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 $p_1(x), p_2(x), \cdots, p_r(x)$ be respectively their density functions with respect to m ($p_i(x) \geqq 0$ and m is Lebesgue, counting, or mixed). Then Matusita [1] has defined the affinity of $F_1, F_2, \cdots, F_r$ as

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

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

$$(2) \qquad \rho_2(F_1, F_2) = \int_R \sqrt{p_1(x)p_2(x)}\, dm .$$

Matusita [1] has shown that

$$(3) \qquad \rho_r(F_1, F_2, \cdots, F_r) \leqq \min_{i,j} [\rho_2(F_i, F_j)]^{2/r} .$$

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

$$(4) \qquad d_2(F_1, F_2) = \left\{ \int_R [\sqrt{p_1(x)} - \sqrt{p_2(x)}\,]^2 dm \right\}^{1/2}$$

and the relation is given by

$$(5) \qquad d_2^2(F_1, F_2) = 2[1 - \rho_2(F_1, F_2)] .$$

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

$$(6) \qquad d_r(F_1, F_2) = \left\{ \int_R |p_1^{1/r}(x) - p_2^{1/r}(x)|^r dm \right\}^{1/r} ,$$

and the relation is given by

$$(7) \qquad \rho_r(F_1, F_2, \cdots, F_r) \geqq 1 - (r-1)\delta ,$$

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

Equation (5) suggests that other distance measures besides $d_2(F_1, F_2)$ are good measures of affinity. One well known measure of distance between two distributions is Kullback-Leibler's information given by

$$I(F_1, F_2) = \int_R p_1(x) \log \frac{p_1(x)}{p_2(x)} dm . \tag{8}$$

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

$$J(F_1, F_2) = \int_R [p_1(x) - p_2(x)] \log \left[\frac{p_1(x)}{p_2(x)} \right] dm . \tag{9}$$

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 $\rho_r(F_1, F_2, \cdots, F_r)$ are derived. In particular, relations are found between $\rho_r(F_1, F_2, \cdots, F_r)$ and $\rho_2(F_i, F_j)$, $d_r(F_i, F_j)$, and $J(F_i, F_j)$, respectively. In addition, a generalized version of Matusita's measure of affinity is proposed and related to the expected value of $J(F_i, F_j)$.

2. Some theorems

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

$$\rho_r(F_1, F_2, \cdots, F_r) \leqq \left[\frac{2}{r(r-1)} \right]^{1/2} \sum_{i<j} \rho_2(F_i, F_j) . \tag{10}$$

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

$$[p_1(x) p_2(x) \cdots p_r(x)]^{1/r} \leqq \left[\frac{2}{r(r-1)} \sum_{i<j} p_i(x) p_j(x) \right]^{1/2} , \tag{11}$$

from which it follows that

$$\rho_r(F_1, F_2, \cdots, F_r) \leqq \left[\frac{2}{r(r-1)} \right]^{1/2} \int_R \left[\sum_{i<j} p_i(x) p_j(x) \right]^{1/2} dm . \tag{12}$$

Now, it is known that

$$\left[\sum_{i<j} p_i(x) p_j(x) \right]^{1/2} \leqq \sum_{i<j} \sqrt{p_i(x) p_j(x)} . \tag{13}$$

Substituting (13) into (12) and interchanging signs yields (10).

COROLLARY. *It holds that*

$$\rho_r(F_1, F_2, \cdots, F_r) \leqq \left[\frac{2}{r(r-1)}\right]^{1/2} \sum_{i<j} \left\{1 - \frac{1}{4}[d_r(F_i, F_j)]^{2r}\right\}^{1/2} . \tag{14}$$

PROOF. $d_1(F_1, F_2)$ 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

$$d_1(F_1, F_2) \leqq 2[1 - \rho_2^2(F_1, F_2)]^{1/2} . \tag{15}$$

Also, Matusita [1] showed that

$$d_1(F_1, F_2) \geqq [d_r(F_1, F_2)]^r . \tag{16}$$

Combining (15) and (16) yields

$$\rho_2(F_i, F_j) \leqq \left\{1 - \frac{1}{4}[d_r(F_i, F_j)]^{2r}\right\}^{1/2} . \tag{17}$$

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*:

$$\rho_r(F_1, F_2, \cdots, F_r) \geqq 1 - \frac{1}{r^2} \sum_{i<j} J(F_i, F_j) \tag{18}$$

and

$$\rho_r(F_1, F_2, \cdots, F_r) \geqq \exp\left[-\frac{1}{r^2} \sum_{i<j} J(F_i, F_j)\right] . \tag{19}$$

