I. What is entropy? (In Information Theory)

World Cup 2022 Predictions — Opta

		QF	SF	Final	WINNER
1	Brazil	63.7%	40.2%	25.3%	15.8%
2	Argentina	55.3%	36.3%	21.1%	12.6%
3	France	58.8%	37.3%	21.9%	12.2%
4	Spain	55.1%	31.6%	17.4%	9.1%
5	England	56%	30.9%	17%	8.7%
6	Germany	51%	28.7%	15.4%	7.8%
7	Netherlands	53.1%	27.5%	14%	7.2%
8	Portugal	47.2%	24.1%	12.2%	5.8%
9	Belgium	41.6%	22.2%	11.3%	5.4%

Group H World Cup 2022 — Opta Analyst

$S \cong 1.592$

		GROUP				KNOCKOUT				
		1st	2nd	3rd	4th	Last 16	QF	SF	Final	WINNER
1	Portugal	55.7%	26.9%	12.1%	5.3%	82.6%	45.2%	23.3%	11.7%	5.4%
2	Ghana	8%	18.1%	31.8%	42.1%	26.1%	7.2%	2.2%	0.6%	0.1%
3	Uruguay	26.9%	35.3%	23.4%	14.4%	62.2%	25.6%	10.4%	4.2%	1.6%
4	Korea Republic	9.4%	19.7%	32.7%	38.3%	29.1%	8.2%	2.5%	0.7%	0.2%

Group E World Cup 2022 — Opta Analyst

$S \cong 1.395$

		GROUP				KNOCKOUT				
		1st	2nd	3rd	4th	Last 16	QF	SF	Final	WINNER
1	Spain	50.5%	34%	12.2%	3.3%	84.5%	52.5%	30.1%	16.4%	8.9%
2	Costa Rica	1.9%	7.9%	27%	63.2%	9.8%	2.7%	0.6%	0.1%	0%
3	Germany	40.8%	39.5%	15.1%	4.6%	80.4%	49%	27.3%	14.7%	7.7%
4	Japan	6.7%	18.6%	45.7%	29%	25.4%	8.9%	2.9%	0.9%	0.2%

- The lower the probability, the higher the Surprise (ness)!

- $S_x = \log(1/P_x)$, P_x : probability of event x.

why "log"? Indep. prob. ⇒ additive
($P_{x,y} = P_x \cdot P_y$) ($S_{x,y} = S_x + S_y$)

(Any basis is o.k. but we will take $\log_2$ or $\ln$)

⇒ Average Surpriseness (or uncertainty)

$$\text{"}S(X)\text{"} = \langle S_x \rangle_{P_x} = \sum_{x \in X} P_x \log(1/P_x)$$

or $H(X)$ or $H(P_x)$

Entropy quantifies

"How well we can predict the system"

II. Basic Properties of Classical (Shannon Entropy)

$$H(P_x) = H(X) = \sum_{x \in X} P_x \log(1/P_x) = -\sum_{x \in X} P_x \log P_x$$

① $0 \leq H(X) \leq \log|X|$

100% certainty $P_x = \begin{cases} 1 & (x = x_0) \\ 0 & (\text{else}) \end{cases}$

of x's (when $P_x = \frac{1}{|X|}$ for all $x \in X$)

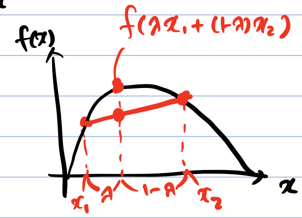
② For two different probability dists, P_x^1, P_x^2

⇒ mixing together : $P_x^\lambda = \lambda P_x^1 + (1-\lambda) P_x^2$

$$H(P_x^\lambda) = H(\lambda P_x^1 + (1-\lambda) P_x^2)$$
$$\geq \lambda H(P_x^1) + (1-\lambda) H(P_x^2)$$

$f(x) = -x \log x$

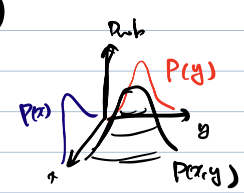
$f(\lambda x_1 + (1-\lambda) x_2)$



"Mixing always increases entropy" (Concavity)

- Entropy of two random variables $x \in X$ and $y \in Y$. $\Rightarrow P(x,y)$
  - Joint entropy : $H(X,Y) = -\sum_{x,y} P(x,y) \log (P(x,y))$

    (Overall uncertainty)



  - When we only look at one variable,

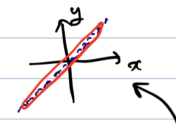
$$\begin{cases} H(X) = -\sum_x P(x) \log P(x), & P(x) = \sum_y P(x,y) \\ H(Y) = -\sum_y P(y) \log P(y), & P(y) = \sum_x P(x,y) \end{cases}$$

  - How closely two variables x & y are related?

Mutual information : $H(X:Y) = H(X) + H(Y) - H(X,Y)$ or $I(X:Y)$

i) if X & Y are indep. $P(x,y) = P(x) \cdot P(y)$

$\Rightarrow H(X:Y) = 0$



ii) if X & Y are perfectly correlated : $P(x,y) = P(x=y) = P(x)\delta_{x,y}$

$\Rightarrow H(X:Y) = H(X)$

$H(X,Y) = -\sum_{x,y} P(x,y) \log P(x,y)$

$= -\sum_{x,y} P(x,y) \log [P(x) \cdot P(y|x)]$

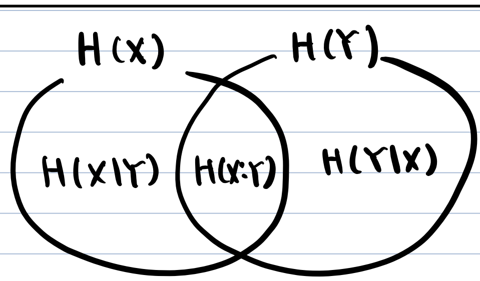
$= +H(X) - \sum_{x,y} P(x,y) \log P(y|x)$ ≥ 0.

