L_1 - DISTANCE IN CLASSIFICATION AND DISCRIMINATION

by **Tai VoVan, Khiem NguyenNgoc**
Department of Mathematics, CanTho University, VietNam

Abstract
In this paper, we propose the expression which is considered as L_1 - distance between more than two functions $g_i(x) = q_1 f_1(x)$, where $\{f_i(x)\}$ are probability density functions, $q_i \in (0,1), \sum_{i=1}^{k} q_i = 1, k > 2$. From this idea, some results, related with L_1 - distance between $\{f_i(x)\}$, $\{g_i(x)\}, i = 1,2,...,k$ are obatained. Beside, relations between L_1 - distance and other measures, used in pattern regconition, are also established.
Key words: L^1 - distance; maximum function; Bayes error; affinity.

1. Introduction

The solution of classification and discrimination is concerned with evaluating measure of affinity between functions, defined as distance between them. According to Glick [5] Mardia [9] and Ben Bassat [3], distance is the key in classification and discrimination. Distance between descrete elements was bulit and popurlaly used. Beside, there are many kinds of distance between two probability densities, such as Chernoff, Bhattacharyya, Divergence, Patrick – Fisher, ... distances [17]. When more than two probability density functions, some of definitions can be considered as distance: the conception of affinity of Matusita [10] and Tousiant [11], the concept of k-point separation measurement of Glick [5] or distance of Webb (2002) [16]. It is not easy to select suitable distance in each problem. Gower [6] had much argument around this matter, but he had no best conclusion of the optimal distance for whole situations. To choose the sensible distance, we can base on some following criterions: first, simple and easy to calculate, second, knowledge concerned with data and finally, speedy to run algorithm. With these criterions, it is interesting to pay attention to L_1 distance for data of pattern regconition problem.

L_1 - distance between two functions are defined and used commonly. However, for general, L_1 - distance between more than two functions has not been paid much attention yet. On basic of maximum function, L_1 - distance between more than two probability density functions was, first time, well - defined by T – Pham Gia [14]. It gives author new criterion to build cluster of probability density functions [15]. In this writing, we propose the expression, considered as L_1 - distance more than two functions $g_i(x) = q_i f_i(x)$, where $\{f_i(x)\}$ are probability density functions, $q_i \in (0,1), \sum_{i=1}^{k} q_i = 1, k > 2$. Beside, some results, related with L_1 - distance between $\{f_i(x)\}, \{g_i(x)\}, i = 1,...,k$ are drawn. More over, the relations between L_1 distance and other measures, as well as, quanlities of classification and discrimination is shown in detail.

2. Relations between L_1 - Distance and Qualities in Classification and Discrimination

2.1 L_1 - Distance

Let $f_1(x), f_2(x),..., f_k(x)$ be the probability density functions,

$f_{\max}(x) = \max\{f_1(x), f_2(x),..., f_k(x)\}$, $g_i(x) = q_i f_i(x)$, $q_i \in (0,1)$, $\sum_{i=1}^{k} q_i = 1$. Then,

L_1 - distance is definied as following:

$$\text{When } k = 2: \|f_1, f_2\|_1 = \int_{R^n} |f_1(x) - f_2(x)| dx, \ \|g_1, g_2\|_1 = \int_{R^n} |g_1 - g_2| dx$$

$$\text{When } k > 2: \|f_1, f_2,..., f_k\|_1 = \int_{R^n} f_{\max}(x) - 1 \tag{1}$$

Definition 1: *Let* $g_{\max}(x) = \max\{g_1(x), g_2(x),..., g_k(x)\}$, *then* L_1 *- distance between k (k>2) functions* $\{g_i(x)\}, i = 1,2,...,k$ *is defined by*

$$\|g_1, g_2,..., g_k\|_1 = \int_{R^n} g_{\max}(x) dx - \frac{1}{k} \tag{2}$$

Remark

i) $2\|f_1, f_2,..., f_k\|_1$ *severs criterions of k-point separation measurement, proposed by Glick* [6] .

ii) *In case of* $q_i = \frac{1}{k}, i = 1, 2, \ldots, k$ *we have* $\|f_1, f_2,..., f_k\|_1 = k\|g_1, g_2,..., g_k\|_1$. *Then,* $2\|g_1, g_2,..., g_k\|_1$ *also severs criterions of k-point separation measurement.*

2.2 Bayes Error

The probability of misclassification in discrimination by Bayesian method is called Bayes error. It is proved that Bayes error is the best error in pattern regconition. In detail, it is defined as follows:

Let $f_1(x), f_2(x),..., f_k(x)$ be probability densities with prior $(q) = (q_1, q_2,..., q_k)$ and $g_i(x) = q_i f_i(x), \quad i = 1,2,...,k$.

a) When $k = 2$

Without attention to prior probability, Bayes error is defined by $Pe = \tau + \delta$, where

$\tau = P(w_2|w_1) = \int_{R_2^n} f_1(x) dx$: probability of assign a pattern in w_2 when it belongs to w_1,

$\delta = P(w_1|w_2) = \int_{R_1^n} f_2(x) dx$: probability of assign a pattern in w_1 when it belongs to w_2.

Where $R_1^n = \{x \mid f_1(x) \ge f_2(x)\}$, $R_2^n = \{x \mid f_1(x) < f_2(x)\}$.

If we have knowledge of prior probability q of w_1, then τ and δ become τ^* and δ^* respectively as follows:

$$\tau^* = \int_{R_2^{n*}} q f_1(x) dx \text{ and } \delta^* = \int_{R_1^{n*}} (1-q) f_2(x) dx$$

Where $R_1^{n*} = \{x \mid q f_1(x) \ge (1-q) f_2(x)\}$, $R_2^{n*} = \{x \mid q f_1(x) < (1-q) f_2(x)\}$.

