数理統計学から見たThermo-Majorization: 統計モデルの比較と情報スペクトル

National Institute of Informatics
Keiji Matsumoto

Thermo Majorization

ρ_G: (Gibbs) state

$[\rho, \rho_G] = 0, \ [\sigma, \rho_G] = 0$

The transition

$$\{\rho, \rho_G\} \rightarrow \{\sigma, \rho_G\}$$

is possible

$\Longleftrightarrow$ $(\rho_{ii}/\rho_{G,ii})_i$ majorizes $(\sigma_{ii}/\rho_{G,ii})_i$

$$\sum_{i=1}^{j} (\rho_{ii}/\rho_{G,ii}) \geq \sum_{i=1}^{j} (\sigma_{ii}/\rho_{G,ii}), \forall j$$

$$\Longleftrightarrow \quad \mathrm{tr}\,(\rho - t\rho_G)_+ \geq \mathrm{tr}\,(\sigma - t\rho_G)_+, \forall t \geq 0$$

$(A)_+$: positive part of the matrix A

Thermo-Majorization : History

The term first appears in Horodecki, M + Oppenheim, J (2014)

But had been known since decades ago:

Mathematical Physics : mapping btw. state families
Uhlmann, Alberiti, Ruch, Schramer, etc …

- Mathematical Statistics : comparison of "statistical experments"
  Blackwell, Le Cam, Hajek, Torgersen, etc …

  Why statisticians are bothered with conversion btw state families ?

Information Spectrum

A framework of information theory free of stochastic assumptions such as IID, Markov, etc

Classical version : invented by Te-Sun Han (韓太舜),
with Shannon Award
Quantum version : Hiroshi Nagaoka (UEC)

$$\overline{D}(\{\rho^n\}||\{\sigma^n\}) = \inf\{c;\ \lim_{n\to\infty} \mathrm{tr}(\rho^n - e^{nc}\sigma^n)_+ = 0\}$$

$$\underline{D}(\{\rho^n\}||\{\sigma^n\}) = \sup\{c;\ \lim_{n\to\infty} \mathrm{tr}(\rho^n - e^{nc}\sigma^n)_+ = 1\}$$

(version by Bowen-Datta)

c.f. Renner, Datta, Bowen give alternative representation of these by Smooth Renyi entropy

Computer science, randomness extraction, privacy amplification

Thermo-Majorization

$$\mathrm{tr}\,(\rho - t\rho_G)_+ \geq \mathrm{tr}\,(\sigma - t\rho_G)_+, \forall t \geq 0$$

Information Spectrum

$$\overline{D}(\{\rho^n\}||\{\sigma^n\}) = \inf\{b;\ \lim_{n\to\infty} \mathrm{tr}\big(\rho^n - e^{nb}\sigma^n\big)_+ = 0\}$$

$$\underline{D}\,(\{\rho^n\}||\{\sigma^n\}) = \sup\{b;\ \lim_{n\to\infty} \mathrm{tr}\big(\rho^n - e^{nb}\sigma^n\big)_+ = 1\}$$

If you are not xxxxxx,

Should notice some relations btw these....

Plan of the talk:

Comparison of statistical experiments

- general theory, historical contexts
- 2-states case
- relation to information measures
- quantum states

Information spectrum (mainly hypothesis test)

- upper & lower divergence rate (classical, quantum)
- hypothesis test
  (- resolvability)
- relation to smooth Renyi entropies

Asymptotic 2-states conversions

- sufficient condition by information spectrum
- characterization of quantum relative entropy (KL-divergence)

Comparison of statistical experiments

References

Blackwell, D. and Girshick, M. A.
Theory of Games and Statistical Decisions (1954)

LeCam, L. (many works since 60’s)
Asymptotic Methods in Statistical Decision Theory, Springer (1986)

Torgersen, E. *Comparison of Statistical Experiments,*
Cambridge University Press(1991)

Goel P K   (concise and comprehensive review)
“When is one experiment ‘always better than another?”
The Statistician 52 , Part 4 , pp. 515–537 (2003)

Statistical decision problems

$x \in \mathrm{X}$: (array of) data $d(x) \in \mathrm{D}$: decision

x obeys probability distribution $p_\theta(x)$, $\theta \in \Theta$ unknown

$g_\theta(d) \in \boldsymbol{R}$: the gain of the decision d when the data source is p_θ

X 等は各自分かりやすいので考えて
（理論は極限まで一般化されてるが気にしないで）

$\mathrm{E} = \{ p_\theta ; \theta \in \Theta \}$ is called an experiment or a statistical model

Optimize d to maximize the expectation

$$G_\theta(d, \mathrm{E}) := E_\theta\, g_\theta(d(X)) = \sum_{x \in \mathrm{X}} g_\theta\big(d(x)\big) p_\theta(x)$$

under various setting min-max/average about θ/constrained

Some notes

Settings are fairly general

X : measurable space D : topological space, with Bair σ-filed,
d: measureable
P_θ: prob measure, may not be majorized by a common measure
Θ: a set

g is usually subject to some reasonable conditions

g_θ: upper-semi continuous/continuous
+ bounded from above/bouded

In general, randomized decision is also allowed

ここでは記述が面倒だから避けるが、どっか破たんしてたらごめんなさい

Usually,
“loss “ is minimized, rather than “gain” is maximized

大人の事情があるので、ここではゲイン。

Example of Statistical Decisions

Ex 1

x : past/present data about economics

θ : parameters determining the market price

d : your strategy of selling/buying

$g_\theta(d)$: how much you earn

Ex 2

$\theta \in \boldsymbol{R}^k$

d : your estimate of ϑ

$g_\theta(d) = -||d - \theta||^2$: how close was your guess

Ex 3

$\theta \in \{0,1\}$

d : your guess about $\theta \in \{0,1\}$

$g_\theta(d) = 1$ if you are right, otherwise =0

On arbitrariness of g

Ex 2 $\theta \in \boldsymbol{R}^k$

d : your estimate about the true value of parameter

$g_\theta(d) = -||d - \theta||^2$: how close was your guess

No overwhelming reason to prefer square-error to other error measures

Statistical Decision Problems : Optimization ?

