Information **2015**, *6*, 361-374; doi:10.3390/info6030361

OPEN ACCESS

information

ISSN 2078-2489

www.mdpi.com/journal/information

Article

A Class of New Metrics Based on Triangular Discrimination

Guoxiang Lu [1,2,*] and Bingqing Li [1]

[1] Department of Risk Management and Insurance, Nankai University, No. 94 Weijin Road, 300071 Tianjin, China; E-Mail: libq@nankai.edu.cn

[2] School of Statistics and Mathematics, Zhongnan University of Economics and Law, No. 182 Nanhu Avenue, 430073 Wuhan, China

* Author to whom correspondence should be addressed; E-Mail: lgxmath@znufe.edu.cn; Tel.: +86-27-8838-5335.

Academic Editor: Willy Susilo

Received: 2 June 2015 / Accepted: 14 July 2015 / Published: 17 July 2015

Abstract: In the field of information theory, statistics and other application areas, the information-theoretic divergences are used widely. To meet the requirement of metric properties, we introduce a class of new metrics based on triangular discrimination which are bounded. Moreover, we obtain some sharp inequalities for the triangular discrimination and other information-theoretic divergences. Their asymptotic approximation properties are also involved.

Keywords: triangular discrimination; metric; triangle inequality; information-theoretic divergence; information inequalities

---

1. Introduction

In many applications such as pattern recognition, machine learning, statistics, optimization and other applied branches of mathematics, it is beneficial to use the information-theoretic divergences rather than the squared Euclidean distance to estimate the (dis)similarity of two probability distributions or positive arrays [1–9]. Among them the Kullback–Leibler divergence (relative entropy), triangular discrimination, variation distance, Hellinger distance, Jensen–Shannon divergence, symmetric Chi-square divergence, J-divergence and other important measures often play a critical role. Unfortunately, most of these divergences do not satisfy the metric properties and unboundedness [10]. As we know, metric properties are the preconditions for numerous convergence properties of iterative algorithms [11]. Moreover, boundedness is also highly concerned in numerical computations and simulations. In paper [12], Endres and Schindelin have proved that the square root of twice Jensen–Shannon divergence is a metric. Triangular discrimination presented by Topsøe in [13] is a non-logarithmic measure and is simple in complex computation. Inspired by [12], we discuss the triangular discrimination. In this paper, the main result is that a class of new metrics derived from the triangular discrimination are introduced. Finally, some new relationships among triangular discrimination, Jensen–Shannon divergence, square of Hellinger distance, variation distance are also obtained.

2. Definition and Auxiliary Results

Definition 1. *Let*

$$\Gamma_n = \left\{ P = (p_1, p_2, \cdots, p_n) | p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}, \quad n \geq 2$$

be the set of all complete finite discrete probability distributions. For all $P, Q \in \Gamma_n$, the triangular discrimination is defined by

$$\Delta(P, Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i}. \tag{1}$$

In the above definition, we use convention based on limitation property that $\frac{0}{0} = 0$.

The triangular discrimination is obviously symmetric, nonnegative and vanishes for $P = Q$, but it does not fulfill the triangle inequality. In the view of the foregoing, the concept of triangular discrimination should be generalized. If $P, Q \in \Gamma_n$, the function $\Delta_\alpha(P, Q)$ is studied:

$$\Delta_\alpha(P, Q) = \left(\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i} \right)^\alpha, \tag{2}$$

where $\alpha \in (0, +\infty)$.

In the following, the $\alpha-$power of the summand in $\Delta(P, Q)$ with all $\alpha \in (0, +\infty)$ are discussed.

Definition 2. *Let the function $L(p, q) : [0, +\infty) \times [0, +\infty) \to [0, +\infty)$ be defined by*

$$L(p, q) = \frac{(p - q)^2}{p + q}. \tag{3}$$

It is easy to see that $L(p, q) \geq 0$ and $L(p, q) = L(q, p)$. To all $\alpha \in (0, +\infty)$, **the issue of whether $(L(p, q))^\alpha$ satisfies the triangle inequality is considered in the following.**

Lemma 1. *If the function $g : [0, a) \cup (a, +\infty) \to (-\infty, +\infty)$ is defined by*

$$g(x) = \frac{\frac{(x-a)(x+3a)}{(x+a)^2}}{\sqrt{L(a, x)}}$$

with $a > 0$, then

$$\lim_{x \to a^+} g(x) = \sqrt{\frac{2}{a}}, \quad \lim_{x \to a^-} g(x) = -\sqrt{\frac{2}{a}}.$$

Proof. As

$$g(x) = \begin{cases} \frac{x+3a}{(x+a)^{\frac{3}{2}}}, & x > a \\ -\frac{x+3a}{(x+a)^{\frac{3}{2}}}, & 0 \leq x < a \end{cases}$$

we can get

$$\lim_{x\to a^+} g(x) = \frac{a+3a}{(a+a)^{\frac{3}{2}}} = \sqrt{\frac{2}{a}}, \quad \lim_{x\to a^-} g(x) = -\frac{a+3a}{(a+a)^{\frac{3}{2}}} = -\sqrt{\frac{2}{a}}.$$

