数理解析研究所講究録
第 1935 巻 2015 年 156-162

行列の相対エントロピーと情報エントロピー
Relative entropies and information ones for matrices

大阪教育大学・教養学科・情報科学　藤井　淳一 (Jun Ichi Fujii)
Departments of Arts and Sciences (Information Science)
Osaka Kyoiku University

Let $X=(x_j)$ and $Y=(y_j)$ be probability vectors, and $P_c=(p_{ij})$ be a probability matrix which is called now a *classical channel*. The standard bases $\{a_j\}$ and $\{b_i\}$ are considered as elementary events for X and Y; $x_j=p(a_j)$, $y_j=p(b_j)$ and $p_{ij}=p(b_i|a_j)$. Then the classical entropies are defined as:

$$\text{compound entropy } H(X,Y)=-\sum_{i,j}p(a_j,b_i)\log p(a_j,b_i),$$

$$\text{conditional entropy } H(X|Y)=-\sum_{i,j}p(a_j,b_i)\log p(a_j|b_i),$$

$$\text{conditional entropy } H(Y|X)=-\sum_{i,j}p(a_j,b_i)\log p(b_i|a_j), \quad \text{and}$$

$$\text{mutual entropy } I(X;Y)=\sum_{i,j}p(a_j,b_i)\log\frac{p(a_j,b_i)}{p(a_j)p(b_i)}.$$

Then the relations between them are:

$$H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y)=I(X;Y)+H(X|Y)+H(Y|X),$$
$$H(X)=I(X;Y)+H(X|Y),\ H(Y)=I(X;Y)+H(Y|X).$$

We try to extend these classical information entropies to matrix ones. They can be expressed by usual sets with the set operations as quantities:

$$A\mapsto H(X),\quad B\mapsto H(Y),\quad A\setminus B\mapsto H(X|Y),\quad B\setminus A\mapsto H(Y|X),$$
$$A\cup B\mapsto H(X,Y),\quad A\cap B\mapsto I(A;B).$$

For this, the key quantity is the *relative entropy* which is initiated as the *Kullback-Leibler one*:

$$s(p|q)=\sum_{ij}p_i\log\frac{p_i}{q_j}$$

for probability vectors p and q.

Let $\eta(x)=-x\log x$ ($\eta(0)=0$) be the entropy function. Then the von Neumann entropy $s(A)=\mathrm{Tr}\eta(A)$ and Nakamura-Umegaki discussed 'an operator entropy' $H(A)=\eta(A)$ [11]. The *Umegaki entropy*, which is expressed by

$$s_U(A|B)=\sum\mathrm{Tr}A(\log A-\log B)$$

for positive-definite matrices A and B, is an extension of $s(p|q)$. Here A and B are often assumed to be *density matrices*, that is, they are positive-semidefinite and $\mathrm{Tr}A = \mathrm{Tr}B = 1$ which are quantum versions for X and Y. The *quantum channel* is a trace-preserving completely positive map Φ.

Based on the Umegaki entropy, Ohya [12] introduced the mutual information for quantum channel and discussed the capacity for the channel: For density operator $A = \sum_k t_k E_k$ with the spectral decomposition for that of identity $E = \{E_k\}$. For compound matrices

$$A_E = \sum_n t_n E_n \otimes \Phi(E_n) \quad \text{and} \quad A_0 = A \otimes \Phi(A),$$

the *Ohya mutual entropy* is defined as

$$I(A;\Phi) = \sup_E s_U(A_E|A_0),$$

which is a nice extension of the classical mutual entropy $I(X; P_c(X))$ for a channel matrix P_c. Also Petz [13] defined a *quantum conditional entropy*

$$h(\rho_{AB}|B) = s(\rho_{AB}) - s(B)$$

and it is related to the Umegaki entropy:

$$h(\rho_{AB}|B) = \log\dim H_A - s_U(\rho_{AB}|\tau_A \otimes \rho_B)$$

where τ_A is a tracial state and ρ_{AB} is a composite matrix as we see later. But unfortunately $h(\rho_{AB}|B)$ is not always positive.

