1 Introduction

Consider any $n$-letter alphabet $Z = \{z_1, z_2, \ldots, z_n\}$ associated with two probability mass functions, $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_n\}$. Chen and Sbert proposed a general divergence measure [1] as follows:

$$D_{CS}(P \parallel Q) = \frac{1}{2} \sum_{i=1}^{n} (p_i + q_i) \log_2 \left( |p_i - q_i|^k + 1 \right)$$

where $k > 0$ is a parameter for moderating the impact of the pairwise-difference between $p_i$ and $q_i$ in relation to other pairwise differences, i.e., $|p_j - q_j|, \forall j \neq i$. Here we focus on the base-2 logarithm in the context of computer science and data science. The transformation to other logarithmic bases is not difficult.

The commonly-used Kullback-Leibler divergence [4] computes first the informative quantity of individual probabilistic values (in $P$ and $Q$) associated with each letter $z_i \in Z$. It then computes the pairwise difference between the informative quantities $\log_2 p_i$ and $\log_2 q_i$ for each letter, and finally computes the probabilistic average of such differences $(\log_2 p_i - \log_2 q_i)$, at the informative scale, across all letters $z_i \in Z$.

Unlike the Kullback-Leibler divergence, $\mathcal{D}_{CS}(P \parallel Q)$ computes first the pairwise difference between $p_i$ and $q_i$ for each letter $z_i \in Z$, then the informative quantity of the pairwise difference $|p_i - q_i|$ with a monotonic transformation $g(|p_i - q_i|) = \log_2 \left( |p_i - q_i|^k + 1 \right)$, and finally the probabilistic average of such informative quantities across all letters $z_i \in Z$.

In comparison with the Kullback-Leibler divergence, $\mathcal{D}_{CS}(P \parallel Q)$ is bounded by 0 and 1 (cf. KL is unbounded); does not suffer from any singularity condition (e.g., $q_i = 0$); and is commutative (cf. KL is not commutative). These properties of $\mathcal{D}_{CS}(P \parallel Q)$ are similar to that of the Jensen-Shannon divergence. In an early report [1], Chen and Sbert found that some empirical similarities between Jensen-Shannon divergence [5] and $\mathcal{D}_{CS}(P \parallel Q)$ (when $k = 2$).

Neither the Kullback-Leibler divergence nor the Jensen-Shannon divergence are a distance metric, but the square root of the Jensen–Shannon divergence is a metric.
This report is concerned with the question whether $\mathcal{D}_{CS}(P||Q)$ is known to be a distance metric. Clearly the measure satisfies most conditions of a distance metric, including:

- **identity of indiscernibles**: $\mathcal{D}_{CS}(P||Q) = 0 \iff P = Q$;
- **symmetry or commutativity**: $\mathcal{D}_{CS}(P||Q) = \mathcal{D}_{CS}(Q||P)$;
- **non-negativity or separation**: $\mathcal{D}_{CS}(P||Q) \geq 0$.

However, it is not clear whether it meets the **triangle inequality** condition (also referred to as the **subadditivity** condition). In other words, if the alphabet $\mathbb{Z}$ is associated with three arbitrary probability mass functions $P$, $Q$, and $R$, do we have the following:

- **triangle inequality**: $\mathcal{D}_{CS}(P||R) \leq \mathcal{D}_{CS}(P||Q) + \mathcal{D}_{CS}(Q||R)$?

## 2 Random Testing and Postulations

When setting $k = 1$ and $k = 0.5$ in Eq. 1 testing the triangle inequality with randomly generated $P$, $Q$, and $R$ indicates such a possibility.