Let $(q) = (q, 1-q)$, we determine Bayes error through out following equation

$$Pe^{(q)} = \tau^* + \delta^*$$

b) When $k > 2$, Bayes error is defined by

$$Pe^{(q)}_{1,2,\ldots,k} = \sum_{i=1}^{k} \int_{R^n \setminus R_i^n} q_i f_i dx = 1 - \sum_{i=1}^{k} \int_{R_i^n} q_i f_i(x) dx$$

2.3 Some Results Concerned with L_1 - Distance between Two Density Functions

a) Let $\lambda_{1,2}$ be the overlapping region's measure of two density functions. Then, $\lambda_{1,2}$ is also minimum probability of misclassification, so we consider it as Bayes error $Pe_{1,2}$. It is not difficult to have some results as follows:

$$\|f_1, f_2\|_1 = 2(1 - Pe_{1,2}) = 2(1 - \lambda_{1,2})$$

$$\|f_1, f_2\|_1 = \int_{R^n} f_{\max}(x)dx - Pe_{1,2} = \int_{R^n} f_{\max}(x)dx - \int_{R^n} f_{\min}(x)dx = 2\left(1 - \int_{R^n} f_{\min}(x)dx\right)$$

b) When $f_1(x)$ and $f_2(x)$ are one dimension normal densities $N(\mu_i, \sigma_i^2)$, $i = 1, 2$. Suppose that $\mu_1 < \mu_2$, then

$$\|f_1, f_2\|_1 = \begin{cases} 2\left(1 - \int_{-\infty}^{x_1} f_2(x)dx - \int_{x_1}^{+\infty} f_1(x)dx\right) & \text{if } \sigma_1 = \sigma_2 \\ 2\left(1 - \int_{-\infty}^{x_2} f_1(x)dx - \int_{x_2}^{x_3} f_2(x)dx - \int_{x_3}^{+\infty} f_1(x)dx\right) & \text{if } \sigma_1 \neq \sigma_2 \end{cases}$$

Where

$$x_1 = \frac{\mu_1 + \mu_2}{2}, \quad x_2 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) - \sigma_1\sigma_2\sqrt{(\mu_1 - \mu_2)^2 + K}}{\sigma_2^2 - \sigma_1^2}$$

$$x_3 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) + \sigma_1\sigma_2\sqrt{(\mu_1 - \mu_2)^2 + K}}{\sigma_2^2 - \sigma_1^2}, \quad K = 2(\sigma_2^2 - \sigma_1^2)\ln\left(\frac{\sigma_2}{\sigma_1}\right) \geq 0.$$

In special case of $\mu_1 = \mu_2$, we have

$$\|f_1, f_2\|_1 = \begin{cases} 0 & \text{if } \sigma_1 = \sigma_2 \\ 2\left(1 - \int_{-\infty}^{x_4} f_1(x)dx - \int_{x_4}^{x_5} f_2(x)dx - \int_{x_5}^{+\infty} f_1(x)dx\right) & \text{if } \sigma_1 \neq \sigma_2 \end{cases}$$

Where $x_4 = \mu - \sigma_1\sigma_2\sqrt{E}$, $x_5 = \mu + \sigma_1\sigma_2\sqrt{E}$ and $E = \dfrac{2}{\sigma_2^2 - \sigma_1^2}\ln\left(\dfrac{\sigma_2}{\sigma_1}\right) \geq 0$.

c) When $f_1(x)$ and $f_2(x)$ are n - dimension normal densities ($n \geq 2$)

$$f_i(x) = \frac{1}{|\Sigma_i|^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} \exp\left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right], i = 1, 2$$

Let $d(x) = \left[\mu_1^T(\Sigma_1)^{-1} - \mu_2^T(\Sigma_2)^{-1}\right]x - \frac{1}{2}x^T\left[(\Sigma_1)^{-1} - (\Sigma_2)^{-1}\right]x - m$, Where $m=\frac{1}{2}\left[\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)+\mu_1^T(\Sigma_1)^{-1}\mu_1-\mu_2^T(\Sigma_2)^{-1}\mu_2\right]$

Now, $\|f_1,f_2\|_1=\int_{R_1} f_1(x)dx+\int_{R_2} f_2(x)dx$, where $R_1=\{x:d(x)\le 0\}$, $R_2=\{x:d(x)>0\}$

If $\Sigma_1=\Sigma_2=\Sigma$, then $d(x)$ is defined by

$$d(x)=(\mu_1-\mu_2)(\Sigma)^{-1}x-\frac{1}{2}(\mu_1-\mu_2)^T(\Sigma)^{-1}(\mu_1+\mu_2)$$

d) With prior probability q and $(1-q)$ are given to $f_1(x)$ and $f_2(x)$ respectively. Let τ and σ be the probability of mis-classifing a pattern to first and second population. According Lissack and Fu [8], relations between Bayes error $Pe_{1,2}=\tau+\delta$ and L^1-distance between qf_1 and $(1-q)f_2$ is determined by $2Pe_{1,2}=1-Z$, in which $Z=\|g_1,g_2\|_1=\|qf_1,(1-q)f_2\|_1$.

Suppose that the prior probability q is random, then, τ, δ and $Pe_{1,2}$ are also Random. When we have no knowledge about $f_1(x)$, $f_2(x)$ and q, but we know that τ and δ are independent, we can find out a density function $h(z)$ of Z. In this case, $h(z)$ is defined by density functions of τ and δ on (0,1/4).

