数理解析研究所講究録
第 1561 巻 2007 年 45-53

Operator-valued divergence and entropy

大阪教育大　藤井 淳一　(Jun Ichi Fujii)
Department of Atrs and Sciences (Information Science)
Osaka Kyoiku University

1. 量子エントロピーとその作用素化

最近、さまざまな情報ダイバージェンスやエントロピーが論じられているが、ここではそれらを概観してそれらの関連について「作用素化」という観点から見つめなおしてみたい。話を単純化するために、作用素は可逆「行列」に限ることにする（A,B は、基本的に（正定値）密度行列としておく)。いわゆる「量子エントロピー」は、代表的なものとして、**von Neumann entropy** [15]

$$s(A) = -\mathrm{Tr}\, A \log A,$$

Kullback-Libler divergence の量子化にあたる **Umegaki (relative) entropy** [21]

$$s(A|B) = \mathrm{Tr}\, A(\log A - \log B),$$

さらに相互エントロピーの量子化として、CP-channel Λ（実際に使われている adjoint も同じ Λ であらわすことにする）を通したときの **Ohya (mutual) entropy** [16]

$$I(A, \Lambda) = \sup_E s\left(\sum t_n E_n \otimes \Lambda E_n \middle| A \otimes \Lambda A\right)$$

（ここで、E は、Shatten 分解 $A = \sum t_n E_n$ とする）などが、挙げられる。この最大値は、量子 channel の通信容量と考えられる。

一方、それらの作用素化として、**Nakamura-Umegaki operator entropy** [14]

$$S(A) = -A \log A,$$

われわれが導入した **relative operator entropy** [8, 9]

$$S(A|B) = A^{1/2} \log \left(A^{-1/2} B A^{-1/2}\right) A^{1/2}$$

などがある。後者については、Umegaki entoropy を考慮するならば、

$$S_U(A|B) = A^{1/2}(\log A - \log B)A^{1/2}$$

という作用素化もある（符号に差異があるので、$-\mathrm{Tr}\,(S_U(A|) = s(A|B)$ であることに注意)。また、Ohya mutual entoropy の作用素化として、Shatten 分解 E: $A = \sum t_n E_n$ に対する

$$M_E(A;\Lambda) = -S_U\left(\sum t_n E_n \otimes \Lambda E_n \middle| A \otimes \Lambda A\right)$$

が考えられる。勿論、$\max_E \mathrm{Tr}\, M_E(A;\Lambda) = I(A;\Lambda)$ である。

mutual entropy は、本来通信容量に関連すべきものであるが、Holevo[12] は（根元）密度作用素 $D = \{D_k\}$ で決まる classical-quantum channel の通信容量が von Neumann entropy s によって、

$$C(D) = \max_{\pi} s\left(\sum_{t_n \in \pi} t_n D_n\right) - \sum_{t_n \in \pi} t_n s(D_n)$$

（ただし、$\pi = \{t_n\}$ は確率分布）となることを指摘した。関連を見るために作用素化を以下のように考えてみよう。作用素凹となるエントロピー関数 $\eta(x) = -x \log x$ で（$S(A) = \eta(A)$ になるが)、Λ を unital completely positive linear map（の adjoint）とすると $\sum \Lambda E_n = I$ と、Jensen type inequality より、

$$H_E(A;\Lambda) = \eta\left(\sum_n t_n \Lambda E_n\right) - \sum_n t_n \eta(\Lambda E_n) \geq 0$$

という非負の量 $H_E(A,\Lambda)$ を Holevo 型の相互作用素エントロピーと見ることができる。このとき、これらのトレースを取ると、Ohya mutual entropy と一致するので、分解についての最大として対応する通信容量も一致することが分かる。実は、このことは、本質的に [17] で示されているが、ここで確認しておこう。

定理. 密度行列 A と、その Shatten 分解 E: $A = \sum_n t_n E_n$ について、

$$\mathrm{Tr}\, H_E(A;\Lambda) = \mathrm{Tr}\, M_E(A;\Lambda), \quad \max_E \mathrm{Tr}\, H_E(A;\Lambda) = I(A;\Lambda).$$