When setting $k = 2$ in Eq. 1 testing the triangle inequality with randomly generated $P$, $Q$, and $R$ finds failures of the triangle inequality. For example, in the case of $n = 3$, $P = \{0.238, 0.013, 0.749\}$, $Q = \{0.253, 0.223, 0.524\}$, and $R = \{0.511, 0.418, 0.071\}$, we have:

\[
\mathcal{D}_{CS}(P||Q) = \mathcal{D}_{CS}(Q||P) = 0.052
\]
\[
\mathcal{D}_{CS}(Q||R) = \mathcal{D}_{CS}(R||Q) = 0.133
\]
\[
\mathcal{D}_{CS}(R||P) = \mathcal{D}_{CS}(P||R) = 0.310
\]
\[
\implies \mathcal{D}_{CS}(P||Q) + \mathcal{D}_{CS}(Q||R) - \mathcal{D}_{CS}(P||R) = -0.124 < 0
\]

Another example of failure is when $n = 4$, $P = \{0.143, 0.282, 0.326, 0.248\}$, $Q = \{0.260, 0.172, 0.300, 0.268\}$, and $R = \{0.040, 0.658, 0.215, 0.088\}$. The calculation shows:

\[
\mathcal{D}_{CS}(P||Q) = \mathcal{D}_{CS}(Q||P) = 0.008
\]
\[
\mathcal{D}_{CS}(Q||R) = \mathcal{D}_{CS}(R||Q) = 0.148
\]
\[
\mathcal{D}_{CS}(R||P) = \mathcal{D}_{CS}(P||R) = 0.102
\]
\[
\implies \mathcal{D}_{CS}(Q||P) + \mathcal{D}_{CS}(P||R) - \mathcal{D}_{CS}(Q||R) = -0.038 < 0
\]

Similar failures have been found when $k = 1.5$ and $k = 50$. One expects to find many cases of failures with other settings of $k > 1$.

Based on the random tests, we propose the following two postulations:

**Postulation 1** When $0 < k \leq 1$, $\mathcal{D}_{CS}(P||R) \leq \mathcal{D}_{CS}(P||Q) + \mathcal{D}_{CS}(Q||R)$.

**Postulation 2** When $k > 1$, $\sqrt[k]{\mathcal{D}_{CS}(P||R)} \leq \sqrt[k]{\mathcal{D}_{CS}(P||Q)} + \sqrt[k]{\mathcal{D}_{CS}(Q||R)}$. 
3 Special Case A: \( D_{\text{CS}} (k = 1) \) for 2-Letter Alphabets

Consider a 2-letter alphabet \( Z = \{z_1, z_2\} \) and three probability mass functions \( P = \{p, 1-p\} \), \( Q = \{q, 1-q\} \), and \( R = \{r, 1-r\} \). When \( k = 1 \), the Chen-Sbert measure is simplified as:

\[
D_{\text{CS-bi}}(P\|Q) = D_{\text{CS-bi}}(Q\|P) = \frac{1}{2} \left( (p+q)\log_2(|p-q|+1) + (2-p-q)\log_2(|p-q|+1) \right) \tag{2}
\]

Similarly, we have

\[
D_{\text{CS-bi}}(Q\|R) = D_{\text{CS-bi}}(R\|Q) = \log_2(|q-r|+1) \tag{3}
\]

\[
D_{\text{CS-bi}}(R\|P) = D_{\text{CS-bi}}(P\|R) = \log_2(|p-r|+1) \tag{4}
\]

**Lemma 1** For any real values \( 0 \leq p, q, r \leq 1 \), the following inequality is true:

\[
T(p, q, r) = \log_2(|p-q|+1) + \log_2(|q-r|+1) - \log_2(|p-r|+1) \geq 0 \tag{5}
\]

**Proof.** The left side of Eq.\( \text{(5)} \) can be rewritten as:

\[
T(p, q, r) = \log_2 \frac{|p-q|+1}{|q-r|+1} \tag{6}
\]

\[
= \begin{cases} 
\log_2 \frac{(a+1)(b+1)}{(a+b+1)} & \text{cases (1) and (6)} \\
\log_2 \frac{(c+d+1)(c+1)}{(c+1)} & \text{cases (2), (3), (4), (5)}
\end{cases} \tag{7}
\]

where the six cases are defined as:

1. When \( p \geq q \geq r \): we set \( a = p-q, b = q-r \);
2. When \( p \geq r \geq q \): we set \( c = p-r, d = r-q \);
3. When \( q \geq p \geq r \): we set \( c = p-r, d = q-p \);
4. When \( q \geq r \geq p \): we set \( c = r-p, d = q-r \);
5. When \( r \geq q \geq p \): we set \( c = r-p, d = p-q \);
6. When \( r \geq q \geq p \): we set \( a = q-p, b = r-q \).

