On the symmetrized $S$-divergence

Slavko Simić\textsuperscript{1,*}

\textsuperscript{1}Mathematical Institute SANU, Kneza Mihaila 36, 11000 Belgrade, Serbia

\textbf{Abstract.} In this paper we worked with the relative divergence of type $s, s \in \mathbb{R}$, which include Kullback-Leibler divergence and the Hellinger and $\chi^2$ distances as particular cases. We give here a study of the symmetrized divergences in additive and multiplicative forms. Some basic properties as symmetry, monotonicity and log-convexity are established. An important result from the Convexity Theory is also proved.

\section{Introduction}

Let

$$\Omega^+ = \{p = \{p_i\} \mid p_i > 0, \sum p_i = 1\},$$

be the set of finite discrete probability distributions.

One of the most general probability measures which is of importance in Information Theory is the famous Csiszár’s $f$-divergence $C_f(p||q)$ ([1]), defined by

\textbf{Definition 1} For a convex function $f : (0, \infty) \to \mathbb{R}$, the $f$-divergence measure is given by

$$C_f(p||q) := \sum q_i f(p_i/q_i),$$

where $p, q \in \Omega^+$.

Some important information measures are just particular cases of the Csiszár’s $f$-divergence.

For example,

(a) taking $f(x) = x^\alpha, \alpha > 1$, we obtain the $\alpha$-order divergence defined by

$$I_\alpha(p||q) := \sum p_i^\alpha q_i^{1-\alpha};$$

\textbf{Remark} The above quantity is an argument in well-known theoretical divergence measures such as Renyi $\alpha$-order divergence $I_\alpha^R(p||q)$ or Tsallis divergence $I_\alpha^T(p||q)$, defined as

$$I_\alpha^R(p||q) := \frac{1}{\alpha - 1} \log I_\alpha(p||q); \quad I_\alpha^T(p||q) := \frac{1}{\alpha - 1} (I_\alpha(p||q) - 1).$$

(b) for $f(x) = x \log x$, we obtain the Kullback-Leibler divergence ([4]) defined by

$$K(p||q) := \sum p_i \log(p_i/q_i);$$

\*e-mail: ssimic@turing.mi.sanu.ac.rs
(c) for \( f(x) = (\sqrt{x} - 1)^2 \), we obtain the Hellinger distance
\[
H^2(p, q) := \sum (\sqrt{p_i} - \sqrt{q_i})^2;
\]

(d) if we choose \( f(x) = (x - 1)^2 \), then we get the \( \chi^2 \)-distance
\[
\chi^2(p, q) := \sum (p_i - q_i)^2 / q_i.
\]

The generalized measure \( K_s(p||q) \), known as the relative divergence of type \( s \) ([8]), or simply \( s \)-divergence, is defined by
\[
K_s(p||q) := \begin{cases} 
(\sum p_i^s q_i^{1-s} - 1) / s(s-1) & , s \in \mathbb{R} \setminus \{0, 1\}; \\
K(q||p) & , s = 0; \\
K(p||q) & , s = 1,
\end{cases}
\]

where \( \{p_i\}_1^n, \{q_i\}_1^n \) are given probability distributions and \( K(p||q) \) is Kullback-Leibler divergence.

It include the Hellinger and \( \chi^2 \) distances as particular cases.

Indeed,
\[
K_{1/2}(p||q) = 4(1 - \sum \sqrt{p_i q_i}) = 2 \sum (p_i + q_i - 2\sqrt{p_i q_i}) = 2H^2(p, q);
\]
\[
K_2(p||q) = \frac{1}{2} \left( \sum \frac{p_i^2}{q_i} - 1 \right) = \frac{1}{2} \sum \frac{(p_i - q_i)^2}{q_i} = \frac{1}{2} \chi^2(p, q).
\]

The \( s \)-divergence represents an extension of Tsallis divergence to the real line and accordingly is of importance in Information Theory. Main properties of this measure are given in [8].

**Theorem A** For fixed \( p, q \in \Omega^+, p \neq q \), the \( s \)-divergence is a positive, continuous and convex function in \( s \in \mathbb{R} \).

We shall use in this article a stronger property.

**Theorem B** For fixed \( p, q \in \Omega^+, p \neq q \), the \( s \)-divergence is a log-convex function in \( s \in \mathbb{R} \).

