King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 044

June 1982

Matusita's Distance with its Application

Ibrahim A. Ahmad

Ibrahim A. Ahmad

MATUSITA'S DISTANCE

INTRODUCTION

In statistical decision problems, many procedures are based on distance functions. The distance between two distributions can be measured in many different ways. One such definition is that introduced by Matusita (1955) and developed further by him in a series of subsequent investigations.

Let F_1 and F_2 be two distribution functions admitting probability densities f_1 and f_2, respectively, with respect to some measure μ. Matusita's distance between F_1 and F_2 is defined by

$$\|F_1 - F_2\|_r = \left|\int (f_1^{\frac{1}{r}}(x) - f_2^{\frac{1}{r}}(x))^r \, d\mu(x)\right|^{\frac{1}{r}}, \quad r \geq 1. \tag{1}$$

Note that if $\rho(F_1, F_2) = \int (f_1(x) f_2(x))^{\frac{1}{2}} d\mu(x)$, then $\|F_1 - F_2\|_2^2 = 2(1 - \rho(F_1, F_2))$. Here and elsewhere, integrals will be taken over the common support of f_1 and f_2. The distance defined in (1) is also known as the Hellinger distance, see Beran (1977) and Rao (1963).

The duality between $\|F_1 - F_2\|_2$ and $\rho(F_1, F_2)$ (also known as the affinity between F_1 and F_2) is one of the most important aspects in the applicability of the distance (1) in statistical inference. The dual notion of affinity

between F_1 and F_2 can be extended to measure the closeness between individual members of a finite family of distributions, $F_1, \ldots, F_m$, all admitting densities $f_1, \ldots, f_m$, respectively, with respect to some measure μ. This can be done as follows:

$$\rho_m(F_1, \ldots, F_m) = \int f_1^{r_1}(x) \ldots f_m^{r_m}(x)d\mu(x), \qquad (2)$$

where $r_i \geq 0$, $i = 1, \ldots, m$ and $\Sigma_{i=1}^m r_i = 1$. When $r_i = \frac{1}{m}$, $i = 1, \ldots, m$ we obtain the notion defined and studied by Matusita (1967, 1971). The extension in (2) was first proposed by Toussaint (1974).

The Matusita distance defined in (1) is proved to be of both mathematical and statistical interest and the duality between $\|\cdot\|_2$ and $\rho(F_1, F_2)$ is of particular interest since one can discuss $\rho(F_1, F_2)$ instead of $\|F_1 - F_2\|_2$.

In the next sections we shall present mathematical, statistical, and related properties of $\rho(F_1, F_2)$ and its extension to more than two distributions stressing the relation with $\|F_1 - F_2\|_r$, $r \geq 1$.

MATHEMATICAL PROPERTIES OF MATUSITA'S AFFINITY AND DISTANCE MEASURES

First, note that with $r_i = \frac{1}{m}$ we can easily see that (Matusita (1967)):

$$0 \le \rho_m^m(F_1, \ldots, F_m) \le \rho_{m-1}^{m-1}(F_{i_1}, \ldots, F_{i_{m-1}}) \le \ldots \le \rho_2^2(F_{i_\ell}, F_{i_p}) \le 1 \tag{3}$$

where $i_1, \ldots, i_{m-1}$ is a permutation subset of $1, \ldots, m$ and $\{i_\ell, i_p\} \subset \ldots \subset \{i_1, \ldots, i_{m-1}\} \subset \{1, \ldots, m\}$. Thus $\rho_m(F_1, \ldots, F_m) = 1$ whenever $F_1 = \ldots = F_m$ and that $\rho_m(F_1, \ldots, F_m) \le \min_{i,j} \rho_2^{2/m}(F_i, F_j)$.

Next, if $\|F_i - F_j\|_r \le \delta$, for all $i, j = 1, \ldots, m$ then $\rho_m(F_1, \ldots, F_m) \ge 1 - (m-1)\delta$. Combining these lower and upper bounds we get

$$1 - (m-1)\delta \le \rho_m(F_1, \ldots, F_m) \le \min_{i,j} \rho^{2/m}(F_i, F_j). \tag{4}$$

Also it is not difficult to prove that for all $r > 1$,

$$\|F_1 - F_2\|_{r-1}^{r-1} \ge \|F_1 - F_2\|_r^r, \tag{5}$$

and also get that $\|F_1 - F_2\| \le r \|F_1 - F_2\|_r$.

More refined upper and lower bounds of $\rho_m(F_1, \ldots, F_m)$ given by Toussaint (1974) are:

$$1 - m^{-2} \Sigma_{i<j} J(F_i, F_j) \le \rho_m(F_1, \ldots, F_m) \le \left[\frac{2}{m(m-1)}\right]^{\frac{1}{2}} \Sigma_{i<j} \rho_2(F_i, F_j), \tag{6}$$

where $J(F_1, F_2) = \int [f_1(x) - f_2(x)] \ell n[f_1(x)/f_2(x)] \, d\mu(x)$.