Conditional entropy : $H(X|Y) = H(X,Y) - H(Y) \geq 0$

"entropy of X conditioned to knowing Y"
(uncertainty)

$\Rightarrow H(X:Y) = H(X) - H(X|Y) \geq 0 \Leftrightarrow H(X,Y) \leq H(X) + H(Y)$
(Sub additivity)

"Summary"



$$H(X) = \sum_{x \in X} P_x \log (1/P_x) = -\sum_{x \in X} P_x \log P_x$$

① $0 \leq H(X) \leq \log |X|$

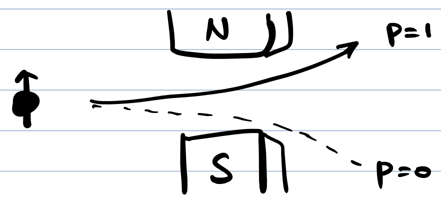
(concavity)

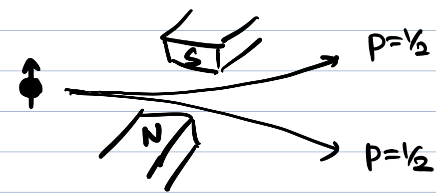
② $H(\lambda P_x^1 + (1-\lambda) P_x^2) \geq \lambda H(P_x^1) + (1-\lambda) H(P_x^2)$

③ $H(X|Y) = H(X,Y) - H(Y) \geq 0$ (conditional entropy)

④ $H(X:Y) = H(X) + H(Y) - H(X,Y) \geq 0$
(Mutual information) (Sub additivity)

Ⅱ. Entropy of quantum states





Entropy = 0 Entropy = log 2

(Entropy varies with the axis of measurement)

· A general description of quantum states : density "matrix" ρ

complex conjugate + transpose

① ρ is Hermitian ($\rho^\dagger = \rho$)

② ρ is positive semi-definite ($\rho \geq 0$)

③ $Tr[\rho] = 1$. (trace is always 1)

(ex)

· Wave function $|\psi\rangle \longrightarrow$ Corresponding density matrix : $\rho = |\psi\rangle\langle\psi|$.

$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \Rightarrow \rho = |\psi\rangle\langle\psi| = |\alpha|^2 |0\rangle\langle 0| + |\beta|^2 |1\rangle\langle 1| + \alpha\beta^* |0\rangle\langle 1| + \alpha^*\beta |1\rangle\langle 0|$

w. $|\alpha|^2 + |\beta|^2 = 1$.
$$= \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}$$

$Tr[\rho] = |\alpha|^2 + |\beta|^2 = 1$.

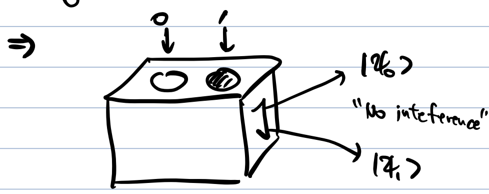
· Probability : $\begin{cases} P_0 = |\langle 0|\psi\rangle|^2 = |\alpha|^2 \\ P_1 = |\langle 1|\psi\rangle|^2 = |\beta|^2 \end{cases}$

$|\langle 0|\psi\rangle|^2 = \langle 0|\psi\rangle\langle\psi|0\rangle$
$= Tr[\, |\psi\rangle\langle\psi| \, |0\rangle\langle 0| \,]$
$= Tr[\, \rho \cdot \Pi_0 \,]$

(More generally, $P_x = Tr[\rho \Pi_x]$)

Measurement Op.
$\sum_x \Pi_x = 1$.
$(\Leftrightarrow \sum_x P_x = 1)$

· Why do we use ρ instead of $|\psi\rangle$?

$\Rightarrow$ 

$\rho = P_0 |\psi_0\rangle\langle\psi_0| + P_1 |\psi_1\rangle\langle\psi_1|$

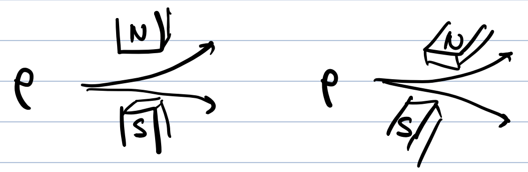
("Classical" mixture of "quantum" states)

· Quantum density matrix.

① ρ is Hermitian ($\rho^\dagger = \rho$)

② ρ is positive semi-definite ($\rho \geq 0$)

③ $\mathrm{Tr}[\rho] = 1$. (trace is always 1)



$\boxed{\text{Q. What is the least uncertain measurement basis?}}$

① : ρ has real-valued eigenvalues : $\rho = U \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} U^\dagger$

② : $\lambda_i \geq 0$ for all $i = 1, \cdots, d$

③ : $\mathrm{Tr}[\rho] = \sum_i \lambda_i = 1$

$$\Updownarrow$$

$$\rho = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|$$

$\Rightarrow \{\lambda_i\}_{i=1}^{d}$ acts as a prob. dist. (P_x) "eigenvalue decomposition"

If we measure the state with the eigenbasis : $\{|\phi_i\rangle\langle\phi_i|\} = \{\Pi_i\}$

$\Rightarrow P_i = \mathrm{Tr}[\rho\, \Pi_i] = \langle\phi_i|\rho|\phi_i\rangle = \lambda_i$

$\Rightarrow S(\rho) = H(\lambda_i) = -\sum_i \lambda_i \log \lambda_i$

$\| $

$-\mathrm{Tr}[\rho \log \rho]$

"von Neumann Entropy"

Function applied to a Hermitian operator?

$f(A) = f\left(\sum_i a_i |\phi_i\rangle\langle\phi_i|\right)$

$= \sum_i f(a_i)\, |\phi_i\rangle\langle\phi_i|$

when $f(a_i)$ are well-defined for $^\forall a_i$

For a pure state $\rho = |\psi\rangle\langle\psi|$,

$S(|\psi\rangle\langle\psi|) = 0$ $\quad (\because \lambda_i = 1, 0, \cdots, 0)$

IV. Basic Properties of Quantum entropy

$$S(\rho) = -\mathrm{Tr}[\rho \log \rho] \leq H(P_x) \qquad P_x = \mathrm{Tr}[\rho \Pi_x]$$

① $0 \leq S(\rho) \leq \log d$ (d: dimension of the system)

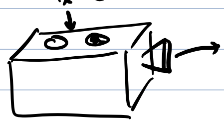
② $S(\lambda \rho^1 + (1-\lambda)\rho^2) \geq \lambda S(\rho^1) + (1-\lambda) S(\rho^2)$

" quantum entropy is also concave "

- **Classical information from quantum states**

(preparation)

x (classical variable)



$\rightarrow \rho_x \rightarrow \boxed{\Pi} \rightarrow P(y|x) = \mathrm{Tr}[\rho_x \Pi_y]$ (measurement) $\{\Pi_y\}$

$P(x,y) = P_x P(y|x) = P_x \mathrm{Tr}[\rho_x \Pi_y]$

: classical joint probability

$\Rightarrow$ Correlation between (preparation) x & (measurement) y ?

$$\Rightarrow H(X:Y) = H(X) + H(Y) - H(X,Y)$$
$$\leq \underbrace{S(\rho) - \sum_x P_x S(\rho_x)}_{\chi \text{ (Holevo information)}}, \quad \rho = \sum_x P_x \rho_x$$

- What is the best measurement $\{\Pi_y\}$?? We do not know the formula!