□

Lemma 2. *If the function* $h : [0, +\infty) \to (0, +\infty)$ *is defined by* $h(x) = \frac{3x+a}{(x+a)^{\frac{3}{2}}}$ *with* $a > 0$*, then* h *is monotonic increasing in* $[0, a)$ *and monotonic decreasing in* $(a, +\infty)$*.*

Proof. Straightforward derivative shows

$$h'(x) = \frac{3(a-x)}{2(x+a)^{\frac{5}{2}}},$$

$h'(x) > 0$ in $[0, a)$ and $h'(x) < 0$ in $(a, +\infty)$. Thus the lemma holds. □

Assuming $0 < p < q$, we introduce function $R_{pq} : [0, +\infty) \to [0, +\infty)$ defined by

$$R_{pq}(r) = \sqrt{L(p,r)} + \sqrt{L(q,r)}.$$

Lemma 3. *The function* $R_{pq}(r)$ *has two minima, one at* $r = p$ *and the other at* $r = q$*.*

Proof. The derivative of the function $R_{pq}(r)$ is

$$R_{pq}{}'(r) = \frac{1}{2}\left(\frac{\frac{(r-p)(r+3p)}{(r+p)^2}}{\sqrt{L(p,r)}} + \frac{\frac{(r-q)(r+3q)}{(r+q)^2}}{\sqrt{L(q,r)}}\right). \tag{4}$$

So $R_{pq}{}'(r) < 0$ for $r \in [0, p)$ and $R_{pq}{}'(r) > 0$ for $r \in (q, +\infty)$. It shows $R_{pq}(r)$ is monotonic decreasing in $[0, p)$ and monotonic increasing in $[q, +\infty)$.

Next consider the monotonicity of $R_{pq}(r)$ in the open interval (p, q).

From Lemma 3, we have

$$\begin{aligned} \lim_{r\to p^+} \frac{\frac{(r-p)(r+3p)}{(r+p)^2}}{\sqrt{L(p,r)}} &= \sqrt{\frac{2}{p}}, \\ \lim_{r\to q^-} \frac{\frac{(r-q)(r+3q)}{(r+q)^2}}{\sqrt{L(q,r)}} &= -\sqrt{\frac{2}{q}}. \end{aligned} \tag{5}$$

From Lemma 2, we have

$$\begin{aligned} \frac{\frac{(p-q)(p+3q)}{(p+q)^2}}{\sqrt{L(p,q)}} &= -\frac{p+3q}{(p+q)^{\frac{3}{2}}} > -\frac{p+3p}{(p+p)^{\frac{3}{2}}} = -\sqrt{\frac{2}{p}}, \\ \frac{\frac{(q-p)(q+3p)}{(p+q)^2}}{\sqrt{L(p,q)}} &= \frac{q+3p}{(p+q)^{\frac{3}{2}}} < \frac{q+3q}{(q+q)^{\frac{3}{2}}} = \sqrt{\frac{2}{q}}. \end{aligned} \tag{6}$$

Using (5) and (6),

$$\lim_{r\to p^+} R_{pq}{}'(r) = \frac{1}{2}\left(\lim_{r\to p^+}\frac{\frac{(r-p)(r+3p)}{(r+p)^2}}{\sqrt{L(p,r)}} + \frac{\frac{(p-q)(p+3q)}{(p+q)^2}}{\sqrt{L(p,q)}}\right) = \frac{1}{2}\left(\sqrt{\frac{2}{p}} - \frac{p+3q}{(p+q)^{\frac{3}{2}}}\right) > 0,$$

$$\lim_{r\to q^-} R_{pq}{}'(r) = \frac{1}{2}\left(\frac{\frac{(q-p)(q+3p)}{(q+p)^2}}{\sqrt{L(p,q)}} + \lim_{r\to q^-}\frac{\frac{(r-q)(r+3q)}{(r+q)^2}}{\sqrt{L(r,q)}}\right) = \frac{1}{2}\left(\frac{q+3p}{(p+q)^{\frac{3}{2}}} - \sqrt{\frac{2}{q}}\right) < 0.$$

Let

$$A(y,r) = \frac{\frac{(r-y)(r+3y)}{(r+y)^2}}{\sqrt{L(y,r)}} = \frac{\frac{(r-y)(r+3y)}{(r+y)^2}}{\sqrt{r}\sqrt{L(\frac{y}{r},1)}} = \frac{1}{\sqrt{r}}B(y,r), \quad y > 0,$$

then

$$\frac{\partial B(y,r)}{\partial r} = -\frac{3y\sqrt{\frac{(r-y)^2}{r+y}}}{2\sqrt{r}(r+y)^2} \le 0.$$

The equality holds if and only if $r = y$. So with respect to variable r in the open interval (p,q), $B(p,r)$ and $B(q,r)$ are both monotonic decreasing, $B(p,r) + B(q,r)$ is also monotonic decreasing. Using (4),

$$R_{pq}{}'(r) = \frac{1}{2}\left(A(p,r) + A(q,r)\right) = \frac{1}{2\sqrt{r}}\left(B(p,r) + B(q,r)\right),$$

this shows $\lim\limits_{r\to p^+} B(p,r) + B(q,r) > 0$, $\lim\limits_{r\to q^-} B(p,r) + B(q,r) < 0$. So we can see $B(p,r) + B(q,r)$ has only one zero point in the open interval (p,q) with respect to variable r. As a consequence, $R_{pq}{}'(r)$ has only one zero point x_0 in the open interval (p,q) with respect to variable r. This means $R_{pq}{}'(r) > 0$ in the interval (p,x_0), $R_{pq}{}'(r) < 0$ in the interval (x_0,q). From the above we know $R_{pq}{}'(r)$ has only one maximum and no minimum in the open interval (p,q).

