From f-divergence to quantum quasi-entropy

Dénes Petz
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences

E-mail: petz@math.bme.hu  Home page: http://www.renyi.hu/~petz
Divergence

Let $\mathcal{X}$ be a finite space with probability measures $p$ and $q$. Their relative entropy or divergence

$$D(p\|q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

was introduced by Kullback and Leibler in 1951. A possible generalization of the relative entropy is the $f$-divergence introduced by Csiszár:

$$D_f(p\|q) = \sum_{x \in \mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right)$$

with a real function $f(x)$ defined for $x > 0$. For $f(x) = x \log x$ the relative entropy is obtained. (This function $f$ is convex.)
Monotonicity

Let $\mathcal{A}$ be a partition of $\mathcal{X}$. If $p$ is a probability distribution on $\mathcal{X}$, then $p_\mathcal{A}(A) := \sum_{x \in A} p(x)$ becomes a probability distribution on $\mathcal{A}$.

**Theorem 1** Let $\mathcal{A}$ be a partition of $\mathcal{X}$ and $p, q$ be probability distributions on $\mathcal{X}$. If $f$ is a continuous convex function, then

$$D_f(p_\mathcal{A}||q_\mathcal{A}) \leq D_f(p||q).$$

The inequality in the theorem is the monotonicity of the $f$-divergence. A particular case is

$$f(1) \leq D_f(p||q).$$
Example

The function

\[ f_\alpha(t) = \frac{1}{\alpha(1 - \alpha)}(1 - t^\alpha), \]

gives the relative \( \alpha \)-entropy defined as

\[ S_\alpha(p||q) = \frac{1}{\alpha(1 - \alpha)} \left( 1 - \sum_x p(x)^\alpha q(x)^{1-\alpha} \right). \]

The limit \( \alpha \to 0 \) leads to the relative entropy.
Characterization

**Theorem 2** Assume that a number $C(p, q) \in \mathbb{R}$ is associated to probability distributions on the same set $\mathcal{X}$ for all finite sets $\mathcal{X}$. If

(a) $C(p, q)$ is invariant under the permutations of the basic set $\mathcal{X}$.

(b) if $A$ is a partition of $\mathcal{X}$, then $C(p_A, q_A) \leq C(p, q)$ and the equality holds if and only if

$$p_A(A)q(x) = q_A(A)p(x)$$

whenever $x \in A \in \mathcal{A}$,

then there exists a convex function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ which is continuous at 0 and $C(p, q) = D_f(p || q)$ for every $p$ and $q$. 
Quantum setting

In the mathematical formalism of quantum mechanics, instead of \(n\)-tuples of numbers one works with \(n \times n\) complex matrices. They form an algebra and this allows an algebraic approach. In this approach, a probability density is replaced by a positive semidefinite matrix of trace 1 which is called density matrix. The eigenvalues of a density matrix give a probability density. However, this is not the only probability density provided by a density matrix. If we rewrite the matrix in a certain orthonormal basis, then the diagonal element \(p_1, p_2, \ldots, p_n\) form a probability density.
Quasi-entropy

Let $\mathcal{M}$ denote the algebra of $n \times n$ matrices with complex entries. For positive definite matrices $\rho_1, \rho_2 \in \mathcal{M}$, for $A \in \mathcal{M}$ and a function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$, the quasi-entropy is defined as

$$S^A_f(\rho_1 \parallel \rho_2) := \langle A\rho_2^{1/2}, f(\Delta(\rho_1/\rho_2))(A\rho_2^{1/2}) \rangle$$

$$= \text{Tr} \rho_2^{1/2} A^* f(\Delta(\rho_1/\rho_2))(A\rho_2^{1/2}),$$

where $\Delta(\rho_1/\rho_2) : \mathcal{M} \rightarrow \mathcal{M}$ is a (self-adjoint) linear mapping acting on matrices:

$$\Delta(\rho_1/\rho_2)B = \rho_1 B \rho_2^{-1}.$$

This concept was introduced by DP in 1984. (The motivation was Csiszár’s $f$-divergence and Araki’s formula for the Umegaki’s relative entropy.)
Matrix concave functions

The monotonicity of the $f$-divergence is the consequence of the Jensen inequality. A function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ is called matrix concave if one of the following two equivalent conditions holds:

$$f(\lambda A + (1 - \lambda) B) \geq \lambda f(A) + (1 - \lambda) f(B)$$

for every number $0 < \lambda < 1$ and for positive definite square matrices $A$ and $B$ (of the same size). Next $\lambda$ is replaced by a matrix:

$$f(CAC^* + DBD^*) \geq Cf(A)C^* + Df(B)D^*$$

if $CC^* + DD^* = I$. 
Matrix monotone functions

A function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ is called **matrix monotone** if for positive definite matrices $A \leq B$ the inequality $f(A) \leq f(B)$ holds.