- Classically we can always take $Y = X \Rightarrow H(X:X) = H(X)$

"Encoding to quantum states does not help transmitting classical information"

- Conditional quantum entropy

classical

"two random variables x, y" $\Longleftrightarrow$ "Quantum state in $\mathcal{H}_A \otimes \mathcal{H}_B$"

Hilbert space

$P(x,y)$ $\qquad\qquad$ $\rho_{AB} = \sum_{\substack{i,j \\ k,l}} \rho_{ijkl} \, |i\rangle\langle j|_A \otimes |k\rangle\langle l|_B$

- $H(X,Y) = -\sum_{x,y} P(x,y) \log P(x,y)$ $\Longleftrightarrow$ $S(\rho_{AB}) = -\mathrm{Tr}[\rho_{AB} \log \rho_{AB}]$

$\begin{cases} P(x) = \sum_y P(x,y) \\ P(y) = \sum_x P(x,y) \end{cases}$ $\Longleftrightarrow$ $\begin{cases} \rho_A = \mathrm{Tr}_B[\rho_{AB}] = \sum_{i,j} \sum_k \rho_{ijkk} |i\rangle\langle j| & \text{"ignoring B"} \\ \rho_B = \mathrm{Tr}_A[\rho_{AB}] & \text{"ignoring A"} \end{cases}$

Partial Trace

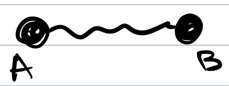
Reduced density matrix

$\rho_B = \sum_{k,l} \sum_i \rho_{iikl} |k\rangle\langle l|$

- $H(X|Y) = H(X,Y) - H(Y) \geq 0$ $\Longleftrightarrow$ $S(A|B) = S(\rho_{AB}) - S(\rho_B) \overset{?}{\geq} 0$ (No!)

"Conditional quantum entropy can be negative"

(ex) Entangled state



$|\Phi\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB})$

$\Rightarrow \rho_{AB} = \frac{1}{2} \begin{array}{c} \begin{array}{cccc} |00\rangle & |01\rangle & |10\rangle & |11\rangle \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{array}{c} \langle 00| \\ \langle 01| \\ \langle 10| \\ \langle 11| \end{array} \Rightarrow S(\rho_{AB}) = 0$

$\Rightarrow \rho_B = \mathrm{Tr}_A[\rho_{AB}] = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow S(\rho_B) = \log 2$

$S(A|B) = S(\rho_{AB}) - S(\rho_B) = -\log 2$ "≤ 0"

- What does it mean? (Quantum correlation can be even stronger than classical correlation.)

- Then, how about the quantum mutual information?

$S(A:B) = S(A) + S(B) - S(AB) \overset{?}{\geq} 0$ (Yes, fortunately)

- If A & B are independent, i.e. $\rho_{AB} = \rho_A \otimes \rho_B$

$S(\rho_{AB}) = S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B) \Rightarrow \underline{\underline{S(A:B) = 0}}$.

(exercise) "Show it"

- Quantum (von Neumann) entropy.

$$p_x = \mathrm{Tr}[\rho \Pi_x]$$

$$S(\rho) = -\mathrm{Tr}[\rho \log \rho] \leq H(p_x)$$

① $0 \leq S(\rho) \leq \log d$ (d: dimension of the system)

② $S(\lambda \rho^1 + (1-\lambda)\rho^2) \geq \lambda S(\rho^1) + (1-\lambda) S(\rho^2)$

③ $S(A|B)$ can be negative   classically, $H(X|Y) \geq 0$

④ $S(A:B) = S(A) + S(B) - S(AB) \geq 0$

V. Entanglement entropy.

A $|\Psi\rangle_{AB}$ B

A pure bipartite state $|\Psi\rangle_{AB}$

is separable $\iff |\Psi\rangle_{AB} = |\phi\rangle_A \otimes |\psi\rangle_B$

entangled : Not separable ($|\Psi\rangle_{AB} \neq |\phi\rangle_A \otimes |\psi\rangle_B$)

- How to quantify entanglement??

$$|\Psi\rangle_{AB} = \sum_i \sqrt{\lambda_i}\, |ii\rangle_{AB}$$ ← Schmidt decomposition (unique upto local basis trans.)

($|i\rangle_A$, $|i\rangle_B$ are both orthogonal bases)

$$\Rightarrow \rho_A = \mathrm{Tr}_B[|\Psi\rangle\langle\Psi|_{AB}]$$

$$= \sum_{i,j} \sqrt{\lambda_i}\sqrt{\lambda_j}\, \mathrm{Tr}_B[\,|ii\rangle_{AB}\langle jj|\,] \quad \delta_{ij}$$

$$= \sum_i \lambda_i |i\rangle_A\langle i| = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix}$$

$$\rho_B = \mathrm{Tr}_A[|\Psi\rangle\langle\Psi|_{AB}]$$

$$= \sum_{i,j} \sqrt{\lambda_i}\sqrt{\lambda_j}\, \mathrm{Tr}_A[\,|ii\rangle_{AB}\langle jj|\,] \quad \delta_{ij}$$

$$= \sum_i \lambda_i |i\rangle_B\langle i| = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix}$$

Same eigenvalues !

$$\Rightarrow S(\rho_A) = S(\rho_B)$$

$S(B|A)$ $S(A|B)$

$\Rightarrow$ Entanglement entropy : $\underbrace{E(\Psi) = S(\rho_A) = S(\rho_B)}$

(More uncertainty in each local party $\Rightarrow$ larger entanglement)

(Exercise) ① $|\Psi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle \Rightarrow E(\Psi) = -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = \log_2 2$

$= 1$: 1-ebit

② $|\Psi'\rangle = \frac{1}{\sqrt{3}}|00\rangle + \frac{\sqrt{2}}{\sqrt{3}}|11\rangle \Rightarrow E(\Psi') = -\frac{1}{3}\log_2(\frac{1}{3}) - \frac{2}{3}\log_2(\frac{2}{3})$

$= +\log_2 3 - \frac{2}{3}\log_2 2 = 0.918$

$\therefore$ $|\Psi\rangle$ is more entangled than $|\Psi'\rangle$

③ $|\Psi''\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$ ← Not a Schmidt decomposition!