As a result, the conclusion in the lemma is obtained. □

Theorem 1. *Let* $p,q,r \in [0,+\infty)$*, then*

$$(L(p,q))^{\frac{1}{2}} \le (L(p,r))^{\frac{1}{2}} + (L(q,r))^{\frac{1}{2}}. \tag{7}$$

Proof. If $p = q$, then $L(p,q) = 0$. The triangle inequality (7) obviously holds.

If $p \ne q$ and one of p,q is equal to 0, it is easy to obtain that (7) holds.

Next we assume $0 < p < q$ without loss of generality. Note that the formula is valid:

$$(L(p,q))^{\frac{1}{2}} = \lim_{r\to p}\left((L(p,r))^{\frac{1}{2}} + (L(q,r))^{\frac{1}{2}}\right) = \lim_{r\to q}\left((L(p,r))^{\frac{1}{2}} + (L(q,r))^{\frac{1}{2}}\right).$$

From Lemma 3 the triangle inequality (7) can be easily proved for any number $r \in [0,+\infty)$. □

Corollary 1. *Let* $p,q,r \in [0,+\infty)$*. If* $0 < \alpha < \frac{1}{2}$*, then*

$$(L(p,q))^{\alpha} \le (L(p,r))^{\alpha} + (L(q,r))^{\alpha}. \tag{8}$$

Proof. Let $a,b > 0$ and $0 < \gamma < 1$, then $a^\gamma + b^\gamma > (a+b)^\gamma$ which follows from the concavity of x^γ. Now a γ which satisfies $\alpha = \frac{1}{2}\gamma$ can be found. Thus from Theorem 1,

$$\begin{aligned}(L(p,r))^{\alpha} + (L(q,r))^{\alpha} &= (L(p,r))^{\frac{1}{2}\gamma} + (L(q,r))^{\frac{1}{2}\gamma} \\ &\ge \left((L(p,r))^{\frac{1}{2}} + (L(q,r))^{\frac{1}{2}}\right)^{\gamma} \ge (L(p,q))^{\frac{1}{2}\gamma} = (L(p,q))^{\alpha}.\end{aligned}$$

This is the triangle inequality (8) for the function $(L(p,q))^{\alpha}$. □

Theorem 2. *Let $p, q, r \in [0, +\infty)$. If $\alpha > \frac{1}{2}$, then the triangle inequality (8) does not hold.*

Proof. Assuming $0 < p < q$, let $l(r) = (L(p, r))^\alpha + (L(q, r))^\alpha$. Firstly the formula is valid:

$$(L(p, q))^\alpha = \lim_{r \to p} \left((L(p, r))^\alpha + (L(q, r))^\alpha\right) = \lim_{r \to q} \left((L(p, r))^\alpha + (L(q, r))^\alpha\right).$$

The derivative of the function l is

$$l'(r) = \alpha \left(\frac{(r-p)(3p+r)}{(p+r)^2} (L(p, r))^{\alpha-1} + \frac{(r-q)(3q+r)}{(q+r)^2} (L(q, r))^{\alpha-1} \right).$$

When $r \in (p, q)$, let

$$m(r) = \left(\frac{(r-p)(3p+r)}{(p+r)^2} (L(p, r))^{\alpha-1} \right)^{\frac{1}{1-\alpha}}.$$

Using l'Hôspital's rule,

$$\lim_{r \to p^+} m(r) = \frac{8p^2}{(1-\alpha)(p+r)^3} \left(\frac{(r-p)(3p+r)}{(p+r)^2} \right)^{\frac{2\alpha-1}{1-\alpha}} = 0.$$

So

$$\lim_{r \to p^+} l'(r) = \frac{(p-q)(3q+p)}{(q+p)^2} (L(p, q))^{\alpha-1} < 0.$$

According to the definition of derivative, there exists a $\delta > 0$ such that for any $s \in (p, p+\delta)$,

$$(L(p, q))^\alpha = \lim_{r \to p^+} \left((L(p, r))^\alpha + (L(q, r))^\alpha\right) > (L(p, s))^\alpha + (L(q, s))^\alpha.$$

This shows the triangle inequality (8) does not hold. □

To sum up the theorems and corollary above, we can obtain the main theorem:

Theorem 3. *The function $(L(p, q))^\alpha$ satisfies the triangle inequality (8) if and only if $0 < \alpha \le \frac{1}{2}$.*

