情報処理学会論文誌　Vol.54　No.11　2344–2348 (Nov. 2013)

テクニカルノート

いくつかの f ダイバージェンス間の不等式について

一森 哲男[1,a)]

受付日 2013年5月9日, 採録日 2013年7月3日

概要：情報理論の分野では，分布間の差異を測るダイバージェンスという情報量が多数存在する．本ノートでは，いくつかの f ダイバージェンス間の不等式について議論する．具体的には，全変動距離と θ ダイバージェンス間の不等式，θ ダイバージェンスと α ダイバージェンス間の不等式，および，α ダイバージェンス間どうしの不等式を導く．θ ダイバージェンスとは新たに本ノートの中で定義した情報量で，Pearson と Neyman のカイ二乗ダイバージェンス，および，三角ダイバージェンスを一般化したものである．

キーワード：情報ダイバージェンス，f ダイバージェンス，α ダイバージェンス，全変動距離

On Inequalities between f-Divergences

Tetsuo Ichimori[1,a)]

Received: May 9, 2013, Accepted: July 3, 2013

Abstract: There exist many measures of the divergence (or distance or discrimination) between probability distributions in the literature on information theory. This note deals with inequalities for some of the f-divergences, specifically, for the total variation distance, the θ-divergences and the α-divergences. The θ-divergences introduced here are generalizations of Pearson's and Neyman's χ^2 divergences and the triangular discrimination.

Keywords: K-L divergence, f-divergence, α-divergence, total variation distance

1. はじめに

情報理論では，2 つの分布間の差異を定量化するために f ダイバージェンスという情報量のクラスが定義されている．本論文では，この f ダイバージェンスのクラスに属するダイバージェンス間で成立する不等式について議論する．つまり，Kullback-Leibler ダイバージェンスと全変動距離との関係を表す Pinsker の不等式のような不等式について議論する．具体的には，全変動距離と θ ダイバージェンス間で成立する不等式，θ ダイバージェンスと α ダイバージェンス間で成立する不等式，最後に，α ダイバージェンス間どうしで成立する不等式を新たに導く．θ ダイバージェンスは新しく定義した情報量で，Pearson のカイ二乗ダイバージェンス，Neyman のカイ二乗ダイバージェンス，および，三角ダイバージェンスを一般化したものである．

導かれた不等式はすべてパラメータを含むが，これらは著者の知る限り新規の不等式である．パラメータに具体的な数値を代入すればいくらでも（パラメータを含まない）不等式を導くことが可能である．実際，θ ダイバージェンスおよび α ダイバージェンスに含まれるパラメータ θ および α にいくつかの値を代入することにより，多くの既知の不等式を導いている．これらは，たとえば，文献 [1] や [2] などに散見される．また，θ ダイバージェンスは新しく提案したものであることもあり，パラメータ θ および α にその他のさまざまな値を代入することにより，未知の関係式を作り出すことも可能である．

1　大阪工業大学
Osaka Institute of Technology, Hirakata, Osaka 573–0196, Japan
a)　ichimori@is.oit.ac.jp

2. f ダイバージェンス

分布間の差異を計測するダイバージェンスとして，Csiszár の f ダイバージェンス [3] は最も一般的なものとして知られている．これは Ali-Silvey 距離 [4] ともいわれているが，f ダイバージェンスは数学で定義される距離関数とは異なり，対称性や三角不等式が成り立たないので，厳密には，距離とはいえない．

1 から n までの整数の集合 $\mathbb{N} = \{1, 2, \ldots, n\}$ を定義する．2 つの離散確率分布 $P = (p_1, p_2, \ldots, p_n)$ と $Q = (q_1, q_2, \ldots, q_n)$ が与えられたとする．ここで，$p_i > 0$ $(i \in \mathbb{N})$ かつ $\sum_{i=1}^n p_i = 1$ であり，$q_i > 0$ $(i \in \mathbb{N})$ かつ $\sum_{i=1}^n q_i = 1$ である．このとき，両分布間の f ダイバージェンスとは

$$D_f(P||Q) = \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right)$$

と定義されている．正の実数の集合を $\mathbb{R}_+$ とし，実数の集合を $\mathbb{R}$ と書く．関数 $f : \mathbb{R}_+ \to \mathbb{R}$ は狭義凸で，さらに，$f(1) = 0$ と正規化されている．しばしば，$D_f(P||Q)$ は P から Q までの（有向）距離という．ただし，これを Q から P までの距離と定義する人もいる．