It is very simple to see that if $\{F_{in}\}$, $i = 1, \ldots, m$ are

sequences of distributions, then $\rho_m(F_{1n}, \ldots, F_{mn}) \to \rho(F_{10}, \ldots, F_{m0})$ provided that $f_{in}(x) \to f_{i0}(x)$, a.e., as $n \to \infty$. Next, if we partition the support of $F_1, \ldots, F_m$ into subsets $E_1, E_2, \ldots$, then

$$\rho_m(F_1, \ldots, F_m) \leq \Sigma_i (\Pi_{j=1}^m \int_{E_i} f_j(x) d\mu(x))^{\frac{1}{m}}, \qquad (7)$$

and also

$$\rho_m(F_1, \ldots, F_m) = \inf \Sigma_i \{\Pi_{j=1}^m \int_{E_i} f_j(x) d\mu(x)\}^{\frac{1}{m}}, \qquad (8)$$

where the inf is taken over all partitions $\{E_i\}$. Other properties of ρ when transformations of variables are sought are in Matusita (1971).

Kirmani (1968, 1971) proved that $\rho_2(F_1, F_2)$ and also $\|F_1 - F_2\|_2$ has similar limiting properties to those of Kullback-Leibler information number $J(F_1, F_2)$ mentioned above.

When $m = 2$, Mathai and Rathie (1972), gave a complete characterization of $\rho_2(F_1, F_2)$ when μ is taken as a finite counting measure, i.e., they proved that $\rho_2(F_1, F_2) = \Sigma_{i=1}^n \sqrt{p_i q_i}$, $\Sigma_{i=1}^n p_i = \Sigma_{i=1}^n q_i = 1$, $p_i, q_i \geq 0$ if and only if the following three conditions hold:

1. $\rho(p_1, \ldots, p_n, q_1, \ldots, q_n) = \rho(p_1 + p_2, p_3, \ldots, p_n, q_1 + q_2, q_3, \ldots, q_n) + [(p_1+p_2)(q_1+q_2)]^{\frac{1}{2}} \times$

$$\times \left[\rho\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}, \frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2}\right) - 1\right],$$

for all $p_1 + p_2 > 0$ and $q_1 + q_2 > 0$.

2. For $n = 3$, ρ is symmetric in pairs (p_i, q_i).

3. For $n = 2$, $\rho\left(\frac{1}{4}, \frac{3}{4}, \frac{3}{4}, \frac{1}{4}\right) = \cos\frac{\pi}{6}$.

The extension of this result to any $m > 2$ is given in Kaufman and Mathai (1973).

STATISTICAL APPLICATIONS AND PROPERTIES

If μ is a finite counting measure many inferential problems have been treated using Matusita's measure of distance or equivalently its dual affinity measure. Matusita (1955) used $\|F_1 - F_2\|_2$ in the one sample goodness-of-fit problem, showed the size of the test based on $\|F_0 - F_n\|_2$ and that its limiting distribution under the null is chi-square and under the alternative is normal, see also Ahmad and VanBelle (1974). He also discussed the two-sample problem and showed that in this case the null distribution is only approximately weighted sum of chi-squares. In this direction similar results were also obtained by Rao (1963). In the decision problem, Matusita and Akaike (1956) and Matusita (1956, 1961) used similar ideas in various statistical problems such as independence, invariance,

two-sample goodness-of-fit, classification and pattern recognition, and interval estimation. Formulation of the statistical decision problem in terms of distance functions in general terminology is presented in Matusita (1964).

Even in this very special case of fintie counting measure μ, one very quickly finds out that some serious drawbacks are evident when basing a statistical procedure on $\rho_2(F_1, F_2)$ or $\|F_1 - F_2\|_2$. This led Ahmad and VanBelle to consider alternative distance or equivalently affinity function. An affinity measure that proved particularly successful is:

$$\lambda(F_1, F_2) = 2\int f_1(x) f_2(x) d\mu(x) / [\int f_1^2(x) d\mu(x) + \int f_2^2(x) d\mu(x)]. \qquad (9)$$

In the case of counting measure μ, $\lambda(F_1, F_2)$ enjoys some very attractive properties and can be used at a much wider scale than $\rho(F_1, F_2)$. Note that $\lambda(F_1, F_2)$ assumes that f_i, $i = 1, 2$ are square integrable; this is not very restrictive assumption and is fulfilled in many situations.

Returning to $\rho(F_1, F_2)$, other statistical applications include the proving that a limit of a sequence of sufficient statistics is also sufficient, see Matusita (1971) and also its use as a measure of discrimination.