Life is so hard that no decision is uniformly optimal usually :

if $G_\theta(d,\mathrm{E}) := \sum_x g_\theta(x)p_\theta(x)$ very good, $G_{\theta'}(d,\mathrm{E})$ is poor

Ex 3 $\theta \in \{0,1\}$ d : your guess about $\theta \in \{0,1\}$
$g_\theta(d) = 1$ if you are right, otherwise =0

If $d(x) = 0$ irrespective of x, $G_0(d,\mathrm{E}) = 1$ but $G_1(d,\mathrm{E}) = 0$

1 Maximize the average about θ (Bayesian gain):

$$G_\pi(d,\mathrm{E}) := \sum_\theta \pi_\theta G_\theta(d,\mathrm{E})$$

2 Maximize the worst case about θ (min-max):

$$G(d,\mathrm{E}) = \inf_{\theta\in\Theta} G_\theta(d,\mathrm{E})$$

3 Maximize e.g. $G_{\theta_1}(d,\mathrm{E})$, subject to the constrain, e.g.

$$G_\theta(d,\mathrm{E}) \geq c, \forall\theta \neq \theta_1$$

And so on …

"optimal decision" ?

- depends on G
- depends on min-max or Bayesian gain, or others

Life is so hard.

You have to solve them one by one.

But, there is a hope ....

➡ "Comparison of statistical experiments"

- asymptotic theory 存在感こっちのほうが圧倒的で

  LAN, bounds on cloning etc
- experimental design

でも説明しやすいのはこっちなので・・・

Experimental design

$$\mathrm{E}_F = \{F(f_\theta); \theta \in \Theta\}$$

Given $\{f_\theta; \theta \in \Theta\}$, can move F (experimental design)

Nice if one can say $\{F(f_\theta); \theta \in \Theta\}$ is more informative than the other

Ex

$$x = Ff_\theta + \xi, \qquad \xi \sim N(0, \sigma I_k)$$

f_θ: l-dim vector, x: k-dim vector, k$\leqq$l, F: k$\times$l real matrix

Optimize F obtain most "informative" $p_\theta(x) = N(Ff_\theta, \sigma I_k)$

If $N(Ff_\theta, \sigma I_k)$ is better than $N(F' f_\theta, \sigma I_k)$, use F rather than F'

Comparison of statistical experiments

$e=(e_\theta)_{\theta\in\Theta}$ （普通はεを使うが、εほかのところでも使うので）

DEF E is e- defficinet relative to E', $\mathrm{E}\geq_e \mathrm{E}'$

$\Leftrightarrow$ $\forall d' \exists d\ \ \forall\theta\in\Theta\ \ G_\theta(d,\mathrm{E})\geq G_\theta\,(d',\mathrm{E}')-\frac{1}{2}e_\theta$

holds for any g with $0\leq g_\theta\leq 1$, on any D

どんなE’上のdecision d’に対しても、それを上回るE上のdecision dがある（e程度のマージンで）。しかも、それが任意の決定空間と利得で。

$\mathrm{E}\sim_e \mathrm{E}'\ \Leftrightarrow \mathrm{E}\geq_e \mathrm{E}'\ \&\ \ \mathrm{E}\leq_e \mathrm{E}'$ E’で最適化考えても Eと大体同じ

If $e_\theta=0, \forall\theta\in\Theta$, $\mathrm{E}\geq \mathrm{E}'$ $\mathrm{E}\sim \mathrm{E}'$

“$\sim$” is an equivalence relation

Distance measure $\Delta(\mathrm{E},\mathrm{E}'):=\min\{\epsilon\,;\mathrm{E}\sim_e \mathrm{E}', e_\theta=\epsilon\}$

“≧” is partial order

Randomization criteria (Blackwell, Le Cam)

Thm $\mathrm{E} \geq_e \mathrm{E}'$

$\Leftrightarrow$ $\exists\Lambda$: transition probability

$$\forall\theta \in \Theta, \quad \|\Lambda(p_\theta) - p'_\theta\|_1 \leq e_\theta$$

$\Leftrightarrow$

$$\sup_d \sum_{\theta\in\Theta} \pi_\theta \, G_\theta(d, \mathrm{E}) \geq \sup_d \sum_{\theta\in\Theta} (\pi_\theta \, G_\theta(d, \mathrm{E}') - \frac{1}{2} e_\theta)$$

holds for any π_θ (non-zero at most finitely many points)

for any g with $0 \leq g_\theta \leq 1$, on any D

This is the reason why statisticians are interested in thermo-majorization-type conditions

Ex. Comparison of linear normal experiments

Ex $x = Ff_\theta + \xi, \qquad \xi \sim N(0, \sigma I_k)$

f_θ: l-dim vector, x: k-dim vector, k≦l, F: k×l real matrix

Hansen, O. and Torgersen, E. : *The Annals of Statistics* 1974, Vol. 2, No. 2, 367-373

$$\{N(Ff_\theta, \sigma I_k); \theta \in \Theta\} \geq \{N(F'f_\theta, \sigma I_k); \theta \in \Theta\}$$

$$\Leftrightarrow \quad FF^T - F'F'^T \geq 0$$

(if σ is unknown, $k - k' - \mathrm{rank}(FF^T - F'F'^T) \geq 0$, in addition)

many extensions,e.g., about covariance matrix, also computation of Δ

Quantum

[M 2010]

ρ_θ: Gaussian state $E_\theta(\mathrm{tr}\rho_\theta X_1, \mathrm{tr}\,\rho_\theta P_1, \cdots)) = Ff_\theta,$

covariance =Σ J:matrix defining CCR

$$\{\rho_\theta; \theta \in \Theta\} \geq^q \{\rho_\theta'; \theta \in \Theta\}$$

$$\Leftrightarrow \quad \exists C, S, \Sigma_\omega \quad C^T JF = JF', \quad \Sigma' = S^T \Sigma_\omega S + C^T \Sigma C, \quad \Sigma_\omega + \mathrm{i}J \geq 0$$