Recall the sesquilinear version for the Uhlmann relative entropy s_{UL} (cf. [15]) which is an extension of the Umegaki one: Let α and β be positive sesquilinear forms and $\gamma(t) = QF_t(\alpha,\beta)$ be their interpolation. Then

$$s_{UL}(\alpha|\beta)(x) = -\liminf_{t\to 0} \frac{QF_t(\alpha,\beta) - \alpha}{t}(x,x).$$

Considering the derivatives A and B for α and β, we have, when they commute,

$$-\liminf_{t\to 0} \mathrm{Tr}\frac{A^{1-t}B^t - A}{t} = -\mathrm{Tr}\lim_{t\to 0}\frac{A^{1-t}B^t - A}{t} = \mathrm{Tr}A(\log A - \log B).$$

It suggests that the relative entropy can be defined as the initial tangent vector for some good path. Though a matrix version of the Umegaki entropy might be $A^{\frac{1}{2}}(\log A - \log B)A^{\frac{1}{2}}$, it might be not suitable from the geometrical viewpoint. In fact, the geodesic of one of the Hiai-Petz geometries ([9]) is $M_t(A,B) = \exp((1-t)\log A + t\log B)$ and hence its initial tangent vector is expressed by

$$\mathfrak{S}_U(A|B) \equiv \frac{dM_t(A,B)}{dt}\Big|_{t=0} = U\left(\left(\frac{1}{\log^{[1]}(d_i,d_j)}\right) \circ U^*(\log B - \log A)U\right)U^*$$

where U is a unitary with $\mathrm{diag}\,(d_i) = U^*AU$ and $f^{[1]}$ is the divided difference $f^{[1]}(x,y) = \frac{f(x)-f(y)}{x-y}$, see [7, 8] . We think it is a matrix version of the Umegaki entropy. In fact, $\mathrm{Tr}\mathfrak{S}_U(A|B) = \mathrm{Tr}A(\log B - \log A) = -s_U(A|B)$. Since the quantum conditional entropy is not positive though it is a numerical quantity and $\mathfrak{S}_U(A|B)$ is somewhat an awkward tool, here we do not use $\mathfrak{S}_U(A|B)$ while we fully use the above idea, in particuler, Ohya's construction.

In [5], we defined another relative entropy for positive operators based on the Kubo-Ando theory of operator means: Let $A\#_tB$ be a weighted geometric operator mean in the sense of Kubo-Ando [10];

$$A\#_tB = A^{\frac{1}{2}}\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)^t A^{\frac{1}{2}}$$

if A is invertible and $A\#_tB = \lim_{n\to\infty}(A+\frac{1}{n})\#_tB$ if not. We introduced in [5, 4] the *relative operator entropy* $S(A|B)$ as a derivative for a differentiable path of geometric operator means $A\#_tB$ if the following limit exists as a bounded operator;

$$\lim_{t\to 0}\frac{A\#_tB - A}{t}.$$

Afterwards, Corach, Porta and Recht [2] shows that the path $A\#_tB$ is the geodesic of their geometry of the positive operators and the realtive operator entropy is its initial tangent vector where the affine connection can be expressed by

$$\nabla_{\dot\gamma}\dot\delta = \ddot\delta - \frac{1}{2}\left(\dot\gamma\gamma^{-1}\dot\delta + \dot\delta\gamma^{-1}\dot\gamma\right).$$

for differential curves γ and δ, see also [3, 7].

If B is invertible, then $S(A|B) = B^{\frac{1}{2}}\eta\left(B^{-\frac{1}{2}}AB^{-\frac{1}{2}}\right)B^{\frac{1}{2}}$. In addition, if A is invertible, then $S(A|B) = A^{\frac{1}{2}}\log\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)A^{\frac{1}{2}}$. $-\mathrm{Tr}S(A|B)$ is the *Belavkin-Staszewski relative entropy.* So it is always exists for invertible operators, or positive-definite matrices. Basic properties are as follows:

Lemma 1. *The relative operator entropy has the following properties if it exists:*

(1) *If* $B \le B'$, *then* $S(A|B) \le S(A|B')$.

