Ann. Inst. Statist. Math.
30 (1978), Part A, 177-179

NOTE ON K-POINT SEPARATION MEASUREMENT

RINYA TAKAHASHI

(Received May 12, 1975; revised Nov. 5, 1976)

Summary

Glick [1] introduced the notion of a separation measurement and showed that for a set $f_1, \cdots, f_K$ of densities, $s_K^*(f_1, f_2, \cdots, f_K)=2\left[\int \max \{f_1, f_2, \cdots, f_K\}-1\right]$ is a K-point separation measurement. This notion is some generalization of Matusita's distance (affinity) of densities $f_1, f_2, \cdots, f_K$, and its interesting applications were shown in Matusita [2], [3]. In this paper we give some statistical remarks on a separation measurement.

If f_1 and f_2 are densities, then the L_1 norm satisfies that

$$||f_1-f_2||=\int |f_1-f_2|$$

$$=2\left[\int \max \{f_1, f_2\}-1\right] \tag{1}$$

$$=2\left[1-\int \min \{f_1, f_2\}\right] \tag{2}$$

$$=\int \max \{f_1, f_2\}-\int \min \{f_1, f_2\} . \tag{3}$$

Therefore, Glick's s_K^* is a natural extension of the L_1 norm in the sense (1). We show that extensions of the L_1 norm in the sense (2) and (3) are K-point separation measurements, too.

DEFINITION (Glick). A symmetric function s will be called a K-point separation measurement ($K \geqq 2$) for a subset S in a vector space with norm $||\cdot||$ if, for any elements $a_1, a_2, \cdots, a_K \in S$, the function s satisfies

$$\max_{i<j} ||a_i-a_j|| \leqq s(a_1, a_2, \cdots, a_K) \leqq \sum_{i<j}\sum ||a_i-a_j|| .$$

We assume that $f_1, f_2, \cdots, f_K$ are densities with respect to some σ-finite measure μ.

THEOREM. *The function* s_*^K *given by*

$$s_*^K(f_1, f_2, \cdots, f_K) = 2\left[1 - \int \min\{f_1, f_2, \cdots, f_K\}\right]$$

is a K-point separation measurement for the probability densities in $L_1[\mu]$.

PROOF. From the definition, we obtain

$$\begin{aligned} s_*^K(f_1, f_2, \cdots, f_K) &= 2\left[1 - \int \min\{f_1, f_2, \cdots, f_K\}\right] \\ &= \max_{i<j} 2\left[1 - \int \min\{f_i, f_j\}\right] \\ &= \max_{i<j} \|f_i - f_j\| . \end{aligned}$$

This completes the proof.

If we define $s_K(f_1, f_2, \cdots, f_K) = \int \max\{f_1, f_2, \cdots, f_K\} - \int \min\{f_1, f_2, \cdots, f_K\} = s_K^*(f_1, f_2, \cdots, f_K)/2 + s_*^K(f_1, f_2, \cdots, f_K)/2$, then we get the following corollary.

COROLLARY. *The function* $s_K(f_1, f_2, \cdots, f_K)$ *is a K-point separation measurement.*

Remarks.

(i) By using the same technique as in Glick [1] we can obtain an upper bound for s_*^K, that is

$$s_*^K(f_1, f_2, \cdots, f_K) \leqq \frac{2}{K} \sum_{i<j}\sum \|f_i - f_j\| .$$

(ii) Glick's s^* and our s_* have the following relationship:

$$\begin{aligned} s_K^*(f_1, f_2, \cdots, f_K) = &\sum_{i_1<i_2}\sum s_*^2(f_{i_1}, f_{i_2}) - \sum_{i_1<i_2<i_3}\sum\sum s_*^3(f_{i_1}, f_{i_2}, f_{i_3}) \\ &+ \cdots + (-1)^K s_*^K(f_1, f_2, \cdots, f_K) , \end{aligned}$$

which follows from the identity:

$$\begin{aligned} \int \max\{f_1, f_2, \cdots, f_K\} = K &- \sum_{i_1<i_2}\sum \int \min\{f_{i_1}, f_{i_2}\} \\ &+ \sum_{i_1<i_2<i_3}\sum\sum \int \min\{f_{i_1}, f_{i_2}, f_{i_3}\} \\ &- \cdots + (-1)^K \int \min\{f_1, f_2, \cdots, f_K\} . \end{aligned}$$

(iii) In contrast with the interpretation of s_K^* given in Glick [1] the function s_*^K may be interpreted as follows: a least favourable classification rule has the unconditional probability of incorrect classification given by $1-\int \min\{g_1, g_2, \cdots, g_K\} = s_*^K(g_1, g_2, \cdots, g_K)/2$, where $g_j(x) = g_j f_j(x)$, g_j standing for the prior probability of f_j.

Acknowledgement

The author wishes to thank Professor M. Okamoto for his kind advice during the completion of this paper.

OSAKA UNIVERSITY

REFERENCES

[1] Glick, N. (1973). Separation and probability of correct classification among two or more distributions, *Ann. Inst. Statist. Math.*, **25**, 373-382.

[2] Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications, *Ann. Inst. Statist. Math.*, **19**, 181-192.

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