有名な Jensen の不等式より

$$\sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right) \geq f\left(\sum_{i=1}^n q_i \frac{p_i}{q_i}\right) = f(1) = 0$$

が導かれ，その値が非負 $D_f(P||Q) \geq 0$ であることが分かる．また，関数の狭義凸性より，$D_f(P||Q) = 0$ となるのは，すべての $i \in \mathbb{N}$ に対し $p_i = q_i$ となるとき，および，そのときのみである．

たとえば，$f(t) = t \log t$ $(t > 0)$ とおけば，f ダイバージェンスは

$$\sum_{i=1}^n p_i \log \frac{p_i}{q_i}$$

となり，Kullback-Leibler（以下，K-L と略する）ダイバージェンス [5] が得られる．これは相対エントロピーや情報ダイバージェンスなどさまざまな名前で呼ばれている [6].

表 1 に関数 $f(t)$ とそれにより定まるダイバージェンスをいくつか与える．表 1 の $D = D(P||Q)$ は K-L ダイバージェンスを，また，$\bar{D} = D(Q||P)$ は逆 K-L ダイバージェンスを示す．$\chi^2 = \chi^2(P||Q)$ は Pearson のカイ二乗ダイバージェンス [7] を，また，$\bar{\chi}^2 = \chi^2(Q||P)$ は Neyman のカイ二乗ダイバージェンスを表している．$H = H(P||Q)$ は Hellinger 距離の二乗（Hellinger ダイバージェンス）[8] と呼ばれ，P と Q に関して対称である．すなわち，$H = H(P||Q)$ は $H(Q||P)$ に等しい．$\Delta = \Delta(P||Q)$ と $V = V(P||Q)$ はそれぞれ三角ダイバージェンス（triangular discrimination）と全変動距離（ℓ_1 距離）といわれ，これらも P と Q に関して対称である．最後は Chernoff の α ダイバージェンス [9] であり，これは 1 つのパラメータ $-\infty < \alpha < +\infty$ を含んでいる．$\alpha = 1$ の場合は極限操作により

$$f(t) = \lim_{\alpha \to 1} \frac{1}{\alpha(\alpha - 1)}(t^\alpha - 1) = t \log t$$

となり，K-L ダイバージェンス D が導かれる．$\alpha = 0$ の場合は極限操作により $f(t) = -\log t$ となり，逆 K-L ダイバージェンス $\bar{D}$ が導かれる．そのため，α ダイバージェンス A_α はすべての $-\infty < \alpha < +\infty$ の値に対して定義される．いくつかのパラメータ α の値に対応するダイバージェンスを**表 2** に与える．

表 1 代表的な f ダイバージェンス
Table 1 Typical examples of f-divergences.

$f(t)$ $(t > 0)$	ダイバージェンス
$t \log t$	$D = D(P\|\|Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i}$
$-\log t$	$\bar{D} = D(Q\|\|P) = \sum_{i=1}^n q_i \log \frac{q_i}{p_i}$
$(t-1)^2$	$\chi^2 = \chi^2(P\|\|Q) = \sum_{i=1}^n \frac{(p_i - q_i)^2}{q_i}$
$\frac{(t-1)^2}{t}$	$\bar{\chi}^2 = \chi^2(Q\|\|P) = \sum_{i=1}^n \frac{(q_i - p_i)^2}{p_i}$
$\frac{(\sqrt{t}-1)^2}{2}$	$H = H(P\|\|Q) = \sum_{i=1}^n \frac{1}{2}\left(\sqrt{p}_i - \sqrt{q}_i\right)^2$
$\frac{(t-1)^2}{t+1}$	$\Delta = \Delta(P\|\|Q) = \sum_{i=1}^n \frac{(p_i - q_i)^2}{p_i + q_i}$
$\|t-1\|$	$V = V(P\|\|Q) = \sum_{i=1}^n \|p_i - q_i\|$
$\frac{t^\alpha - 1}{\alpha(\alpha-1)}$	$A_\alpha = A_\alpha(P\|\|Q) = \frac{1}{\alpha(\alpha-1)}\left(\sum_{i=1}^n \frac{p_i^\alpha}{q_i^{\alpha-1}} - 1\right)$

表 2 代表的な α ダイバージェンス
Table 2 Typical examples of α-divergences.

