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 \( f_1, \ldots, f_K \) of densities, \( s^*_K(f_1, f_2, \ldots, f_K) = 2 \left[ \max \{f_1, f_2, \ldots, f_K\} - 1 \right] \) is a \( K \)-point separation measurement. This notion is some generalization of Matusita's distance (affinity) of densities \( f_1, f_2, \ldots, f_K \), 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

\[
\|f_1 - f_2\| = \int |f_1 - f_2|
\]

(1)

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

(2)

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

(3)

\[
= \int \max \{f_1, f_2\} - \int \min \{f_1, f_2\}.
\]

Therefore, Glick's \( s^*_K \) is a natural extension of the \( L_1 \) norm in the sense (1). We show that extensions of the \( L_i \) 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 (\( K \geq 2 \)) for a subset \( S \) in a vector space with norm \( \| \cdot \| \) if, for any elements \( a_1, a_2, \ldots, a_K \in S \), the function \( s \) satisfies

\[
\max_{i<j} \|a_i - a_j\| \leq s(a_1, a_2, \ldots, a_K) \leq \sum_{i<j} \|a_i - a_j\|.
\]

We assume that \( f_1, f_2, \ldots, f_K \) are densities with respect to some \( \sigma \)-finite measure \( \mu \).
THEOREM. The function $s^*_k$ given by

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

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

PROOF. From the definition, we obtain

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

$$= \max_{i<j} 2\left[1 - \min \{f_i, f_j\}\right]$$

$$= \max_{i<j} \|f_i - f_j\|.$$ 

This completes the proof.

If we define $s_k(f_1, f_2, \cdots, f_K) = \max \{f_1, f_2, \cdots, f_K\} - \min \{f_1, f_2, \cdots, f_K\} = s^*_k(f_1, f_2, \cdots, f_K)/2 + s^*_k(f_1, f_2, \cdots, f_K)/2$, 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

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

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

$$s^*_k(f_1, f_2, \cdots, f_K) = \sum \sum_{i_1 < i_2} s^*_k(f_{i_1}, f_{i_2}) - \sum \sum \sum s^*_k(f_{i_1}, f_{i_2}, f_{i_3})$$

$$+ \cdots + (-1)^K s^*_k(f_1, f_2, \cdots, f_K),$$

which follows from the identity:

$$\max \{f_1, f_2, \cdots, f_K\} = K - \sum \sum_{i_1 < i_2} \min \{f_{i_1}, f_{i_2}\}$$

$$+ \sum \sum \sum \min \{f_{i_1}, f_{i_2}, f_{i_3}\}$$

$$- \cdots + (-1)^K \min \{f_1, f_2, \cdots, f_K\}.$$ 

(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 $1 - \int \min\{g_1, g_2, \ldots, g_k\} = s_k^2(g_1, g_2, \ldots, g_k)/2$, 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