$= \left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right) \otimes \left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)$

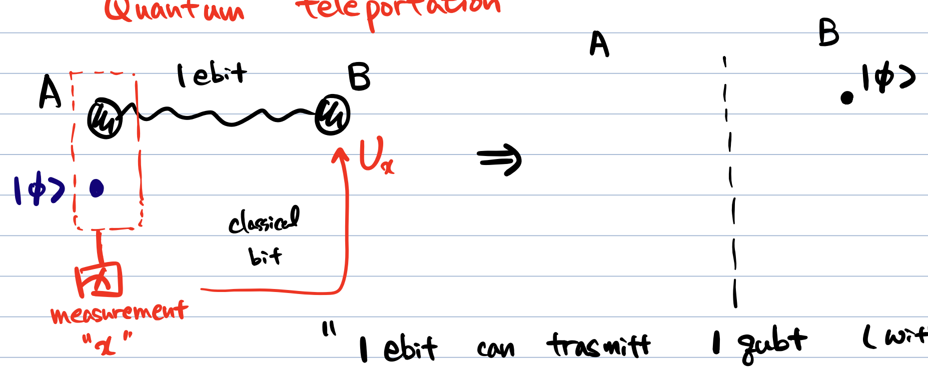
$= |+\rangle|+\rangle \Rightarrow E(\Psi'') = 1\cdot\log 1 + 0\cdot\log 0 = $ "0"

"No entanglement"

- Operational meaning of entanglement entropy

- What we can do with 1 ebit ($|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$) ?

"Quantum teleportation"



" 1 ebit can trasmitt 1 qubit (with classical commun.) "

Elementary protocol in quantum information processing

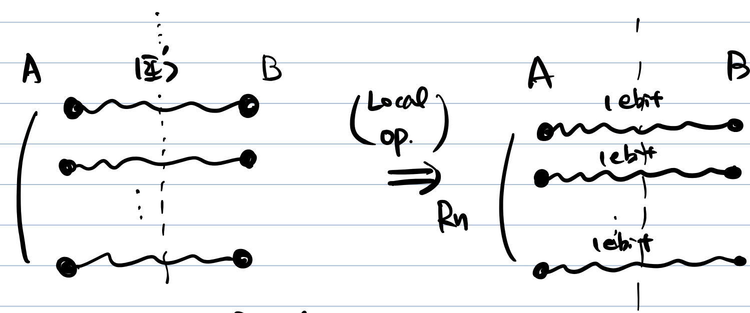
$\therefore$ Sharing as many copies of ebits as possible

$\Rightarrow$ Higher Q. communication Capacity.

· What happens if we have imperfect entanglement?

ex) $|\Phi'\rangle = \frac{1}{\sqrt{3}}|00\rangle + \frac{\sqrt{2}}{\sqrt{3}}|11\rangle \Rightarrow E = 0.918$

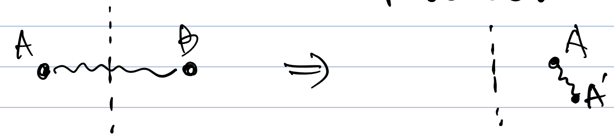
⇒ We cannot directly use for teleportation, but we can "distill" it to "perfect" entanglement.



n-copies of Φ' Rn-copies of Φ ⇒ "Rn ebits"

"$R = E(\Phi)$" : Entanglement entropy = distillation rate

(cf. 1) Interpretation in Q. Communication



"$S(A|B)$"

"Negative entanglement cost for state merging"

↓ entanglement gain

[Horodecki, Oppenhien, Winter] Nature 436, 673 (2005)

(cf 2) What is the average amount of entanglement of Randomly chosen states?

⇒ Almost all states are (nearly) maximally entangled!

4

Lemma II.4 *Let $|\varphi\rangle$ be chosen according to the unitarily invariant measure on a bipartite system $A \otimes B$ with local dimensions $d_A \leq d_B$, i.e. $\varphi \in_R \mathcal{P}(A \otimes B)$. Then*

$$\mathbb{E}S(\varphi_A) = \frac{1}{\ln 2}\left(\sum_{j=d_B+1}^{d_A d_B} \frac{1}{j} - \frac{d_A - 1}{2d_B}\right) > \log d_A - \tfrac{1}{2}\beta\,,$$

where $\beta = \frac{1}{\ln 2}\frac{d_A}{d_B}$. □

[PRL. 71, 1291] (1993)

However, for many many-body kinds of Hamiltonian, (e.g.) Local-gapped

the ground state's entanglement scales by Area.

$E(\Phi) \sim |\partial A|$

"Area law"

(Cf 3) What is the "maximum" entropy state with fixed energy?

Hamiltonian

$\langle E\rangle = \mathrm{Tr}[\rho H] = \text{const}$

Ans: $\rho = \frac{1}{Z} e^{-\beta H}$

VI. Entropy in classical information Processing

"Classical" information processing (Markov Process)

$$X \xrightarrow{T} Y \xrightarrow{T'} Z$$

$P(x)$ $\quad P'(y) = \sum_x T(y|x) P(x)$ $\quad P''(z) = \sum_z T'(z|y) P'(y)$

Transition Matrix
(conditional probability)

"How does information change during the process?"

$\Rightarrow$ Correlation between random variables before/after the process.

- Mutual information : $H(X:Y) = H(X) + H(Y) - H(X,Y)$ or $P(x,y) = T(y|x)P(x)$

Maximum? If we do nothing!

$P(x,y) = P(x)\,\delta(x-y)$
$\Rightarrow P(y) = P(x)$

$H(X:X) = H(X) + H(X) - H(X,X)$
$= H(X)$

$$\therefore \boxed{H(X) \geq H(X:Y)}$$

- What happen when another process is applied to Y?

$$H(X) \geq H(X:Y) \geq H(X:Z)$$

"Data processing inequality"

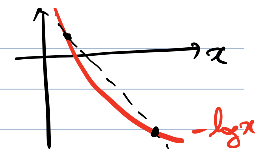
"Information does not increase by post-processing"

How do we prove this??

