On the Conjugate Measure of an Information Divergence using the Hermite-Hadamard Inequality

Hiroyuki SHIOYA

Muroran Institute of Technology
Address: 27-1 Mizumoto Muroran, 050-8585, JAPAN
Email: shioya@csse.muroran-it.ac.jp
Tel: +81-143-46-5436

Abstract

There are many types of information divergences with used two probability distributions. A new information divergence measure using the Hermite-Hadamard inequality was derived as the type similar to Csisár's f-divergence. In this paper, we analyze the property concerning to the conjugate of its new divergence and derive some properties of the sequence of convex functions, which is generated by the transformation obtained from the conjugate of its divergence.
Keywords : Hermite-Hadamard inequality, Information divergence.

1 Introduction

There are some various information divergences used in information theory, statistics and engineering applications. Csiszár's f-divergence is a generalized version of usual information divergences, and it's defined as [5]:

$$D_f(p\|q) \overset{\text{def}}{=} \int_{\boldsymbol{X}} p(x) f\Big(\frac{q(x)}{p(x)}\Big)\mu(dx) \tag{1}$$

where p, q $(\in \mathcal{P})$,

$$\mathcal{P} \overset{\text{def}}{=} \{p| \int_{\boldsymbol{X}} p(x)\mu(dx) = 1, p(x) \geq 0 \ \forall x \in \boldsymbol{x}\}, \tag{2}$$

and $f(u)$ is a convex function on $]0, \infty[$, $f(1) = 0$ and strictly convex at $u = 1$. Let $\mathcal{F}$ be a set of such all convex functions, and we will use the following convention:$0f(0/0) = 0$ and $0f(a/0) = \lim_{\epsilon\to+0} \epsilon f(a/\epsilon) = a \lim_{u\to\infty} f(u)/u$.

The theoretical framework of the f-divergence is based on the properties of convex functions. Jensen's inequality is important for Csiszar's f-divergence in order to satisfy the axiom of information divergences. Its refinement version is called the Hermite-Hadamard inequality [12]:

$$f(\frac{a+b}{2}) \leq \frac{\int_a^b f(t)dt}{b-a} \leq \frac{f(a)+f(b)}{2} \tag{3}$$

where f is a convex function on I, $a, b \in I$ and $a < b$.

We will treat some typical information divergence defined by using the Hermite-Hadamard inequality. It is called the Hermite-Hadamard divergence (HH-divergence in short) which was derived in [9]:

$$D^f_{HH}(p\|q) \stackrel{\text{def}}{=} \int_{\boldsymbol{X}} p(x) \frac{\int_1^{\frac{q(x)}{p(x)}} f(t)dt}{\frac{q(x)}{p(x)} - 1} \mu(dx) \tag{4}$$

where $f(u) \in \mathcal{F}$. The HH-divergence has the following properties:

(i) $D^f_{HH}(p\|q) \geq 0$, with the equality if and only if $p = q$.

(ii) $D^f_{HH}(p\|q) = D^g_{HH}(p\|q)$, where $g(u) = f(u) + c(u-1)$, $c \in \boldsymbol{R}$,

(iii) $D_f(p\|\frac{p+q}{2}) \leq D^f_{HH}(p\|q) \leq \frac{1}{2} D_f(p\|q)$

The property (ii) is also satisfied in the case of the f-divergence. Thus we use the equivalence $p \sim q$, if $g(u) = c_1 f(u) + c_2(u-1)$ and $c_1 \in \boldsymbol{R}^+, c_2 \in \boldsymbol{R}$. The conjugate measure of the HH-divergence has not been discussed.

In this paper, we analyze the conjugate of the HH-divergence, then we obtain some transformation on $\mathcal{F}$. We find the algebraic structure in the sequence of convex functions generated by its transformation. In conclusion, we have that the structure of the sequence is relate to the symmetric HH-divergence.