Theorem 1 *Let τ and δ be independent random variable, characterized by two density functions $f_1(x)$ and $f_2(x)$ respectively, on interval (0;1/4). Then, the density function of Z on (0, 1) is well - defined by*

$$h(z)=\begin{cases}\frac{1}{2}\int_{\frac{1-2z}{4}}^{\frac{1}{4}} f_1(t)f_2\left(\frac{1-z-2t}{2}\right)dt & \text{if } 0<z<\frac{1}{2}\\ \frac{1}{2}\int_{0}^{\frac{1-z}{2}} f_1(t)f_2\left(\frac{1-z-2t}{2}\right)dt & \text{if } \frac{1}{2}\le z<1\end{cases} \quad (3)$$

Proof

Suppose that $f_1(x)$, $f_2(x)$ are probability density functions of τ and δ respectively. Let $y=\tau+\delta$, we have

$$g(y)=\int_{-\infty}^{+\infty} f_1(y-x)f_2(x)dx$$

Since δ is random *variable* on (0,1/4), meaning $f_2(x)=0\ \ \forall x\notin(0,1/4)$, we obtain

$$g(y)=\int_{0}^{\frac{1}{4}} f_1(y-x)f_2(x)dx$$

Let $t=y-x\Rightarrow dt=-dx;\ x=0\ \Rightarrow t=y;\ x=\frac{1}{4}\Rightarrow t=y-\frac{1}{4}$

Then, $g(y)=\int_{y}^{y-\frac{1}{4}} f_1(t)f_2(y-t)(-dt)=\int_{y-\frac{1}{4}}^{y} f_1(t)f_2(y-t)dt$

As τ and $\delta\in(0,1/4)$, $y\in(0,1/2)$, we investigate two cases

i) If $0 < y \le \frac{1}{4}$, then $y - \frac{1}{4} \le 0$. From that,

$$g(y) = \int_{y-\frac{1}{4}}^{y} f_1(t) f_2(y-t)dt = \int_{0}^{y} f_1(t) f_2(y-t)dt \tag{4}$$

ii) If $\frac{1}{4} < y \le \frac{1}{2}$, then $y - \frac{1}{4} > 0$. Hence,

$$g(y) = \int_{y-\frac{1}{4}}^{y} f_1(t) f_2(y-t)dt = \int_{y-\frac{1}{4}}^{\frac{1}{4}} f_1(t) f_2(y-t)dt \tag{5}$$

Because $y = \frac{1-z}{2}$ and $y'_z = -\frac{1}{2}$, density function of z is defined by

$$h(z) = | y'_z | g\left(\frac{1-z}{2}\right) = \frac{1}{2} g\left(\frac{1-z}{2}\right) \tag{6}$$

Replae (6) by (4) and (5), we obtain (3). ■

Corollary

a) Let τ and δ be inpendent random variables which have Beta distributions on (0;1/4), result about $h(z)$ is descrided in [13].

b) Let τ and δ be inpendent random variables which receive normal distributions on (0;1/4): $\tau \sim N(\mu_1, \sigma_1^2; 0, 1/4))$, $\delta \sim N(\mu_2, \sigma_2^2; 0, 1/4)$

i) If $0 < z < \frac{1}{2}$, then

$$h(z) = \frac{1}{2} K_1 \exp\left(\frac{2C-B}{4}\right) \exp\left(-\frac{B}{4} z^2 + \frac{B-C}{2} z\right) \left[\Phi\left(-\frac{\sigma_1(1-z)}{2\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}} + K_2 + \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{4\sigma_1\sigma_2} \right) \right.$$
$$\left. - \Phi\left(\frac{\sigma_2(1-z)}{2\sigma_1\sqrt{\sigma_1^2+\sigma_2^2}} + K_2 - \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{4\sigma_1\sigma_2} \right) \right] \tag{7}$$

Where $B = \frac{1}{2(\sigma_1^2+\sigma_2^2)^2}$, $C = \frac{\mu_1+\mu_2}{\sigma_1^2+\sigma_2^2}$, $K_2 = \frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_1\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$.

ii) If $\frac{1}{2} \le z < 1$, then

$$h(z) = \frac{1}{2} K_1 \exp\left(\frac{2C-B}{4}\right) \exp\left(-\frac{B}{4} z^2 + \frac{B-C}{2} z\right) -$$
$$\left[\Phi\left(\frac{\sigma_2(1-z)}{2\sigma_1\sqrt{\sigma_1^2+\sigma_2^2}} + K_2 \right) - \Phi\left(\frac{-\sigma_1(1-z)}{2\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}} + K_2 \right) \right] \tag{8}$$

c) On other hand, let τ and δ be inpendent random variables which are characterized by exponential distributions on $(0,1/4)$: $\tau \sim \exp(b_1; 0, 1/4)$; $\delta \sim \exp(b_2; 0, 1/4)$

i) If $b_1 > b_2$

$$h(z)=\begin{cases}\dfrac{b_1b_2}{2ab(b_1-b_2)}\exp\left(-\dfrac{b_2+b_2}{2}\right)\left[\exp\left(\dfrac{b_1z}{2}\right)-\exp\left(\dfrac{b_2z}{2}\right)\right] & \text{if } 0<z<\dfrac{1}{2}\\ \dfrac{b_1b_2}{2ab(b_1-b_2)}\left[\exp\left(-\dfrac{b_2(1-z)}{2}\right)-\exp\left(-\dfrac{b_1(1-z)}{2}\right)\right] & \text{if } \dfrac{1}{2}\le z<1\end{cases} \quad (9)$$

Where $a=\int_0^{\frac{1}{4}} f_1(x)dx = 1-e^{-\frac{b_1}{4}}$; $b=\int_0^{\frac{1}{4}} f_2(x)dx = 1-e^{-\frac{b_2}{4}}$

ii) In special, if $b_1 = b_2 = c$, then

$$h(z)=\begin{cases}\left(\dfrac{c}{2d}\right)^2 z\exp\left(\dfrac{-c(1-z)}{2}\right) & \text{if } 0<z<\dfrac{1}{2}\\ \left(\dfrac{c}{2d}\right)^2 (1-z)\exp\left(\dfrac{-c(1-z)}{2}\right) & \text{if } \dfrac{1}{2}\le z<1\end{cases} \quad (10)$$

In which $d=\int_0^{\frac{1}{4}} f_1(x)dx = \int_0^{\frac{1}{4}} ce^{-cx}dx = 1-e^{-\frac{c}{4}}$

To obtain (7), (8), (9) from (3), we have to run complex calculation. We do not describe it in this writing.

