Superadditivity of Fisher Information: Classical vs. Quantum

Shunlong Luo

Academy of Mathematics and Systems Science
Chinese Academy of Sciences
luosl@amt.ac.cn

Information Geometry and its Applications IV
in honor of Shun-Ichi Amari on the Occasion of His 80th Birthday

June 13-17, 2016, Liblice, Czech

A story about

- a conjecture of more than 50 years old
- strange difference between classical and quantum statistics
- Implications for clock synchronization

Outline

1. Classical Fisher Information

- Fisher, 1922, 1925

Fisher information of a probability density $p(x) = p(x_1, x_2, \cdots, x_n)$ (with respect to the location parameters) is defined as

$$I_{\mathrm{F}}(p) = 4 \int_{R^n} |\nabla \sqrt{p(x)}|^2 dx.$$

∇: gradient

$|\cdot|$: Euclidean norm in R^n

More generally, the Fisher information matrix of a parametric densities $p_\theta(x)$ on R^n with parameter $\theta = (\theta_1, \theta_2, \cdots, \theta_m) \in R^m$ is the $m \times m$ matrix

$$\mathbf{I}_F(p_\theta) = (I_{ij})$$

defined as

$$I_{ij} = 4 \int_{R^n} \frac{\partial \sqrt{p_\theta(x)}}{\partial \theta_i} \frac{\partial \sqrt{p_\theta(x)}}{\partial \theta_j} dx$$

with $i, j = 1, 2, \cdots, m$.

In particular, if $n = m$ and $p_\theta(x) = p(x - \theta)$ is a translation family, then $\mathbf{I}_F(p_\theta) = (I_{ij})$ is independent of the parameter θ, and

$$I_{ij} = 4 \int_{R^n} \frac{\partial \sqrt{p(x)}}{\partial x_i} \frac{\partial \sqrt{p(x)}}{\partial x_j} dx.$$

In this case, we may simply denote $\mathbf{I}_{\mathrm{F}}(p_\theta)$ by $\mathbf{I}_{\mathrm{F}}(p)$. We see that

$$I_{\mathrm{F}}(p) = \mathrm{tr}\mathbf{I}_{\mathrm{F}}(p).$$

Statistical Origin of Fisher Information

Data: n samples $x_1, x_2, \cdots, x_n \sim p_\theta(x)$.

Aim: Estimate the parameter θ.

- Cramér-Rao: Unbiased estimate $\widehat{\theta}$

$$\Delta\widehat{\theta} \geq \frac{1}{nI(p_\theta)}.$$

- Maximum Likelihood: $\widehat{\theta}(x_1, \cdots, x_n)$

$$\sqrt{n}(\widehat{\theta} - \theta) \to N(0, 1/I(p_\theta)).$$

Fisher Information vs. Shannon Entropy

- For a probability density p, its Shannon entropy is $S(p) = -\int p(x) \ln p(x) dx$.
- de Bruijin identity:

$$\frac{\partial}{\partial t} S(p * g_t)\Big|_{t=0} = \frac{1}{2} I(p),$$

where $g_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$.

Basic Properties of Fisher Information

(a). Fisher information is convex:

$$I_{\mathrm{F}}(\lambda_1 p_1 + \lambda_2 p_2) \leq \lambda_1 I_{\mathrm{F}}(p_1) + \lambda_2 I_{\mathrm{F}}(p_2).$$

Here p_1 and p_2 are two probability densities and $\lambda_1 + \lambda_2 = 1,\ \lambda_j \geq 0,\ j = 1, 2.$

Informational meaning: Mixing decreases information.

(b). Fisher information is additive:

$$I_{\mathrm{F}}(p_1 \otimes p_2) = I_{\mathrm{F}}(p_1) + I_{\mathrm{F}}(p_2).$$

Here p_1 and p_2 are two probability densities, and $p_1 \otimes p_2(x, y) := p_1(x)p_2(y)$ is the independent product density (which is a kind of tensor product).

(c). Fisher information is invariant under location translation, that is, for any fixed $y \in R^n$, if we put $p_y(x) := p(x-y)$, then $I_{\mathrm{F}}(p_y) = I_{\mathrm{F}}(p)$.

Superadditivity

(d). Fisher information $I_{\mathrm{F}}(p)$ is superadditive:

$$I_{\mathrm{F}}(p) \geq I_{\mathrm{F}}(p_1) + I_{\mathrm{F}}(p_2).$$

Here $p(x) = p(x_1, x_2)$ is a bivariate density with marginal densities p_1 and p_2.