(2) $T^*S(A|B)T \le S(T^*AT|T^*BT)$ (*the equality holds for invertible* T).

(3) $S(A_1|B_1) + S(A_2|B_2) \le S(A_1 + A_2|B_1 + B_2)$.

(3′) $(1-t)S(A_1|B_1)+tS(A_2|B_2) \le S((1-t)A_1+tA_2|(1-t)B_1+tB_2)$ *for all* $t \in [0,1]$.

(4) $S\left(\bigoplus_k A_k|\bigoplus_k B_k\right) = \bigoplus_k S(A_k|B_k)$.

(5) $S(A|B) \le B - A$.

(6) $S(A|\alpha B) = (\log\alpha)A + S(A|B)$ *for* $\alpha > 0$.

Based on this relative matrix entropy, we discuss basic matrix entropies in the information theory.

Assume that $A \in \mathcal{M}_n^+$, the $n \times n$ positive-definite matrices and $B \in \mathcal{M}_m^+$, the $m \times m$ ones. Let $\{E_k\}$ be the (fixed) decomposition of the identity, that is, each E_k be a projection and $\sum_k E_k = I_n$. A set $\{E_k\}$ is considered as that of elementary probability events. Let $A = \sum_k t_k E_k$ be a spectral decomposition of of an invertible density matrix, that is, A is positive-definite and $\mathrm{tr}A = 1$. Then, we can observe that the probability $p(E_k)$ is given by $\mathrm{Tr}(t_k E_k) = t_k \mathrm{Tr}(E_k)$.

Let Φ be a quantum channel from $\mathcal{M}_n$ to $\mathcal{M}_m$. Then $F_k = \Phi(E_k)$ is considered as an elementary event, but it is no longer a projection. So we take a fixed set of positive-semidefinite matrices $\{F_\ell\}$ with $\sum_\ell F_\ell = I_m$, which is also called a POVM (positive operator-valued measure), and consider a density matrix $B = \sum_\ell s_\ell F_\ell$. Assume $s_\ell > 0$. Then note that B is invertible since $B \geq \sum_\ell \min_j\{s_j\} F_\ell = \min_j\{s_j\} I_m$.

In this situation, we define a composite matrix W_{AB} for A and B by

$$W_{AB} = \sum_{k,\ell} w_{k\ell} E_k \otimes F_\ell \quad \text{where } w_{k\ell} \geqq 0,\ \sum_k w_{k\ell} \mathrm{tr} E_k = s_\ell,\ \sum_\ell w_{k\ell} \mathrm{tr} F_\ell = t_k.$$

A typical example for composite matrices is $\sum_{k,\ell} t_k s_\ell E_k \otimes F_\ell$. In this case, A and B are called *independent*.

If all E_k and F_ℓ are of rank 1, then every (entrywise-)positive matrix $\{w_{k\ell}\}$ with $\sum_{k,\ell} w_{k\ell} = 1$ may induce the composite matrix as in the following example:

Example 1. Let $E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $E_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $(w_{k\ell}) = \frac{1}{12}\begin{pmatrix} 6 & 1 \\ 2 & 3 \end{pmatrix}$ and

$$A = \frac{1}{12}\begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}, \quad B = \frac{8}{12}F_1 + \frac{4}{12}F_2 = \frac{2}{3}F_1 + \frac{1}{3}F_2.$$

Then,

$$W_{AB} = \begin{pmatrix} \frac{6}{12}F_1 + \frac{1}{12}F_2 & O \\ O & \frac{2}{12}F_1 + \frac{3}{12}F_2 \end{pmatrix}.$$

The *composite matrix entropy* is defined by $H(W_{AB}) = \eta(W_{AB})$, the *mutual matrix entropy* by $I(A;B)$, and the *conditional entropies* $H(W_{AB}|A)$, $H(W_{AB}|B)$ by

$$I(A;B) = -S(W_{AB}|A \otimes B),\ H(W_{AB}|A) = S(W_{AB}|A \otimes I),\ H(W_{AB}|B) = S(W_{AB}|I \otimes B).$$

Immediately we have $H(W_{AB}) \geq 0$ and $H(W_{AB}|B) \geq 0$, while $I(A;B)$ is not always positive. But its trace is positive.