**Proof.** This is a corollary of an assertion proved in [6]. It says that for arbitrary positive sequence \( \{x_i\} \) and associated weight sequence \( q \in Q \) (see Appendix), the quantity \( \lambda_s \) defined by
\[
\lambda_s := \frac{\sum q_i x_i^s - (\sum q_i x_i)^s}{s(s-1)}
\]
is logarithmically convex in \( s \in \mathbb{R} \).

Putting there \( x_i = p_i / q_i \), we obtain that \( \lambda_s = K_s(p||q) \) is log-convex in \( s \in \mathbb{R} \).

Hence, for any real \( s, t \) we have that
\[
K_s(p||q)K_t(p||q) \geq K_{s+t}(p||q).
\]

Among all mentioned measures, only Hellinger distance has a symmetry property \( H^2 = H^2(p, q) = H^2(q, p) \). Our aim in this paper is to investigate some global properties of the symmetrized measures \( U_s := U_s(p, q) = U_s(q, p) := K_s(p||q) + K_s(q||p) \) and \( V_s := V_s(p, q) = V_s(q, p) := K_s(p||q)K_s(q||p) \). Since S. Kullback and R. Leibler themselves in their fundamental paper [4] (see also [3]) worked with the symmetrized variant \( J(p, q) := K(p||q) + K(q||p) = \sum (p_i - q_i) \log(p_i / q_i) \), our results can be regarded as a continuation of their ideas.
2 Results and Proofs

We shall give firstly some properties of the symmetrized divergence \( V_s = K_s(p||q)K_s(q||p) \).

**Proposition 2.1.** 1. For arbitrary, but fixed probability distributions \( p, q \in \Omega^+, p \neq q \), the divergence \( V_s \) is a positive and continuous function in \( s \in \mathbb{R} \).
2. \( V_s \) is a log-convex (hence convex) function in \( s \in \mathbb{R} \).
3. The graph of \( V_s \) is symmetric with respect to the line \( s = 1/2 \), bounded from below with the universal constant \( 4H^4 \) and unbounded from above.
4. \( V_s \) is monotone decreasing for \( s \in (-\infty, 1/2) \) and monotone increasing for \( s \in (1/2, +\infty) \).
5. The inequality

\[
V_{s-r}^{t-s} \leq V_{r-s}^{s-t}
\]

holds for any \( r < s < t \).

**Proof.** The Part 1. is a simple consequence of Theorem A above. The proof of Part 2. follows by using the result from Theorem B. Namely, for any \( s, t \in \mathbb{R} \) we have

\[
V_sV_t = [K_s(p||q)K_s(q||p)][K_t(p||q)K_t(q||p)] = [K_s(p||q)K_t(p||q)][K_s(q||p)K_t(q||p)]
\]

\[
\geq [K_{s+t}(p||q)]^2[K_{s+t}(q||p)]^2 = [V_{s+t}]^2.
\]

3. Note that

\[
K_s(p||q) = K_{1-s}(q||p); K_s(q||p) = K_{1-s}(p||q).
\]

Hence \( V_s = V_{1-s} \), that is \( V_{1/2-s} = V_{1/2+s} \), \( s \in \mathbb{R} \).

Also,

\[
V_s = K_s(p||q)K_s(q||p) = K_s(p||q)K_{1-s}(p||q) \geq K_{1/2}^2(p||q) = 4H^4.
\]

4. We shall prove only the "increasing" assertion. The other part follows from graph symmetry.

Therefore, for any \( 1/2 < x < y \) we have that

\[
1 - y < 1 - x < x < y.
\]

Applying Proposition X (see Appendix) with \( a = 1 - y, b = y, s = 1 - x, t = x; f(s) := \log K_s(p||q) \), we get

\[
\log K_x(p||q) + \log K_{1-x}(p||q) \leq \log K_y(p||q) + \log K_{1-y}(p||q),
\]

that is \( V_x \leq V_y \) for \( x < y \).

5. From the parts 1 and 2, it follows that \( \log V_s \) is a continuous and convex function on \( \mathbb{R} \). Therefore we can apply the following alternative form [2]:

**Lemma 2.2.** If \( \phi(s) \) is continuous and convex for all \( s \) of an open interval \( I \) for which \( s_1 < s_2 < s_3 \), then

\[
\phi(s_1)(s_3 - s_2) + \phi(s_2)(s_1 - s_3) + \phi(s_3)(s_2 - s_1) \geq 0.
\]
Hence, for \( r < s < t \) we get
\[
(t - r) \log V_s \leq (t - s) \log V_r + (s - r) \log V_t,
\]
which is equivalent to the assertion of Part 5.