Amusing and Remarkable

1925: Fisher information was introduced.
1991: Superadditivity was discovered and proved by Carlen.

Statistical meaning:
When a composite system is decomposed into two subsystems, the correlation between them is missing, and thus the Fisher information decreases.

Superadditivity

•Analytical Proof

E. A. Carlen

Superadditivity of Fisher's information and logarithmic Sobolev inequalities

Journal of Functional Analysis, 101 (1991), 194-211.

- Statistical Proof

A. Kagan and Z. Landsman
Statistical meaning of Carlen's superadditivity of the Fisher information
Statist. Probab. Lett. 32 (1997), 175-179.

3. Quantum Fisher Information

Analogy between Classical and Quantum:

•Probability p_θ $\longrightarrow$ Density operator (non-negative matrix with unit trace) ρ_θ

•Integral $\int$ $\longrightarrow$ Trace operation tr

Quantum Mechanics as a Framework of Calculating Probabilities, a Statistical Theory

E. Schrödinger

Quantum mechanics began with statistics, and will end with statistics.

• In classical statistics, probabilities are given *a priori*: $(\Omega, \mathcal{F}, P)$.

• In quantum physics, probabilities are generated from the pairing:
(density operators ρ, observable H)

$$p_i = \mathrm{tr} \rho E_i$$

where $H = \sum_i \lambda_i E_i$ is the spectral decomposition of the self-adjoint operator H.

- H. Araki, M. M. Yanase

Measurement of Quantum Mechanical Operators

Phys. Rev. 120, 1960

Wigner-Araki-Yanase Theorem

The existence of a conservation law imposes limitation on the measurement of an observable. An operator which does not commute with a conserved quantity cannot be measured exactly.

• E. P. Wigner and M. M. Yanase
Information content of distribution
Proc. Nat. Acad. Sci., 49, 910-918 (1963)
Skew information

$$I(\rho, H) = -\frac{1}{2}\mathrm{tr}[\sqrt{\rho}, H]^2$$

where
ρ: density operator
H: any self-adjoint operator
$[\cdot, \cdot]$: commutator

- Wigner-Yanase-Dyson information

$$I_\alpha(\rho, H) = -\frac{1}{2}\mathrm{tr}[\rho^\alpha, H][\rho^{1-\alpha}, H]$$

where $\alpha \in (0, 1)$.

Basic Properties of Skew Information

1. $I(\rho, H) \leq \Delta_\rho H := \mathrm{tr}\rho H^2 - (\mathrm{tr}\rho H)^2$.
2. Invariance: $I(U\rho U^\dagger, H) = I(\rho, H)$ if unitary U satisfying $UH = HU$.
3. Additivity

$$I(\rho_1 \otimes \rho_2, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2)$$

$$= I(\rho_1, H_1) + I(\rho_2, H_2).$$

4. Convexity

$$I(\lambda_1\rho_1+\lambda_2\rho_2, H) \leq \lambda_1 I(\rho_1, H)+\lambda_2 I(\rho_2, H).$$

Four Interpretations of Skew Information

- As information content of ρ with respect to observable not commuting with H

Wigner and Yanase, 1963

• As a measure of non-commutativity between ρ and H

Connes, Stormer, J. Func. Anal. 1978

- As a kind of quantum Fisher information

D. Petz, H. Hasegawa, On the Riemannian metric of α-entropies of density matrices, Lett. Math. Phys. 1996

S. Luo
Phys. Rev. Lett. 2003
IEEE Trans. Inform. Theory, 2004
Proc. Amer. Math. Soc. 2004

• As the quantum uncertainty of H in the state ρ

S. Luo, Phys. Rev. A, 2005, 2006

Skew Information as Quantum Fisher Information

Generalizing classical Fisher information

$$I_F(p_\theta) := 4\int\left(\frac{\partial\sqrt{p_\theta(x)}}{\partial\theta}\right)^2 dx$$

to the quantum scenario, we may define

$$I_F(\rho_\theta) := 4\mathrm{tr}\left(\frac{\partial\sqrt{\rho_\theta}}{\partial\theta}\right)^2$$

as a kind of quantum Fisher information.
Here ρ_θ is a family of density operators.