2.4 Some results related with L_1- distance between more than two densities

Determine k $n-$dimension densities $f_i(x)$, $k \ge 3$, let $d_{ij} = f_i - f_j$ and

$R_1^n = \{x: d_{1p}(x) > 0\}$, $R_l^n = \{x: d_{lm}(x) \ge 0 \cap d_{nl}(x) \le 0\}$, $R_k^n = \{x: d_{qk}(x) < 0\}$

where $p = 2, \ldots, k$; $q = 1, \ldots, k-1$; $l = 2, \ldots, k-1$; $m = l+1, \ldots, k$; $n = 1, 2, \ldots, l-1$. We obtain $\|f_1, f_2, \ldots, f_k\|_1 = \int_{R_1^n} f_1(x)dx + \sum_{l=2}^{k-1}\int_{R_l^n} f_l(x)dx + \int_{R_k^n} f_k(x)dx - 1$.

If random variables receive $n-$dimension distributions, discriminant function $d_{ij}(x)$ becomes $d_{ij}(x) = \left[\mu_i^T(\Sigma_i)^{-1} - \mu_j^T(\Sigma_j)^{-1}\right]x - \frac{1}{2}x^T\left[(\Sigma_i)^{-1} - (\Sigma_j)^{-1}\right]x - m$

in which $m = \dfrac{1}{2}\left[\ln\left(\dfrac{|\Sigma_i|}{|\Sigma_j|}\right) + \mu_i^T(\Sigma_i)^{-1}\mu_i - \mu_j^T(\Sigma_j)^{-1}\mu_j\right]$

Theorem 2 *The relations between Bayes error and overlapping coefficience are shown as follows*:

a) $\|g_1, g_2, \ldots, g_k\|_1 = \dfrac{1}{k}\left[k-1-kPe_{1,2,\ldots,k}^{(q)}\right]$, $k \ge 3$ (12)

b) $\|g_1, g_2, \ldots, g_k\|_1 = 1 - \dfrac{1}{k} - \left[\sum_{k<l}\lambda_{kl}^{(q)} - \sum_{k<l<m}\lambda_{klm}^{(q)} + \ldots + (-1)^{k-1}\lambda_{12\ldots k}^{(q)}\right]$ (13)

Where $\lambda_{ij...l}^{(q)}$ *is the measure of overlapping of probability density functions* $g_i(x), g_j(x), ..., g_l(x)$.

Proof

a) We define $\max_{1\le l\le k}\{q_l f_l(x)\} = q_j f_j$ on R_j^n. Then, Bayes error in pattern regconition is given by

$$Pe_{1,2,...,k}^{(q)} = \sum_{j=1}^{k} \int_{R^n \setminus R_j^n} q_j f_j(x)dx$$

$$= \sum_{j=1}^{k}\left[\int_{R^n} q_j f_j(x)dx - \int_{R_j^n}\max_{1\le l\le k}\{q_l f_l(x)\}dx\right]$$

$$= \int_{R^n}\sum_{j=1}^{k} q_j f_j(x)dx - \sum_{j=1}^{k}\int_{R_j^n}\max_{1\le l\le k}\{q_l f_l(x)\}dx$$

$$= 1 - \int_{R^n}\max_{1\le l\le k}\{q_l f_l(x)\}dx$$

Otherwise, $\|g_1, g_2, ..., g_k\|_1 = \int_{R^n} g_{\max}(x)dx - \frac{1}{k}$. Thus,

$$\|g_1, g_2, ..., g_k\|_1 = 1 - Pe^{(q)} - \frac{1}{k} = \frac{1}{k}\left(k - 1 - kPe^{(q)}\right)$$

b) Since

$$g_{\max}(x) = \sum_{i=1}^{k} g_k - \sum_{i<j}\min(g_i, g_j) + \sum_{i<j<l}\min(g_i, g_j, g_l) + \cdots + (-1)^{k-1}\min(g_1, g_2, \cdots g_k)$$

We reach $\int_{R^n} g_{\max}(x)dx = 1 - \sum_{i<j}\lambda_{i,j}^{(q)} + \sum_{i<j<l}\lambda_{i,j,l}^{(q)} + \cdots + (-1)^{k-1}\lambda_{1,2,...,k}^{(q)}$

Therefore, $k\int_{R^n} g_{\max}(x)dx = k - k\left[\sum_{i<j}\lambda_{i,j}^{(q)} + \sum_{i<j<l}\lambda_{i,j,l}^{(q)} + \cdots + (-1)^{k-1}\lambda_{1,2,...,k}^{(q)}\right]$

The above expression means that

$$\|g_1, g_2, ..., g_k\|_1 = 1 - \frac{1}{k} - \left[\sum_{i<j}\lambda_{i,j}^{(q)} + \sum_{i<j<l}\lambda_{i,j,l}^{(q)} + \cdots + (-1)^{k-1}\lambda_{1,2,...,k}^{(q)}\right] \quad \blacksquare$$

In case of $q_i = \frac{1}{k}$, the equations (12) and (13) becomes

$\|f_1, f_2, ..., f_k\|_1 = k - 1 - kPe_{1,2,...,k}^{(1/k)}, \quad k \ge 3$

$\|f_1, f_2, ..., f_k\|_1 = (k-1)\left(1 - \frac{k}{2}\right) + \frac{1}{2}\sum_{i<j}\|f_i, f_j\|_1 + \sum_{k<l}\lambda_{kl} - \sum_{k<l<m}\lambda_{klm} + ... + (-1)^{k-1}\lambda_{12...k}$

These equations are shown in [14].

Theorem 3 *The relations between* $\|g_1, g_2, ..., g_k\|_1$ *and other measures is described by results*:

a) $(k+1)\|g_1, g_2,\ldots,g_{k+1}\|_1 - k\|g_1, g_2,\ldots,g_k\|_1 = 1 - \int_{R^n} \min\{h_1(x), f_{k+1}(x)\}dx$ (14)

where $h_1(x) = \max\{f_1(x), f_2(x),\ldots,f_k(x)\}$, $k \geq 3$

b) $k\|g_1, g_2,\ldots,g_k\|_1 = n\|g_1, g_2,\ldots,g_n\|_1 + (k-n)\|g_{n+1}, g_{n+2},\ldots,g_k\|_1 + 1 - A$ (15)

in which $n, k \geq 3, n < k$, $A = \int_{R^n} \min\{k_1(x), k_2(x)\}dx$ *and*

$k_1(x) = \max\{f_1(x), f_2(x),\ldots,f_n(x)\}$, $k_2(x) = \max\{f_{n+1}(x), f_{n+2}(x),\ldots,f_k(x)\}$.