Properties of the symmetrized measure \( U_s := K_s(p||q) + K_s(q||p) \) are very similar; therefore some analogous proofs will be omitted.

**Proposition 2.3.** 1. The divergence \( U_s \) is a positive and continuous function in \( s \in \mathbb{R} \).

2. \( U_s \) is a log-convex function in \( s \in \mathbb{R} \).

3. The graph of \( U_s \) is symmetric with respect to the line \( s = 1/2 \), bounded from below with \( 4H^2 \) and unbounded from above.

4. \( U_s \) is monotone decreasing for \( s \in (-\infty, 1/2) \) and monotone increasing for \( s \in (1/2, +\infty) \).

5. The inequality
\[
U_s^{t-r} \leq U_r^{t-s} U_t^{s-r}
\]
holds for any \( r < s < t \).

**Proof.** 1. Omitted.

2. Since both \( K_s \) and \( V_s \) are log-convex functions, we get
\[
U_s U_t - U_{s+t}^2 = [K_s(p||q) + K_s(q||p)][K_t(p||q) + K_t(q||p)] - [K_{s+t}(p||q) + K_{s+t}(q||p)]^2
\]
\[
= [K_s(p||q)K_t(p||q) - K_{s+t}(p||q)]^2 + [K_s(q||p)K_t(q||p) - K_{s+t}(q||p)]^2
\]
\[
+ [K_s(p||q)K_t(q||p) + K_s(q||p)K_t(p||q) - 2K_{s+t}(p||q)K_{s+t}(q||p)]
\]
\[
\geq [K_s(p||q)K_t(p||q) - K_{s+t}(p||q)]^2 + [K_s(q||p)K_t(q||p) - K_{s+t}(q||p)]^2
\]
\[
+ 2[\sqrt{V_s V_t} - V_{s+t}] \geq 0.
\]

3. The graph symmetry follows from the fact that \( U_s = U_{1-s}, s \in \mathbb{R} \).

We also have, due to arithmetic-geometric inequality, that
\[
U_s \geq 2\sqrt{V_s} \geq 4H^2.
\]

Finally, since \( p \neq q \) yields \( \max\{p_i/q_i\} = p_s/q_s > 1 \), we get
\[
K_s(p||q) > \frac{q_s(p_s/q_s)^s - 1}{s(s-1)} \to \infty (s \to \infty).
\]

It follows that both \( U_s \) and \( V_s \) are unbounded from above.

4. Omitted.

5. The proof is obtained by another application of Lemma 2.2 with \( \phi(s) = \log U_s \).

**Remark 2.4.** We worked here with the class \( \Omega^+ \) for the sake of simplicity. Obviously that all results hold, after suitable adjustments, for arbitrary probability distributions and in the continuous case as well.

**Remark 2.5.** It is not difficult to see that the same properties are valid for normalized divergences \( U^*_s = \frac{1}{2}(K_s(p||q) + K_s(q||p)) \) and \( V^*_s = \sqrt{K_s(p||q)K_s(q||p)} \), with
\[
2H^2 \leq V^*_s \leq U^*_s.
\]
3 Conclusion

In this paper we consider symmetrized divergences

\[ U_s = U_s(p, q) = U_s(q, p) := K_s(p||q) + K_s(q||p), \]

and

\[ V_s = V_s(p, q) = V_s(q, p) := K_s(p||q)K_s(q||p), \]

where \( K_s(p||q) \) is the \( s \)-divergence given by

\[
K_s(p||q) := \begin{cases} 
(p_s q_1^{-s} - 1)/s(s - 1), & s \in \mathbb{R} \setminus \{0, 1\}; \\
K(q||p), & s = 0; \\
K(p||q), & s = 1.
\end{cases}
\]

Also, well known Kullback-Leibler divergence \( K(p||q) \) is defined as

\[ K(p||q) := \sum p_i \log(p_i/q_i). \]

It is proved here that both \( U_s \) and \( V_s \) are log-convex for \( s \in \mathbb{R} \), monotone decreasing for \( s \in (-\infty, 1/2) \) and monotone increasing for \( s \in (1/2, +\infty) \).

Also, they are unbounded from above with \( U_s \geq 4H^2 \) and \( V_s \geq 4H^4 \), where \( H \) denotes well known Hellinger distance.