Since \( 0 \leq a, b, c, d \leq 1 \), both parts of Eq.\( \text{(7)} \) are non-negative. For cases (1) and (6), we have \((a+1)(b+1) = (ab+a+b+1) \geq (a+b+1)\). For cases (2), (3), (4), (5), we have \((c+d+1)(d+1) \equiv (c+1)\). Both fractions inside the logarithmic function in Eq.\( \text{(7)} \) are thus \( \geq 1 \), and therefore \( T(p, q, r) \geq 0 \). According to Eq.\( \text{(2)} \) and Eq.\( \text{(5)} \) we have:

\[
D_{\text{CS-bi}}(P\|Q) + D_{\text{CS-bi}}(Q\|R) - D_{\text{CS-bi}}(P\|R) = \frac{1}{2} T(p, q, r) \geq 0
\]

\(D_{\text{CS-bi}}(P\|Q)\) therefore satisfies the triangle inequality condition. \(\blacksquare\)
Theorem 1 For any 2-letter alphabet, the Chen-Sbert divergence measure (when \( k = 1 \)) is a metric.

Proof. The proof can easily be derived from Lemma 2 together with the other properties of a distance metric discussed in Section 3.

4 Special Case B: \( \mathcal{D}_{CS} (k \leq 1) \) for 2-Letter Alphabets

We can extend the above special case to any \( 0 < k \leq 1 \) for any 2-letter alphabet.

Lemma 2 For any real values \( 0 \leq p, q, r \leq 1 \) and \( 0 < k \leq 1 \), the following inequality is true:

\[
T(p, q, r) = \log_2(|p - q|^k + 1) + \log_2(|q - r|^k + 1) - \log_2(|p - r|^k + 1) \geq 0
\]  

(8)

Proof. The left side of Eq. 8 can be rewritten as:

\[
T(p, q, r) = \log_2 \left( \frac{|p - q|^k + 1}{|p - r|^k + 1} \right) \left( \frac{|q - r|^k + 1}{|q - p|^k + 1} \right) = \log_2 \left( \frac{(a+b)^k + 1}{(a+b)^k + 1} \right)
\]  

(9)

where the six cases are the same as those in Section 3. Since \( 0 \leq c, d \leq 1 \), it is straightforward to conclude that \( ((c+d)^k + 1)(a+b)^k \geq (c^k + 1) \) for cases (2), (3), (4), and (5). Meanwhile, for cases (1) and (6), we consider two terms, \( X = a^k b^k + a^k + b^k \) and \( Y = (a+b)^k \). If we can show that \( X \geq Y \), we will be able to prove the following:

\[
\frac{X}{Y} \geq 1 \implies \frac{X + 1}{Y + 1} \geq \frac{a^k b^k + a^k + b^k + 1}{(a+b)^k + 1} = \frac{(a^k + 1)(b^k + 1)}{(a+b)^k + 1} \geq 1
\]  

(11)

We attempt a proof by contradiction. Supposing that \( X < Y \), we would have

\[
X < Y \implies a^k b^k + a^k + b^k < (a+b)^k
\]

\[
\implies (a^k b^k + a^k + b^k)^{1/k} < (a+b)
\]

(12)

where \( t = 1/k > 1 \) because \( k < 1 \). Since \( 0 \leq a, b \leq 1 \), we would have:

\[
(ab + a + b) \leq (\sqrt{ab} + \sqrt{a} + \sqrt{b})^t < (a+b)
\]

