MATUSITA'S GENERALIZED MEASURE OF AFFINITY

J. N. KAPUR

Matusita's measure of affinity for m multivariate probability distributions is generalized and by using Holder's inequality, it is shown that the generalized measure always lies between 0 and 1. A generalized average distance is also defined and it is shown that this measure also always lies between 0 and 1 and is always $\geq$ the corresponding Matusita's generalized measure. Expressions for these measures for the case of multivariate normal distributions are also obtained and on the basis of these, a generalized Cauchy-Schwarz inequality is obtained. An alternative proof of the well-known result that the logarithm of a positive definite matrix is a concave function is also given.

1. MATUSITA'S GENERALIZED MEASURE OF AFFINITY

Matusita's [6] measure of affinity for m probability distributions with density functions $f_1(X), f_2(X), \ldots, f_m(X)$, where X is an n dimensional random vector, is defined by

$$\rho_m = \int [f_1(X) f_2(X) \ldots f_m(X)]^{1/m} \, dX. \tag{1}$$

It is easily seen that

$$0 \leq \rho_m \leq \int \frac{f_1(X) + f_2(X) + \ldots + f_m(X)}{m} \, dX = 1. \tag{2}$$

Now $\rho_m = 1$ when $f_1(X) = f_2(X) = \ldots = f_m(X)$ almost every where and $\rho_m = 0$, if at least one of $f_1(X), f_2(X), \ldots, f_m(X)$ vanishes at every point of the domain of X. Thus $\rho_m = 1$ when the distributions are identical almost everywhere i.e. when there is no separation or divergence between the distributions and $\rho_m = 0$ when the distributions are distinct. Thus $1 - \rho_m$ can be regarded as a measure of 'mutual divergence' or as a measure of 'multiclass probabilistic distance'. Instead of $1 - \rho_m$ which varies between 0 and 1, we can use $-\ln \rho_m$, which varies between 0 and ∞, as a measure of multiclass probabilistic distance.

A large number of such multiclass probabilistic distance measures are available and these have been recently reviewed by Ray and Turner [7].

For $m = 2$,

$$\rho_i = \int \sqrt{f_1(X) f_2(X)} \, dX, \tag{3}$$

which is the well-known Bhattacharya [2] coefficient. This coefficient has been generalized to give generalized Bhattacharya coefficient [4]

$$\nu_2 = \int f_1^s(X) f_2^{1-s}(X)\, dX \quad , \quad 0 \le s \le 1\,. \tag{4}$$

When $s = 1/2$, ν_2 reduces to p_2. By using Holder's inequality [5]

$$\nu_2 = \int f_1^s(X) f_2^{1-s}(X)\, dX \ge \left[\int f_1(X)\, dX \right]^s \left[\int f_2(X)\, dX \right]^{1-s} = 1, \tag{5}$$

so that when $0 \le s \le 1$, $0 \le \nu_2 \le 1$. If howewer $s > 1$, then Holder's inequality gives

$$\nu_2 = \int f_1^s(X) f_2^{1-s}(X)\, dX \ge \left[\int f_1(X)\, dX \right]^s \left[\int f_2(X)\, dX \right]^{1-s} = 1\,. \tag{6}$$

Of course in this case we assume that either $f_2(X)$ is not zero anywhere or $f_2(X) = 0$ on a set of measure zero, such that ν_2 exists and is finite or whenever $f_2(X) = 0$, $f_1(X)$ is also zero and $f_1^s(X) f_2^{1-s}(X) = 0$. When $s > 1$, ν_2 has not been very much used in pattern recognition and signal processing, but it is used extensively in information theory in the form of Havrda and Charvats' [3] measure of directed divergence

$$(\alpha - 1)^{-1} \left[\int f_1^\alpha(X) f_2^{1-\alpha}(X)\, dX - 1 \right] \tag{7}$$

or in terms of Renyi's [8] measure of directed divergence

$$(\alpha - 1) \ln \int f_1^\alpha(X) f_2^{1-\alpha}(X)\, dX\,. \tag{8}$$

We can similarly generalize p_n to get Matisuta's generalized measure of affinity

$$\nu_m = \int f_1^{a_1}(X) f_2^{a_2}(X) \ldots f_m^{a_m}(X)\, dX\,. \tag{9}$$

If $\quad 0 \le a_1 \le 1, \quad 0 \le a_2 \le 1, \ldots, 0 \le a_m \le 1\,; a_1 + a_2 + \ldots + a_m = 1\,, \qquad (10)$

then by using Holder's inequality we get

$$\nu_m \leqq \left[\int f_1(X)\, dX) \right]^{a_1} \left[\int f_2(X)\, dX \right]^{a_2} \ldots \left[\int f_m(X)\, dX \right]^{a_m} = 1 \tag{11}$$