α	ダイバージェンス
-1	$\bar{\chi}^2 = 2A_{-1}$
0	$\bar{D} = A_0$
$1/2$	$H = 1/4\, A_{\frac{1}{2}}$
1	$D = A_1$
2	$\chi^2 = 2A_2$

3. 全変動距離と θ ダイバージェンス間の不等式

この節では，カイ二乗ダイバージェンス χ^2，$\bar{\chi}^2$ および三角ダイバージェンス Δ を一般化した θ ダイバージェンスを定義し，全変動距離 V の二乗との大小関係を導く．

補題 1 $0 \leq \theta \leq 1$ ならば, 関数 $(x-1)^2/(\theta x + 1 - \theta)$ $(x > 0)$ は狭義凸である．

証明 $f(x) = (x-1)^2/(\theta x + 1 - \theta)$ とするとき $f'(x) = 2\{\theta x^2 + 2(1-\theta)x + \theta - 2\}/(\theta x + 1 - \theta)^2$ となり，さらに，$f''(x) = 4/(\theta x + 1 - \theta)^3$ となる．この分母に含まれる項 $\theta x + 1 - \theta$ を $\theta x + 1 - \theta = \theta x + (1-\theta)1$ と見

表 3　代表的な θ ダイバージェンス
Table 3　Typical examples of θ-divergences.

θ	ダイバージェンス
0	$\chi^2 = \Theta_0$
1/2	$\Delta = 1/2\,\Theta_{\frac{1}{2}}$
1	$\bar{\chi}^2 = \Theta_1$

なせば，$0 < \min\{x, 1\} \leq \theta x + 1 - \theta \leq \max\{x, 1\}$ となることから，$\theta x + 1 - \theta > 0$ が得られる．ゆえに，$f''(x) > 0$ となり，$f(x)$ は狭義凸となる．□

この狭義凸関数 $(x-1)^2/(\theta x + 1 - \theta)$ $(x > 0)$ により定まる f ダイバージェンスを θ ダイバージェンスと呼ぶことにする．すなわち，

$$\Theta_\theta = \Theta_\theta(P||Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{\theta p_i + (1-\theta) q_i} \tag{1}$$

を定義する．順に，$\theta = 0, 1/2, 1$ を代入すると，$\Theta_0 = \chi^2$（Pearson のカイ二乗ダイバージェンス），$\Theta_{\frac{1}{2}} = 2\Delta$（三角ダイバージェンス），$\Theta_1 = \bar{\chi}^2$（Neyman のカイ二乗ダイバージェンス）が導かれる．この結果を**表 3** に与える．

定理 1　$0 \leq \theta \leq 1$ ならば，不等式 $V^2 \leq \Theta_\theta$ が成り立つ．

証明　定義より，$V^2 = \{\sum_{i=1}^{n} |p_i - q_i|\}^2 = \left\{\sum_{i=1}^{n} \sqrt{(p_i - q_i)^2}\right\}^2$ となる．ここで，$\theta p_i + (1-\theta) q_i > 0$ なので，この式は

$$V^2 = \left\{\sum_{i=1}^{n} \sqrt{\frac{(p_i - q_i)^2}{\theta p_i + (1-\theta) q_i}} \sqrt{\theta p_i + (1-\theta) q_i}\right\}^2$$

と書き直せる．ここで，Cauchy-Schwarz の不等式 $(\sum_i x_i y_i)^2 \leq (\sum_i x_i^2)(\sum_i y_i^2)$ を用いると，すなわち，$x_i = \sqrt{(p_i - q_i)^2/(\theta p_i + (1-\theta) q_i)}$ および $y_i = \sqrt{\theta p_i + (1-\theta) q_i}$ とおくと，

$$V^2 \leq \left(\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{\theta p_i + (1-\theta) q_i}\right) \left(\sum_{i=1}^{n} (\theta p_i + (1-\theta) q_i)\right)$$