Based on the supposition of \( X < Y \) and Eq. 12, this would lead to the following conclusion:

\[
(ab + a + b) \leq (\sqrt{ab} + \sqrt{a} + \sqrt{b})^t < (a+b) \implies ab < 0
\]

This contradicts the fact that \( 0 \leq a, b \leq 1 \). Hence \( X < Y \) cannot be true. Because \( X \geq Y \) is true, the fraction in the second part of Eq. 10 is also \( \geq 1 \). For \( T(p, q, r) \) in Eq. 8 we can now conclude \( T(p, q, r) \geq 0 \), and \( \mathcal{D}_{CS-bi} (0 < k \leq 1) \) satisfies the triangle inequality condition. \( \blacksquare \)
Theorem 2 For any 2-letter alphabet, the Chen-Sbert divergence measure (with $0 < k \leq 1$) is a metric.

Proof. The proof can easily be derived from Lemma 2 together with the other properties of a distance metric discussed in Section 1.

5 Special Case C: $D_{CS}(k = 1)$ for n-Letter Alphabets

When $k = 1$, the might-be triangle inequality of the Chen-Sbert divergence can be expressed as:

$$D_{CS}(P||Q) + D_{CS}(Q||R) - D_{CS}(P||R) = \frac{1}{2} \left( \sum_{i=1}^{n} (p_i + q_i) \log_2(|p_i - q_i| + 1) + \sum_{i=1}^{n} (q_i + r_i) \log_2(|q_i - r_i| + 1) - \sum_{i=1}^{n} (p_i + r_i) \log_2(|p_i - r_i| + 1) \right)$$

$$= \frac{1}{2} \log_2 \frac{\prod_{i=1}^{n} (|p_i - q_i| + 1)^{p_i+q_i} (|q_i - r_i| + 1)^{q_i+r_i}}{\prod_{i=1}^{n} (|p_i - r_i| + 1)^{p_i+r_i}} \geq 0$$

Proving or falsifying this special case appears to be more difficult.

5.1 Unsuccessful Pathway: The Individual Term for Each Letter

This pathway attempts to find a simple proof by examining whether the individual term associated with each letter in the overall summation satisfies the triangle inequality. In other words, we ask a question: Is the following inequality always true?

$$(p + q) \log_2(|p - q| + 1) + (q + r) \log_2(|q - r| + 1) - (p + r) \log_2(|p - r| + 1)$$

$$= \log_2 \frac{(|p - q| + 1)^{(p+q)} (|q - r| + 1)^{(q+r)}}{(|p - r| + 1)^{(p+r)}} \geq 0 \quad (13)$$

Random tests show that this does not seem to be always true. For example, when $p = 0.05, q = 0.01, r = 0.85$, Eq. 13 yields $-0.0016$. In fact, random tests also find instances of negativity when $k < 1$.

However, this is only one term associated with an individual letter (among $n > 2$ letters). The negativity may be cancelled by the terms associated with other letters. For example, for a 2-letter alphabet, when the first letter is associated with $p = 0.05, q = 0.01, r = 0.85$, and the second letter must be associated with $p = 0.95, q = 0.99, r = 0.15$. Eq. 13 yields 0.0899 for the second letter. The sum of $-0.0016$ and 0.0899 yields 0.0883, which is still positive.

Hence, the possible negativity for an individual letter is not sufficient for concluding that $D_{CS}(k = 1)$ is not a metric for $n$-letter alphabets ($n > 2$). Nevertheless it at least indicates that if there is a proof, it may not be simple.
5.2 Unsuccessful Pathway: Combining the Terms of Two Letters

This pathway attempts to find a proof using induction by examining whether there is a positivity/negativity pattern when combining the terms associated with any two letters. In other words, let

$$T(p, q, r) = (p + q) \log_2(|p - q| + 1) + (q + r) \log_2(|q - r| + 1) - (p + r) \log_2(|p - r| + 1)$$

and we ask a question: Is the following inequality always true?

$$T(p_a + p_b, q_a + q_b, r_a + r_b) \leq T(p_a, q_a, r_a) + T(p_b, q_b, r_b)$$

where $0 \leq p_a, p_b, q_a, q_b, r_a, r_b \leq 1$ and $0 \leq (p_a + p_b), (q_a + q_b), (r_a + r_b) \leq 1$.