It is interesting that a matrix monotone function is matrix concave and a matrix concave function is matrix monotone if it is bounded from below.

The function

$$-f_\alpha(t) = -\frac{1}{\alpha(1 - \alpha)}(1 - t^\alpha),$$

is matrix monotone, therefore it is matrix concave.
Monotonicity of quasi-entropy

Let $\alpha : \mathcal{M}_0 \rightarrow \mathcal{M}$ be a mapping between two matrix algebras. The dual $\alpha^* : \mathcal{M} \rightarrow \mathcal{M}_0$ with respect to the Hilbert-Schmidt inner product is positive if and only if $\alpha$ is positive.

$\alpha : \mathcal{M}_0 \rightarrow \mathcal{M}$ is called a **Schwarz mapping** if

$$\alpha(B^*B) \geq \alpha(B^*)\alpha(B) \quad (B \in \mathcal{M}_0).$$

**Theorem 3** Assume that $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ is an operator monotone function with $f(0) \geq 0$ and $\alpha : \mathcal{M}_0 \rightarrow \mathcal{M}$ is a unital Schwarz mapping. Then

$$S^A_f(\alpha^*(\rho_1), \alpha^*(\rho_2)) \geq S^{\alpha(A)}_f(\rho_1, \rho_2)$$

holds for $A \in \mathcal{M}_0$ and for invertible density matrices $\rho_1$ and $\rho_2$ from the matrix algebra $\mathcal{M}$. 
Classical-quantum relation

Let $\rho_1$ and $\rho_2$ be density matrices in $\mathcal{M}$. If in certain basis they have diagonal $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$, then the monotonicity theorem gives the inequality

$$D_f(p\|q) \leq S_f(\rho_1\|\rho_2)$$

for a matrix convex function $f$.

The Pinsker-Csiszár inequality

$$(\|p - q\|_1)^2 \leq 2D(p\|q).$$

extends to the quantum case as

$$(\|\rho_1 - \rho_2\|_1)^2 \leq 2S(\rho_1\|\rho_2).$$

**Question:** Can Csiszár’s characterization theorem be extended?
Generalized covariance

If $\rho_2 = \rho_1 = \rho$ and $A, B \in \mathcal{M}$ are arbitrary, then one can approach to the **generalized covariance**:

$$\text{qCov}_\rho^f(A, B) := \langle A\rho^{1/2}, f(\Delta(\rho/\rho))(B\rho^{1/2}) \rangle - \langle \text{Tr}\rho A^* \rangle \langle \text{Tr}\rho B \rangle.$$  

If $\rho$, $A$ and $B$ commute, then this becomes

$$f(1)\text{Tr}\rho A^* B - \langle \text{Tr}\rho A^* \rangle \langle \text{Tr}\rho B \rangle.$$  

If $\text{Tr} A = \text{Tr} B = 0$ and $f(x) = (1 + x)/2$, then

$$\text{qCov}_\rho^f(A, B) = \frac{1}{2} \text{Tr}\rho (A^* B + B A^*).$$
Quadratic cost function

The normalization $f(1) = 1$ and the symmetry

$$q\text{Cov}_\rho^f(A, A) = q\text{Cov}_\rho^f(A^*, A^*)$$

are natural. The latter property is equivalent to the symmetry $xf(x^{-1}) = f(x)$ of $f$.