の関係が得られる．$\sum_i p_i = \sum_i q_i = 1$ なので，上式の右辺第 2 項は 1 に等しくなり，θ ダイバージェンスの定義式 (1) より $V^2 \leq \Theta_\theta$ が得られる．□

系 1　$V^2 \leq \chi^2$，$V^2 \leq 2\Delta$，$V^2 \leq \bar{\chi}^2$ が成り立つ．

証明　順に，$\theta = 0, 1/2, 1$ を定理 1 の不等式に代入すると結果が得られる．□

4. α ダイバージェンスと θ ダイバージェンス間の不等式

パラメータ $-1 \leq \alpha \leq 2$ を持つ関数 $f_\alpha(x)$ $(x > 0)$ を以下のように定義する．すなわち，$\alpha \neq 0, 1$ の場合，

$$f_\alpha(x) = \left(\frac{x^\alpha - 1}{\alpha(\alpha - 1)} - \frac{x - 1}{\alpha - 1}\right) \times \left(-\frac{\alpha - 2}{3} x + \frac{\alpha + 1}{3}\right) - \frac{(x-1)^2}{2}$$

と定義し，$\alpha = 0, 1$ の場合は，それぞれ，

$$f_0(x) = (-\log x + x - 1)\left(\frac{2}{3} x + \frac{1}{3}\right) - \frac{(x-1)^2}{2}$$

および

$$f_1(x) = (x \log x - x + 1)\left(\frac{1}{3} x + \frac{2}{3}\right) - \frac{(x-1)^2}{2}$$

と定義する．

補題 2　パラメータが $-1 \leq \alpha \leq 2$ のとき，$x > 0$ に対し $f_\alpha(x) \geq 0$ となる．

証明　$\alpha \neq 0, 1$ の場合を考える．$f'_\alpha(x)$ および $f''_\alpha(x)$ を求めると，それぞれ，

$$f'_\alpha(x) = \frac{x^{\alpha-1} - 1}{\alpha - 1}\left(-\frac{\alpha - 2}{3} x + \frac{\alpha + 1}{3}\right) - \frac{\alpha - 2}{3}\left(\frac{x^\alpha - 1}{\alpha(\alpha - 1)} - \frac{x - 1}{\alpha - 1}\right) - (x - 1)$$

および

$$f''_\alpha(x) = x^{\alpha-2}\left(-\frac{\alpha - 2}{3} x + \frac{\alpha + 1}{3}\right) - \frac{2(\alpha - 2)}{3}\left(\frac{x^{\alpha-1} - 1}{\alpha - 1}\right) - 1$$

が得られる．さらに，$g(x) = f''_\alpha(x)$ とおくと，

$$g'(x) = -\frac{1}{3}(\alpha + 1)(\alpha - 2) x^{\alpha-3}(x - 1)$$

が得られる．$-1 \leq \alpha \leq 2$ なので，$(\alpha+1)(\alpha-2) \leq 0$ となることから，$g(x)$ は $0 < x < 1$ の範囲で減少，$x > 1$ で増加となる．さらに，$g(1) = 0$ となることから，$g(x) = f''_\alpha(x) \geq 0$ となり，$f_\alpha(x)$ が凸となる．$f_\alpha(1) = f'_\alpha(1) = 0$ であることから，$f_\alpha(x) \geq 0$ $(x > 0)$ が得られる．$\alpha = 0, 1$ の場合も，同様に証明ができる．よって，$-1 \leq \alpha \leq 2$ のとき，$f_\alpha(x) \geq 0$ $(x > 0)$ となる．□

