable measure of affinity. The divergence, which has been frequently used as a measure of distance between distributions, [2], [3], [4] is given by

\[
J(F_1, F_2) = \int_R [p_1(x) - p_2(x)] \log \left( \frac{p_1(x)}{p_2(x)} \right) \, dm.
\]

Recently, Matusita [5] considered the affinity of several distributions in detail and derived certain properties in addition to those found in [1]. In this paper some additional properties of \(\rho_s(F_1, F_2, \cdots, F_r)\) are derived. In particular, relations are found between \(\rho_s(F_1, F_2, \cdots, F_r)\) and \(d_s(F_1, F_2), d_s(F_1, F_3), \ldots, J(F_1, F_r)\), respectively. In addition, a generalized version of Matusita's measure of affinity is proposed and related to the expected value of \(J(F_1, F_r)\).

2. Some theorems

THEOREM 1. The affinity of several distributions is bounded above by the following inequality:

\[
\rho_s(F_1, F_2, \cdots, F_r) \leq \frac{2}{r(r-1)} \sum_{i < j} \rho_s(F_i, F_j).
\]

PROOF. The affinity can be considered to be the geometric mean of \(p_1(x), p_2(x), \cdots, p_r(x)\). From the inequality of symmetric means it follows that

\[
[p_1(x)p_2(x) \cdots p_r(x)]^{1/r} \leq \left[ \frac{2}{r(r-1)} \sum_{i < j} p_i(x)p_j(x) \right]^{1/r},
\]

from which it follows that

\[
\rho_s(F_1, F_2, \cdots, F_r) \leq \left[ \frac{2}{r(r-1)} \int_R \left[ \sum_{i < j} p_i(x)p_j(x) \right] \, dm \right]^{1/r}.
\]

Now, it is known that

\[
\left[ \sum_{i < j} p_i(x)p_j(x) \right]^{1/r} \leq \sum_{i < j} \sqrt{p_i(x)p_j(x)}.
\]
Substituting (13) into (12) and interchanging signs yields (10).

**COROLLARY.** It holds that

\[ \rho_\tau(F_1, F_2, \ldots, F_r) \leq \left[ \frac{2}{r(r-1)} \right]^{1/2} \sum_{i<j} \left( 1 - \frac{1}{4} [d_\tau(F_i, F_j)]^{r} \right)^{1/2}. \]

**Proof.** \( d_\tau(F_1, F_2) \) has been called by many the Kolmogorov variational distance [2] and has been related to minimum error probability in the multihypothesis decision problem [6]. It has been shown by Kraft [7] referenced in [3] that

\[ d_\tau(F_1, F_2) \leq 2[1 - \rho_\tau(F_1, F_2)]^{1/2}. \]

Also, Matusita [1] showed that

\[ d_\tau(F_1, F_2) \geq [d_\tau(F_1, F_2)]^\tau. \]

Combining (15) and (16) yields

\[ \rho_\tau(F_1, F_2) \leq \left( 1 - \frac{1}{4} [d_\tau(F_1, F_2)]^{r} \right)^{1/2}. \]

Finally, combining (17) with the result of Theorem 1 yields (14), the desired result.

**Theorem 2.** The affinity of several distributions is bounded below by the following inequalities:

\[ \rho_\tau(F_1, F_2, \ldots, F_r) \geq 1 - \frac{1}{r^2} \sum_{i<j} J(F_i, F_j) \]

and

\[ \rho_\tau(F_1, F_2, \ldots, F_r) \geq \exp \left[ - \frac{1}{r^2} \sum_{i<j} J(F_i, F_j) \right]. \]

**Proof.** First we prove (18). Let \( \bar{J} \) stand for the average divergence, i.e.,

\[ \bar{J} = \frac{1}{r^2} \sum_{i=1}^{r} \sum_{j=1}^{r} J(F_i, F_j). \]

Equation (20) can be written as follows:

\[ \bar{J} = \frac{2}{r} \sum_{i=1}^{r} \int_R p_i(x) \log p_i(x) dm - \frac{2}{r} \sum_{i=1}^{r} K_i(x) \]

where

\[ K_i(x) = \int_R p_i(x) \log g(x) dm \]
and

\[ g(x) = [p_1(x)p_2(x) \cdots p_r(x)]^{1/r}. \]