Random tests show that this is often not true. For example, when $p_a = 0.3907, p_b = 0.2422, q_a = 0.1134, q_b = 0.0358, r_a = 0.3525, r_b = 0.4558$, Eq. 14 yields 0.4043 on the left and 0.2055 on the right. Hence, this attempt is not successful.

5.3 Possible Pathway: Replacing the Terms of Three Letters

Using the same definition of $T(p, q, r)$ as in the previous subsection:

$$T(p, q, r) = (p + q) \log_2(|p - q| + 1) + (q + r) \log_2(|q - r| + 1) - (p + r) \log_2(|p - r| + 1)$$

$$= \log_2 \left( \frac{(p - q)^{p+q} (q - r)^{q+r}}{|p - r|^{p+r}} \right)$$

this pathway attempts to reduce any three terms $T(p_1, q_1, r_1), T(p_2, q_2, r_2), T(p_3, q_3, r_3)$ associated with three letters to two terms $T(p_s, q_s, r_s)$ and $T(p_y, q_y, r_y)$, such that

$$T(p_1, q_1, r_1) + T(p_2, q_2, r_2) + T(p_3, q_3, r_3) = T(p_s, q_s, r_s) + T(p_y, q_y, r_y)$$

where

$$0 \leq p_1, p_2, p_3, p_s, p_y \leq 1, \quad 0 \leq q_1 + p_2 + p_3 = p_s + p_y \leq 1$$

$$0 \leq q_1, q_2, q_3, q_s, q_y \leq 1, \quad 0 \leq q_1 + q_2 + q_3 = q_s + q_y \leq 1$$

$$0 \leq r_1, r_2, r_3, r_s, r_y \leq 1, \quad 0 \leq r_1 + r_2 + r_3 = r_s + r_y \leq 1$$

(16)

Experiments suggest that this may be achieved by using the following algorithmic steps:

1. Initiate three parameters $\alpha = 0, \beta = 0$, and $\gamma = 0$.

2. Initiate $p_x = p_1 + \frac{1}{2}p_3 + \alpha, \quad q_x = q_1 + \frac{1}{2}q_3 + \beta, \quad r_x = r_1 + \frac{1}{2}r_3 + \gamma$.

3. Initiate $p_y = p_2 + \frac{1}{2}p_3 - \alpha, \quad q_y = q_2 + \frac{1}{2}q_3 - \beta, \quad r_y = r_2 + \frac{1}{2}r_3 - \gamma$.

4. Use an optimisation algorithm to adjust $\alpha, \beta, \gamma$ and to obtain optimised $(p_x, p_y), (q_x, q_y)$, and $(r_x, r_y)$ such that the requirements in Eqs. 15 and 16 are met.
For example, given:

\( (p_1, p_2, p_3) = (0.5, 0.1, 0.2); (q_1, q_2, q_3) = (0.1, 0.2, 0.4); (r_1, r_2, r_3) = (0.3, 0.3, 0.1) \)

\[ \implies T(p_1, q_1, r_1) + T(p_2, q_2, r_2) + T(p_3, q_3, r_3) = 0.4967 \]

Following the above algorithmic steps, we can obtain:

\[ p_x = p_1 + \frac{1}{2}p_3 + \alpha = 0.6 + \alpha; \quad p_y = p_2 + \frac{1}{2}p_3 - \alpha = 0.2 - \alpha \]
\[ q_x = q_1 + \frac{1}{2}q_3 + \beta = 0.3 + \beta; \quad q_y = q_2 + \frac{1}{2}q_3 - \beta = 0.4 - \beta \]
\[ r_x = r_1 + \frac{1}{2}r_3 + \gamma = 0.35 + \gamma; \quad r_y = r_2 + \frac{1}{2}r_3 - \gamma = 0.35 - \gamma \]

\[ \implies T(p_x, q_x, r_x) + T(p_y, q_y, r_y) = 0.4967 \]