4 Appendix

A convexity property

Most general class of convex functions is defined by the inequality

\[ \frac{\phi(x) + \phi(y)}{2} \geq \phi\left(\frac{x+y}{2}\right). \]  \hspace{1cm} (4.1)

A function which satisfies this inequality in a certain closed interval \( I \) is called convex in that interval. Geometrically it means that the midpoint of any chord of the curve \( y = \phi(x) \) lies above or on the curve.

Denote now by \( Q \) the family of weights i.e., positive real numbers summing to 1. If \( \phi \) is continuous, then much more can be said i.e., the inequality

\[ p\phi(x) + q\phi(y) \geq \phi(px + qy) \]  \hspace{1cm} (4.2)

holds for any \( p, q \in Q \). Moreover, the equality sign takes place only if \( x = y \) or \( \phi \) is linear (cf. [2]).

We shall prove here an interesting property of this class of convex functions.

**Proposition X** Let \( f(\cdot) \) be a continuous convex function defined on a closed interval \( [a, b] := I \). Denote

\[ F(s, t) := f(s) + f(t) - 2f\left(\frac{s + t}{2}\right). \]

Then

\[ \max_{s, t \in I} F(s, t) = F(a, b). \] \hspace{1cm} (1)
Proof. It suffices to prove that the inequality

\[ F(s, t) \leq F(a, b) \]

holds for \( a < s < t < b \).

In the sequel we need the following assertion (which is of independent interest).

**Lemma 4.3.** Let \( f(\cdot) \) be a continuous convex function on some interval \( I \subseteq \mathbb{R} \). If \( x_1, x_2, x_3 \in I \) and \( x_1 < x_2 < x_3 \), then

(i) \[ \frac{f(x_2) - f(x_1)}{2} \leq f\left(\frac{x_2 + x_3}{2}\right) - f\left(\frac{x_1 + x_3}{2}\right); \]

(ii) \[ \frac{f(x_3) - f(x_2)}{2} \geq f\left(\frac{x_2 + x_3}{2}\right) - f\left(\frac{x_1 + x_2}{2}\right). \]

We shall prove the first part of the lemma; the proof of second part goes along the same lines.

Since \( x_1 < x_2 < \frac{x_2 + x_3}{2} < x_3 \), there exist \( p, q; \ 0 < p, q < 1, p + q = 1 \) such that \( x_2 = px_1 + qx_2 + x_3 \).

Hence,

\[
\frac{f(x_1) - f(x_2)}{2} + f\left(\frac{x_2 + x_3}{2}\right) \geq \frac{1}{2} [f(x_1) - (pf(x_1) + qf\left(\frac{x_2 + x_3}{2}\right))] + f\left(\frac{x_2 + x_3}{2}\right) \\
= \frac{q}{2} f(x_1) + \frac{2 - q}{2} f\left(\frac{x_2 + x_3}{2}\right) \geq f\left(\frac{q}{2} x_1 + \frac{2 - q}{2} x_2 + \frac{x_2 + x_3}{2}\right) = f\left(\frac{x_1 + x_3}{2}\right).
\]

Now, applying the part (i) with \( x_1 = a, x_2 = s, x_3 = b \) and the part (ii) with \( x_1 = s, x_2 = t, x_3 = b \), we get

\[
\frac{f(s) - f(a)}{2} \leq f\left(\frac{s + b}{2}\right) - f\left(\frac{a + b}{2}\right); \tag{2}
\]

\[
\frac{f(b) - f(t)}{2} \geq f\left(\frac{s + b}{2}\right) - f\left(\frac{s + t}{2}\right), \tag{3}
\]

respectively.

Subtracting (2) from (3), the desired inequality follows.

**Corollary 4.4.** Under the conditions of Proposition X, we have that the double inequality

\[ 2f\left(\frac{a + b}{2}\right) \leq f(t) + f(a + b - t) \leq f(a) + f(b) \tag{4} \]

holds for each \( t \in I \).

Proof. Since the condition \( t \in I \) is equivalent with \( a + b - t \in I \), applying Proposition X with \( s = a + b - t \) we obtain the right-hand side of (4). The left-hand side inequality is obvious.

**Remark 4.5.** The relation (4) is a kind of pre-Hermite-Hadamard inequalities. Indeed, integrating both sides of (4) over \( I \), we obtain the famous H-H inequality

\[
f\left(\frac{a + b}{2}\right) \leq \frac{1}{b - a} \int_a^b f(t) dt \leq f\left(\frac{a + f(b)}{2}\right),
\]

since \( \int_a^b f(a + b - t) dt = \int_a^b f(t) dt \).
References