（証明）分解 E: $A = \sum_n t_n E_n$ について、

$$\begin{aligned}
\operatorname{Tr} M_E(A;\Lambda) &= \operatorname{Tr} A_E \left(\log\left(\sum_n E_n \otimes t_n \Lambda E_n\right) - \log\left(\sum_n E_n \otimes t_n \Lambda A\right)\right) \\
&= \operatorname{Tr} A_E \left(\sum_n E_n \otimes (\log t_n \Lambda E_n - \log t_n \Lambda A)\right) \\
&= \operatorname{Tr} \left(\sum_n E_n \otimes t_n \Lambda E_n\right)\left(\sum_n E_n \otimes (\log \Lambda E_n - \log \Lambda A)\right) \\
&= \operatorname{Tr} \left(\sum_n E_n \otimes (t_n \Lambda E_n \log \Lambda E_n - t_n \Lambda E_n \log \Lambda A)\right) \\
&= \sum_n \operatorname{Tr} \, (t_n \Lambda E_n \log \Lambda E_n) - \operatorname{Tr} \left(\sum_n t_n \Lambda E_n\right) \log \Lambda A \\
&= \sum_n t_n \operatorname{Tr} \, (\Lambda E_n \log \Lambda E_n) - \operatorname{Tr} \, (\Lambda A) \log \Lambda A \\
&= -\sum_n t_n s(\Lambda E_n) + s\,(\Lambda A) = \operatorname{Tr} \, (H_E(A;\Lambda)).
\end{aligned}$$

したがって、結論が分かる。 □

2. path と相対作用素エントロピー

Uhlmann[20] は、quadratic interpolation という平均の path を使って、Umegaki entropy を拡張した。その手法と同様に、われわれ [9] は、作用素幾何平均 $(0 \le t \le 1)$ の A から B への path

$$\begin{aligned}
A \#_t B &= A^{1/2}\left(A^{-1/2} B A^{-1/2}\right)^t A^{1/2} \\
&= B^{1/2}\left(B^{-1/2} A B^{-1/2}\right)^{1-t} B^{1/2}
\end{aligned}$$

として、相対作用素エントロピー

$$S(A|B) = A^{1/2} \log\left(A^{-1/2} B A^{-1/2}\right) A^{1/2}$$

が、両端での微分係数として得られることを示した：

$$\lim_{t\to 0} \frac{A\#_t B - A}{t-0} = S(A|B), \quad \lim_{t\to 1} \frac{A\#_t B - B}{t-1} = -S(B|A)$$

この path を測地線に持つような幾何学は、Corach-Porta-Recht らが [4] で論じている。

一方、いわゆる power mean と呼ばれているものの作用素平均族

$$A\#_{t,s}B = A^{1/2}\left(1-t+t(A^{-1/2}BA^{-1/2})^s\right)^{1/s}A^{1/2}$$

は、加重をもった平均として、$s=-1$ のとき調和平均、$s=0$ のとき幾何平均 $\#_t$、$s=1$ のとき算術平均 ∇_t となっている。最初のものを「横断 path」と呼ぶならば、これは作用素平均の「縦断 path」といえるだろう。これらの微分係数は、

$$\lim_{t\to 0}\frac{A\#_{t,s}B-A}{t-0} = \frac{A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^s A^{1/2}-A}{s} \equiv T_s(A|B)$$

となり、Yanagi-Kuriyama-Furuichi[19] によって、**Tsallis relative operator entropy** と名づけられたが、実際は既にわれわれが [9] で考察していた量であった。