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

$$\bar{J} = \frac{1}{r^2} \sum_{i=1}^{r} \sum_{j=1}^{r} J(F_i, F_j) . \tag{20}$$

Equation (20) can be written as follows:

$$\bar{J} = \frac{2}{r} \sum_{i=1}^{r} \int_R p_i(x) \log p_i(x) dm - \frac{2}{r} \sum_{i=1}^{r} K_i(x) \tag{21}$$

where

$$K_i(x) = \int_R p_i(x) \log g(x) dm \tag{22}$$

and

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

By definition

$$\int_R \left[\frac{1}{r}\sum_{j=1}^{r} p_j(x)\right]dm=1 . \tag{23}$$

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

$$[p_1(x)p_2(x)\cdots p_r(x)]^{1/r}\leqq\frac{1}{r}\sum_{j=1}^{r} p_j(x) . \tag{24}$$

From (24) and (23) it follows that

$$\rho_r(F_1, F_2,\cdots, F_r)\leqq 1 . \tag{25}$$

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

$$g(x)=p_i(x)+c_i(x)=p_i(x)\left[1+\frac{c_i(x)}{p_i(x)}\right] . \tag{26}$$

Integrating (26) and using (25) yields

$$\int_R c_i(x)dm=\rho_r(F_1, F_2,\cdots, F_r)-1\leqq 0 . \tag{27}$$

Substituting (26) into (22) yields

$$K_i(x)=\int_R p_i(x)\log p_i(x)dm+\int_R p_i(x)\log\left[1+\frac{c_i(x)}{p_i(x)}\right]dm . \tag{28}$$

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

$$z\geqq\log(1+z) . \tag{29}$$

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

$$K_i(x)\leqq\int_R p_i(x)\log p_i(x)dm+\int_R c_i(x)dm . \tag{30}$$

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

$$\int_R p_i(x)\log\left[\frac{p_i(x)}{g(x)}\right]dm\geqq 1-\rho_r(F_1, F_2,\cdots, F_r) , \tag{31}$$

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 distributions [8]. We now prove (19).

The average divergence can be written as

$$\bar{J}=\frac{2}{r}\sum_{i=1}^{r}\int_R p_i(x)\log\left\{\frac{p_i(x)}{\prod_{j=1}^{r}[p_j(x)]^{1/r}}\right\}dm\ . \tag{32}$$

Also,

$$\int_R p_i(x)\log\left\{\frac{\prod_{j=1}^{r}[p_j(x)]^{1/r}}{p_i(x)}\right\}dm\leqq\log\left[\rho_r(F_1, F_2,\cdots, F_r)\right] \tag{33}$$

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

$$\bar{J}\geqq -2\log\left[\rho_r(F_1, F_2,\cdots, F_r)\right]\ ,$$

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:

$$\rho_r^*(F_1, F_2,\cdots, F_r)=\int_R\prod_{i=1}^{r}[p_i(x)]^{\omega_i}dm\ , \tag{34}$$

where $\omega_i\geqq 0$, $i=1, 2,\cdots, r$ and $\sum_{i=1}^{r}\omega_i=1$. Similarly, the expected divergence can be defined as

$$E_{ij}\{J(F_i, F_j)\}=\sum_{i=1}^{r}\omega_i\sum_{j=1}^{r}\omega_j J(F_i, F_j)\ . \tag{35}$$

THEOREM 3. *It holds that*

$$E_{ij}\{J(F_i, F_j)\}\geqq 2[1-\rho_r^*(F_1, F_2,\cdots, F_r)]$$

and

$$E_{ij}\{J(F_i, F_j)\}\geqq -2\log\left[\rho_r^*(F_1, F_2,\cdots, F_r)\right]\ .$$

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

[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] Kobayashi, H. and Thomas, J. B. (1967). Distance measures and related criteria, *Proc. 5th Annual Allerton Conf. Circuit and System Theory*, 491–500.

[3] Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection, *I.E.E.E. Trans. Commun. Technol.*, COM-**15**, 52–60.

[4] Toussaint, G. T. (1972). Comments on 'The divergence and Bhattacharyya distance measures in signal selection', *I.E.E.E. Trans. Commun. Technol.*, COM-**20**, 485.

[5] Matusita, K. (1971). Some properties of affinity and applications, *Ann. Inst. Statist. Math.*, **23**, 137–155.

[6] Toussaint, G. T. (1971). Some upper bounds on error probability for multiclass pattern recognition, *I.E.E.E. Trans. Comput.*, C-**20**, 943–944.

[7] Kraft, C. H. (1955). *Some conditions for consistency and uniform consistency of statistical procedures*, University of California Publications on Statistics.

[8] Toussaint, G. T. (1972). Some inequalities between distance measures for feature evaluation, *I.E.E.E. Trans. Comput.*, C-**21**, 409–410.

[9] Toussaint, G. T. (1971). Some functional lower bounds on the expected divergence for multihypothesis pattern recognition, communication, and radar systems, *I.E.E.E. Trans. Systems, Man, and Cybernetics*, SMC-**1**, 384–385.