Proof

a) We have

$$Pe^{[(1/k)]}_{1,2,..,k} = 1 - \frac{1}{k}\int_{R^n} h_1(x)dx \quad \text{and} \quad Pe^{[(1/k+1)]}_{1,2,..,k+1} = 1 - \frac{1}{k+1}\int_{R^n} h_2(x)dx,$$

where $h_1(x) = \max\{f_1(x), f_2(x),\ldots,f_k(x)\}$, $h_2(x) = \max\{f_1(x), f_2(x),\ldots,f_{k+1}(x)\}$.

Since $h_2(x) = h_1(x) + f_{k+1}(x) - \min\{h_1(x), f_{k+1}(x)\}$

We reach

$$Pe^{[(1/k+1)]}_{1,2,..,k+1} = \frac{1}{k+1}\int_{R^n}\left[h_1(x) + f_{k+1}(x) - \min\{h_1(x), f_{k+1}(x)\}\right]dx$$

$$= \frac{1}{k+1} + \frac{k}{k+1}\left[1 - \frac{1}{k}\int_{R^n} h_1(x)dx\right] - \frac{1}{k+1}\int_{R^n} f_{k+1}(x)dx + \mathrm{B}$$

$$= \frac{k}{k+1} Pe^{(1/k)}_{1,2,\ldots,k} + B,$$

In which $B = \dfrac{1}{k+1}\displaystyle\int_{R^n} \min\{h_1(x), f_{k+1}(x)\}dx$.

Replace $Pe^{(1/k)}_{1,2,..,k} = 1 - \frac{1}{k} - \|g_1, g_2,\ldots,g_k\|_1$ and $Pe^{(1/k)}_{1,2,..,k+1} = 1 - \frac{1}{k+1} - \|g_1, g_2,\ldots,g_{k+1}\|_1$ into above expression, we obtain equation (14).

b) In case of equal prior, we have

$$n\|g_1, g_2,\ldots,g_n\|_1 = n - 1 - nPe^{(1/n)}$$

$$m\|g_{n+1}, g_{n+2},\ldots,g_k\|_1 = m - 1 - mPe^{(1/m)} \quad \text{where } m = k - n.$$

Therefore, $n\|g_1, g_2,\ldots,g_n\|_1 + m\|g_{n+1}, g_{n+2},\ldots,g_k\|_1 = k - 2 - \left[nP^{(1/n)}_{e1,2,\ldots,n} + mP^{(1/m)}_{e1,2,\ldots,m}\right]$ (16)

Other hand, $Pe^{(1/k)}_{1,2,..,k} = 1 - \frac{1}{k}\int_{R^n} \max\{k_1(x), k_2(x)\}dx$

$$= 1 - \frac{1}{k}\int_{R^n}\left[k_1(x) + k_2(x) - \min\{k_1(x), k_2(x)\}\right]dx$$

$$= \frac{1}{k}\left[m - \int_{R^n} k_1(x)dx + n - \int_{R^n} k_2(x)dx + \int_{R^n} \min\{k_1(x), k_2(x)\}dx\right]$$

$$= \frac{1}{k}\left[mP^{(1/m)}_{e1,2,\ldots,m} + nP_e{}^{(1/n)}_{1,2,..n} + A\right]$$

The above equation means that

$$\left[mP_{e1,2,\ldots,m}^{(1/m)}+nP_{e\ 1,2,\ldots,n}^{(1/n)}\right]=kPe_{1,2,\ldots,k}^{(1/k)}-A=k\left(1-\frac{1}{k}-\|g_1,g_2,\ldots,g_k\|_1\right)-A$$

Substitute this result in (16), the proof can be finished. ■

Theorem 4 *We have some results about bounds for distance between k functions* $g_1, g_2, \ldots, g_k$ as following:

a) $$\max\{q_i\}-\frac{1}{k}\le\|g_1,g_2,\ldots,g_k\|_1\le 1-\frac{1}{k},\quad k\ge 2 \qquad (17)$$

The equality, in left side, happens if $g_i(x)$ coincide and does in the right if $g_i(x)$ disjoint.

b) $$\frac{1}{2}\max\left\{\|g_i,g_j\|_1\right\}+\min\{q_i\}-\frac{1}{k}\le\|g_1,g_2,\ldots,g_k\|_1\le\frac{1}{k}\sum_i\sum_j\|g_i,g_j\|_1 \qquad (18)$$

Proof

a) We have $q_i f_i(x)\le\max\{q_1f_1(x),q_2f_2(x),\ldots,q_kf_k(x)\}\le\sum_{i=1}^{k}q_if_i(x)$ and

$$\int_{R^n}f_i(x)dx=1$$

Hence, $q_i\le\int_{R^n}g_{\max}(x)dx\le 1\Leftrightarrow\max\{q_i\}\le\int_{R^n}g_{\max}(x)dx\le 1$

$$\Leftrightarrow\max\{q_i\}-\frac{1}{k}\le\int_{R^n}g_{\max}(x)dx-\frac{1}{k}\le 1-\frac{1}{k}$$

From that, we obtain (17).

b) We also have

$$\begin{aligned}\int_{R^n}\max\{g_1(x),g_2(x),\ldots,g_k(x)\}dx &\ge\max_{i<j}\int_{R^n}\max\{g_i(x),g_j(x)\}dx\\ &=\max_{i<j}\left\{\frac{1}{2}\|g_i,g_j\|_1+\frac{1}{2}(q_i+q_j)\right\}\\ &\ge\max_{i<j}\left\{\frac{1}{2}\|g_i,g_j\|_1\right\}+\min_{i<j}\left\{\frac{1}{2}(q_i+q_j)\right\}\\ &\ge\max_{i<j}\left\{\frac{1}{2}\|g_i,g_j\|_1\right\}+\min_{i<j}\{(q_1,q_2,\ldots,q_k)\}\end{aligned}$$

So, $\int_{R^n}g_{\max}(x)dx-\frac{1}{k}\ge\frac{1}{2}\max_{i<j}\left\{\|g_i,g_j\|_1\right\}+\min_{i<j}\{(q_1,q_2,\ldots,q_k)\}-\frac{1}{k}$, meaning that we finish the proof of left side.