3. Metric Properties of $\Delta_\alpha(P, Q)$

In this section, we mainly prove the following theorem:

Theorem 4. *The function $\Delta_\alpha(P, Q)$ is a metric on the space Γ_n if and only if $0 < \alpha \le \frac{1}{2}$.*

Proof. From (2) we can get $\Delta_\alpha(P, Q) = \left(\sum_{i=1}^{n} L(p_i, q_i) \right)^\alpha$. It is easy to see that $\Delta_\alpha(P, Q) \ge 0$ with equality only for $P = Q$ and $\Delta_\alpha(P, Q) = \Delta_\alpha(Q, P)$. So what we concern is whether the triangle inequality

$$\Delta_\alpha(P, Q) \le \Delta_\alpha(P, R) + \Delta_\alpha(Q, R) \tag{9}$$

holds for any $P, Q, R \in \Gamma_n$.

When $P = Q$, $\Delta_\alpha(P, Q) = 0$, the triangle inequality (9) holds apparently. So we assume $P \ne Q$ in the following.

Next we consider the value of α in two cases respectively:

(i) $0 < \alpha \le \frac{1}{2}$:

From Theorem 3, the inequality $(L(p_i, q_i))^\alpha \leq (L(p_i, r_i))^\alpha + (L(q_i, r_i))^\alpha$ holds. Applying Minkowski's inequality we have

$$\begin{aligned}
\left(\sum_{i=1}^{n} L\left(p_i, q_i\right)\right)^\alpha &= \left\{\sum_{i=1}^{n} \left((L(p_i, q_i))^\alpha\right)^{\frac{1}{\alpha}}\right\}^\alpha \\
&\leq \left\{\sum_{i=1}^{n} \left((L(p_i, r_i))^\alpha + (L(q_i, r_i))^\alpha\right)^{\frac{1}{\alpha}}\right\}^\alpha \\
&\leq \left\{\sum_{i=1}^{n} \left((L(p_i, r_i))^\alpha\right)^{\frac{1}{\alpha}}\right\}^\alpha + \left\{\sum_{i=1}^{n} \left((L(q_i, r_i))^\alpha\right)^{\frac{1}{\alpha}}\right\}^\alpha \\
&= \left(\sum_{i=1}^{n} L\left(p_i, r_i\right)\right)^\alpha + \left(\sum_{i=1}^{n} L\left(q_i, r_i\right)\right)^\alpha .
\end{aligned}$$

So the triangle inequality (9) holds.

(ii) $\alpha > \frac{1}{2}$:

Let

$$F(x_1, \cdots, x_n) = F_1(x_1, \cdots, x_n) + F_2(x_1, \cdots, x_n),$$

where

$$\begin{aligned}
F_1(x_1, \cdots, x_n) &= \left(\sum_{i=1}^{n} \frac{(p_i - x_i)^2}{p_i + x_i}\right)^\alpha, \\
F_2(x_1, \cdots, x_n) &= \left(\sum_{i=1}^{n} \frac{(q_i - x_i)^2}{q_i + x_i}\right)^\alpha.
\end{aligned}$$

Then $F(p_1, \cdots, p_n) = F(q_1, \cdots, q_n) = \Delta_\alpha(P, Q)$.

Next we prove $(p_1, \cdots, p_n)$ and $(q_1, \cdots, q_n)$ are not the extreme points of the function $F(x_1, \cdots, x_n)$. By the symmetry we only need to prove $(p_1, \cdots, p_n)$ is not the extreme point.

By partial derivative,

$$\left.\frac{\partial F}{\partial x_i}\right|_{(x_1,\cdots,x_n)=(p_1,\cdots,p_n)} = \left.\frac{\partial F_1}{\partial x_i}\right|_{(x_1,\cdots,x_n)=(p_1,\cdots,p_n)} + \left.\frac{\partial F_2}{\partial x_i}\right|_{(x_1,\cdots,x_n)=(p_1,\cdots,p_n)}. \tag{10}$$

Since $P \neq Q$, we might as well assume $p_1 \neq q_1$ and $p_1 > 0$.

$$\begin{aligned}
\left.\frac{\partial F_2}{\partial x_1}\right|_{(x_1,\cdots,x_n)=(p_1,\cdots,p_n)} &= \frac{\alpha(p_1 - q_1)(p_1 + 3q_1)}{(p_1 + q_1)^2} \cdot \left(\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i}\right)^{\alpha - 1} \\
&\neq 0.
\end{aligned} \tag{11}$$

$$\begin{aligned}
\left.\frac{\partial F_1}{\partial x_1}\right|_{(x_1,\cdots,x_n)=(p_1,\cdots,p_n)} &= \lim_{\Delta x_1 \to 0} \frac{1}{\Delta x_1} \left(F_1(p_1 + \Delta x_1, \cdots, p_n) - F_1(p_1, \cdots, p_n)\right) \\
&= \lim_{\Delta x_1 \to 0} \frac{1}{\Delta x_1} \left(\frac{\Delta x_1^2}{2p_1 + x_1}\right)^\alpha \\
&= \lim_{\Delta x_1 \to 0} \frac{\Delta x_1^{2\alpha - 1}}{(2p_1 + x_1)^\alpha} \\
&= 0.
\end{aligned} \tag{12}$$

Then taking (11) and (12) into (10), we have

$$\left.\frac{\partial F}{\partial x_1}\right|_{(x_1,\cdots,x_n)=(p_1,\cdots,p_n)} \neq 0.$$

Therefore, $(p_1, \cdots, p_n)$ is not the extreme point of the function $F(x_1, \cdots, x_n)$. For the same reason, $(q_1, \cdots, q_n)$ is also not the extreme point.