Ex Statistical Inference on Unital qubit channels

Λ_θ: a unital channel with the unknown parameter θ $\Lambda_\theta(I) = I$

F: input state + measurement

Many papers, each dealing with each own setting

[Fujiwara02] [Fujiwara03][Sacchi05-1][Sacchi05-2][Sacchi05-3] etc

But all of them say maximally entangled state is optimal

Def $\{\rho_\theta\}_{\theta\in\Theta} \geq^c \{\rho'_\theta\}_{\theta\in\Theta}$

$$\Leftrightarrow \quad \forall M' \exists M \; \{P^M_{\rho_\theta}\}_{\theta\in\Theta} \geq \{P^{M}_{\rho_\theta'}\}_{\theta\in\Theta}$$

Fact [M2013]

Φ: max-ent state

$$\forall\rho \quad \{\Lambda_\theta \otimes I(\Phi)\}_{\theta\in\Theta} \geq^c \{\Lambda_\theta \otimes I(\rho)\}_{\theta\in\Theta}$$

Local asymptotic normality (LAN) $\boldsymbol{\theta} \in \boldsymbol{R}$

- Once upon a time, A. Fisher insisted that maximally likelihood estimate （MLE,最尤推定量) should be the best estimate, and gave a proof, using “≒”, “～” , which would satisfy most of us, but not mathematicians

- Counter example found out, even for Gaussian distributions (Hodge estimator etc)
- Introduction of regularity conditions, on models and estimators
- To simplify the argument, invented was LAN …

$$\log p^n_{\theta+\frac{h}{\sqrt{n}}}(x^n) - \log p^n_\theta(x^n) \sim (J_\theta)^{-1} \dot{\ell}^n_\theta(x^n)\, h + h^2 J_\theta^{-1}$$

If $\theta \to \sqrt{p_\theta(\cdot)}$ has the first derivative (as a function onto L^2)

Local asymptotic normality (LAN) $\boldsymbol{\theta} \in \boldsymbol{R}$

$$\mathrm{E}_\theta^n := \left\{ p_{\theta+\frac{h}{\sqrt{n}}}^n (x^n); h \in \boldsymbol{R} \right\} \qquad \mathrm{E}_\theta := \left\{ N\left(h, J_\theta^{-1}\right) \right\}_{h \in \boldsymbol{R}} \qquad \boldsymbol{\theta} \in \boldsymbol{R}$$

Thm

If $\theta \to \sqrt{p_\theta(\cdot)}$ has the first derivative (as a function on L^2)

And some mild conditions on p_θ

$\Delta(\mathrm{E}_\theta^n, \mathrm{E}_\theta) \to 0$, uniformly $\forall \theta$ in arbitray compact subset of Θ

Estimation of θ on $\{p_\theta^n\}$ reduces to the one on $\left\{ N\left(h, J_\theta^{-1}\right) \right\}_{h \in \boldsymbol{R}}$

If you believe x ($\sim N\left(h, J_\theta^{-1}\right)$) is the best estimate of the mean h , should also believe that MLE is the best

反例は全てガウスでもなので、ガウスに落としてから排除するよう条件つける

Shiryaev, Spokoiny, Statistical Experiments and Decisions: Asymptotic Theory, World Scientific 2000

Ibragimov. Khasminskii, Statistical Estimation: Asymptotic Theory, Springer,1981

LANの面白いところ

- First reduce to the problem to the estimation of the mean of Gaussian random variable, which is easy
- Gaussian shift model $\{N(h, J_\theta^{-1})\}_{h\in\boldsymbol{R}^k}$ functions as a "standard form" of parameter family

- when not iid, other "standard forms , such as LAMN etc
- non-parametric

こういった、「問題の帰着」は情報科学だといろんなところに現れる
計算量理論 (comparison of hardness by reduction)
量子情報のエンタングルメントの理論 (resource theory)

$|\boldsymbol{\Theta}| = 2, \ \ \mathbf{E} = \{p_0, p_1\}$

Thm [Torgersen70]

To check $\mathrm{E} \ \geq_e \mathrm{E}'$, only have to check
Binary decision problem, $d \in \mathrm{D} = \{0,1\}$

$$\Leftrightarrow \quad \sup_{d} \textstyle\sum_{\theta\in\Theta} \pi_\theta \, G_\theta(d, \mathrm{E}) \geq \sup_{d'} \textstyle\sum_{\theta\in\Theta} (\pi_\theta \, G_\theta(d', \mathrm{E}') - \frac{1}{2}\pi_\theta \, e_\theta)$$

holds for any π_θ
for any g with $0 \leq g_\theta \leq 1$, on $\mathrm{D} = \{0,1\}$

w.l.g., $\pi_0 = 1, \ \pi_1 = t,$ $\qquad g_1(d) = 1 - g_0(d),$

$|\Theta| = 2, \quad \mathrm{E} = \{p_0, p_1\}$

Thm [Torgersen70]

To check $\mathrm{E} \geq_e \mathrm{E}'$, one only have to check
Binary decision problem, $d \in \mathrm{D} = \{0,1\}$

$$\Leftrightarrow \quad \sup_d \textstyle\sum_{\theta\in\Theta} \pi_\theta\, G_\theta(d, \mathrm{E}) \geq \sup_{d'} \textstyle\sum_{\theta\in\Theta} (\pi_\theta\, G_\theta(d', \mathrm{E}') - \frac{1}{2}\pi_\theta\, e_\theta)$$

holds $\pi_0 = 1, \ \pi_1 = t,$

for any g with $0 \leq g_\theta \leq 1, \quad g_1(d) = 1 - g_0(d),$

$$\sup_d \sum_{\theta\in\Theta} \pi_\theta\, G_\theta(d, \mathrm{E}) = \sup_d \sum_x g_0\big(d(x)\big)[p_0(x) - t\, p_1(x)] + t$$

$$\leq \sum_x \big(p_0(x) - t\, p_1(x)\big)_+ + t$$ "=" if $g_0(0) = g_1(1) = 1$

$|\boldsymbol{\Theta}| = \mathbf{2}, \ \ \mathbf{E} = \{p_0, p_1\}$

Thm [Torgersen70]