According to Glick (1973),

$$\begin{aligned}\sum_i\sum_j|g_i-g_j|&\ge\sum_i^k\left[\max\{g_1,g_2,\cdots g_k\}-g_j\right]\\ &=k[\max\{g_1,g_2,\cdots g_k\}]-\sum_{j=1}^k g_j\end{aligned}$$

Hence, $\max\{g_1,g_2,\cdots g_k\}\le\frac{1}{k}\sum_{i<}\sum_j|g_i-g_j|+\frac{1}{k}\sum_{j=1}^k g_j$

As $\int\limits_{R^n} g_i(x)dx = 1$, the above inequality turns to $\int\limits_{R^n} g_{\max}(x) \le \frac{1}{k}\sum_{i<}\sum_{j}\int\limits_{R^n}\|g_i, g_j\|_1 + \frac{1}{k}$

It is synonym with right side is proved. ■

3. The Relations between L_1 - Distance and Other Measures

According [14], the relation between Bhattacharrya $D_B(f_1, f_2)$ and $Pe_{1,2}$ is defined by $\frac{1}{2}\left(1-\sqrt{1-4D_B{}^2(f_1, f_2)}\right) \le Pe_{1,2} \le D_B(f_1, f_2)$. (19)

Substitute $Pe_{1,2} = \int\limits_{R^n} f_{\max}(x)dx - \|f_1, f_2\|_1$ in (19), we obtain the relation between D_B and $\|f_1, f_2\|_1$:

$$\int\limits_{R^n} f_{\max}(x)dx - D_B(f_1, f_2) \le \|f_1, f_2\|_1 \le \int\limits_{R^n} f_{\max}(x)dx - \frac{1}{2}\left(1-\sqrt{1-4D_B^2(f_1, f_2)}\right)$$

Theorem 5 *The theorem demonstrates the relation between L_1 - distance and Affinity, proposed by Toussaint*

Let $f_i(x)$ be probability density functions and $g_i(x) = q_i f_i(x)$, $q_i \in (0, 1)$ $\sum_{i=1}^{k} q_i = 1$, $k > 2$,

we reach

$$\text{a) } \|g_1, g_2, \ldots, g_k\|_1 \ge \frac{1}{k-1}\left(1-\prod_{j=1}^{k} q_j^{\alpha_j} D_T(f_1, f_2, \ldots, f_k)^{\alpha}\right) - \frac{1}{k} \quad (20)$$

$$\text{b) } \|g_1, g_2, \ldots, g_k\|_1 \ge 1 - \frac{1}{k} - \left(\sum_{i<j} q_i^{\alpha_\beta} q_j^{1-\beta} D_T(f_i, f_j)^{(\beta, 1-\beta)}\right) \quad (21)$$

in which $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, $\beta, \alpha_j \in (0,1)$, $\sum_{j=1}^{k} \alpha_j = 1$, $i, j = 1, 2, \ldots, k$ \

$$D_T(f_1, f_2, \ldots, f_k)^{\alpha} = \int\limits_{R^n}\prod_{j=1}^{k}[f_j(x)]^{\alpha_j} dx$$

Proof

a) For each $j = 1, 2, \ldots,\ k$, we have $\left(\sum_{j=1}^{k} q_j f_j\right)^{\alpha_i} \ge (q_i f_i)^{\alpha_i}$, $i = 1, 2, \ldots, k$.

$$\text{Therefore, } \left(\sum_{j=1}^{k} q_j f_j\right)^{\alpha_1+\alpha_2+\ldots+\alpha_k} \ge \prod_{j=}^{k}(q_j f_j)^{\alpha_j} \Leftrightarrow \sum_{j=1}^{k} q_j f_j \ge \prod_{j=}^{k}(q_j f_j)^{\alpha_j} \quad (22)$$

Other wise, $\left(\min_{1\le j\le k}\{q_j f_j\}\right)^{\alpha_1} \le (q_1 f_1)^{\alpha_1}, \ldots\ldots, \left(\min_{1\le j\le k}\{q_j f_j\}\right)^{\alpha_k} \le (q_k f_k)^{\alpha_k}$

Hence, $\left(\min_{1\le j\le k}\{q_j f_j\}\right)^{\alpha_1+\cdots+\alpha_k} \le \prod_{j=1}^{k}(q_j f_j)^{\alpha_j}$

$$\text{It means that } \min_{1\le j\le k}\{q_j f_j\} \le \prod_{j=1}^{k}(q_j f_j)^{\alpha_j} \quad (23)$$

Combine (22) with (23), we obtain $0 \le \sum_{j=1}^{k} q_j f_j - \prod_{j=1}^{k} (q_j f_j)^{\alpha_j} \le \sum_{j=1}^{k} q_j f_j - \min_{1\le j\le k}\{q_j f_j\}$

Since $\sum_{j=1}^{k} q_j f_j - \min_{1\le j\le k}\{q_j f_j\}$ consists of $k-1$ terms $q_j f_j$,

$$\sum_{j=1}^{k} q_j f_j - \min_{1\le j\le k}\{q_j f_j\} \le (k-1)\max_{1\le j\le k}\{q_j f_j\}$$

Therefore, we obtain $\quad 0 \le \sum_{j=1}^{k} q_j f_j - \prod_{j=1}^{k} (q_j f_j)^{\alpha_j} \le (k-1)\max_{1\le j\le k}\{q_j f_j\}$