2 Conjugate HH-divergences

The conjugate measure of the f-divergence $D_f(p\|q)$ is represented by $D_{f^*}(p\|q)$. Thus the conjugate of f-divergences is given by the mapping from f to f^* (where $f^*(u) = uf(1/u)$). Using the function "g", the conjugate of D^f_{HH} will be given by:

$$D^f_{HH}(q\|p) = D^g_{HH}(p\|q) \tag{5}$$

If "g" is not convex for $\forall f \in \mathcal{F}$, D^f_{HH} does not have the property of the conjugate. Thus, we will show the theorem concerning to the function "g", which gives the conjugate HH-divergence.

Theorem 1 *If $f(\in \mathcal{F})$ satisfies $f''(u) > 0$ for $\forall u$, the function "f_{dh}" in eq.(6) is also convex.*

$$D^f_{HH}(q\|p) = D^{f_{dh}}_{HH}(p\|q). \tag{6}$$

where,

$$f_{dh} = f\left(\frac{1}{u}\right) - 2u \int_1^{\frac{1}{u}} f(t)dt. \tag{7}$$

Let $H_f(u)$ be $\frac{\int_1^u f(t)dt}{u-1}$. The conjugate $H_f(u)$ is

$$H^*_f(u) = -u^2 \frac{\int_1^{1/u} f(t)dt}{u-1}. \tag{8}$$

Using some function $h(u)$,

$$D^h_{HH}(p\|q) = D_{H^*_f}(p\|q). \tag{9}$$

So we have,

$$h(u) = f(\frac{1}{u}) - 2u \int_1^{\frac{1}{u}} f(t)dt. \tag{10}$$

In the next, we show that the mapping from f to f_{dh} is divided by two types of transformations on $\mathcal{F}$.

Lemma 1

$$f_{dh}(u) = f^{\oplus *}(u) \tag{11}$$

where,

$$f^{\oplus}(u) = uf(u) - 2\int_1^u f(t)dt. \tag{12}$$

We easily have

$$f_{dh}(u) = f^{\oplus *}(u). \tag{13}$$

If $f(u) > 0$ for $u \neq 1$, $f^{\oplus}$ is convex, because:

$$[f^{\oplus}]''(u) = uf''(u) > 0 \tag{14}$$

Thus, $f^{\oplus}(u)$ is convex if $f'' > 0$. Using Lemma 1, we have $f_{dh}(u) = f^{\oplus *}(u)$ and $f^{\oplus *}(u)$ is convex. So Theorem 1 is showed.

From Theorem 1, we restrict a class of convex functions $\mathcal{F}$ satifying $f''(u) > 0$ ($\forall f \in \mathcal{F}$). Using the transformation "$\oplus *$", we are able to have the conjugate HH-divergence. If $g \sim g^{\oplus *}$ for $g \in \mathcal{F}$, we have

$$D^g_{HH}(p\|q) = D^g_{HH}(q\|p), \tag{15}$$

Moreover we notice that "$\oplus * \oplus *$" is an identity transformation, so we have the symmetric HH-divergence using any convex function $f \in \mathcal{F}$ as the following.

Corollary 1 $\forall f \in \mathcal{F}$,

$$D^{f+f^{\oplus *}}_{HH}(p\|q) = D^{f+f^{\oplus *}}_{HH}(q\|p), \tag{16}$$

$$D^f_{HH}(q\|p) = D^{f^{\oplus *}}_{HH}(p\|q) \leq \frac{1}{2}D_{f^*}(p\|q) \tag{17}$$

As the remarkable example, if $f(u) = -\frac{1}{2}\ln u$,

$$D^{-\ln t}_{HH}(p\|q) = D^{-\ln t}_{HH}(q\|p) \leq \frac{1}{2}D_k(p\|q). \tag{18}$$

The conjugate measure of D^f_{HH} is given by the product of the transformations $\oplus$ and $*$. Especially $D^{f_k}_{HH}$ is a typical example of symmetric HH-divergence. So we examine the detail properties of the transformation "$\oplus$", which was obtained by symmetric HH-divergences.