and $1 - \nu_m$ or $-\ln \nu_m$ can be used as a measure of multiclass probabilistic distance or of mutual divergence or generalized divergence. If one or more of of a_i's are greater than unity, then some others have to be negative and we have to see that the condition for the existence and finiteness for ν_m are satisfied. However in these cases, unlike the case when $m = 2$, ν_m is not necessarily always greater than or equal to unity. We shall discuss these cases in a separate communication. For the present we assume that conditions (10) are satisfied. As a special case, we may consider the case when $a_i = \pi_i$, where π_i is the a priori probability of the i th class, so that

$$\nu_m = \int f_1^{\pi_1}(X) f_2^{\pi_2}(X) \ldots f_m^{\pi_m}(X)\, dX\,. \tag{12}$$

2. AN ALTERNATIVE MEASURE OF AFFINITY

We can also use the following measure as a measure of affinity

$$\mu_m = \sum_{j=1}^{m} \sum_{i=1}^{m} a_i a_j \int \sqrt{f_i(X) f_j(X)}\, dX, \tag{13}$$

where a_i's satisfy (10). Since

$$0 \leq \int \sqrt{f_i(X) f_j(X)\, dX} \leq 1 \tag{14}$$

we get

$$0 \leq \mu'_m \leq \sum_{j=1}^{m} \sum_{i=1}^{m} a_i a_j = \sum_{j=1}^{m} a_j \sum_{i=1}^{m} a_i = 1, \tag{15}$$

so that μ_m also lies between 0 and 1, so that $1 - \mu_m$ or $-\ln \mu_m$ can also be used as a measure of generalized divergence. Now

$$\begin{aligned}\mu_m &= \int \left(a_1 \sqrt{f_1(X)} + a_2 \sqrt{f_2(X)} + \ldots + a_m \sqrt{f_m(X)}\right)^2 dX \\ &\leq \int (f_1^{a_1/2}(X) f_2^{a_2/2}(X) \ldots f_m^{a_m/2}(X))^2\, dX \\ &= \int f_1^{a_1}(X) f_2^{a_2}(X) \ldots f_m^{a_m}(X)\, dX = \nu_m,\end{aligned} \tag{16}$$

so that

$$0 \leq \mu_m \leq \nu_m \leq 1. \tag{17}$$

Now $\mu_m = \nu_m$ if $f_1(X) = f_2(X) = \ldots = f_m(X)$ almost everywhere and in that case

$$\mu_m = \nu_m = 1. \tag{18}$$

Instead of (13) we can also use

$$\mu_m = \sum_{j=1}^{m} \sum_{i=1}^{m} a_i a_j \int f_i^{s_{ij}}(X) f_j^{1-s_{ij}}(X)\, dX, \quad 0 \leq s_{ij} \leq 1. \tag{19}$$

3. GENERALIZED AFFINITY MEASURE FOR MULTIVARIATE NORMAL DISTRIBUTIONS

Let $N(M_i, V_i)$ be the i th multivariate normal distribution, then

$$\nu_m = \int \prod_{i=1}^{m} \left\{ \frac{1}{(2\pi)^{n a_i/2} \,|V_i|^{a_i/2}} \exp\left[-\frac{a_i}{2}(X - M_i)' V_i^{-1} (X - M_i)\right] \right\} dX \tag{20}$$

$$= \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{m} |V_i|^{a_i/2}} \int \exp\left[-\frac{1}{2}(X - M)' V^{-1} (X - M) \frac{1}{2}\right] dX, \tag{21}$$

where

$$V^{-1} = \sum_{i=1}^{m} a_i V_i^{-1} \tag{22}$$

$$V^{-1} M = \sum_{i=1}^{m} a_i V_i^{-1} M_i \tag{23}$$

$$K = \sum_{i=1}^{m} M_i' a_i V_i^{-1} M_i - M' V^{-1} M \tag{24}$$

so that

$$\nu_m = \frac{|V|^{\frac{1}{2}}}{\prod_{i=1}^{m} |V_i|^{a_i/2}} e^{-\frac{1}{2}K}, \tag{25}$$

where V is determined by (22), then M is determined from (23) and finally K is determined from (24).