Using the definition of extreme point, there exists a point $R = (r_1, \cdots, r_n)$ such that $F(r_1, \cdots, r_n) < F(p_1, \cdots, p_n) = \Delta_\alpha(P, Q)$. As $F_1(r_1, \cdots, r_n) = \Delta_\alpha(P, R)$, $F_2(r_1, \cdots, r_n) = \Delta_\alpha(Q, R)$, then $\Delta_\alpha(P, R) + \Delta_\alpha(Q, R) < \Delta_\alpha(P, Q)$. The inequality is not consistent with the triangle inequality (9).

From what has been discussed above, the conclusion in the theorem is obtained. □

The generalization of this result to continuous probability distributions is straightforward. Consider a measurable space $(\mathcal{X}, \mathcal{A})$, and P, Q are probability distributions with Radon-Nykodym densities $p = \frac{\mathrm{d}P}{\mathrm{d}\mu}$, $q = \frac{\mathrm{d}Q}{\mathrm{d}\mu}$ w.r.t. a dominating σ-finite measure μ. Then

$$\Delta_\alpha(P, Q) = \left(\int_{\mathcal{X}} \frac{(p-q)^2}{p+q}\mathrm{d}\mu\right)^\alpha \tag{13}$$

is a metric if and only if $0 < \alpha \leq \frac{1}{2}$.

Next we will discuss the maxima and minima of $\Delta_\alpha(P, Q)$. It is obvious that $\Delta_\alpha(P, Q) = 0$ is the minima, if and only if $P = Q$. Because $\Delta(P, Q)$ can rewrite in the form

$$\begin{aligned}\Delta(P, Q) &= \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i} \\ &= \sum_{i=1}^{n}\left(p_i + q_i - \frac{4p_iq_i}{p_i + q_i}\right) \\ &= 2 - \sum_{i=1}^{n} \frac{4p_iq_i}{p_i + q_i} \leq 2.\end{aligned} \tag{14}$$

$\Delta(P, Q)$ obtains the maxima 2 when P, Q are two distinct deterministic distributions, namely $p_iq_i = 0$. Then the metric $\Delta_\alpha(P, Q)$ achieves its maximum value 2^α.

4. Some Inequalities among the Information-Theoretic Divergences

Definition 3. *For all $P, Q \in \Gamma_n$, the Jensen–Shannon divergence is defined by*

$$JS(P, Q) = \frac{1}{2}\sum_{i=1}^{n}\left[p_i \ln\left(\frac{2p_i}{p_i + q_i}\right) + q_i \ln\left(\frac{2q_i}{p_i + q_i}\right)\right].$$

The square of the Hellinger distance is defined by

$$H^2(P, Q) = \frac{1}{2}\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2.$$

The variance distance is defined by

$$V(P, Q) = \sum_{i=1}^{n} |p_i - q_i|.$$

Next we introduce the Csiszár's f-divergence[14].

Definition 4. *Let $f : [0, +\infty) \to (-\infty, +\infty)$ be a convex function satisfying $f(1) = 0$, the f-divergence measure introduced by Csiszár is defined as*

$$C_f(P, Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \tag{15}$$

for all $P, Q \in \Gamma_n$.

The triangular discrimination, Jensen–Shannon divergence, the square of the Hellinger distance, variance distance are all f-divergence.

Example 1. *(Triangular Discrimination) Let us consider*

$$f_\Delta(x) = \frac{(x-1)^2}{x+1}, \quad x \in [0, +\infty)$$

in (15). Then we can verify $f_\Delta(x)$ is convex because $f''_\Delta(x) = \frac{8}{(x+1)^3} \geq 0$, $f_\Delta(1) = 0$, $f_\Delta(x) \geq 0$ and $C_{f_\Delta}(P, Q) = \Delta(P, Q)$.

Example 2. *(Jensen–Shannon divergence) Let us consider*

$$f_{JS}(x) = \frac{x}{2} \ln \frac{2x}{x+1} + \frac{1}{2} \ln \frac{2}{x+1}, \quad x \in [0, +\infty)$$

in (15). Then we can verify $f_{JS}(x)$ is convex because $f''_{JS}(x) = \frac{1}{2x^2+2x} \geq 0$, $f_{JS}(1) = 0$ and $C_{f_{JS}}(P, Q) = JS(P, Q)$. By standard inequality $\ln x \geq 1 - \frac{1}{x}$, $f_{JS}(x) \geq \frac{x}{2}(1 - \frac{x+1}{2x}) + \frac{1}{2}(1 - \frac{x+1}{2}) = 0$ holds.