3 Sequences of Convex Functions and Conjugate HH-divergences

3.1 Transformation $\oplus$

The transformation $\oplus$ is obtained by the derivation of the symmetric HH-divergence. But however, it was proposed in [6]. In this subsection, we review the fundamental properties of $\oplus$ from [6].

$$f^{\oplus}(u) \stackrel{\text{def}}{=} uf(u) - 2\int_1^u f(t)dt. \tag{19}$$

And the inverse of $\oplus$ has been given by the following $\ominus$:

$$f^{\ominus}(u) \stackrel{\text{def}}{=} \frac{f(u)}{u} + 2\int_0^u \frac{f(t)}{t^2}dt + 2\int_1^u \int_1^s \frac{f(t)}{t^2}dt\, ds. \tag{20}$$

Lemma 2 *([6])* $\forall w \in \mathcal{F}C$

$$\lim_{n\to\infty} w^{\oplus^n} = w_0(t) \tag{21}$$

$$\lim_{n\to\infty} w^{\ominus^n} = w_0^*(t) \tag{22}$$

Using $\oplus$*, we have* $w^{*\ominus*\ominus} = w^{\ominus*\ominus*} = w$*, where,*

$$w_0(t) \stackrel{\mathrm{def}}{=} \begin{cases} 0 & \textit{if} \quad 0 < t \leq 1 \\ +\infty & \textit{otherwise} \end{cases} \tag{23}$$

and $\oplus^n$ *denotes nth-composed mapping of* $\oplus$*.*

We have the following divergence using w_0.

$$D_{w_0}(p\|q) = D_{w_0^*}(p\|q) = \begin{cases} 0 & \text{if } p(x) = q(x) \quad \forall x \in \boldsymbol{x} \\ +\infty & \text{otherwise} \end{cases} \tag{24}$$

3.2 Sequences of Convex Functions

If $f \sim g \in \mathcal{F}$, then $f^* \sim g^*$ and $f^\oplus \sim g^\oplus$ are established. Using the symmetric HH-divergence, we have the identity transformation "$\oplus * \oplus *$". So the inverse $\oplus^{-1}$ is obtained in the following.

Lemma 3 *The inverse* $\oplus^{-1}$ *is* $* \oplus *$*.*

For $\forall w \in \mathcal{F}$, we have the sequence of convex functions $\{w^{\oplus^m}\}_{m\in Z}$ by repeatedly using $\oplus$ and $\oplus^{-1}$. We will show a new result representing the relation between $\oplus$ and $*$.

Theorem 2 *For* $\forall w \in \mathcal{F} \setminus \{w_0, w_0^*\}C$ *The sequence* $\{w^{\oplus^n}\}_{n\in Z}$ *is:*

$$w_0^*, \cdots, w^{\oplus^{-n}}, \cdots, w^{\oplus^{-1}}, w, w^\oplus, \cdots, w^{\oplus^n}, \cdots, w_0 \tag{25}$$

For some $w \in \{w^{\oplus^n}\}_{n\in Z}$ *and a finite integer* m*, if* $w^* \sim w^{\oplus^m}$*, the sequence contains* f^* *for any* $f \in \{w^{\oplus^n}\}_{n\in Z}$*, moreover there exists* w_c *such that* $w_c^* \sim w_c^\oplus$ *or* $w_c^* \sim w_c$*.*

ProofF By using the assumption in this theorem, there exists the integer "m" such that $w^* \sim w^{\oplus^m}$. By replacing w with $w^\oplus$ and using $w^{\oplus*\oplus*} \sim w$, we have

$$(w^\oplus)^* \sim (w^*)^{\oplus^{-1}} \sim (w^{\oplus^m})^{\oplus^{-1}} \sim w^{\oplus^{m-1}}. \tag{26}$$

$\forall v \in \{w^{\oplus^n}\}_{n\in Z} \setminus \{w_0, w_0^*\}$, there exists the integer k such that $v \sim w^{\oplus^k}$ and

$$w^{\oplus^k *} \sim w^{\oplus^{m-k}}. \tag{27}$$

$m - k$ is an integer, thus v^* is involved in the sequence.