To check $\mathrm{E} \ \geq_e \mathrm{E}'$, one only have to check
Binary decision problem, $d \in \mathrm{D} = \{0,1\}$

$$\mathrm{E} \ \geq_e \mathrm{E}' \ \Leftrightarrow$$

$$\forall t \geq 0 \ \sum_x \big(p_0(x) - t\, p_1(x)\big)_+ \ \geq \sum_x \big(p_0{}'(x) - t\, p_1{}'(x)\big)_+ \ + e_0 - te_1$$

Thermo-majorization with error term
When $e_\theta = 0$, thermo-majorization

… had been known since almost half-a-century ago

Relation to f-divergence

$$\sup_{d(\cdot)} \sum_{\theta=0,1} \pi_\theta\, G_\theta(d, \mathrm{E}) = \sup_{d(\cdot)} \sum_{\theta=0,1} \pi_\theta \sum_x g_\theta\big(d(x)\big) p_\theta(x)$$

$$= \sup_{d(\cdot)} \sum_x \sum_{\theta=0,1} \pi_\theta g_\theta\big(d(x)\big)\, p_\theta(x) \; = \sum_x \sup_d \sum_{\theta=0,1} \pi_\theta g_\theta(d) p_\theta(x)$$

$$= \sum_x p_1(x) f\left(\frac{p_0(x)}{p_1(x)}\right) =: D_f\,(p_0||p_1)$$

$$f(z) \coloneqq \sup_d\{\pi_0 g_0(d) z + \pi_1 g_1(d)\}$$

f is convex and lower semi-continuous

Any such f can be written in above form (use Legendre transform)

$f(z) = z\mathrm{log}z$: KL-divergence (or relative entropy)

z^s : Tsalis-Renyi-like quantity

$|1 - z|$: L1-norm

$(1 - tz)_+$: Thermo-Majorization

Relation to f-divergence

Thm

For any g and π, there is f s.t.

$$\sup_{d(\cdot)} \sum_{\theta=0,1} \pi_\theta \, G_\theta(d, \mathrm{E}) = D_f(p_0||p_1) := \sum_x p_1(x) f(p_0(x)/p_1(x))$$

And vice versa !

Any f-divergence is optimized Bayes gain

$$\mathrm{E} \ \geq_e \mathrm{E}' \qquad \Leftrightarrow \qquad D_f(p_0||p_1) \geq D_f(p_0{}'||p_1{}')$$

holds for any lower semicont. convex function f

Lem [CohenKempermanZbaganu98]

Any lower-semicontinuous convex f can be :

$$f(z) = a + bz + \int (z-t)_+ d\mu(t)$$

Combination of above facts leads to another proof of Thermo-Majorization

fact

(i) $D_f(tp_0||tp_1) = tD_f(p_0||p_1)$

(ii) $(p_0, p_1) \to D_f(p_0||p_1)$ is convex

(iii) $D_f(p_0||p_1) \geq D_f(\Lambda(p_0)||\Lambda(p_1))$ Λ: transition probability

Quantum version I

$\mathrm{E} = \{\rho_\theta\}_{\theta\in\Theta}$ $\quad\quad G_\theta(\Lambda, \mathrm{E}) \coloneqq \mathrm{tr} g_\theta \Gamma(\rho_\theta)$

g_θ: bounded self-adjoint, on H_D

Γ: Completely positive trace preserving map

DEF $\mathrm{E} \geq_e^q \mathrm{E}'$ $\;$ E is e- defficinet relative to E',

$$\Leftrightarrow \quad \forall\Gamma'\exists\Gamma \;\; \forall\theta \in \Theta \;\; G_\theta(\Gamma, \mathrm{E}) \geq G_\theta\,(\Gamma', \mathrm{E}') - \frac{1}{2} e_\theta$$

holds for any g with $0 \leq g_\theta \leq I$, on any H_D

Thm [M10] $\quad \mathrm{E} \geq_e \mathrm{E}'$

$\Leftrightarrow$ $\exists\Lambda$: CPTP map

$$\forall\theta \in \Theta, \quad \|\Lambda(p_\theta) - p'_\theta\|_1 \leq e_\theta$$

$$\Leftrightarrow \quad \sup_\Gamma \textstyle\sum_{\theta\in\Theta} \pi_\theta\, G_\theta(\Gamma, \mathrm{E}) \geq \sup_\Gamma \sum_{\theta\in\Theta} (\pi_\theta\, G_\theta(\Gamma, \mathrm{E}') - \frac{1}{2} e_\theta)$$

holds for any π_θ (non-zero at most finitely many points)

for any g with $0 \leq g_\theta \leq I$, on any H_D

Quantum version II

Def

$$\mathrm{E} \geq_e^c \mathrm{E}'$$

$$\Leftrightarrow \quad \forall M' \exists M \ \{P_{\rho_\theta}^M\}_{\theta\in\Theta} \geq_e \ \{P_{\rho_{\theta'}}^M\}_{\theta\in\Theta}$$

Fact

$\mathrm{E} \geq_e^c \mathrm{E}' \ \Leftarrow \quad \exists\Lambda$: positive (may not be CP) trace preserving map

$$\forall\theta \in \Theta, \ \ \|\Lambda(p_\theta) - p'_\theta\|_1 \leq e_\theta$$

$\Rightarrow$ is not true. Counter example by [M13]

No good necessary and sufficient condition for $\geq_e^c$

$$\mathrm{E} \geq_e^c \mathrm{E}' \ \Leftarrow \mathrm{E} \geq_e^q \mathrm{E}'$$

The opposite not true. But if e=0, some good relation.