Then, by Lemma 1 (4), we express these entropies:

Lemma 2. *Matrix entropies have the following decompositions:*

(1) $H(W_{AB}) = \sum_k E_k \otimes \eta\left(\sum_\ell w_{k\ell} F_\ell\right)$.

(2) $I(A;B) = -\sum_k E_k \otimes S\left(\sum_\ell w_{k\ell} F_\ell | t_k B\right)$.

(3) $H(W_{AB}|A) = \sum_k E_k \otimes S\left(\sum_\ell w_{k\ell} F_\ell | t_k I\right) = \sum_k t_k E_k \otimes \eta\left(\sum_\ell \frac{w_{k\ell}}{t_k} F_\ell\right)$.

(4) $H(W_{AB}|B) = \sum_k E_k \otimes S\left(\sum_\ell w_{k\ell} F_\ell | B\right)$.

Thus, the latter case where ($\{F_\ell\}$ is PVM (projection-valued measure) shows the entropy values in the classical (commutative) case.

In the context for the composite elementary events $\{E_k \otimes F_\ell\}$, the entropy $\eta(A)$, $\eta(B)$ should be extended to

$$H_F(A) = -\sum_{k,\ell} \log(t_k) w_{k\ell} E_k \otimes F_\ell, \text{ and } H_E(B) = -\sum_{k,\ell} \log(s_\ell) w_{k\ell} E_k \otimes F_\ell.$$

In fact, we obtain by taking the partial trace

$$\mathrm{Tr}_2(H_F(A)) = -\sum_k \mathrm{Tr}\left(w_{k\ell} F_\ell\right) \log(t_k) E_k = -\sum_k t_k \log(t_k) E_k = \sum_k \eta(t_k) E_k = \eta(A)$$

and similarly $\mathrm{Tr}_1(H_E(B)) = \eta(B)$. Then we have the following relations similar to the classical cases:

Theorem 3. *The following equalities hold:*

(1) $H(W_{AB}|B) + I(A;B) = H_F(A)$, (2) $H_F(A) + H(W_{AB}|A) = H(W_{AB})$.

Example 2. If $\{F_\ell\}$ is a PVM, then

$$I(A;B) = -\begin{pmatrix} \frac{6}{12}\log\left(\frac{7}{9}\right) F_1 + \frac{1}{12}\log\left(\frac{7}{3}\right) F_2 & O \\ O & \frac{2}{12}\log\left(\frac{5}{3}\right) F_1 + \frac{3}{12}\log\left(\frac{5}{9}\right) F_2 \end{pmatrix},$$

$$H(W_{AB}|B) = \begin{pmatrix} \frac{6}{12}\log\left(\frac{8}{6}\right) F_1 + \frac{1}{12}\log(4) F_2 & O \\ O & \frac{2}{12}\log\left(\frac{8}{2}\right) F_1 + \frac{3}{12}\log\left(\frac{4}{3}\right) F_2 \end{pmatrix},$$

$$H_F(A) = -\begin{pmatrix} \frac{6}{12}\log\left(\frac{7}{12}\right) F_1 + \frac{1}{12}\log\left(\frac{7}{12}\right) F_2 & O \\ O & \frac{2}{12}\log\left(\frac{5}{12}\right) F_1 + \frac{3}{12}\log\left(\frac{5}{12}\right) F_2 \end{pmatrix} \quad \text{and}$$

$$H(W_{AB}) = \begin{pmatrix} \eta\left(\frac{6}{12}\right) F_1 + \eta\left(\frac{1}{12}\right) F_2 & O \\ O & \eta\left(\frac{2}{12}\right) F_1 + \eta\left(\frac{3}{12}\right) F_2 \end{pmatrix}.$$

The following example shows that the matrix entropies include the classical ones as diagonal matrices:

Example 3. For the case $F_k = E_k$, we have $W_{AB} = \frac{1}{12}\begin{pmatrix} 6 & & & \\ & 1 & & \\ & & 2 & \\ & & & 3 \end{pmatrix}$ and

$$A \otimes B = \frac{1}{12}\begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} \otimes \frac{1}{3}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then we obtain

$$
\begin{aligned}
I(A|B) &= -\frac{1}{36} S\left(\begin{pmatrix} 18 & & & \\ & 3 & & \\ & & 6 & \\ & & & 9 \end{pmatrix} \middle| \begin{pmatrix} 14 & & & \\ & 7 & & \\ & & 10 & \\ & & & 5 \end{pmatrix} \right) \\
&= \frac{1}{36} \begin{pmatrix} 18\log\frac{18}{14} & & & \\ & 3\log\frac{3}{7} & & \\ & & 6\log\frac{6}{10} & \\ & & & 9\log\frac{9}{5} \end{pmatrix}.
\end{aligned}
$$

Unlike the classical case, another equalities for the conditional matrix entropies do not always hold. But if F_ℓ are projections, they hold:

Proposition. *If $\{F_\ell\}$ is a PVM, then the equalities*

$$H(W_{AB}|A) + I(A;B) = H_E(B) \qquad \textit{and} \quad H_E(B) + H(W_{AB}B) = H(W_{AB})$$

hold.

The following example shows the above inequalities do not hold for POVMs:

Example 4. Let $(w_{k\ell}) = \frac{1}{12}\begin{pmatrix} 6 & 1 \\ 2 & 3 \end{pmatrix}$, $A = \frac{1}{12}\begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$,

$$P_1 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad P_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad F_1 = \frac{1}{4}P_1 + \frac{3}{4}P_2 \quad \text{and} \quad F_2 = \frac{3}{4}P_1 + \frac{1}{4}P_2.$$

Then we have

$$
\begin{aligned}
H_E(B) &= \begin{pmatrix} \frac{1}{16}\log\frac{3^3}{2^2}P_1 + \frac{1}{48}\log\frac{3^{19}}{2^{18}}P_2 & \\ & \frac{1}{48}\log\frac{3^{11}}{2^2}P_1 + \frac{1}{16}\log\frac{3^3}{2^2}P_2 \end{pmatrix}, \\
H(W_{AB}|A) &= \begin{pmatrix} \frac{9}{48}\log\frac{28}{9}P_1 + \frac{19}{48}\log\frac{28}{19}P_2 & \\ & \frac{11}{48}\log\frac{20}{11}P_1 + \frac{9}{48}\log\frac{20}{9}P_2 \end{pmatrix} \quad \text{and} \\
I(A;B) &= -\begin{pmatrix} \frac{9}{48}\log\frac{35}{27}P_1 + \frac{19}{48}\log\frac{49}{57}P_2 & \\ & \frac{11}{48}\log\frac{25}{33}P_1 + \frac{9}{48}\log\frac{35}{27}P_2 \end{pmatrix}.
\end{aligned}
$$

Thus the desired equality does not hold.

参考文献

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

[2] G.Corach, H.Porta and L.Recht, Geodesics and operator means in the space of positive operators. Internat. J. Math. **4** (1993), 193–202.

[3] G.Corach and A.L.Maestripieri, Differential and metrical structure of positive operators, Positivity **3** (1999), 297–315.

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

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

[6] J.I.Fujii and E.Kamei, Interpolational paths and their derivatives, Math. Japon. **39** (1993), 557–560.

[7] J.I.Fujii, Structure of Hiai-Petz parametrized geometry for positive definite matrices, Linear Algebra. Appl. **432** (2010), 318–326.

[8] J.I.Fujii, Path of quasi-means as a geodesic, Linear Algebra. Appl. **434** (2011), 542–558.

[9] F.Hiai and D.Petz, Riemannian metrics on positive definite matrices related to means, Linear Alg. Appl. **430** (2009), 3105–3130.

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

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

[12] M.Ohya, On compound state and mutual information in quantum information theory, IEEE trans. IT, **29**(1983), 770-774.

[13] D.Petz, "Quantum Information Theory and Quantum Statistics", Springer, 2008.

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

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

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