4. CONCAVITY OF LOGARITHM OF THE DETERMINANT OF A POSITIVE DEFINITE MATRIX

If

$$M_1 = M_2 = \ldots = M_m = M_0, \tag{26}$$

then from (23) and (24), $M = M_0$ and from (22) and (24) $K = 0$, whether V_i's are equal or not. In this case also $\nu_m \geq 1$ so that

$$|V|^{\frac{1}{2}} \geq \prod_{i=1}^{m} |V_i|^{a_i/2}$$

or

$$\ln |V|^{-\frac{1}{2}} \geq \sum_{i=1}^{m} a_i \ln |V_i|^{-1} \tag{27}$$

or

$$\ln \sum_{i=1}^{m} a_i |V_i|^{-1} \geq \sum_{i=1}^{m} a_i \ln |V_i|^{-1}. \tag{28}$$

Since V_i is a non-singular positive definite matrix, we can write $V_i^{-1} = W_i$, where W_i is a non-singular positive definite matrix and

$$\ln \sum_{i=1}^{m} a_i |W_i| \geq \sum_{i=1}^{m} a_i \ln |W_i|, \; a_i \geq 0, \; \sum_{i=1}^{m} a_i = 1, \tag{29}$$

which shows that the logarithm of the determinant of a non-singular positive definite matrix is concave. This result is true for every such matrix since given any matrix, we can have a multivariate normal distribution corresponding to it.

Thus we have got an information-theoretic proof of an important result in matrix theory. An alternative proof of this result is available in Bellman [1] .

5. GENERALIZATION OF CAUCHY-SCHWARZ INEQUALITY

If each distribution has the same variance-covariance matrix V_0, then $\nu_m = \exp\left(-\frac{1}{2}K\right)$, where

$$K = \sum_{i=1}^{m} a_i M_i^t V_0 M_i - M^t V_0 M \tag{30}$$

and since $\nu_m \geq 1$, we get $K \geq 0$ or

$$\sum_{i=1}^{m} a_i M_i^t V_0 M_i \geq \left(\sum_{i=1}^{m} a_i M_i\right)^t V_0 \left(\sum_{i=1}^{m} a_i M_i\right). \tag{31}$$

If we write

$$a_i = \frac{b_i}{\sum_{i=1}^{m} b_i}, \tag{32}$$

this gives

$$\sum_{i=1}^{m} b_i \sum_{i=1}^{m} b_i M_i^t V_i M_i \geq \left(\sum_{i=1}^{m} b_i M_i\right)^t V_0 \left(\sum_{i=1}^{m} b_i M_i\right). \tag{33}$$

When $n = 1$, this gives

$$\sum_{i=1}^{m} b_i \sum_{i=1}^{m} b_i m_i^2 \geq \left(\sum_{i=1}^{m} b_i m_i\right)^2 \tag{34}$$

or

$$\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i^2 \geq \left(\sum_{i=1}^{m} x_i y_i\right)^2, \tag{35}$$

which is Cauchy-Schwarz inequality. This is a special case of (33). As such (33) can be regarded as a generalization of Cauchy-Schwarz inequality. If c_{jk}'s are the elements of the matrix V_0, then (33) can be written as

$$\sum_{i=1}^{m} b_i \sum_{i=1}^{m} b_i \left(\sum_{k=1}^{n} \sum_{j=1}^{n} c_{jk} m_{ij} m_{ik}\right) \geq \sum_{k=1}^{n} \sum_{j=1}^{n} c_{jk} \left(\sum_{i=1}^{m} b_i m_{ij}\right)\left(\sum_{i=1}^{m} b_i m_{ik}\right). \tag{36}$$

If

$$b_i = x_i^2, \quad \sqrt{b_i}\, m_{ij} = y_{ij}, \quad b_i m_{ij} = x_i y_{ij}, \tag{37}$$

then (36) gives

$$\Big(\sum_{i=1}^{m} x_i^2\Big)\Big[\sum_{i=1}^{m}\sum_{k=1}^{n}\sum_{j=1}^{n} c_{jk}\, y_{ij}\, y_{ik}\Big] \geq \sum_{k=1}^{n}\sum_{j=1}^{n} c_{jk}\Big(\sum_{i=1}^{m} x_i\, y_{ij}\Big)\Big(\sum_{i=1}^{m} x_i\, y_{ik}\Big). \qquad (38)$$

Here the equality sign holds iff all the mean vectors are equal i.e. if m_{ij} is independent of i or if $m_{ij} = d_j$ or

$$x_i\, y_{ij} = b_i\, d_j \quad \text{or} \quad y_{ij} = x_i\, d_j\,. \qquad (39)$$