Thm [Buscemi 12]

$$\forall \mathrm{E}_0 \ \ \mathrm{E} \otimes \mathrm{E}_0 \geq^c \ \mathrm{E}' \otimes \mathrm{E}_0 \Leftrightarrow \ \mathrm{E} \geq^q \ \mathrm{E}'$$

Quantum local asymptotic normality (LAN)

It had been noted that $\rho_\theta^{\otimes n}$ and its tangent space can be approximated by Gaussian states, showing the achievable lower bound to asymptotic error bound of asymptotically unbiased estimators
[M 98] [HayashiM 2002] [Hayashi 2002]

$$\mathrm{E}_\theta^n := \left\{ \rho_{\theta+\frac{h}{\sqrt{n}}}^{\otimes n} ; h \in \boldsymbol{R}^k \right\}$$

E_θ: Gaussian shift (multi-mode)

$h \in \boldsymbol{R}^k$

Thm

[Kahn Guta 2005] $\dim H < \infty$

$$\Delta(\mathrm{E}_\theta^n, \mathrm{E}_\theta) \to 0, \text{ uniformly } \forall\theta$$

$|\boldsymbol{\Theta}| = \mathbf{2}, \ \ \mathbf{E} = \{\rho_0, \rho_1\}$

Thm [AlbertiUhlmann85] Suppose $\dim H = 2$

$$\mathrm{E} \ \geq^q \ \mathrm{E}' \quad \Leftrightarrow \quad \mathrm{E} \ \geq^c \ \mathrm{E}'$$

$$\Leftrightarrow \quad \forall t \geq 0, \ \ \mathrm{tr}(\rho_0 - t\,\rho_1)_+ \geq \ \mathrm{tr}(\rho_0{}' - t\,\rho_1{}')_+$$

[M10] [Jencova10] Suppose $[\rho_0, \rho_1] = 0$

$$\mathrm{E} \ \geq^c_e \mathrm{E}' \quad \Leftrightarrow$$

$$\forall t \geq 0, \ \ \mathrm{tr}(\rho_0 - t\,\rho_1)_+ \geq \ \mathrm{tr}(\rho_0{}' - t\,\rho_1{}')_+ - \frac{1}{2}(e_0 - te_1)$$

Relation to quantum divergence

$D^Q(\rho_0||\rho_1) := \sup_{\Gamma} \sum_{\theta=0,1} \pi_\theta \mathrm{tr} g_\theta \Gamma(\rho_\theta)$ satisfies

(i) $D^Q(t\rho_0||t\rho_1) = D^Q(t\rho_0||t\rho_1)$

(ii) $(\rho_0, \rho_1) \to D^Q(\rho_0||\rho_1)$ is convex

(iii) $D^Q(\rho_0||\rho_1) \geq D^Q(\Lambda(\rho_0)||\Lambda(\rho_1))$, Λ: CPTP

Also, any function satisfying above three is written in the RHS form

If restricted to commutative ops, $D^Q(p_0||p_1) = D_f(p_0||p_1)$, for some f

D^Q: quantum version of f-divergence

$$D_f^{\max}(\rho_0||\rho_1) \geq D^Q(\rho_0||\rho_1) \geq D_f^{\min}(\rho_0||\rho_1),$$

Fact

$\mathrm{E} \geq^q \mathrm{E}' \Leftrightarrow$

$D^Q(\rho_0||\rho_1) \geq D^Q(\rho_0'||\rho_1')$ holds for all D^Q with above conditions

Classical and Quantum Information Spectrum,

Especially on divergence rate

Classical Information Spectrum

Founded by Te-Sun Han (韓太舜)

Novel frame work of information theory :
no need to assume iid, memoryless, Markov, Ergodic etc.

What is done:
Rewrite the solutions of information theoretic problems into "spectrum formulas", which is still abstract.
Thus, if more concrete form is necessary, have to evaluate them further.
Yet, information theoretic part finishes at this stage: the rest of the job is mathematical.

Also, abstract "spectrum formulas" help understanding relations btw various information theoretic problems.

As such, if you want to appreciate its full value, have to learn various aspects of information theory.

Here treat only hypothesis test

Hypothesis test and divergence rate

$\mathrm{E}^n = \{p_0^n, p_1^n\}$: Guess which from the data $x^n \sim p_\theta^n$

$P^n\{1\,|0\}$: Prob of choosing 1 while 0 is true
$P^n\{0|1\}$: Prob of choosing 0 while 1 is true