If $Tr A = Tr B = 0$ and $\rho = \text{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Then

$$q\text{Cov}_\rho^f(A, B) = \sum_{i, j} \lambda_i f(\lambda_j / \lambda_i) A_{i, j}^* B_{i, j}.$$

The interpretation of the covariances is not at all clear. It might be better to call it **quadratic cost function**. (It turns out that there is a one-to-one correspondence between quadratic cost functions and Fisher informations.)
Inequality of Cramér-Rao type

Let $\varphi_0[K, L]$ be an inner product (or quadratic cost function) on the linear space of self-adjoint matrices. When $\rho_\theta$ is smooth in $\theta$, then

$$\frac{\partial}{\partial \theta} \text{Tr} \rho_\theta B \bigg|_{\theta=0} = \varphi_0[B, L]$$

with some $L = L^*$. If $\text{Tr} \rho_\theta A = \theta$, then $\varphi_0[A, L] = 1$ and the Schwarz inequality yields

$$\varphi_0[A, A] \geq \frac{1}{\varphi_0[L, L]}.$$  

This is the celebrated inequality of Cramér-Rao type for the locally unbiased estimator. $\varphi_0[L, L]$ depends on the tangent of $\rho_\theta$, it is called Fisher information.
Characterization

Theorem 4 If for every invertible density matrix $\rho$ a positive definite bilinear form $\gamma_{\rho}$ is given such that

1. $\gamma_{\rho}(A, A) \geq \gamma_{\beta(\rho)}(\beta(A), \beta(A))$ holds for all completely positive coarse grainings $\beta$,

2. $\gamma_{\rho}(A, A)$ is continuous in $\rho$ for every fixed $A$,

3. $\gamma_{\rho}(A, A) = \gamma_{\rho}(A^*, A^*)$,

4. $\gamma_{\rho}(A, A) = \text{Tr} \rho^{-1} A^2$ if $A\rho = \rho A$,

then there exists a unique standard operator monotone function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that

$$\gamma_{\rho}^{f}(A, A) = \text{Tr} A J_{\rho}^{-1}(A) \quad \text{and} \quad J_{\rho} = \mathbb{R}^{1/2}_\rho f(L_{\rho} R_{\rho}^{-1}) R_{\rho}^{1/2}.$$
Remarks

1. $\gamma^f_\rho(A, A)$ is formally a quasi-entropy

$$S_{1/f}^A \rho^{-1} (\rho, \rho).$$

2. If $\rho = \text{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Then

$$\gamma^f_\rho(A, A) = \sum_{ij} \frac{1}{\lambda_i f(\lambda_j / \lambda_i)} |A_{ij}|^2.$$

3. There is a one-to-one correspondence between Fisher informations and quadratic cost functions:

$$\gamma^f_\rho(A, B) = \text{Tr} A A^\rho \rho^{-1}(B^*), \quad \varphi_\rho[A, A] := \text{Tr} A A^\rho(A).$$
Example 1

\( f_a(t) = (1 + t)/2 \) is the maximal standard function. This leads to the minimal Fisher information:

\[ \gamma_{\rho}^{\min}(A, A) = \text{Tr}AL = \text{Tr}\rho L^2, \quad \text{where} \quad \rho L + L\rho = 2A \]

and the quadratic cost function is

\[ \varphi_\rho[A, B] = \frac{1}{2} \text{Tr}\rho(AB + BA). \]

We have

\[ \mathbb{J}_\rho(A) = \frac{1}{2}(\rho A + A\rho) \quad \text{and} \quad \mathbb{J}_\rho^{-1}(A) = L = 2 \int_0^\infty e^{-t\rho} Ae^{-t\rho} \, dt \]

for the operator \( \mathbb{J} \).
Example 2

The function

\[ f_\beta(t) = \beta (1 - \beta) \frac{(x - 1)^2}{(x^\beta - 1)(x^{1-\beta} - 1)} \]

is operator monotone if \( 0 < |\beta| < 1 \).

If \( A = [\rho, B] \), then

\[ \gamma^\beta_\rho(A, A) = \frac{1}{2\beta(1 - \beta)} \text{Tr}(\rho^\beta B[\rho^{1-\beta}, B]). \]

This is the skew information proposed by bf Wigner and Yanase.
Example 3

If $\beta \to 0$, then

$$f_0(x) = \frac{x - 1}{\log x}$$

and the corresponding Fisher information

$$\gamma_\rho(A, B) := \int_0^\infty \text{Tr} A(\rho + t)^{-1} B(\rho + t)^{-1} dt$$

is named after Kubo, Mori, Bogoliubov etc. In this case

$$\mathbb{J}^{-1}(B) = \int_0^\infty (\rho + t)^{-1} B(\rho + t)^{-1} dt \quad \text{and} \quad \mathbb{J}(A) = \int_0^1 \rho^t A \rho^{1-t} dt .$$

Therefore the corresponding quadratic cost functional is

$$\varphi_\rho[A, B] = \int_0^1 \text{Tr} A \rho^t B \rho^{1-t} dt .$$
Fisher information matrix

Let $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$ be a smooth $m$-dimensional manifold of invertible density matrices. When a quadratic cost function $\varphi_0$ is fixed, the corresponding Fisher information is a Riemannian metric on the manifold.