We have found the following solution to meet the requirements of Eqs.\[15\] and \[16\]

\[-0.1668334 < \alpha < -0.1668333, \quad \beta = -0.125, \quad \gamma = -0.04 \]

Similarly, we have found solutions for many other instances, e.g.,:

\( (p_1, p_2, p_3) = (0.1, 0.2, 0.2); (q_1, q_2, q_3) = (0.0, 0.0, 1.0); (r_1, r_2, r_3) = (0.2, 0.7, 0.1) \)

\[ \implies \alpha = -0.14, \quad 0.4161126 < \beta < 0.4161127, \quad \gamma = -0.14 \]

\( (p_1, p_2, p_3) = (0.1, 0.9, 0.0); (q_1, q_2, q_3) = (0.9, 0.1, 0.0); (r_1, r_2, r_3) = (0.2, 0.2, 0.5) \)

\[ \implies \alpha = -0.05, \quad \beta = 0.05, \quad 0.0442517 < \gamma < 0.0442518 \]

If an acceptable solution for \( \alpha, \beta, \) and \( \gamma \) can always be found, we can use induction to prove that \( \mathcal{D}_{CS} \) (when \( k = 1 \)) is a metric for any \( n \)-letter alphabet \( (n > 0) \).

However, we have not made adequate progress beyond that point.

6 General Case: \( \mathcal{D}_{CS} (0 < k \leq 1) \) for \( n \)-Letter Alphabets

The strategy for replacing the triangle inequality terms for three letters with equivalent terms of two letters can also be applied to the general case of \( \mathcal{D}_{CS} (0 < k \leq 1) \) for \( n \)-letter alphabets \( (n > 0) \). Our experiments on a number of instances when \( k = 0.2, 0.5, 0.8 \) have successfully found solutions. Nevertheless, a formal mathematical proof will be necessary.

If we can find a proof for special case C \( (k = 1) \) in Section\[5,6\] i.e., there is always a solution of \( \alpha, \beta, \) and \( \gamma \) that meets the requirements of Eqs.\[15\] and \[16\] it is likely that such a proof can be extended to the general case. In general, it is rather hopeful that Postulation\[1\] can proved.

7 Interim Conclusions

• For Postulation\[1\] in Section\[2\] we have managed to obtain a partial proof that \( \mathcal{D}_{CS}(P\parallel Q) (0 < k \leq 1) \) is a metric for 2-letter alphabets.
• Our attempts to obtain a full proof for the general case of $n$-letter alphabets have only found a possible but unconfirmed pathway.

• We have not yet made any solid progress in proving or falsifying Postulation 2 in Section 2.

We shall appreciate any effort to prove or falsify the two postulations by colleagues in the international scientific communities.

References

[1] M. Chen and M. Sbert. On the upper bound of the kullback-leibler divergence and cross entropy. arXiv:1911.08334, 2019.

[2] D. M. Endres and J. E. Schindelin. A new metric for probability distributions. IEEE Transactions on Information Theory, 49(7):1858–1860, 2003.

[3] B. Fuglede and F. Topsoe. Jensen-shannon divergence and hilbert space embedding. In Proc. International Symposium on Information Theory, pages 31–36, 2004.

[4] S. Kullback and R. A. Leibler. On information and sufficiency. Annals of Mathematical Statistics, 22(1):79–86, 1951.

[5] J. Lin. Divergence measures based on the shannon entropy. IEEE Transactions on Information Theory, 37:145–151, 1991.

[6] F. Österreicher and I. Vajda. A new class of metric divergences on probability spaces and its statistical applications. Annals of the Institute of Statistical Mathematics, 55(3):639–653, 2003.