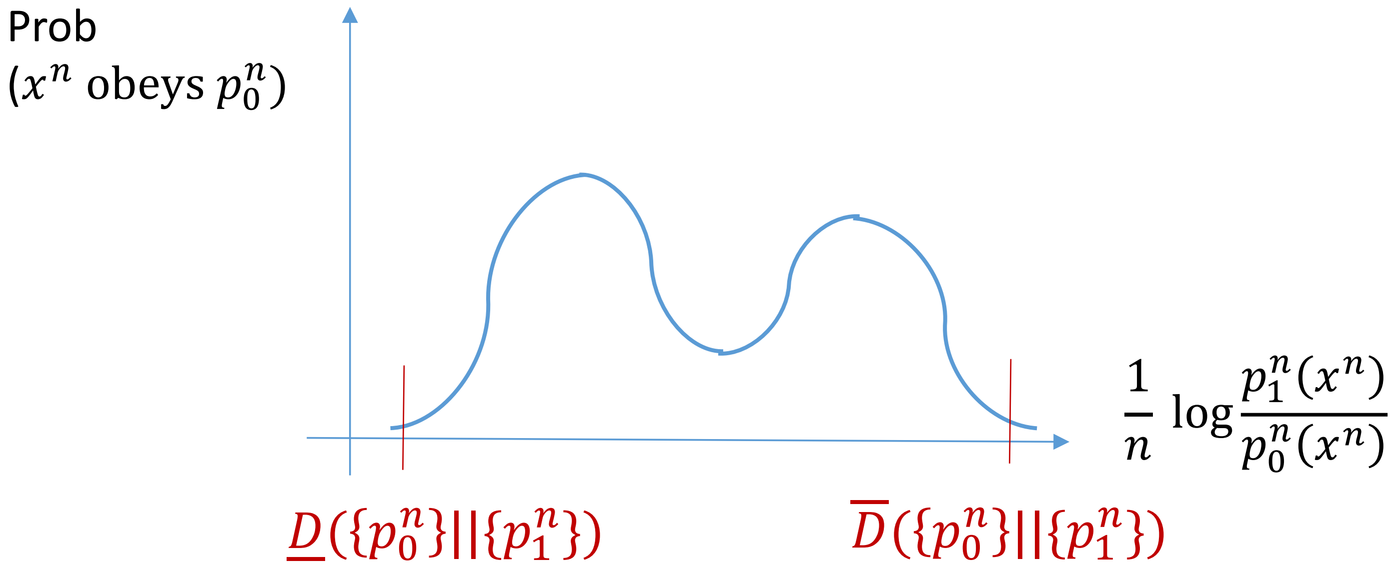
Make $P^n\{0|1\}$ small, while keeping $P^n\{1|0\}$ reasonably small

$e^{-nb} \sim P^n\{0\,|1\}$, $\quad b = \limsup_{n\to\infty} \frac{1}{n} \log P^n\{0{,}1\}$

1. $\sup\{b\,; P^n\{1|0\} \to 0\}$

2. $\sup\{b\,; \limsup P^n\{1|0\} < 1\}$

Divergence rate and hypothesis test



When iid $\frac{1}{n}\sum \log\left(\frac{p_1(X_i)}{p_0(X_i)}\right) \to \sum_x p_0(x)\log\frac{p_0(x)}{p_1(x)} = D(p_0||p_1)$

$$e^{-nb} \sim P^n\{0\,|1\}, \qquad b = \limsup_{n\to\infty}\frac{1}{n}\log P^n\{0,1\}$$

$$\sup\{b\,;P^n\{1|0\} \to 0\} = \underline{D}(\{p_0^n\}||\{p_1^n\})$$

$$\sup\{b\,;\limsup P^n\{1|0\} < 1\} = \overline{D}(\{p_0^n\}||\{p_1^n\})$$

Quantum version

By Hirohoshi Nagaoka（長岡浩司）in about 1998

普通、log p/q のオペレーター版を考えたくなるが、そこにハマらなかった

仮説検定の最適化の途中経過で何がポイントかを考えた。「比の非可換版」

$$\overline{D}(\{\rho_0^n\}||\{\rho_1^n\}) = \inf\{b;\ \lim_{n\to\infty} \mathrm{tr}\left(\rho_0^n - e^{nb}\rho_1^n\right)_+ = 0\}$$

$$\underline{D}\ (\{\rho_0^n\}||\{\rho_1^n\}) = \sup\{b;\ \lim_{n\to\infty} \mathrm{tr}\left(\rho_0^n - e^{nb}\rho_1^n\right)_+ = 1\}$$

1. When commutative, coincide with classical version

2. equals error exponent, just as its classical version

$$\sup\{b\,;P^n\{1|0\} \to 0\} = \underline{D}(\{\rho_0^n\}||\{\rho_1^n\})$$

$$\sup\{b\,;\mathrm{limsup}\ P^n\{1|0\} < 1\} = \overline{D}(\{\rho_0^n\}||\{\rho_1^n\})$$

3. If $\rho_\theta^n = \rho_\theta^{\otimes n}$ both $\overline{D}$ and $\underline{D}$ coincide with $D(\rho_0||\rho_1) := \mathrm{tr}\rho_0\ (\log\rho_0 - \log\rho_1)$

$$D(\rho_0||\rho_1) = \sup\{b\,;P^n\{1|0\} \to 0\} = \sup\{b;\mathrm{limsup}\ P^n\{1|0\} < 1\}$$

Had been known [日合Petz 80]

was not known [OgawaNagaoka 00]

Another representation of $\overline{D}$: smooth Renyi

In fact, spectrum-like quantity have been also used in computer science, for analysis of random number generation, security of cryptography (randomness extractor)

Its quantum version by R. Renner

$$D_{\max}(\rho_0||\rho_1) = \min\{b; \rho_0 \le e^b \rho_1\}$$

$$D_{\max}^{\epsilon}(\rho_0||\rho_1) = \min\{D_{\max}(\rho_0{}'||\rho_1); \|\rho_0' - \rho_1\| \le \epsilon\}$$

$$\lim_{\epsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} D_{\max}^{\epsilon}(\rho_0^n||\rho_1^n) = \overline{D}(\{\rho_0^n\}||\{\rho_1^n\})$$

[RennarDatta 05]

Yet some more representations …

$$\lim_{\epsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} D^{\epsilon}_{\max}\left(\rho_0^n || \rho_1^n\right) = \overline{D}(\{\rho_0^n\}||\{\rho_1^n\})$$

[RennarDatta 05]

$$D^{\epsilon}_{\max}\left(\rho_0 || \rho_1\right) = \min\{D_{\max}\left(\rho_0{}' || \rho_1\right); \|\rho_0' - \rho_1\| \le \epsilon\}$$

$$\lim_{\epsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} D^{\epsilon}_{\alpha}\left(\rho_0^n || \rho_1^n\right) = \overline{D}(\{\rho_0^n\}||\{\rho_1^n\})$$

$$D^{\epsilon}_{\alpha}\left(\rho_0 || \rho_1\right) = \min\{D_{\alpha}\left(\rho_0{}' || \rho_1\right); \|\rho_0' - \rho_1\| \le \epsilon\}$$