Example 3. *(Square of Hellinger distance) Let us consider*

$$f_h(x) = \frac{1}{2}(\sqrt{x} - 1)^2, \quad x \in [0, +\infty)$$

in (15). Then we can verify $f_h(x)$ is convex because $f''_h(x) = \frac{1}{4x\sqrt{x}} \geq 0$, $f_h(1) = 0$, $f_h(x) \geq 0$ and $C_{f_h}(P, Q) = H^2(P, Q)$.

Example 4. *(Variation distance) Let us consider*

$$f_V(x) = |x - 1|, \quad x \in [0, +\infty)$$

in (15). Then we can easily get $f_V(x)$ is convex, $f_V(1) = 0$, $f_V(x) \geq 0$ and $C_{f_V}(P, Q) = V(P, Q)$.

Theorem 5. *Let f_1, f_2 be two nonnegative generating functions and there exists the real constants k, K such that $k < K$ and if $f_2(x) \neq 0$ then*

$$k \leq \frac{f_1(x)}{f_2(x)} \leq K,$$

if $f_2(x) = 0$, then $f_1(x) = 0$. We have the inequalities:

$$kC_{f_2}(P, Q) \leq C_{f_1}(P, Q) \leq KC_{f_2}(P, Q).$$

Proof. The conditions can be rewritten as $kf_2(x) \le f_1(x) \le Kf_2(x)$. So from the formula (15),

$$\begin{aligned} C_{f_1}(P,Q) &= \sum_{i=1}^{n} q_i f_1(\frac{p_i}{q_i}) \geq \sum_{i=1}^{n} q_i \left(k f_2 \left(\frac{p_i}{q_i} \right) \right) \\ &= k \sum_{i=1}^{n} q_i f_2 \left(\frac{p_i}{q_i} \right) = k C_{f_2}(P,Q). \end{aligned}$$

and

$$\begin{aligned} C_{f_1}(P,Q) &= \sum_{i=1}^{n} q_i f_1(\frac{p_i}{q_i}) \leq \sum_{i=1}^{n} q_i \left(K f_2 \left(\frac{p_i}{q_i} \right) \right) \\ &= K \sum_{i=1}^{n} q_i f_2 \left(\frac{p_i}{q_i} \right) = K C_{f_2}(P,Q). \end{aligned}$$

□

We have shown that f_Δ, f_{JS}, f_h, f_V are all nonnegative. In the following we will have some inequalities.

Theorem 6.

$$\frac{1}{4}\Delta(P,Q) \leq JS(P,Q) \leq \frac{\ln 2}{2}\Delta(P,Q).$$

Proof. When $x \neq 1$, both $f_\Delta(1)$ and $f_{JS}(1)$ are not equal to 0. We consider the function:

$$\phi(x) = \frac{f_{JS}(x)}{f_\Delta(x)} = \frac{\frac{x}{2}\ln\frac{2x}{x+1} + \frac{1}{2}\ln\frac{2}{x+1}}{\frac{(x-1)^2}{x+1}}.$$

The derivative of the function $\phi(x)$ is

$$\phi'(x) = \frac{(1+3x)\ln x + 4(1+x)\ln\frac{2}{x+1}}{2(1-x)^3}.$$

Let

$$\psi(x) = (1+3x)\ln x + 4(1+x)\ln\frac{2}{x+1}. \tag{16}$$

Straightforward derivative shows

$$\begin{aligned} \psi'(x) &= 3\ln x + 4\ln\frac{2}{1+x} + \frac{1}{x} - 1, \\ \psi''(x) &= -\frac{(x-1)^2}{x^2(x+1)} < 0. \end{aligned}$$

So $\psi(x)$ is concave function when $x \in [0,+\infty)$ and $\psi'(1) = \psi(1) = 0$. This means $\psi(x)$ gets the maximum 0 at the point $x = 1$. Accordingly $\psi(x) < 0$ when $x \neq 1$. From (16), we find

$$\begin{cases} \phi'(x) < 0, & 0 < x < 1 \\ \phi'(x) > 0, & x > 1 \end{cases}$$

and

$$\lim_{x\to 0^+}\phi(x)=\frac{\frac{1}{2}\ln 2}{1}=\frac{\ln 2}{2}.$$

Using l'Hôspital's rule (differentiate twice),

$$\lim_{x\to 1}\phi(x)=\lim_{x\to 1}\frac{\frac{1}{2}\left(\frac{1}{x}-\frac{1}{x+1}\right)}{\frac{8}{(x+1)^3}}=\frac{1}{4}.$$

Using l'Hôspital's rule (differentiate once),

$$\lim_{x\to +\infty}\phi(x)=\frac{\frac{1}{2}\ln\frac{2x}{x+1}}{\frac{(x-1)(x+3)}{(x+1)^2}}=\frac{\ln 2}{2}.$$