Fisher information appears not only as a Riemannian metric but as an information matrix as well. The quantum score operators (or logarithmic derivatives) are defined as

$$L_i(\theta) := J_{\rho(\theta)}^{-1}(\partial_{\theta_i}\rho(\theta)) \quad (1 \leq i \leq m)$$

and

$$I_{ij}^Q(\theta) := Tr L_i(\theta) J_{\rho(\theta)}(L_j(\theta)) \quad (1 \leq i, j \leq m)$$

is the quantum Fisher information matrix.
Monotonicity

**Theorem 5** Let $\beta$ be a coarse-graining sending density matrices on the Hilbert space $\mathcal{H}_1$ into those acting on the Hilbert space $\mathcal{H}_2$ and let $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$ be a smooth $m$-dimensional manifold of invertible density matrices on $\mathcal{H}_1$. For the Fisher information matrix $I^{1Q}(\theta)$ of $\mathcal{M}$ and for Fisher information matrix $I^{2Q}(\theta)$ of $\beta(\mathcal{M}) := \{\beta(\rho(\theta)) : \theta \in G\}$ we have the monotonicity relation

$$I^{2Q}(\theta) \leq I^{1Q}(\theta).$$

Particular case: LHS is classical (due to a measurement), RHS is quantum. The later is minimal for the symmetric logarithmic derivative.
Conjecture

The Kubo-Mori (or Bogoliubov) inner product is given by

\[ \gamma_\rho(A, B) = \text{Tr}(\partial_A \rho)(\partial_B \log \rho). \]

Conjecture: The scalar curvature is monotone under coarse graining.

Another form of the conjecture is to consider the scalar curvature along curves of Gibbs states:

\[ \frac{e^{-\beta H}}{\text{Tr}e^{-\beta H}} \]

Conjecture: The scalar curvature is monotone decreasing function of \( \beta \).
Von Neumann algebras

Let $\mathcal{M}$ be a von Neumann algebra. Assume that it is in standard form, it acts on a Hilbert space $\mathcal{H}$, $\mathcal{P} \subset \mathcal{H}$ is the positive cone and $J : \mathcal{H} \rightarrow \mathcal{H}$ is the modular conjugation.

Let $\varphi$ and $\omega$ be normal states with representing vectors $\Phi$ and $\Omega$ in the positive cone. For the sake of simplicity, assume that $\varphi$ and $\omega$ are faithful. This means that $\Phi$ and $\Omega$ are cyclic and separating vectors.

The closure of the unbounded operator $A\Omega \mapsto A^*\Phi$ has a polar decomposition $J\Delta(\varphi, \omega)^{1/2}$ and $\Delta(\varphi, \omega)$ is called relative modular operator. $A\Omega$ is in the domain of $\Delta(\varphi, \omega)^{1/2}$ for every $A \in \mathcal{M}$. 
Quasi-entropy

For $A \in \mathcal{M}$ and $F : \mathbb{R}^+ \to \mathbb{R}$, the quasi-entropy is

$$S_F^A(\omega, \varphi) := \langle A\Omega, F(\Delta(\varphi, \omega))A\Omega \rangle.$$  

(The right-hand-side can be understood via the spectral decomposition of the positive operator $\Delta(\varphi, \omega)$.)

**Theorem 6** Assume that $F : \mathbb{R}^+ \to \mathbb{R}$ is an operator monotone function with $F(0) \geq 0$ and $\alpha : \mathcal{M}_0 \to \mathcal{M}$ is a unital normal Schwarz mapping. Then

$$S_F^A(\omega \circ \alpha, \varphi \circ \alpha) \geq S_F^{\alpha(A)}(\omega, \varphi)$$

holds for $A \in \mathcal{M}_0$ and for normal states $\omega$ and $\varphi$ of the von Neumann algebra $\mathcal{M}$. 
Quadratic cost