Introduce a useful quantity to
compare two probability distributions

a.k.a. Relative entropy

- Relative entropy (Kullback - Leibler divergence)



  - For two probability distribution $p(x)$ & $q(x)$,

$$D(p\|q) := \sum_x p(x) \log\left[\frac{p(x)}{q(x)}\right] \geq 0 \quad \left(\because D(p\|q) = -\sum_x p(x) \log\left[\frac{q(x)}{p(x)}\right] \geq -\log\left(\sum_x p(x) \frac{q(x)}{p(x)}\right) = 0 \right)$$

"=" hold iff. $p=q$.

$D(p\|q) \neq D(q\|p)$

  - A distance quantifier (but not measure as being asymmetric)

$\Rightarrow$ Many entropic quantities can be redefined using it.

① $H(X) = -\underbrace{D(p\,\|\,\mathrm{id.})}_{}\; + \log|X| \quad \left(\mathrm{id}(x) = \frac{1}{|X|} \text{ for all } x \in X\right)$

: uniform dist.

$$\left(\sum_x p(x) \log p(x) - \sum_x p(x) \log\left(\tfrac{1}{|X|}\right) = -H(X) + \log|X|\right)$$

How far the distribution deviates from the uniform distribution!

$\leq \log|X|$! $(\because D(p\|q) \geq 0)$

② $H(X:Y) = D(\,p(x,y)\,\|\,p(x)p(y)\,)$, $\begin{cases} p(x) = \sum_y p(x,y) \\ p(y) = \sum_x p(x,y) \end{cases}$

$$= \sum_{x,y} p(x,y) \log\left[\frac{p(x,y)}{p(x)p(y)}\right]$$

$$= -H(X,Y) - \sum_{x,y} p(x,y) \log p(x) - \sum_{x,y} p(x,y) \log p(y)$$

$$= -H(X,Y) + H(X) + H(Y)$$

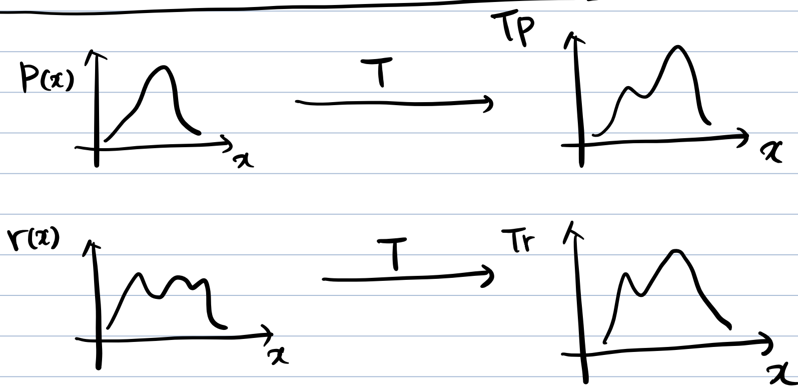
≥ 0 ! (by $D(p\|q) \geq 0$)

$\Rightarrow$ Mutual information : How close the distribution $p(x,y)$ is from independent distribution $p(x) \cdot p(y)$.

③ $H(X|Y) = -D(\,p(x,y)\,\|\,\mathrm{id}(x) \cdot p(y)\,) + \log|X|$

- Monotonicity of relative entropy

  - Suppose two different distributions $p(x)$ and $r(x)$.



Statistical distance

$$D(p \| r) \geq D(Tp \| Tr)$$

"After data processing distributions get closer"

$\Longleftrightarrow$ Hard to distinguish

$\Longleftrightarrow$ Loss in information

$X \xrightarrow{T} Y \xrightarrow{T'} Z$

$H(X) \geq H(X:Y) \geq H(X:Z)$

"Equivalent notion of Data processing inequality"

($\Rightarrow$) Take $p = p(x,y)$ and $r = p(x)p(y)$

$(T' = T'(z|y))$

Then, $D(p(x,y) \| p(x)p(y)) = H(X:Y)$

After applying the process T' $\quad X \xrightarrow{T} Y \xrightarrow{T'} Z$

$$T'p(x,y) = p(x,z) \quad \text{and} \quad T'p(x)p(y) = p(x)p(z)$$

Then, $D(T'p(x,y) \| T'p(x)p(y)) = D(p(x,z) \| p(x)p(z)) = H(X:Z)$

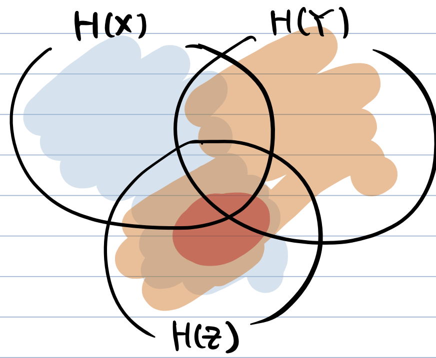
Hence, $H(X:Y) \geq H(X:Z)$ from the monotonicity of relative entropy.

The first inequality

"=" iff X can be (re)obtained from Y (No information loss)

$H(X|Y) \geq 0 \Longleftrightarrow H(X) \geq H(X:Y)$ is the special case of $X \to X \xrightarrow{T} Y$.

(c.f.1) Strong subadditivity of entropy



Data processing inequality

$$\left(\begin{array}{c} X \xrightarrow{T} Y \xrightarrow{T'} Z \\ H(X:Y) \geq H(X:Z) \end{array} \right)$$

$$H(X,Y,Z) + H(Z) \leq H(X,Z) + H(Y,Z)$$

Equivalent to the monotonicity of relative entropy

$$D(P(x,y,z) \| P(x)P(y,z)) \geq D(P(x,z) \| P(x)P(z))$$

T : "Ignoring y"

(C.f.2)

(Monotonicity proof) $\quad (x \xrightarrow{T} r)$

$$D(p \| r) \geq D(Tp \| Tr)$$

Let $Tp = p'$ & $Tr = r'$

$$\Rightarrow \sum_y p'(y) = 1$$

$$\Leftrightarrow \sum_y p'(y) \cdot \frac{r'(y)}{r'(y)} = 1$$

$$\Leftrightarrow \sum_y p'(y) \cdot \frac{\sum_x T(y|x) r(x)}{r'(y)} = 1$$

$$\Leftrightarrow \sum_{x,y} p'(y)\, T(y|x) \frac{r(x)}{r'(y)} \cdot \frac{p(x)}{p(x)} = 1$$

$$\Leftrightarrow \sum_{x,y} T(y|x) \cdot P(x) \cdot \left[\frac{p'(y)}{p(x)} \cdot \frac{r(x)}{r'(y)} \right] = 1$$

$$\Leftrightarrow \sum_{x,y} \underbrace{T(y|x)\, P(x)}_{p(x,y)} \cdot \underbrace{e^{-\log\left[\frac{p(x)}{r(x)} \cdot \frac{r'(y)}{p'(y)}\right]}}_{e^{-\Sigma}} = 1$$