Take intergral on R^n both two sides of above inequality, we reach

$$1 - \prod_{j=1}^{k} q_j^{\alpha_j} D_T(f_1, f_2, \ldots, f_k)^{\alpha} \le (k-1) \int_{R^n} g_{\max}(x)dx$$

Substitute $\int_{R^n} g_{\max}(x) = \frac{1}{k}\left(k\|g_1, g_2, \ldots, g_k\|_1 + 1\right)$ in above result, we obtain (20)

b) We have

$$Pe_{1,2\ldots,k}^{[(q)]} = 1 - \int_{R^n} g_{\max} dx = \sum_{j=1}^{k} \int_{R^n \setminus R_j} q_j f_j dx$$

$$= \sum_{j=1}^{k} \sum_{j \ne i} \int_{R_j} \min\{q_i f_i, q_j f_j\} dx = \sum_{i<j} \int_{R_i \cup R_j} \min\{q_i f_i, q_j f_j\} dx$$

As $[\min\{q_i f_i, q_j f_j\}]^{\beta} \le (q_i f_i)^{\beta}$ and $[\min\{q_i f_i, q_j f_j\}]^{1-\beta} \le (q_j f_j)^{1-\beta}$, we attain $\min\{q_i f_i, q_j f_j\} \le (q_i f_i)^{\beta}(q_j f_j)^{1-\beta}$. Thence, $Pe_{1,2\ldots,k}^{[(q)]} \le \sum_{i<j} \int_{R^n} (q_i f_i)^{\beta}(q_j f_j)^{1-\beta} dx$

Once more, substitute $Pe^{[(q)]} = 1 - \int_{R^n} g_{\max}(x)dx = \frac{1}{k}\left[k - 1 - k\|g_1, g_2, \ldots, g_k\|_1\right]$,

we obtain (21). ■

In instance, $\alpha_1 = \alpha_2 = \ldots = \alpha_k = \frac{1}{k}$,on basic of (20) and (21), we obtain relation between $\|g_1, g_2, \ldots, g_k\|_1$ and Matusita distance. In special, when $k = 2$, relation between $\|g_1, g_2\|_1$ and Hellinger distance is also drawn.

4. Computing L_1-Distance and Bayes Error

To calculate L_1- distance between $\{f_i(x)\}$ and $\{g_i(x)\}$, $i = 1, 2, \ldots, k$, as well as Bayes error $Pe_{1,2,\ldots,k}^{(q)}$, two essential tasks which have to be done are finding out maximum functions $f_{\max}(x)$, $g_{\max}(x)$ and evaluating integral of maximum functions on R^n.

In case of one dimension and $k = 2$, we can find out, for exponentail, beta and normal distributions, the expression of maximum function and formula of Bayes error in detail. When $k > 2$, the task becomes more complex. This difficulty comes from the very varied forms of interaction space curves between the density surfaces, which are either normal or not, although they are project into conic curves in the horizonetal plan, in the normal case. On basic of Maple software, a program is described by us in [13] in order to calculate L_1 - distance for k one – dimension densities and Bayes error in classification and discrimination.

Otherwise, in case of multi - dimension, the task gets extremely complex, even in case of two dimension. To solve the task, intergration of max function can be done by quasi-Monte Carlo method. A program by Matlab software which be used to calculate in example 2 this section is also written by us.

Explicitly, we determine two following examples:

Example 1: Let examine seven normal densities with parameters μ_i, σ_i

$$\mu_1 = 0.3, \mu_2 = 4.0, \mu_3 = 9.1, \mu_4 = 1.9, \mu_5 = 5.3, \mu_6 = 8, \mu_7 = 4.8$$
$$\sigma_1 = 1.0, \sigma_2 = 1.3, \sigma_3 = 1.4, \sigma_4 = 1.6, \sigma_5 = 2, \sigma_6 = 1.9, \sigma_7 = 2.3$$

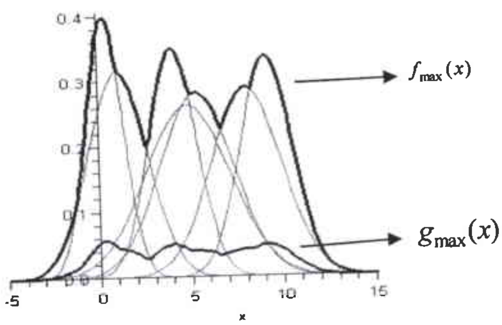

Figure 1. Graph, $f_{\max}(x)$ and $g_{\max}(x)$ for 6 densities

Max function can be found as following:

$$f_{\max}(x) = \begin{cases} f_1 & \text{if} \quad -1.2831 < x \le 0.9856 \\ f_2 & \text{if} \quad 2.5835 < x \le 4.8932 \\ f_3 & \text{if} \quad 8.2961 < x \le 12.5172 \\ f_4 & \text{if} \quad \{-7.8585 < x \le -1.2831\} \cup \{0.9856 < x \le 2.5835\} \\ f_5 & \text{if} \quad 4.8932 < x \le 6.6485 \\ f_6 & \text{if} \quad \{6.6485 < x \le 8.2961\} \cup \{12.5171 < x \le 23.3294\} \\ f_7 & \text{if} \quad \{x \le -7.8585\} \cup \{x > 23.3294\} \end{cases}$$

L_1 disance between f_i is $\|f_1, f_2, \dots, f_7\|_1 = 2.6946$.