The natural extension of the covariance (from probability theory) is

$$qCov^f_\omega(A, B) = \langle \sqrt{f(\Delta(\omega, \omega))}A\Omega, \sqrt{f(\Delta(\omega, \omega))}B\Omega \rangle - \omega(A)\omega(B),$$

where $\Delta(\omega, \omega)$ is actually the modular operator. Motivated by the application, we always assume that the function $f$ is standard. For such a function $f$, the inequalities

$$\frac{2x}{x + 1} \leq f(x) \leq \frac{1 + x}{2}$$

holds. Therefore $A\Omega$ is in the domain of $\sqrt{f(\Delta(\omega, \omega))}$ and the covariance $qCov^f_\omega(A, B)$ is a well-defined sesquilinear form.
Skew information

For a standard function $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ and for a normal unital Schwarz mapping $\beta : \mathcal{N} \rightarrow \mathcal{M}$ the inequality

$$q\text{Cov}^f_\omega(\beta(X), \beta(X)) \leq q\text{Cov}^f_{\omega \circ \beta}(X, X) \quad (X \in \mathcal{N})$$

is a particular case of Theorem 6 and it is the monotonicity of the generalized covariance under coarse-graining.

Following Hansen, the skew information (as a bilinear form) can be defined as

$$I^f_\omega(X, Y) := \text{Cov}_\omega(X, Y) - q\text{Cov}^f_\omega(X, Y)$$

if $\omega(X) = \omega(Y) = 0$. (Then $I^f_\omega(X) = I^f_\omega(X, X)$.)
Theorem 7  Assume that $f, g : \mathbb{R}^+ \to \mathbb{R}$ are standard functions and $\omega$ is a faithful normal state on a von Neumann algebra $\mathcal{M}$. Let $A_1, A_2, \ldots, A_m \in \mathcal{M}$ be self-adjoint operators such that $\omega(A_1) = \omega(A_2) = \ldots = \omega(A_m) = 0$. Then the determinant inequality

$$\text{Det} \left( \left[ q \text{Cov}_D^g (A_i, A_j) \right]_{i,j=1}^m \right) \geq \text{Det} \left( \left[ 2g(0) I_\omega^f (A_i, A_j) \right]_{i,j=1}^m \right)$$

holds.
Proof for $m = 1$

Let $E(\cdot)$ be the spectral measure of $\Delta(\omega, \omega)$. Then for $m = 1$ the inequality is

$$\int g(\lambda)\,d\mu(\lambda) \leq g(0) \left( \int \frac{1 + \lambda}{2}\,d\mu(\lambda) - \int \tilde{f}(\lambda)\,d\mu(\lambda) \right),$$

where $d\mu(\lambda) = d\langle A\Omega, E(\lambda)A\Omega \rangle$. Since the inequality

$$f(x)g(x) \geq f(0)g(0)(x - 1)^2$$

holds for standard functions, we have

$$g(\lambda) \geq g(0) \left( \frac{1 + \lambda}{2} - f(0)\tilde{f}(\lambda) \right)$$

and this implies the integral inequality.
Proof for \( m \)

Consider the finite dimensional subspace \( \mathcal{N} \) generated by the operators \( A_1, A_2, \ldots, A_m \). On \( \mathcal{N} \) we have the inner products

\[
\langle A, B \rangle := \text{Cov}_\omega^g(A, B)
\]

and

\[
\langle A, B \rangle := 2g(0)I^f_\omega(A, B).
\]

Since \( \langle A, A \rangle \leq \langle A, A \rangle \), the determinant inequality holds. \( \square \)

This theorem is interpreted as quantum uncertainty principle. In the earlier works the function \( g \) from the left-hand-side was

\[(x + 1)/2\]

and the proofs were more complicated.
References

Recent references

- P. Gibilisco, F. Hiai and D. Petz, Quantum covariance, quantum Fisher information and the uncertainty principle, to be published

- D. Petz and V.E. S. Szabó, From quasi-entropy to skew information, to be published

Information and Communication

e-mail: infocom@renyi.hu

Scientific committee: Rudolph Ahlswede, Charles H. Bennett, Thomas M. Cover, Imre Csiszár, Dénès Petz, Andreas Winter

Local organizers: Villó Csiszár, György Michaletzky, László Györfi, Dénès Petz


The conference is an international meeting bringing together mathematicians, computer theorists, mathematical physicists, engineers and scientists in related fields: Algorithmic information theory; channel coding; coded modulation; complexity, information and entropy; complexity measures; convolutional coding; error-correcting codes; information theory and statistics; iterative decoding; quantum information theory; quantum-theoretical aspects of coding; randomness and pseudo-randomness; relationships between codes and complexity; rate distortion theory; soft-decision decoding; source coding; source-channel coding; turbo codes.