$(\because)$ Convexity of exp. ft.

$$\Rightarrow e^{-\sum_{x,y} p(x,y) \log\left[\frac{p(x)}{r(x)} \cdot \frac{r'(y)}{p'(y)}\right]} \leq 1$$

$$\Rightarrow \sum_{x,y} p(x,y) \log\left[\frac{p(x)}{r(x)} \cdot \frac{r'(y)}{p'(y)}\right] \geq 0$$

$$\left(\begin{array}{l} \sum_x p(x,y) = p'(y) \\ \sum_y p(x,y) = p(x) \end{array} \right)$$

$$\therefore \ D(p \| r) - D(p' \| r') \geq 0$$

entropy production $(\Sigma = \Delta S - Q)$

"Fluctuation theorem"

$$\langle e^{-\Sigma} \rangle = 1 \Rightarrow \langle \Sigma \rangle \geq 0$$

VII. Entropy in quantum information processing.

- We can also imagine a quantum process

$$\rho \xrightarrow{\mathcal{N}} \rho' \xrightarrow{\mathcal{N}'} \rho''$$

- Unitary evolution (closed system)

$$\begin{cases} \dfrac{\partial|\psi\rangle}{\partial t} = (-i/\hbar)\,\hat{H}|\psi\rangle \;\Rightarrow\; \dot{\rho} = -i/\hbar\,[H,\rho] \\ |\psi(t)\rangle = \underbrace{\mathcal{T}\left[e^{-\frac{i}{\hbar}\int_0^t \hat{H}\,dt}\right]}_{=\,U(t)}|\psi_0\rangle \;\Rightarrow\; \rho(t) = U(t)\,\rho_0\,U^\dagger(t) \end{cases}$$

- How does the entropy change?

$$S(\rho(t)) = S(U(t)\rho_0 U^\dagger(t)) = S(\rho_0)$$ "No change"

($\because$ eigenvalues do not change)

$$\rho(t) = U(t)\rho_0 U^\dagger(t) = \sum_i \lambda_i\, U(t)|\phi_i\rangle\langle\phi_i|U^\dagger(t) = \sum_i \lambda_i\, |\phi_i(t)\rangle\langle\phi_i(t)|$$

- Then, when does the entropy change?

Interaction with environment.
(open quantum system)
"Yesterday's lecture"

Env.
$\Updownarrow$
Sys.

"ignore environment"

$$\dot{\rho}_S = (-i/\hbar)\,\mathrm{Tr}_E\left([H_{SE}, \rho_{SE}]\right)$$

"Measurement" is also interaction with environment.

- <u>Quantum channel</u>

(Recall) <u>Quantum density matrix</u>

$$\begin{cases} ① \ \rho \text{ is Hermitian } (\rho^\dagger = \rho) \\ ② \ \rho \text{ is positive semi-definite } (\rho \geq 0) \\ ③ \ Tr[\rho] = 1. \ \text{(trace is always 1)} \end{cases}$$

$\rho \xmapsto{\mathcal{N}} \rho' = \mathcal{N}(\rho)$ then ρ' should also satisfy ①-③

But, this is not enough to consider ② $\rho' \geq 0$

ex) Transpose : $\rho_{ij} \xmapsto{T} (\rho^T)_{ij} = \rho_{ji}$ (satisfy ①-③)

But when considering an extended state, $\mathcal{H}_A \otimes \mathcal{H}_B$,

only applying the transpose to A makes $(T \otimes \mathbb{1})(\rho_{AB}) \ngeq 0$

Let us consider an entangled state

$$|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \Rightarrow \rho_{AB} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Partial transposed state :

$$\rho^{T_A} = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|)$$

$$= \frac{1}{2}(|00\rangle\langle 00| + |10\rangle\langle 01| + |01\rangle\langle 10| + |11\rangle\langle 11|)$$

$$= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \rightsquigarrow \text{eigenvalue} : +\tfrac{1}{2}, +\tfrac{1}{2}, +\tfrac{1}{2}, -\tfrac{1}{2}$$

$\therefore \ \rho_{AB}^{T_A} \ngeq 0$, "T_A is not a proper process"

- Positive map : $\mathcal{N}(\rho) \geq 0$ for all ρ
- Completely positive map : $(\mathcal{N}_A \otimes \mathbb{1}_B)(\rho_{AB}) \geq 0$ for all ρ_{AB}

  ↑ proper condition,,

- Quantum channel : Completely Positive & Trace Preserving (CPTP map)

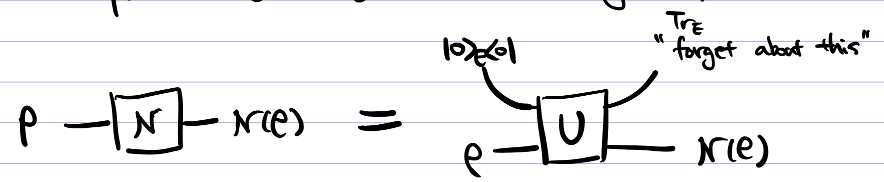
"Important property" (but will not show here)

[Stinespring Dilation Theorem]

⇒ Any CPTP map can be written as

$$\mathcal{N}(\rho) = \mathrm{Tr}_E\left[U \left(\rho \otimes |0\rangle\langle 0|_E\right) U^\dagger \right]$$

for some pure state $|0\rangle_E$ and unitary operator U



(c.f.) Equivalent expression of a quantum channel (CPTP map)

① $\mathcal{N}(\rho) = \sum_\mu K_\mu \rho K_\mu^\dagger$ with $\sum_\mu K_\mu^\dagger K_\mu = \mathbb{1}$ (Kraus)

② $\mathcal{N}(\rho) = \mathrm{Tr}_E\left[U (\rho \otimes |0\rangle\langle 0|_E) U^\dagger \right]$ (Stinespring Dilation)

③ $(\mathcal{N} \otimes \mathbb{1})(|\Phi^+\rangle\langle\Phi^+|_{SE}) = \rho^{\mathcal{N}}_{SE} \geq 0$, $|\Phi^+\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d-1} |ii\rangle_{SE}$

: Maximally entangled

A quantum channel can be converted into a quantum state and vice versa.