In case of $q_i = \frac{1}{7}, i = 1,2,..,7$, we have $\|g_1, g_2, \dots, g_7\|_1$ = 0.3849, and Bayes error is $Pe^{(1/7)} = 0.4722$

Example 2: Let w_1, w_2 and w_3 be populations, characterized by two – dimension normal densities with known parameters

$$\Sigma_1 = \begin{bmatrix} 0.706 & -0.251 \\ -0.251 & 0.507 \end{bmatrix}, \ \mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \ \Sigma_2 = \begin{bmatrix} 0.792 & -0.298 \\ -0.298 & 0.507 \end{bmatrix}$$
$$\mu_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 0.397 & -0.200 \\ -0.200 & 0.706 \end{bmatrix}, \ \mu_3 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

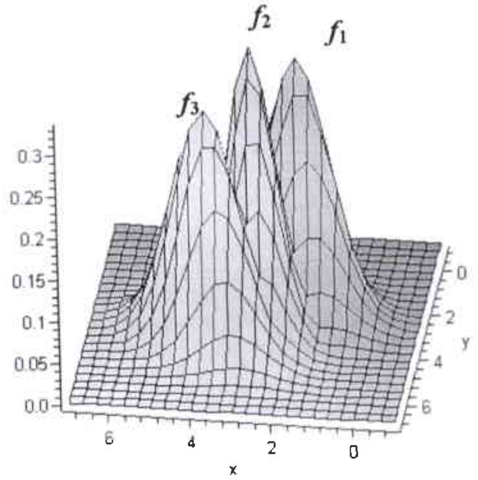

Figure 2. Graph for 3 two – dimension normal densities

Max function of them is defined by

$$f_{\max}(x,y)=\begin{cases} f_1(x,y) & \text{if } (x,y)\in R_1 \\ f_2(x,y) & \text{if } (x,y)\in R_2 \\ f_3(x,y) & \text{if } (x,y)\notin (R_1\cup R_2) \end{cases}$$

In which,

$$R_1=\{(h_1-y<0\cup h_2-y>0)\cap(h_3-y>0\cap h_4-y<0)\},$$

$$R_2=\{(h_1-y>0\cap h_2-y<0)\cap(h_5-y>0\cap h_6-y<0)\},$$

$$h_1=-0.0421x-1.0956+1.2787.10^{-10}\sqrt{9.5067.10^{18}x^2-9.54027.10^{19}x+2.61776.10^{21}}$$

$$h_2=-0.0421x-1.0956-1.2787.10^{-10}\sqrt{9.5067.10^{18}x^2-9.54027.10^{19}x+2.61776.10^{21}}$$

$$h_3=-0.7292x+52.2358+6.8626.10^{-10}\sqrt{2.5348.10^{18}x^2-9.5629.10^{18}x+4.7005.10^{21}}$$

$$h_4=-0.7292x+52.2358-6.8626.10^{-10}\sqrt{2.5348.10^{18}x^2-9.5629.10^{18}x+4.7005.10^{21}}$$

$$h_5=-0.1500x+7.2805+1.0778.10^{-10}\sqrt{1.2354.10^{20}x^2-3.5745.10^{20}x+6.2027.10^{20}}$$

$$h_6=-0.1500x+7.2805-1.0778.10^{-10}\sqrt{1.2354.10^{20}x^2-3.5745.10^{20}x+6.2027.10^{20}}$$

we obtain $\|f_1, f_2, f_3\|_1 = 0.95$ and $Pe^{(1/3)}_{1,2,3}=0.35$.

5. Conclusion

In this article, we have supplemented some results about using L^1- distance between these k probability density functions. Furthermore, we propose the expression which is considered as L^1- distance between more than two functions $g_i(x)$. From here, some new results have present: Bounds for distance, relations with affinity of Toussaint, Bayes error and overlapping coefficience. Relations concerning this distance of two consecutive distances (which only differ by one element) and those of two distances and their union have established. Problems of computing are also considered in article. These results will be used in classification, discriminantion and cluster analysis.

REFERENCES

[1] Anderson, T.W. (1984), *An introduction to multivariate statistical analysis*, Wiley, New York.

[2] Andrew Webb (2002), *Statistical partten recoginition*, JonhWiley and Sons, England.

[3] Ben Bassat, M. (1982), "Use of distance measure, iformation meaure and error bounds in feature evalution. In: Krishnaiah, P. R. and Kanal, L.N.", *Handbook of Statistics*, North Holland. Co. 2, pp.773–791.

[4] Devroye, L., Gyorfi, L. and Lugosi, G., (1996), *A probabilistic theory of pattern recognition*, Springer, New York.

[5] Glick, N. (1973). Separation and Probability of Correct Classification among two or more Distributions. *Annals Inst Stat Math*. 25: 373 –382.

[6] Gower, J.C. A (1971), "General coefficent of similarity and some of its I properties", *Biometrics* 27, pp.857–874.

[7] Johnson, R.A. and Wichern, D.W.(1998), *Applied Multivariate Statistical Analysis*, 4th Ed., Prentice–Hall, Upper Saddle River, N.J.

[8] Lissack TSVI, Fu Ks, (1976), *Error estimation in pattern recognitions*, IEEE rans Info Theory 22, pp.34–45.

[9] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979), *Multivariate analysis*, Academic Press, New York.

[10] Matusita, K. (1967), On the Notion of Affinity of Several Distributions and Some of its Applications, *Annals Inst. Stat. Math*, 19, 181–192.

[11] Parzen, E.(1962), "On estimation of a probability density function and mode", Annals of Mathematical statiststics 33, pp. 1065 – 1076.

[12] Pham–Gia, T., Turkkan, .(2006). Bayesian Analysis in the L^1–Norm of the Mixing Proportion using Discriminant Analysis, *Metrika*, 64 (91), 1–22.

[13] Pham-Gia,T., Turkkan, N., Bekker, A (2007). Bounds for the Bayes error in classification: a Bayesian approach using discriminant analysis. Metrika 64(1): 1-22.

[14] Pham–Gia, T., Turkkan N and Vovan, T. (2008). Statistical Discrimination Analysis Using the Maximum Function, *Communications in Statistics, Simulation and Computation*, 37(2): 320–336.

[15] Tai, Vovan, Pham–Gia, T., (2009), "Clustering probability distributions", *Journal of Applied statistics*, 37(11): 1891 – 1910.

[16] Toussaint G.T. (1972), "Some inequalities between distance measures for feature evaluation", *I.EEE Trans Comput* 21, pp.409–410.

[17] Webb, A. (2002), *Statistical Pattern Recognition*, 2nd Ed., John Wiley & Sons, New York.