話は逆行するが、梅垣相対エントロピーと平均 path の関係を改めてみるならば、以下のようになる:

$$\begin{aligned}
&\text{横断 path} \quad M_t(A,B) = \mathrm{Tr}\exp((1-t)\log A + t\log B)\\
&\text{横断 path'} \quad \tilde{M}_t(A,B) = \mathrm{Tr}\,A^{1-t}B^t \quad (\text{Uhlmann 型})\\
&\text{縦断 path} \quad M_{t,s}(A,B) = \mathrm{Tr}\,((1-t)A^s + tB^s)^{1/s}
\end{aligned}$$

$$\begin{aligned}
\lim_{t\to 0}\frac{M_t(A,B)-\mathrm{Tr}\,A}{t-0} &= \lim_{t\to 0}\frac{\tilde{M}_t(A,B)-\mathrm{Tr}\,A}{t-0} = -s(A|B)\\
\lim_{t\to 0}\frac{M_{t,s}(A,B)-\mathrm{Tr}\,A}{t-0} &= \mathrm{Tr}\,\frac{d(A^s+t(B^s-A^s))^{1/s}}{dt}\Big|_{t=0}\\
&= \frac{1}{s}\mathrm{Tr}\,(A^s+t(B^s-A^s))^{1/s-1}(B^s-A^s)\Big|_{t=0}\\
&= \mathrm{Tr}\,\frac{A^{1-s}(B^s-A^s)}{s} = \mathrm{Tr}\,\frac{A^{1-s}B^s-A}{s} \quad : \textbf{Tsallis entropy}
\end{aligned}$$

ここで、相対作用素エントロピーと梅垣相対エントロピーの関連について述べておく。相対作用素エントロピーは作用素平均から生まれたため、符号が逆転しているが、そのトレースを取ったものは、いわゆる、BS entropy である [2]:

Belavkin-Staszewski entropy $\qquad s_{\mathrm{BS}}(A|B) = -\mathrm{Tr}\,S(A|B).$

この関係について、Hiai-Petz[11] は次の不等式を示した:

$$s_{\mathrm{BS}}(A|B) \geq s(A|B).$$

[11] ではさらに次の結果を実質得ていることを、Bebiano-Lemos-da Providência[2] が指摘した:

$$\frac{s_{\mathrm{BS}}(A^p|B^p)}{p} \downarrow s(A|B) \qquad \text{as} \quad p \downarrow 0.$$

また、相対作用素エントロピー自身も operator divergence といえるが、さらに α-operator divergence を [6] で導入した:

$$\begin{aligned} D_\alpha(A,B) &= \frac{4}{1-\alpha^2}\left(A\nabla_{(1+\alpha)/2}B - A\#_{(1+\alpha)/2}B\right) \\ &= \frac{4}{1-\alpha^2}\left(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B - A\#_{(1+\alpha)/2}B\right). \end{aligned}$$

$$\begin{aligned} D_1(A|B) &\equiv \lim_{\alpha\uparrow 1} D_\alpha(A|B) = -S(B|A) + A - B \\ D_{-1}(A|B) &\equiv \lim_{\alpha\downarrow -1} D_\alpha(A|B) = -S(A|B) + B - A. \end{aligned}$$

特に、A, B が密度行列なら、

$$\operatorname{Tr} D_{-1}(A|B) = s_{\mathrm{BS}}(A|B), \quad \operatorname{Tr} D_1(A|B) = s_{\mathrm{BS}}(B|A)$$

となっている。

3. Bregman operator divergence

最近、Petz[18] は、Bregman divergence について論じている:

$$D_\Psi(A,B) = \Psi(A) - \Psi(B) - \lim_{t\to+0}\frac{\Psi(B+t(A-B)) - \Psi(B)}{t}.$$

Ψ が微分可能凸関数の場合には、この量は非負となり、Ψ が作用素凸関数の場合には、正作用素になる。また、A, B が可換で、$\Psi = h$ の場合には、

$$D_f(A,B) = h(A) - h(B) - (A-B)\frac{dh(B+t(A-B))}{dt}\Big|_{t=0}$$

となっている。特に、$\Psi(x) = x\log x = h(x)$ のとき、

$$D_h(A,B) = (B - EA) + h(A) - (\log B)EA - [h(B), W]$$

(E は、$W^*(B)$ への条件付期待値、W は $[f(B), W]$ が E で消えるよううまく選ばれた skew-Hemitian) と表すことができる。また、$\Psi = \mathrm{Tr} \circ f$、$A$,$B$ を密度行列とすれば、梅垣エントロピーが得られることも示している:

$$D_\Psi(A, B) = s(A|B).$$

さらに、この方向で、次の operator divergence も提唱している：

$$S_{FK}(A, B) = B - A - A^{1/2}(\log A^{-1/2} B A^{-1/2}) A^{1/2} = D_{-1}(A|B).$$

この Bregman divergence については、次のように書き直したほうが見やすい：

$$D_\Psi(A, B) = \lim_{t \to +0} \frac{\Psi(A)\nabla_{1-t}\Psi(B) - \Psi(A\nabla_{1-t}B)}{t}.$$

これは双対として、

$$\lim_{t \to 1} \frac{\Psi(A)\nabla_{1-t}\Psi(B) - \Psi(A\nabla_{1-t}B)}{1-t} = D_\Psi(B, A)$$

このように見ると、Ψ が作用素凸なら、定義から正値となることがわかる。

そこで、対称的な量として、**Bregman t-operator divergence** を以下のように提唱したい: $0 \leqq t \leqq 1$ と、作用素凹関数 f について、

$$\mathcal{D}_{f,t}(A, B) = \frac{f(A\nabla_t B) - f(A)\nabla_t f(B)}{t(1-t)}.$$

$$\lim_{t \to 0} \mathcal{D}_{f,t}(A, B) = D_{-f}(A, B)$$
$$\lim_{t \to 1} \mathcal{D}_{f,t}(A, B) = D_{-f}(A, B).$$

こう定義すれば、$f = \log,\ \eta,\ -1/x$ 等について、正定値な量となる。

4. さまざまな divergence

この章では、さまざまな divergence とその作用素化について触れておく。まず、かなり包括的な divergence として、**Csiszár f-divergence (Ali-Silvey distance)** が知られている。(Csiszár[4] によって、前者の名前で知られているが、その1年前に既に Ali-Silvey[1] が導入していたようである)：凸関数 f について、

$$D_f^c(p, q) = \int p(t) f\left(\frac{q(t)}{p(t)}\right) dt.$$

しばしば1での強凸性と、$f(1)=0$ が仮定されていることに注意しておく。一方、われわれは作用素単調関数（作用素凹）f について、solidarity という作用素平均を一般化した量を [7] で導入していた:

$$A\,s_f B = A^{1/2} f(A^{-1/2} B A^{-1/2}) A^{1/2}.$$

また、[9] で、Uhlmann 式に作用素平均の path の 0 での微係数の solidarity なら $f(1)=0$ となることを示した。これがちょうど Csiszár f-divergence (Ali-Silvey distance) の作用素化にあたっていることは、問題ないであろう。Bregman t-operator divergence はこのタイプではないが、勿論、α-operator divergence は含まれる。
その傍系として、Chernoff divergence [3] というのもある：

$$-\log \int \sqrt{p(x)^{1-t} q(x)^t}\,dx = -\log D^c_{x-t}(p^{-1/2}, q^{-1/2})$$

これは、次のように作用素化できる：

$$-\log \sqrt{A \#_t B}$$

これの数値化は、$-\mathrm{Tr}\,\log\sqrt{A\#_t B} = -\dfrac{1}{2}\mathrm{Tr}\,\log A\#_t B$ よりも、$\log \mathrm{Tr}\sqrt{A\#_t B}$ の方が原意に近いといえるだろう。

最後に、J.Lin [13] によって導入された **Jensen-Shannon divergence** について述べる。これは、KL-divergence を weight 付きで対称化したもので、そのまま Umegaki entoropy に適用でき、しかも、von Neumann entropy で表現できる:

$$\begin{aligned} JS(A,B) &\equiv s(A|A\nabla_t B)\,\nabla_t\, s(B|A\nabla_t B) \\ &= s((1-t)A + tB) - (1-t)s(A) - ts(B). \end{aligned}$$

この作用素化の2表現は等しくないので、どちらを採用すべきか分からない:

$$-\left(S(A|A\nabla_t B)\,\nabla_t\, S(B|A\nabla_t B)\right) \neq \eta((1-t)A + tB) - (1-t)\eta(A) - t\eta(B).$$