If (39) is satisfied, then the L.H.S. of (38)

$$= \Big(\sum_{i=1}^{m} x_i^2\Big)\Big[\sum_{i=1}^{m} x_i^2\Big(\sum_{k=1}^{n}\sum_{j=1}^{n} c_{jk}\, d_j\, d_k\Big)\Big]$$

$$= \Big(\sum_{i=1}^{n} x_i^2\Big)^2\Big(\sum_{k=1}^{n}\sum_{j=1}^{n} c_{jk}\, d_j\, d_k\Big) \qquad (40)$$

and the R.H.S. of (38)

$$= \sum_{k=1}^{n}\sum_{j=1}^{n} c_{jh}\Big(\sum_{i=1}^{m} x_i^2\, d_j\Big)\Big(\sum_{i=1}^{m} x_i^2\, d_k\Big)$$

$$= \Big(\sum_{i=1}^{m} x_i^2\Big)^2 \sum_{k=1}^{n}\sum_{j=1}^{n} c_{jk}\, d_k \qquad (41)$$

so that when (39) is satisfied, equality sign holds in (38).

Thus (33) or (38) gives generalized Cauchy-Schwarz inequality in which the equality sign holds only when y_{ij}/x_i is independent of i.

If V_0 is the unit matrix, (38) gives

$$\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} (y_{i1}^2 + y_{i2}^2 + \ldots + y_{in}^2) \geq \sum_{j=1}^{n}\Big(\sum_{i=1}^{m} x_i\, y_{ij}\Big)^2. \qquad (42)$$

If $n = 1$, this again gives Cauchy-Schwarz inequality. However (42) can be at once deduced from Cauchy-Schwarz inequalities.

6. CALCULATION OF THE ALTERNATIVE MEASURE OF AFFINITY

For $m = 2$, we have [7]

$$p_2 = \exp\Big[-\frac{1}{8}(M_1 - M_2)\; W^{-1}(M_1 - M_2)\Big]\frac{|V_1|^{\frac{1}{4}}\,|V_2|^{\frac{1}{4}}}{|W|^{\frac{1}{2}}}, \qquad (43)$$

where

$$W = \frac{V_1 + V_2}{2}. \qquad (44)$$

From (13)

$$\mu_m = \sum_{j=1}^{m} \sum_{i=1}^{m} a_i a_j \frac{|V_i|^{\frac{1}{4}} |V_j|^{\frac{1}{4}}}{\left|\frac{V_i + V_j}{2}\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{8}(M_i - M_j)\left[\frac{V_i + V_j}{2}\right]^{-1} \times \times (M_i - M_j)\right\}. \quad (45)$$

If the mean vectors are equal, (17), (25) and (45) give the inequality

$$0 \leq \sum_{j=1}^{m} \sum_{i=1}^{m} a_i a_j \frac{|V_i|^{\frac{1}{4}} |V_j|^{\frac{1}{4}}}{\left|\frac{V_i - V_j}{2}\right|^{\frac{1}{2}}} \leq \frac{|V|^{\frac{1}{2}}}{\prod_{i=1}^{m} |V_i|^{a_i/2}} = 1 \,. \quad (46)$$

It is observed that the calculation of ν_m is much simpler than that of μ_m.

REFERENCES

[1] BELLMAN, R. : **Introduction to Matrix Analysis**, second edition, Tata McGraw Hill, New Delhi, 1979.

[2] BHATTACHARYA, A. : *On the measure of divergence between two statistical populations defined by their probability distribution*, Bull. Cal. Math. Soc. **35** (1943), 99-109.

[3] HAVRDA, J. and CHARVAT, F. : *Quantification Methods in Classification Processes: Concept of Structural α Entropy*, Kybernetics **3** (1967), 30-35.

[4] KAILATH, T. : *The divergence and Bhattacharya distance measeres in digital selection*, IEEE Trans. Commu. Tech. **15** (1967), 52-69.

[5] MARSHALL, A.W. and OLKIN, I. : **Inequalities, Theory of Majorisation and its applications**, Academic Press, New York, 1979.

[6] MATUSITA, K. : *On the notion of affinity of several distribution and some of its applications*, Ann. Inst. Stat. Math. 19 (1967), 181-192.

[7] RAY, S. and TURNER, L.F. : *Probabilistic Distance Measure as criteria for feature extraction*, Advances in Information Science and Technology (Indian Statistical Institute Calcutta), **1** (1984), 28-45.

[8] BENYI, A. : *On Measures of Entropy and Information*, Proc. 4 th Berkeley Symp. Math. Stad. Prob. **1** (1961), 547-561.

INDIAN INSTITUTE
OF TECHNOLOGY
KANPUR-INDIA 208016

ÖZET

Bu çalışmada Matusita'nın, *m* değişkenli olasılık dağılımları için verdiği afinite ölçümü genelleştirilmekte ve Holder eşitsizliği kullanılarak, genelleştirilmiş ölçümün daima 0 ile 1 arasında olduğu gösterilmektedir.