定理 2　$-1 \leq \alpha \leq 2$ ならば，不等式

$$A_\alpha \geq \frac{1}{2} \Theta_{-\frac{\alpha-2}{3}}$$

が成り立つ．

証明　$\theta = -(\alpha - 2)/3$ とおくと，$-1 \leq \alpha \leq 2$ ならば $0 \leq \theta \leq 1$ となる．よって，$x > 0$ のとき $\theta x + 1 - \theta > 0$ となる．$1 - \theta = (\alpha+1)/3$ より，$(-(\alpha-2)/3)x + (\alpha+1)/3 = \theta x + (1 - \theta) > 0$ に注意すると，補題 2 の不等式 $f_\alpha(x) \geq 0$ は，$\alpha \neq 0, 1$ のとき，

$$\frac{x^\alpha - 1}{\alpha(\alpha - 1)} - \frac{x - 1}{\alpha - 1} \geq \frac{1}{2} \frac{(x-1)^2}{\theta x + 1 - \theta}$$

と書き直せる．左辺は狭義凸関数であり，これにより定まる f ダイバージェンスは，

$$\sum_{i=1}^{n} \frac{1}{\alpha - 1}\left(\frac{p_i}{q_i} - 1\right) q_i = \sum_{i=1}^{n} \frac{p_i - q_i}{\alpha - 1} = 0$$

に注意すると，A_α に等しくなる．また，右辺も補題 1 より狭義凸関数であり，これにより定まる f ダイバージェンスは，定義式 (1) より，$\frac{1}{2}\Theta_\theta$ である．$\alpha = 0, 1$ の場合も，同様に証明ができる．よって，この定理が成り立つ． □

系 2 不等式 $H \geq \frac{1}{4}\Delta$ が成り立つ．

証明 $\alpha = 1/2$ を定理 2 の不等式に代入すると結果が得られる． □

系 3 $-1 \leq \alpha \leq 2$ ならば，不等式 $A_\alpha \geq \frac{1}{2}V^2$ が成り立つ．

証明 定理 1 と定理 2 の不等式から導かれる． □

系 4 不等式 $\bar{\chi}^2 \geq V^2$, $\bar{D} \geq \frac{1}{2}V^2$, $H \geq \frac{1}{8}V^2$, $D \geq \frac{1}{2}V^2$, $\chi^2 \geq V^2$ が成り立つ．

証明 順に，$\alpha = -1, 0, 1/2, 1, 2$ を系 3 の不等式に代入すると結果が得られる． □

$\alpha = 1$ のときの不等式 $D \geq \frac{1}{2}V^2$ は Pinsker の不等式として知られている [10].

5. α ダイバージェンス間どうしの不等式

定理 3 $\alpha \leq \beta$ のとき，不等式

$$\alpha A_\alpha \leq \beta A_\beta$$

が成り立つ．

証明 これを証明するためには，関数

$$F(\alpha) = \alpha A_\alpha \quad (-\infty < \alpha < +\infty)$$

が増加関数であることを示せばよい．最初に，$F(\alpha)$ が $\alpha = 1$ で連続であることを示す．$\alpha = 1$ では $F(1) = A_1 = D$ となる．一方，$\alpha \neq 1$ のとき，

$$F(\alpha) = \frac{1}{\alpha - 1}\left(\sum_{i=1}^{n} \frac{p_i^\alpha}{q_i^{\alpha-1}} - 1\right)$$

となり，$\alpha \to 1$ のときの極限を調べてみと，

$$\lim_{\alpha \to 1} F(\alpha) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} = D$$

となる．このことから，$F(\alpha)$ が $\alpha = 1$ で連続であることが分かる．

次に，$\alpha \neq 1$ と仮定して，$F'(\alpha)$ を求めると，

$$F'(\alpha) = \frac{\sum_{i=1}^{n} \frac{p_i^\alpha}{q_i^{\alpha-1}} \left\{(\alpha - 1)\log \frac{p_i}{q_i} - 1\right\} + 1}{(\alpha - 1)^2}$$

となる．この式の分子の最後の項の 1 を $\sum_{i=1}^{n} p_i$ で置き換えると，この式の分子は

$$\sum_{i=1}^{n} \left[\frac{p_i^\alpha}{q_i^{\alpha-1}} \left\{(\alpha - 1)\log \frac{p_i}{q_i} - 1\right\} + p_i\right]$$