因みに作用素版でも、$t=1/2$ では、どちらも対称になっている：実際、

$$\begin{aligned}
&-(S(A|A\nabla_t B)\nabla_t S(B|A\nabla_t B)) \\
&\quad = -\frac{1}{2}A^{1/2}\log\left(\frac{1+A^{-1/2}BA^{-1/2}}{2}\right)A^{1/2} - \frac{1}{2}B^{1/2}\log\left(\frac{1+B^{-1/2}AB^{-1/2}}{2}\right)B^{1/2}.
\end{aligned}$$

$$\begin{aligned}
&\eta((1-t)A+tB)-(1-t)\eta(A)-t\eta(B) \\
&\quad = \eta\left(\frac{A+B}{2}\right)-\frac{\eta(A)+\eta(B)}{2}
\end{aligned}$$

である。

参考文献

[1] S.M.Ali and S.D.Silvey: A general class of coefficients of divergence of one distribution from another, J. Royal Stat. Soc. Ser. B, **28**(1966), 131–142.

[2] V.P.Belavkin and P.Staszewski: C*-algebraic generalization of relative entropy and entropy, Ann. Inst. H. Poincaré Sect. A., **37**(1982), 51–58.

[3] H. Chernoff: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Stat., **23**(1952), 493–507.

[4] G.Corach, H.Porta and L.Recht: Geodesics and operator means in the space of positive operators. Internat. J. Math. **4**(1993), no. 2, 193–202.

[5] I. Csiszár: Information type measures of differences of probability distribution and indirect observations, Studia Math. Huagarica, **2**(1967), 299–318.

[6] J.I.Fujii: 相対作用素エントロピーをめぐって, 京都大学数理解析研究所講究録 **903**(1995), 49–56.

[7] J.I.Fujii, M.Fujii and Y.Seo: An extension of the Kubo-Ando theory: Solidarities, Math. Japon., **35**(1990), 387–396.

[8] J.I.Fujii and E.Kamei: Relative operator entropy in noncommutative information theory, Math. Japon., **34** (1989), 341–348.

[9] J.I.Fujii and E.Kamei: Uhlmann's interpolational method for operator means. Math. Japon., **34** (1989), 541–547.

[10] F.Kubo and T.Ando: Means of positive linear operators, Math. Ann., **246**(1980), 205–224.

[11] F.Hiai and D.Petz: The proper formula for relative entropy and its asymptotics in quantum probability, Cooum. Math. Phys., **143**(1991), 99–114.

[12] A.S.Holevo: *The capacity of quantum channel with general signal states*, IEEE Trans. Inform. Theory, **44** (1998), 269–273.

[13] J. Lin: Divergence measures based on the shannon entropy, IEEE Trans. Inform. Theory, **37**(1991), 145–151.

[14] M.Nakamura and H.Umegaki: A note on the entropy for operator algebras, Proc. Jap. Acad., **37** (1961), 149–154.

[15] J. von Neumann: "Mathematische Grundlagen der Quantummechanik", Springer-Verlag, 1932.

[16] M.Ohya: *On compound state and mutual information in quantum information theory*, IEEE Trans. Inform. Theory **29**(1983), 770–774.

[17] M.Ohya and I.Volovich: On quantum capacity and its bound, Infin. Dimens. Anal. Quantum Probab. Relat. Top., **6**(2003), 301–310.

[18] D.Petz: Bregman divergence as relative operator entropy, Preprint (`http://www.renyi.hu/~petz/pdf/112bregman.pdf`)

[19] K.Yanagi, K.Kuriyama and S.Furuichi: Generalized Shannon inequalities based on Tsallis relative operator entropy. Linear Algebra Appl., **394**(2005), 109–118.

[20] A.Uhlmann: Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, Commun. Math. Phys., **54** (1977), 22–32.

[21] H.Umegaki: *Conditional expectation in an operator algebra IV*, Kodai Math. Sem. Rep. **14** (1962), 59–85.