Thus

$$\frac{1}{4}\leq\phi(x)=\frac{f_{JS}(x)}{f_\Delta(x)}\leq\frac{\ln 2}{2}.$$

When $x=1$, $f_\Delta(1)=f_{JS}(1)=0$. As a consequence of Theorem 5, we obtain the result

$$\frac{1}{4}C_{f_\Delta}(P,Q)\leq C_{f_{JS}}(P,Q)\leq\frac{\ln 2}{2}C_{f_\Delta}(P,Q).$$

Thus the theorem is proved. □

Theorem 7.

$$JS(P,Q)\leq H^2(P,Q)\leq\frac{1}{\ln 2}JS(P,Q).$$

Proof. When $x\neq 1$, both $f_h(1)$ and $f_{JS}(1)$ are not equal to 0. We consider the function:

$$\xi(x)=\frac{f_{JS}(x)}{f_h(x)}=\frac{\frac{x}{2}\ln\frac{2x}{x+1}+\frac{1}{2}\ln\frac{2}{x+1}}{\frac{1}{2}(\sqrt{x}-1)^2}.$$

The derivative of the function $\phi(x)$ is

$$\xi'(x)=\frac{\ln\frac{2}{x+1}+\sqrt{x}\ln\frac{2x}{x+1}}{\sqrt{x}(1-\sqrt{x})^3}.$$

By standard inequality $\ln x\geq 1-\dfrac{1}{x}$,

$$\begin{aligned}\ln\frac{2}{x+1}+\sqrt{x}\ln\frac{2x}{x+1}&\geq 1-\frac{x+1}{2}+\sqrt{x}\left(1-\frac{x+1}{2x}\right)\\&=\frac{(\sqrt{x}-1)^2(\sqrt{x}+1)}{2\sqrt{x}}>0\end{aligned}$$

So

$$\begin{cases}\xi'(x)>0, & 0<x<1\\ \xi'(x)<0, & x>1\end{cases}$$

and

$$\lim_{x\to 0^+}\xi(x)=\frac{\frac{1}{2}\ln 2}{\frac{1}{2}}=\ln 2.$$

Using l'Hôspital's rule (differentiate twice),

$$\lim_{x\to 1}\xi(x)=\lim_{x\to 1}\frac{\frac{1}{2}\left(\frac{1}{x}-\frac{1}{x+1}\right)}{\frac{1}{4\sqrt{x^3}}}=1.$$

Using l'Hôspital's rule (differentiate once),

$$\lim_{x\to +\infty}\xi(x)=\frac{\frac{1}{2}\ln\frac{2x}{x+1}}{\frac{\sqrt{x}-1}{2\sqrt{x}}}=\ln 2.$$

Thus

$$\ln 2\leq \phi(x)=\frac{f_{JS}(x)}{f_h(x)}\leq 1,$$

or

$$1\leq \frac{1}{\phi(x)}=\frac{f_h(x)}{f_{JS}(x)}\leq \frac{1}{\ln 2}.$$

When $x=1$, $f_h(1)=f_{JS}(1)=0$. As a consequence of Theorem 5, we obtain the result

$$C_{f_{JS}}(P,Q)\leq C_{f_h}(P,Q)\leq \frac{1}{\ln 2}C_{f_{JS}}(P,Q).$$

Thus the theorem is proved. □

Theorem 8.

$$\frac{1}{2}V^2(P,Q)\leq \Delta(P,Q)\leq V(P,Q).$$

Proof. When $x\neq 1$, both $f_\Delta(1)$ and $f_V(1)$ are not equal to 0. We consider the function:

$$\frac{f_\Delta(x)}{f_V(x)}=\frac{\frac{(x-1)^2}{x+1}}{|x-1|}=\frac{|x-1|}{x+1}\leq 1.$$

When $x=1$, $f_\Delta(1)=f_V(1)=0$. As a consequence of Theorem 5, we obtain the result $C_{f_\Delta}(P,Q)\leq C_{f_V}(P,Q)$. This means $\Delta(P,Q)\leq V(P,Q)$. Next,

$$\begin{aligned}\frac{1}{2}V^2(P,Q)&=\frac{1}{2}\left(\sum_{i=1}^{n}|p_i-q_i|\right)^2\\&\leq \frac{1}{2}\left(\sum_{i=1}^{n}(p_i+q_i)\right)\left(\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}\right)\text{(Cauchy–Schwarz inequality)}\\&=\frac{1}{2}\cdot 2\cdot\left(\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}\right)\\&=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}=\Delta(P,Q)\end{aligned}$$

Thus the theorem is proved. □

From the above theorems, inequalities among these measures are given by

$$\begin{aligned}&\frac{1}{8}V^2(P,Q)\leq \frac{1}{4}\Delta(P,Q)\leq JS(P,Q)\leq H^2(P,Q)\leq \frac{1}{\ln 2}JS(P,Q)\\&\leq \frac{1}{2}\Delta(P,Q)\leq \frac{1}{2}V(P,Q)\end{aligned}\tag{17}$$

These inequalities are sharper than the inequalities in [13] Theorem 2 and [15] (Section 3.1).