と書ける．

この式全体を $p_i > 0$ で除し，さらに，添え字 i を省略し，見やすいように $x = p_i/q_i$ と書き換えた第 i 項の式を

$$f(x) = x^{\alpha-1}\{(\alpha - 1)\log x - 1\} + 1, \quad x > 0$$

と定義する．以下では，$f(x)$ $(x > 0)$ が非負であることを示す．

$f(x)$ の右辺で，$t = x^{\alpha-1} > 0$ とおいた式を $g(t)$ とすると

$$g(t) = t \log t - t + 1$$

が得られる．$g'(t) = \log t$，$g''(t) = 1/t$ より $g(1) = g'(1) = 0$ かつ $g''(t) > 0$ より，$g(t) \geq 0$ すなわち $f(x) \geq 0$ が示される．よって，$F'(\alpha) \geq 0$ となり，結果，$F(\alpha)$ は増加関数となる． □

系 5 $2H \leq D$，$2H \leq \chi^2$，$D \leq \chi^2$ が成り立つ．

証明 順に，$(\alpha, \beta) = (1/2, 1), (1/2, 2), (1, 2)$ を定理 3 の不等式に代入すると結果が得られる． □

系 6 $2H \leq \bar{D}$，$2H \leq \bar{\chi}^2$，$\bar{D} \leq \bar{\chi}^2$ が成り立つ．

証明 系 5 の P と Q を入れ替えて，$H = H(P||Q) = H(Q||P)$ に注意すればよい． □

6. あとがき

新たに θ ダイバージェンスを導入し，全変動距離と θ ダイバージェンスの間の大小関係（定理 1），θ ダイバージェンスと α ダイバージェンスの間の大小関係（定理 2），最後に，α ダイバージェンス間どうしの大小関係（定理 3）を明らかにした．著者の知る限り，これらの関係は新規である．さらに，パラメータにいくつかの値を代入することにより，多くの既知の不等式を導いた．

参考文献

[1] Topsøe, F.: Some Inequalities for Information Divergence and Related Measures of Discrimination, *IEEE Trans. Inform. Theory*, Vol.46, No.4, pp.1602–1609 (2000).

[2] Taneja, I.J. and Kumar, P.: Relative Information of Type s, Csiszár's f-divergence, and Information Inequalities, *Information Sciences*, Vol.166, pp.105–125 (2004).

[3] Csiszár, I.: Information-type Measures of Difference of Probability Distributions and Indirect Observation, *Studia Sci. Math. Hungarica*, Vol.2, pp.299–318 (1967).

[4] Ali, S.M. and Silvey, S.D.: A General Class of Coefficients of Divergence of One Distribution from Another, *Journal of the Royal Statistical Society Series B*, Vol.28, pp.131–142 (1966).

[5] Kullback, S. and Leibler, R.A.: On Information and Sufficiency, *Annals of Mathematical Statistics*, Vol.22, pp.79–86 (1951).

[6] Cover, T.M. and Thomas, J.A.: *Elements of Information Theory, 2nd Edition*, Wiley, Hoboken, New Jersey (2006).

[7] Pearson, K.: On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such That it can be Reasonably Supposed to Have Arisen from Random Sampling, *Philosophical Magazine*, Vol.50, No.302, pp.157–172 (1900).

[8] Hellinger, E.: Neue Begründung der Theorie der Quadratischen Formen von Unendlichenvielen Veränderlichen, *Journal für die Reine und Angewandte Mathematik*, Vol.136, pp.210–271 (1909).
[9] Chernoff, H.: A Measure of Asymptotic Efficieny for Tests of a Hypothesis Based on a Sum of Observations, *Annals of Mathematical Statistics*, Vol.23, No.4, pp.493–507 (1952).
[10] Pinsker, M.S.: *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco (1964).

一森 哲男　（正会員）

昭和 28 年生．昭和 57 年大阪大学大学院工学研究科応用物理学専攻博士後期課程修了．同年広島大学工学部助手．昭和 60 年より大阪工業大学工学部専任講師．昭和 63 年より大阪工業大学工学部助教授．平成 8 年より大阪工業大学情報科学部教授．システムの最適化に関する研究に従事．工学博士．昭和 62 年日本オペレーションズ・リサーチ学会事例研究奨励賞受賞．日本オペレーションズ・リサーチ学会，日本応用数理学会各会員