Using the correspondence between w and w^*, we remove the pairs (w, w^*), $(w^\oplus, w^{\oplus^{m-1}})$, $\cdots$ from the finite sequence $\{w^{\oplus^k}\}_{k=-m,\cdots,-1,0,1,\cdots,m}$. If m is even, $w^{\oplus^{\frac{m}{2}}}$ remainds in its sequence. Let $w^{\oplus^{\frac{m}{2}}}$ be w_c,

$$w_c^* \sim w^{\oplus^{\frac{m}{2}}*} \sim w^{\oplus^m \oplus^{-\frac{m}{2}}*} \sim w^{*(*\oplus*)\cdots(*\oplus*)*} \sim w^{\oplus^{\frac{m}{2}}} \sim w_c \tag{28}$$

If m is odd, the pair $(w^{\oplus^{\frac{m-1}{2}}}, w^{\oplus^{\frac{m+1}{2}}})$ remained after removing the pairs, and the following is established.

$$w^{\oplus^{\frac{m-1}{2}}\oplus} \sim w^{\oplus^{\frac{m+1}{2}}} \tag{29}$$

Using $w^{\oplus^{\frac{m-1}{2}}} \sim w_c$, we have $w_c^* \sim w_c^\oplus$. [Q.E.D.]

Definition 1 *w_c is called the center of the sequence $\{w^{\oplus^m}\}_{m\in Z}$, if $w_c^{\oplus} \sim w^*$ or $w_c^* \sim w_c$.*

Theorem 3 *The center of the sequence is unique, if it excists.*

Proof: We suppose that there are two centers $w_1^{\oplus} \sim w_1^*$ and $w_2^{\oplus} \sim w_2^*$ in the sequence. Also both ends of the sequence are w_0 and w_0^*. Without loss of generality, we suppose that:

$$w_1^{\oplus^2} \sim w_2. \tag{30}$$

Concretely the sequence $\{w_1^{\oplus^n}\}_{n\in Z}$ is the following.

$$\cdots, w_2^*, w_2, w_1, w_1^*, w_2^*, w_2, w_1, w_1^*, \cdots \tag{31}$$

Thus we can see the replacement of $w_2^*, w_2, w_1, w_1^*, w_2^*, w_2, w_1, w_1^*$ in the sequence.

In the case of the centers $w_1^* \sim w_1$ and $w_2^* \sim w_2$, we suppose $w_1^{\oplus} \sim w_2$ without loss of generality. Then we can see the replacement of $\cdots w_1, w_2, w_1, w_2 \cdots$ in the sequence. If the sequence has more than two centers, it is periodic. This does not depend on a kind of the center. But however both ends of the sequence are w_0 and w_0^*. This contradicts the assumption. [Q.E.D.]

Example 1 *Using a sequence of convex functions, the relation between the Kullback divergence $D_{f_k}(p\|q)$ and the χ^2-divergence $D_{f_{\chi^2}}(p\|q)$ is the following:*

$$\cdots f_{\chi^2} \xrightarrow{\oplus} f_k \xrightarrow{\oplus} f_k^* \xrightarrow{\oplus} f_{\chi^2}^* \cdots \tag{32}$$

Theorem 4 *If there excists w_k in the sequence $\{w^{\oplus^m}\}_{m\in Z} \setminus \{w_0, w_0^*\}$ such that $w_k^* \notin \{w^{\oplus^m}\}_{m\in Z}$, g^* is not in $\{w^{\oplus^m}\}_{m\in Z}$ for $\forall g \in \{w^{\oplus^m}\}_{m\in Z} \setminus \{w_0, w_0^*\}$. And there is not any center in the sequence $\{w^{\oplus^m}\}_{m\in Z}$.*