$D_\alpha\left(\rho_0 || \rho_1\right)$ = any CPTP-monotone quantum version of α-Renyi relative entropy (α>1)

e.g. $\frac{1}{1-\alpha} \log \mathrm{tr} \rho_0^{1-\alpha} \rho_1^{\alpha}$

on “n”

“n” is merely an index. Can be any number

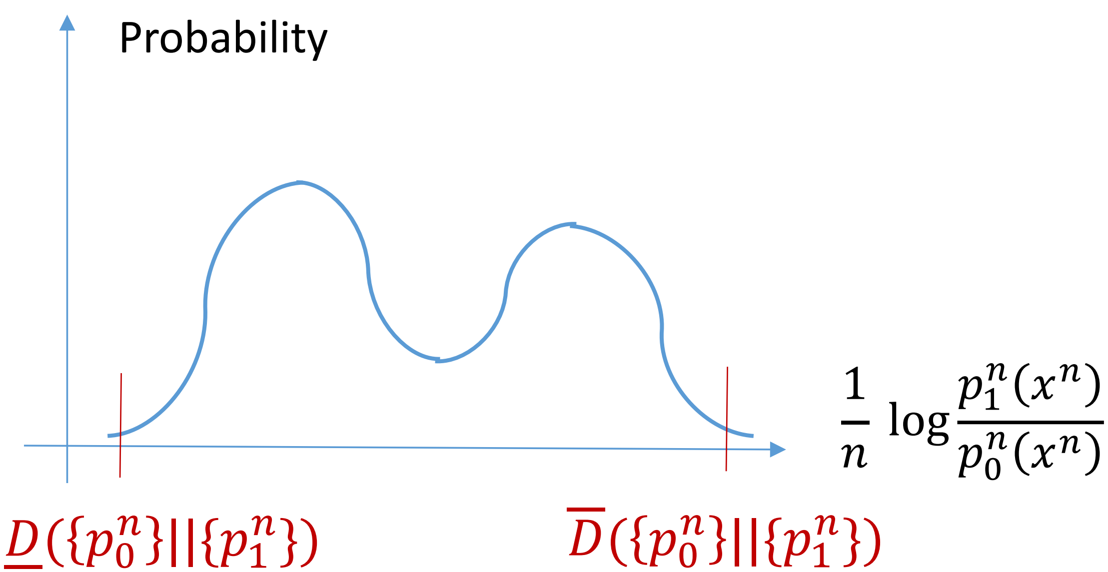
may not be system size....

It could be size of reservoir, for example

When $\overline{D}$ and $D_\alpha\ ({\rho_0}'||\rho_1)$ coincide with D?

- Obviously, if iid.
- In classical case, when n is the system size, and the system is Markovian etc,

(Maybe, when the systems are not globally correlated)



An application : resolvability [Ogawa M]

$x_1^{\mathrm{n}}, x_2^{\mathrm{n}} \cdots x_{L_n}^{n} \sim p^n$

$\frac{1}{L_n}\sum_{i=1}^{L_n} \rho_{x_i^n}^{n}$ will be very close to $E^n \rho_{x_1^n}^{n}$ if L_n is large enough.

How large is enough large ?

$$E^{\mathrm{n}} \left\| \frac{1}{L_n} \sum_{i=1}^{L_n} \rho_{x_i^n}^{n} - E^n \rho_{x_1^n}^{n} \right\|_1 \to 0$$

holds if $\limsup_{\mathrm{n}\to\infty} \frac{1}{n} \log L_n \geq \sup_{x^n} \overline{D}(\{\rho_{x^n}^{n}\}||\{E^n \rho_{x_1^n}^{n}\})$

Putting altogether :

Divergence rate and Thermo-Majorization

Classical case: $|\Theta| = 2, \ \ \mathrm{E}^n = \{p_0^n, p_1^n\}$

By Thermo-Majorization with error

$$\mathrm{E}^n = \{p_0^n, p_1^n\} \geq_{e_n} \mathrm{E}'^n = \{q_0^n, q_1^n\} \quad \lim e_{n,0} = 0, e_{n,1} = 0$$

$$(\Lambda(p_0^n) \approx q_0^n, \ \ \Lambda(p_1^n) = q_1^n)$$

$$\Leftrightarrow \quad \forall b \sum_x \left(p_0^n(x) - \mathrm{e}^{bn}\, p_1^n(x)\right)_+ \geq \sum_x \left(q_0^n(x) - \mathrm{e}^{bn}\, q_1^n(x)\right)_+ + e_{n,0}$$

もしも寝ていなければ以下が見える筈

Thm

$$\mathrm{E}^n \geq_{e_n} \mathrm{E}'^n \quad \Leftarrow \quad \underline{D}(\{p_0^n\}||\{p_1^n\}) \geq \overline{D}(\{q_0^n\}||\{q_1^n\})$$

$$\Rightarrow \quad \overline{D}(\{p_0^n\}||\{p_1^n\}) \geq \overline{D}(\{q_0^n\}||\{q_1^n\})$$

$$\underline{D}(\{p_0^n\}||\{p_1^n\}) \geq \underline{D}(\{q_0^n\}||\{q_1^n\})$$

$$\overline{D}(\{\rho_0^n\}||\{\rho_1^n\}) = \inf\{b;\ \lim_{n\to\infty} \mathrm{tr}\left(\rho_0^n - \mathrm{e}^{nb}\rho_1^n\right)_+ = 0\}$$

$$\underline{D}(\{\rho_0^n\}||\{\rho_1^n\}) = \sup\{b;\ \lim_{n\to\infty} \mathrm{tr}\left(\rho_0^n - \mathrm{e}^{nb}\rho_1^n\right)_+ = 1\}$$

Classical case: $|\boldsymbol{\Theta}| = \mathbf{2}, \quad \mathbf{E}^n = \{p_0^n, p_1^n\}$