In particular, if ρ_θ satisfies the Landau-von Neumann equation

$$i\frac{\partial \rho_\theta}{\partial \theta} = [H, \rho_\theta], \qquad \rho_0 = \rho$$

then

$$I_F(\rho_\theta) = -4\mathrm{tr}[\rho^{1/2}, H]^2 = 8I(\rho, H)$$

S. Luo, Phys. Rev. Lett. 2003

4. Superadditivity in Quantum case

Conjecture: For bipartite density operator ρ,

$$I_\alpha(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2).$$

Here

$\rho_1 = \mathrm{tr}_2\rho$, $\rho_2 = \mathrm{tr}_1\rho$: marginals of ρ

H_1, H_2: selfadjoint operators over subsystems

$\mathbf{1}$: identity operator

$\otimes$: tensor product of operators

Comments

This conjecture was reviewed by Lieb. The only non-trivial confirmed case is for pure states with $\alpha = \frac{1}{2}$.

Wigner-Yanase, 1963: Necessary requirement
Lieb, 1973: Absolute requirement

Disproof
F. Hansen, Journal of Statistical Physics, 2007

Numerical counterexample!

Counterintuitive!

Surprising!

A Simple Counterexample. Let $n > 2$ and take

$$\rho = \frac{1}{n}\begin{pmatrix} n-2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, H_1 = H_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$I(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) < I(\rho_1, H_1) + I(\rho_2, H_2)$$

for large n.

Partial Results
S. Luo, Journal of Statistical Physics, 2007

- Let $H = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2$. If $\rho = |\Psi\rangle\langle\Psi|$ is a pure state, then superadditivity holds, that is

$$I_\alpha(\rho, H) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2).$$

- Let $H = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2$, and ρ be a diagonal density matrix. Then superadditivity holds, that is

$$I_\alpha(\rho, H) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2).$$

Partial Results
S. Luo and Q. Zhang
Journal of Statistical Physics, 2008

- For any classical-quantum state, the superadditivity holds.

Tripartite case
R. Seiringer
Lett. Math. Phys. 2007

Failure of superadditivity of the Wigner-Yanase skew information for tripartite pure states. The following inequality may be violated by certain pure states:

$$I_\alpha(\rho, H) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2) + I_\alpha(\rho_3, H_3)$$

where $\rho = |\Psi_{123}\rangle\langle\Psi_{123}|$,
$H = H_1 \otimes \mathbf{1}_2 \otimes \mathbf{1}_3 + \mathbf{1}_1 \otimes H_2 \otimes \mathbf{1}_3 + \mathbf{1}_1 \otimes \mathbf{1}_2 \otimes H_3.$

5. Weak Superadditivity in Quantum Case

- Though neither

$$I(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) \geq I(\rho_1, H_1) + I(\rho_2, H_2)$$

nor

$$I(\rho, H_1 \otimes \mathbf{1} - \mathbf{1} \otimes H_2) \geq I(\rho_1, H_1) + I(\rho_2, H_2)$$

is always true, their sum is true:

$$I(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) + I(\rho, H_1 \otimes \mathbf{1} - \mathbf{1} \otimes H_2) \geq 2\Big(I(\rho_1, H_1) + I(\rho_2, H_2)\Big).$$

- It holds that

$$I(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) \geq \frac{1}{2}\Big(I(\rho_1, H_1) + I(\rho_2, H_2)\Big).$$

6. Physical Implications of Superadditivity: Clock Synchronization

- Classical clock: (p, Q) ($Q = -i\frac{d}{dx}$ is the moment observable)

$$p_t(x) = e^{-itQ}p(x).$$

Quality: classical Fisher information $I_F(p)$.

- Quantum clock: (ρ, H)

$$\rho_t = e^{-itH}\rho e^{itH}.$$

Quality: quantum Fisher information $I_\alpha(\rho, H)$.

Clock Synchronization

- A quantum clock shared by two parties:

$$(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2)$$

The violation of the superadditivity means that the sum of the quality of the component clock will be better than the overall quality:

$$I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2) > I_\alpha(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2).$$

- Curious property of quantum clock!

A Conjecture

There does not exit nontrivial quantum clocks such that

$$I_\alpha(\rho_1, H_1) = I_\alpha(\rho_2, H_2) = I_\alpha(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2).$$

Intuition: Otherwise, we could copy quantum timing information.

7. Problems

1. Conditions for superadditivity?
2. Intuitive meaning of the failure of superadditivity
3. Difference between classical and quantum from the perspective of Fisher information
4. Quantum logarithmic Sobolev inequalities?

Thank you!