Proof: We suppose that there excists w_k in the sequence $\{w^{\oplus^m}\}_{m\in Z}$ such that $w_k^* \notin \{w^{\oplus^m}\}_{m\in Z}$. Thus w_k^* is in another sequence $T_{w_k^*} = \{(w_k^*)^{\oplus^n}\}_{n\in Z}$. The we have

$$(w^{\oplus^n})^* \sim w^{*(*\oplus*)(*\oplus*)\cdots(*\oplus*)} \sim w^{*\oplus^{-n}} \sim (w^*)^{\oplus^{n'}}, \tag{33}$$

where $n' = -n$. $(w^{\oplus^n})^*$ (for $\forall n \in Z$) is in the sequence T_{w^*}. This indicates that there is not any center in the sequence.[Q.E.D.]

Theorem 5 *There are three types of the sequences using $\oplus$ and $\mathcal{F}$. (a) Center1 ($w_c \sim w_c^*$), (b) Center2 ($w_c^{\oplus} \sim w_c^*$), (c) No center.*

Proof: (a), (b) and (c) are followed by Theorem 2, 3 and 4.

As concerning the conjugate divergence of the HH-divergence $D_{HH}^f(p\|q)$, the convex function "f" is relate to the center of the sequence type (b) in Theorem 5. More precisely,

Example 2 *If w_c is the center type (b), then we have*

$$D_{HH}^{w_c}(p\|q) = D_{HH}^{w_c}(q\|p). \tag{34}$$

4 Conclusion

In this paper, we examine the conjugate of the HH-divergence derived by using the Hermite-Hadamard inequality, and we obtain the transformation $\oplus$ on $\mathcal{F}$. This transformation has been proposed in [6], but however we discuss the significance of the transformation from the view point of the conjugate HH-divergence and f-divergence. We find the algebraic structure in the sequence and classify a kind of the sequences. In conclusion, there are three types of the structure of the sequence, and the convex function corresponding to the symmetric HH-divergence is relate to the center $w^{\oplus} \sim w^*$ of the sequence.

References

[1] Burbea J., & Rao C. R., "On the Convexity of Some Divergence Measures Based on Entropy Functions",IEEE Trans. IT,Vol. 28, No. 3, pp.489-495, May 1982.

[2] Chernoff,H,"A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations", Ann. Math. Ats.,vol. 23, pp. 493-507, 1952.

[3] Cover, T. M.,& Thomas, J. A., "Element of Information Theory", Wiley-Interscience publication,(1991)

[4] Csiszár, I., "On information-type measure of difference of probability distributions and indirect observations", Studia Sci. Math. Hungar., 2, pp. 299-318(1967).

[5] Csiszár, I., "On topological property of f-divergence" Studia Sci. Math. Hungar.,2, pp. 330-339(1967).

[6] Eguchi, S.,: "Constract from one probability measure to another, the associated geometry", Math. Sci. Lab. of Kyoto Univ. No 623 (1987)

[7] Lin, J.,"Divergence Measures based on the Shannon Entropy" IEEE Trans. IT,vol. 37,No. 1,pp,145-151 (1991)

[8] H. Shioya, H. Nagaoka, T. Da-te "A New Inequality and maximum of f-divergence and its Application of Learning Problrem". IEICE, Vol. J77-A,No. 4,pp.720-726(1994)

[9] Shioya,H.,Da-te,T.,: A Generalization of Lin Divergence and the Derivation of a New Information Divergence, Electronics and Communications in Japan, Part III: Fundamental Electronic Science, John Wiley & Sons Vol 78, No. 7, pp. 34-40,July 1995

[10] H. Shioya, T. Da-te, "Transformation of convex functions and Inequalities of Divergences" IEICE, Vol. J80-A, March, No. 3,pp.509-515(1997)

[11] H. Shioya, "f-divergence" Sympo. of Information Geometry, RIKEN (1997)

[12] Pečarić, J. E., "Notes on convex functions" General Inequalities 6, International Series of Numerical Mathematics, Vol.103,Brikhäuser Basel (1992)