(Choi - Jamiolkowski Isomorphism).

$$\rho^{\mathcal{N}}_{SE} = \frac{1}{d} \sum_{ij} \mathcal{N}(|i\rangle\langle j|) \otimes |i\rangle\langle j| = \frac{1}{d} \sum_{\substack{ij \\ kl}} \langle k| \mathcal{N}(|i\rangle\langle j|) |l\rangle \; |k\rangle\langle l| \otimes |i\rangle\langle j|$$

- Quantum Data-processing inequality.

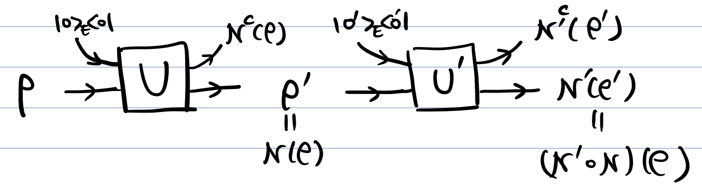
- Classical : $X \xrightarrow{T} Y \xrightarrow{T'} Z$

$$H(X) \geq H(X:Y) \geq H(X:Z)$$

can be shown from $D(p\|q) \geq D(Tp\|Tq)$ "Monotonicity"

- Quantum : $\rho \xrightarrow{\mathcal{N}} \rho' \xrightarrow{\mathcal{N}'} \rho''$

$[S(\rho) \geq \ ? \ \geq \ ?]$



$$S(\rho) \geq \underbrace{S(\mathcal{N}(\rho)) - S(\mathcal{N}^c(\rho))}_{= I(\rho, \mathcal{N})} \geq I(\rho, \mathcal{N}' \circ \mathcal{N})$$

"coherent information"

"Quantum Data processing inequality": Coherent information always decreases under Q.I.P.

$$S(\rho) \geq I(\rho, \mathcal{N}) \geq I(\rho, \mathcal{N}' \circ \mathcal{N})$$

"What does it mean? & How to prove it?"

"Let's first check the monotonicity of quantum relative entropy"

$$S(\rho\|\sigma) = \mathrm{Tr}[\rho(\log\rho - \log\sigma)]$$

$$\left(\begin{array}{ll} ① \ S(\rho\|\sigma) \geq 0 & \\ ② \ S(\rho\|\sigma) = 0 & \text{iff. } \rho = \sigma \end{array}\right)$$

$$\Rightarrow \begin{cases} S(\rho) = -S(\rho\|\mathbb{1}/d) + \log d \\ S(A:B) = S(\rho_{AB}\|\rho_A \otimes \rho_B) \\ S(A|B) = -S(\rho_{AB}\|\mathbb{1}/d_A \otimes \rho_B) + \log d_A \end{cases}$$

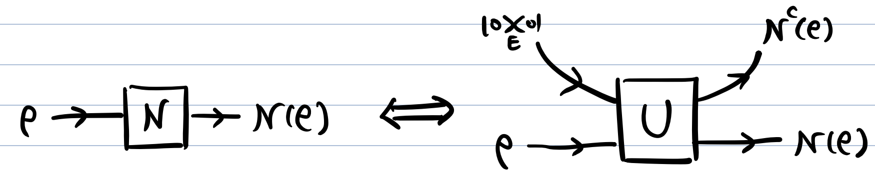
• Monotonicity of quantum relative entropy

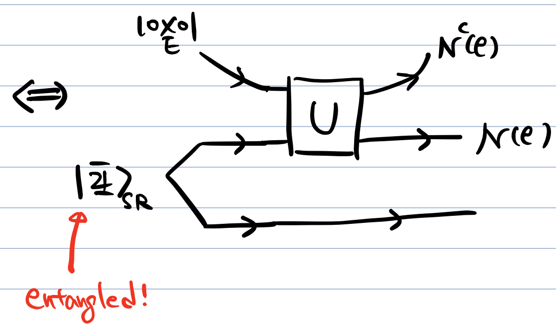
For any CPTP map $\mathcal{N}$,

$$S(\rho\|\sigma) \overset{?}{\geq} S(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$$

Yes!

(But proof is more difficult)

• 



$$|\Psi\rangle_{SR} = \sum_i \sqrt{p_i}\,|ii\rangle_{SR}$$

$\Downarrow$ "Purification of ρ_S"

$$\rho_S = \mathrm{Tr}_R\left[|\Psi\rangle\langle\Psi|_{SR}\right] = \sum_i p_i |i\rangle\langle i|_S$$

Correlation (entanglement) between S & R $\Rightarrow$ Conditional entropy

$$S(R|S) = S(|\Psi\rangle\langle\Psi|_{SR}) - S(\rho_S) = -S(\rho_S) \leq 0$$

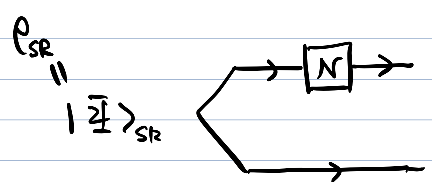
More negativity $\Rightarrow$ stronger corr.

$\Rightarrow$ Let us define "$-$"$S(R|S) \equiv I(S>R) = S(\rho_S) - S(\rho_{SR})$

"coherent information"

Using the quantum relative entropy,

$$I(S>R) = S(\rho_{SR}\|\rho_S \otimes \mathbb{1}/d_R) - \log d_R$$



$$\rho'_{SR} = (\mathcal{N}\otimes\mathbb{1})(|\Psi\rangle_{SR}\langle\Psi|)$$

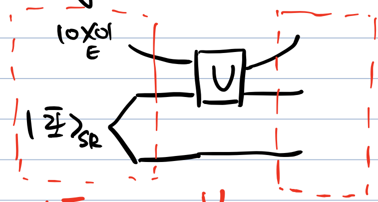
Let us apply the monotonicity of the quantum relative entropy

$$S(\rho_{SR}\|\rho_S\otimes\mathbb{1}/d_R) \geq S((\mathcal{N}\otimes\mathbb{1})(\rho_{SR})\|\mathcal{N}(\rho_S)\otimes\mathbb{1}/d_R)$$

$$\Rightarrow \quad I(S>R) - \log d_R \geq I(S'>R') - \log d_R$$

$\Rightarrow \quad I(S>R) \geq I(S'>R')$: "Coherent information is non-increasing under quantum channel"