When μ is the Lebesgue measure, the use of $\rho(F_1, F_2)$

in statistical inference is much more scarce. In parametric estimation, Beran (1977) shows that an estimate of a vector of parameters $\underset{\sim}{\theta}$ of $f_{\underset{\sim}{\theta}}$ which minimizes $\|F_{\underset{\sim}{\theta}} - F_n\|_2$ always exists (under certain conditions on the parameter space Θ), is unique, and is such that $T(f_{\underset{\sim}{\theta}}) = \underset{\sim}{\theta}$ uniquely for any $\underset{\sim}{\theta} \in \Theta$. Conditions for its consistency (in the weak sense) and approximate normality are also given in Theorems 3 and 4 of Beran (1977). More interestingly, these estimates of location parameters are robust against moderate perturbations. The sister problem of using $\rho(F_1, F_2)$ in hypothesis testing situation is not yet explored and many intractable difficulties seem to stand in the way. Ahmad (1980b) has proven, however, that if one is to estimate $\rho(F_1, F_2)$ using suitable density estimation method $\hat{f}_1$ and $\hat{f}_2$, then one can find conditions under which $\hat{\rho}(F_1,F_2)=\int[\hat{f}_1(x)\hat{f}_2(x)]^{\frac{1}{2}}dx$ is consistent in the mean and is strongly consistent estimate of $\rho(F_1, F_2)$. On the other hand, the affinity measure $\lambda(F_1, F_2)$ is more suitable for statistical inference when we can assume that both F_1 and F_2 admit square integrable densities. This has been demonstrated in Ahmad (1980a) where the two-sample and one-sample goodness-of-fit problems are discussed as well as the problems of independence and symmetry. Parametric estimation using a method similar to that of Beran (1977) but based on $\lambda(F_1, F_2)$ seems within reach but no formal results have been published.

CONCLUDING REMARKS

The Matusita's distance and its dual concept of affinity between two distributions is a method of statistical inference that provides in many situations plausible answers. While, in case of two distributions, there is complete duality between the distance measure and the affinity measure, this is not very clear when there are more than two distributions. The affinity between several distributions was investigated by Matusita (1967, 1971) but the distance between two sets of distributions is still open. The usage of the affinity or distance in statistical inference needs further investigation and in particular as a method for providing robust estimation.

REFERENCES

1. Ahmad, I. and VanBelle, G. (1974). Measuring affinity of distributions. Reliability and Biometry, Statistical Analysis of Lifetesting, F. Proschan and R.J. Sarfling, eds., SIAM, Philadelphia, 651-668.

2. Ahmad, I. (1980a). Nonparametric estimation of an affinity measure between two absolutely continuous distributions with hypothesis testing applications. Ann. Instit. Statist. Math., 32, 223-240.

3. Ahmad, I. (1980b). Nonparametric estimation of Matusita's measure of affinity between absolutely continuous distributions. Ann. Instit. Statist. Math., 32, 241-245.

4. Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. Ann. Statist., 5, 445-463.

5. Kirmani, S.N. (1968). Some results on Matusita's measure of distance. J. Indian Statist. Assoc., 6, 89-98.

6. Kirmani, S.N. (1971). Some limiting properties of Matusita's measure of distance. Ana. Instit. Statist. Math., 23, 157-162.

7. Kaufman, H. and Mathai, A.M. (1973). An axiomatic foundation for a multivariate measure of affinity among a number of distributions. J. Mult. Analysis, 3, 236-242.

8. Mathai, A.M. and Rathei, P.N. (1973). Characterization of Matusita's measure of affinity. Ann. Instit. Statist. Math., 25, 473-483.

9. Matusita, K. (1955). Decision rules based on distance, for problems of fit, two samples and estimation. Ann. Math. Statist., 26, 631-640.

10. Matusita, K. (1956). Decision rule based on the distance for the classification problem. Ann. Instit. Statist. Math., 8, 67-77.

11. Matusita, K. (1961). Interval estimation based on the notion of affinity. Bull. Inter. Statist. Instit., 38, 241-244.

12. Matusita, K. (1964). Distance and decision rules. Ann. Instit. Statist. Math., 16, 305-315.

13. Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications. Ann. Instit. Statist. Math., 19, 181-192.

14. Matusita, K. (1971). Some properties of affinity and applications. Ann. Instit. Statist. Math., 23, 137-155.

15. Matusita, K. and Akaike, H. (1956). Decision rules based on distance for the problem of independence, invariance, and two samples, Ann. Instit. Statist. Math., 7, 67-90.

16. Rao, C.R. (1963). Criteria of estimation in large samples. Sankhya, 25, 189-206.

17. Toussaint, G.T. (1974). Some properties of Matusita's measure of affinity of several variables. Ann. Instit. Statist. Math., 26, 389-394.