Fact

$\mathrm{E}^n = \{p_0^n, p_1^n\} \geq_{e_n} \mathrm{E}'^n = \{q_0^n, q_1^n\} \quad \liminf e_{n,0} < 2, e_{n,1} = 0$

$\Leftrightarrow \quad \overline{D}(\{p_0^n\}||\{p_1^n\}) \geq \underline{D}(\{q_0^n\}||\{q_1^n\})$

$\|\Lambda(p_0^n) - q_0^n\|_1 < 2,$ 完全に区別できるほどには悪くない

$\Lambda(p_1^n) = q_1^n$

$$\overline{D}(\{\rho_0^n\}||\{\rho_1^n\}) = \inf\{b;\ \lim_{n\to\infty} \mathrm{tr}\left(\rho_0^n - \mathrm{e}^{nb}\rho_1^n\right)_+ = 0\}$$
$$\underline{D}(\{\rho_0^n\}||\{\rho_1^n\}) = \sup\{b;\ \lim_{n\to\infty} \mathrm{tr}\left(\rho_0^n - \mathrm{e}^{nb}\rho_1^n\right)_+ = 1\}$$

Standard form: $|\Theta| = 2$, $\mathbf{E}^n = \{\rho_0^n, \rho_1^n\}$

古典のときみたいに安直じゃないので頭捻る

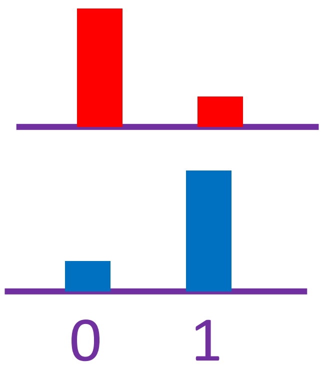
Standard form of binary experiments (dichotomies)

$\{r_0^n, r_1^n\}$: probability distributions over {0,1}

$\lim r_0^n(0) = 1$

$\lim r_1^n(0) = 0$, and exponentially

$$r_1^n(0) \sim \mathrm{e}^{-nb}$$



hypothesis test = conversion from $\mathrm{E}^n = \{\rho_0^n, \rho_1^n\}$ to $\{r_0^n, r_1^n\}$

Among various conversions,
Choose the one maximizing *b*

$$\max b = \underline{D}(\{\rho_0^n\}||\{\rho_1^n\})$$

Standard form: $|\boldsymbol{\Theta}| = \mathbf{2},\ \ \mathrm{E}^n = \{\rho_0^n, \rho_1^n\}$

It is natural to consider conversion from $\{r_0^n, r_1^n\}$ to $\mathrm{E}^n = \{\rho_0^n, \rho_1^n\}$

Lem If $c = \exp(-D_{\max}^{\epsilon}(\rho_0^n || \rho_1^n)$, There is a states τ_0, τ_1 with

$$\Leftrightarrow \quad \|\tau_0 - \rho_0^n\|_1 \leq \epsilon \qquad c\tau_0 + (1-c)\tau_1 = \rho_1^n$$

Proof : obvious from the def. $D_{\max}(\rho_0 || \rho_1) = \min\{b; \rho_0 \leq e^b \rho_1\}$

$$D_{\max}^{\epsilon}(\rho_0 || \rho_1) = \min\{D_{\max}(\rho_0{}' || \rho_1); \|\rho_0' - \rho_1\| \leq \epsilon\}$$

strategy By mixing τ_0, τ_1 with probability $r_\theta^n(0)$, $r_\theta^n(1)$, obtain ρ_0^n, ρ_1^n

Thm There is a CPTP map Λ_n with $\|\Lambda_n(r_0^n) - \rho_0^n\|_1 \to 0, \Lambda_n(r_1^n) = \rho_1^n$

If $r_0^n(0) \to 0,\ \ r_1^n(0) = e^{-nb}$,

$$b > \lim_{\epsilon \downarrow 0} \limsup_{n\to\infty} \frac{1}{n} D_{\max}^{\epsilon}(\rho_0^n || \rho_1^n) = \overline{D}(\{\rho_0^n\} || \{\rho_1^n\})$$

$|\Theta| = 2, \ \mathbf{E}^n = \{\rho_0^n, \rho_1^n\}$

$$\mathrm{E}^n \geq_{e_n} \mathrm{E}'^n \quad \lim e_{n,0} = 0, e_{n,1} = 0$$

$$(\Lambda(\rho_0^n) \approx \sigma_0^n, \ \Lambda(\rho_1^n) = \rho_1^n)$$

$$\mathrm{E}^n = \{\rho_0^n, \rho_1^n\} \to \{r_0^n, r_1^n\} \to \ \mathrm{E}'^n = \{\sigma_0^n, \sigma_1^n\}$$

Thm

$$\mathrm{E}^n \geq_{e_n} \mathrm{E}'^n \quad \Leftarrow \quad \underline{D}(\{\rho_0^n\}||\{\rho_1^n\}) \geq \overline{D}(\{\sigma_0^n\}||\{\sigma_1^n\})$$

$$\Rightarrow \quad \overline{D}(\{\rho_0^n\}||\{\rho_1^n\}) \geq \overline{D}(\{\sigma_0^n\}||\{\sigma_1^n\})$$

$$\underline{D}(\{\rho_0^n\}||\{\rho_1^n\}) \geq \underline{D}(\{\sigma_0^n\}||\{\sigma_1^n\})$$

If $D = \overline{D} = \underline{D}$, only have to compare D

Can prove uniqueness of quantum extension of relative entropy, which is "smooth" and CPTP monotone

Stability

Thm

If $D^Q(\Lambda(\rho)||\Lambda(\sigma)) \leq D^Q(\rho||\sigma)$

$$D(\rho||\sigma) = \operatorname*{clim}_{\epsilon\downarrow 0} \lim_{n\to\infty} \frac{1}{n} \inf\{\, D^Q(\rho' \parallel \sigma^{\otimes n});\ \|\rho' - \rho^{\otimes n}\|_1 \leq \epsilon \,\}$$

おしまい