5. Asymptotic Approximation

Definition 5. *For all $P, Q \in \Gamma_n$, the Chi-square divergence is defined by*

$$\chi^2(P,Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}.$$

In [12],

$$JS(P,Q) = \frac{1}{2} D_{PQ}^2 \approx \frac{1}{2} \sum_{i=1}^{n} \frac{1}{4q_i}(p_i - q_i)^2 = \frac{1}{8}\chi^2(P,Q).$$

In this section, we will discuss the asymptotic approximation of $\Delta(P,Q)$ and $H^2(P,Q)$ when $P \to Q$ in L^2 norm.

Theorem 9. *If $\|P - Q\|_2 \to 0$, then*

$$\Delta(P,Q) \to \frac{1}{2}\chi^2(P,Q), \quad H^2(P,Q) \to \frac{1}{8}\chi^2(P,Q).$$

Proof. From Taylor's series expansion at q, we have

$$\begin{aligned} \frac{(x-q)^2}{x+q} &= \frac{(x-q)^2}{2q} + o\left((x-q)^2\right) \\ \frac{1}{2}(\sqrt{x} - \sqrt{q})^2 &= \frac{(x-q)^2}{8q} + o\left((x-q)^2\right) \end{aligned}$$

Hence

$$\begin{aligned} \Delta(P,Q) &= \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i} = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{2q_i} + o\left(\|P - Q\|_2^2\right) \\ &= \frac{1}{2}\chi^2(P,Q) + o\left(\|P - Q\|_2^2\right) \end{aligned}$$

$$\begin{aligned} H^2(P,Q) &= \sum_{i=1}^{n} \frac{1}{2}(\sqrt{p_i} - \sqrt{q_i})^2 = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{8q_i} + o\left(\|P - Q\|_2^2\right) \\ &= \frac{1}{8}\chi^2(P,Q) + o\left(\|P - Q\|_2^2\right) \end{aligned}$$

□

Equivalently, $JS(P,Q) \approx H^2(P,Q) \approx \frac{1}{4}\Delta(P,Q) \approx \frac{1}{8}\chi^2(P,Q)$ when $P \to Q$. So in some cases, one of the information-theoretic divergences can be substituted for another. The asymptotic property can also interpret the boundedness of triangular discrimination and, on the other hand, the new metrics.

Acknowledgments

The authors would like to thank the editor and referees for their helpful suggestions and comments on the manuscript. This manuscript is supported by China Postdoctoral Science Foundation (2015M571255), the National Science Foundation of China (the NSF of China) Grant No. 71171119, the Fundamental Research Funds for the Central Universities (FRF-CU) Grant No. 2722013JC082, and the Fundamental Research Funds for the Central Universities under grant number NKZXTD1403.

Author Contributions

Wrote the paper: Guoxiang Lu and Bingqing Li. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Basseville, M. Divergence measures for statistical data processing—An annotated bibliography. *Signal Process.* **2013**, *93*, 621–633.
2. Csiszár, I.; Shields, P.C. Information theory and statistics: A tutorial. *Found. Trends Commun. Inf. Theory* **2004**, *1*, 417–528.
3. Dragomir, S.S.; Glušćević, V. Some inequalities for the Kullback–Leibler and χ^2-distances in information theory and applications. *Tamsui Oxf. J. Math. Sci.* **2001**, *17*, 97–111.
4. Reid, M.D.; Williamson, R.C. Information, divergence and risk for binary experiments. *J. Mach. Learn. Res.* **2011**, *12*, 731–817.
5. Liese, F.; Vajda, I. On divergences and informations in statistics and information theory. *IEEE Trans. Inf. Theory* **2006**, *52*, 4394–4412.
6. Vajda, I. *Theory of Statistical Inference and Information*; Kluwer Academic Press: London, UK, 1989.
7. Csiszár, I. Axiomatic characterizations of information measures. *Entropy* **2008**, *10*, 261–273.
8. Cichocki, A.; Cruces, S.; Amari, S. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy* **2011**, *13*, 134–170.
9. Taneja, I.J. Seven means, generalized triangular discrimination, and generating divergence measures. *Information* **2013**, *4*, 198–239.
10. Arndt, C. *Information Measures: Information and its Description in Science and Engineering*; Springer Verlag: Berlin, Germany, 2004.
11. Brown, R.F. *A Topological Introduction to Nonlinear Analysis*; Birkhäuser: Basel, Switzerland, 1993.
12. Endres, D.M.; Schindelin, J.E. A new metric for probability distributions. *IEEE Trans. Inf. Theory* **2003**, *49*, 1858–1860.
13. Topsøe, F. Some inequalities for information divergence and related measures of discrimination. *IEEE Trans. Inf. Theory* **2000**, *46*, 1602–1609.
14. Csiszár, I. Information type measures of differences of probability distribution and indirect observations. *Studia Sci. Math. Hungar* **1967**, *2*, 299–318.
15. Taneja, I.J. Refinement inequalities among symmetric divergence measures. *Austr. J. Math. Anal. Appl.* **2005**, *2*, Available online: http://ajmaa.org/cgi-bin/paper.pl?string=v2n1/V2I1P8.tex (accessed on 14 July 2015).