By definition

\[ \int_R \left[ \frac{1}{r} \sum_{j=1}^{r} p_j(x) \right] dm = 1. \]  

Also, from the arithmetic-mean-geometric-mean inequality it follows that

\[ [p_1(x)p_2(x) \cdots p_r(x)]^{1/r} \leq \frac{1}{r} \sum_{j=1}^{r} p_j(x). \]

From (24) and (23) it follows that

\[ \rho_r(F_1, F_2, \cdots, F_r) \leq 1. \]

Let \( r \) correction functions \( c_i(x), i = 1, 2, \cdots, r \), be defined such that for \( i = 1, 2, \cdots, r \)

\[ g(x) = p_i(x) + c_i(x) = p_i(x) \left[ 1 + \frac{c_i(x)}{p_i(x)} \right]. \]

Integrating (26) and using (25) yields

\[ \int_R c_i(x) dm = \rho_i(F_1, F_2, \cdots, F_r) - 1 \leq 0. \]

Substituting (26) into (22) yields

\[ K_i(x) = \int_R p_i(x) \log p_i(x) dm + \int_R p_i(x) \log \left[ 1 + \frac{c_i(x)}{p_i(x)} \right] dm. \]

Now, it can easily be verified that, for any real \( z \),

\[ z \geq \log (1 + z). \]

Applying (29) to (28) where \( z = c_i(x)/p_i(x) \), yields

\[ K_i(x) \leq \int_R p_i(x) \log p_i(x) dm + \int_R c_i(x) dm. \]

Substituting (22) and (27) into (30) yields

\[ \int_R p_i(x) \log \left[ \frac{p_i(x)}{g(x)} \right] dm \geq 1 - \rho_r(F_1, F_2, \cdots, F_r), \]

for \( i = 1, 2, \cdots, r \). Substituting (31) into (21) and using the fact that \( J(F_i, F_i') = 0, i = 1, 2, \cdots, r \), yields (18), the desired result. It should be mentioned here that (18) was recently proved for the case of two dis-
tributions [8]. We now prove (19).

The average divergence can be written as

\[ J = \frac{2}{r} \sum_{i=1}^{r} \int \! \frac{p_i(x)}{\prod_{j=1}^{r} [p_j(x)]^{1/r}} \log \left( \frac{p_i(x)}{\prod_{j=1}^{r} [p_j(x)]^{1/r}} \right) \, dm. \]  

Also,

\[ \int \! p_i(x) \log \left( \frac{\prod_{j=1}^{r} [p_j(x)]^{1/r}}{p_i(x)} \right) \, dm \leq \log [\rho_i(F_1, F_2, \ldots, F_r)] \]

from Jensen's inequality. Substituting (33) into (32) yields

\[ J \geq -2 \log [\rho_i(F_1, F_2, \ldots, F_r)], \]

which in turn yields (19).

3. A generalization of Matusita's affinity

In many situations, especially in the decision problem, the concept of weighted distance is useful. In the decision problem the weight represents the a priori probability that a sample comes from a certain distribution. In this section it is proposed to generalize Matusita's affinity as follows:

\[ \rho^*_i(F_1, F_2, \ldots, F_r) = \int \! \prod_{i=1}^{r} [p_i(x)]^{\omega_i} \, dm, \]

where \( \omega_i \geq 0, \ i = 1, 2, \ldots, r \) and \( \sum_{i=1}^{r} \omega_i = 1. \) Similarly, the expected divergence can be defined as

\[ E_{ij} [J(F_i, F_j)] = \sum_{i=1}^{r} \omega_i \prod_{j=1}^{r} \omega_j J(F_i, F_j). \]

**Theorem 3.** It holds that

\[ E_{ij} [J(F_i, F_j)] \geq 2[1 - \rho^*_i(F_1, F_2, \ldots, F_r)] \]

and

\[ E_{ij} [J(F_i, F_j)] \geq -2 \log [\rho^*_i(F_1, F_2, \ldots, F_r)]. \]

**Proof.** The proof is similar to that of Theorem 2. Several inequalities between \( E_{ij} [J(F_i, F_j)] \) and other distance measures have been derived in [9].
Acknowledgement

The work reported here was supported by the National Research Council of Canada under grant NRC A-3308 through Dr. Robert W. Donaldson. The author would like to thank Miss Linda Morris for typing the original manuscript.

UNIVERSITY OF BRITISH COLUMBIA

REFERENCES