* Finally, we show that $I(S'\rangle R') = I(\rho, \mathcal{N})$



$$I(S'\rangle R') = S(\rho_S') - S(\rho_{SR}')$$

But we

$$= S(\rho_S') - S(\rho_E')$$
$$= S(\mathcal{N}(\rho_S)) - S(\mathcal{N}^c(\rho_S))$$
$$= I(\rho, \mathcal{N})$$

$|\Psi\rangle_{SRE} \xrightarrow{U} |\Psi'\rangle_{SRE}$

"No entropy change"

$\Downarrow$

$S(\rho_{SR}) = S(\rho_E)$ $\quad$ $S(\rho_{SR}') = S(\rho_E')$

Therefore, we have $I(S\rangle R) \geq I(S'\rangle R') \geq I(S''\rangle R'')$

$$\boxed{\Longleftrightarrow \quad S(\rho) \geq I(\rho, \mathcal{N}) \geq I(\rho, \mathcal{N}'\circ\mathcal{N})}$$ ∎

" What happens if $S(\rho) = I(\rho, \mathcal{N})$? "

$$\rho \to \boxed{N} \to \mathcal{N}(\rho) \to \boxed{R} \to \rho$$

$\Rightarrow$ We can find the reverse process R s.t. $(R\circ\mathcal{N})(\rho) = \rho$.

Furthermore, for any $|\phi\rangle$ s.t. $\langle\phi|\rho|\phi\rangle > 0$,

$$(R\circ\mathcal{N})(|\phi\rangle\langle\phi|) = |\phi\rangle\langle\phi|$$

Quantum information is restored after interaction with environment

$\Rightarrow$ Quantum Error-correction condition.

<Summary>

Classical	Quantum
• $H(X) = -\sum_x P(x)\log P(x) \geq 0$	• $S(\rho) = -\mathrm{Tr}[\rho\log\rho]$
• $H(X:Y) = H(X)+H(Y)-H(X,Y) \geq 0$	• $S(A:B) = S(\rho_A)+S(\rho_B)-S(\rho_{AB}) \geq 0$
• $H(X\|Y) = H(X,Y) - H(Y) \geq 0$	• $S(A\|B) = S(\rho_{AB}) - S(\rho_B) \ngeq 0$
$X \xrightarrow{T} Y \xrightarrow{T'} Z$	$\rho \xrightarrow{\mathcal{N}} \rho' \xrightarrow{\mathcal{N}'} \rho''$
• $H(X) \geq H(X:Y) \geq H(X:Z)$	• $S(\rho) \geq I(\rho,\mathcal{N}) \geq I(\rho, \mathcal{N}'\circ\mathcal{N})$
• $D(p\|\|q) \geq D(Tp\|\|Tq)$	• $S(\rho\|\|\sigma) \geq S(\mathcal{N}(\rho)\|\|\mathcal{N}(\sigma))$

VIII. (Quantum) Information & Thermodynamics.

① Increasing on entropy under a random (permutation) process

- Suppose a coin (Head : 0, Tail : 1) initially has a p, $1-p$ probability.

0 1 0 1 0 0 1 0 ⋯ (for prob "q" we flip the coin) 0 1 1 1 0 0 0 1 ⋯

$\left(\dfrac{\#\text{ of heads}}{\text{total }\#}\right) \to p$

$\left(\dfrac{\#\text{ of tails}}{\text{total }\#}\right) \to 1-p$

$$\begin{pmatrix} p' \\ 1-p' \end{pmatrix} = q \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} + (1-q) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}$$

Entropy : $H(p) = -p\log p - (1-p)\log(1-p)$, $H(p') = -p'\log p' - (1-p')\log(1-p')$

$H(p) \leq H(p')$ (entropy always increases)

Why? Consider a stationary dist $(\tfrac{1}{2}, \tfrac{1}{2}) \xrightarrow{\text{flipping}} (\tfrac{1}{2}, \tfrac{1}{2})$, with $(\tfrac{1}{2}, \tfrac{1}{2}) = (r, 1-r)$

$$D(p\|r) \geq D(\underbrace{Tp}_{=p'} \| \underbrace{Tr}_{=r}) \quad : \text{Monotonicity}$$

$$\Rightarrow -H(p) + \log|X| \geq -H(p') + \log|X|$$

$$\Rightarrow H(p) \leq H(p') \quad \blacksquare.$$

- Corresponding Quantum Process : Unital Channel.

Φ : unital channel $\Longleftrightarrow$ $\Phi(\tfrac{1}{d}) = \tfrac{1}{d}$ (Identity is preserved)

$$S(\rho \| \tfrac{1}{d}) \geq S(\Phi(\rho) \| \Phi(\tfrac{1}{d}))$$

$$\Rightarrow -S(\rho) + \log d \geq -S(\Phi(\rho)) + \log d$$

$$\Rightarrow S(\rho) \leq S(\Phi(\rho))$$

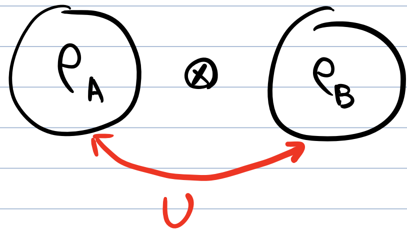
(ex) Mixing unitaries

$$\rho \longrightarrow \boxed{U_k}\ (p_k) \longrightarrow \sum_k p_k U_k \rho U_k^\dagger$$

$$\tfrac{1}{d} \longrightarrow \sum_k p_k U_k \tfrac{1}{d} U_k^\dagger = \sum_k p_k U_k U_k^\dagger = \tfrac{1}{d}$$

"Entropy is non-decreasing by Unital channel."

② Entropy change under unitary interaction



$$\Rightarrow \quad \rho'_{AB} = U(\rho_A \otimes \rho_B)U^\dagger$$

For unitary Process,

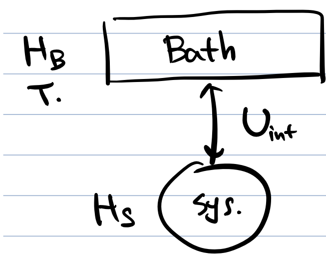
$$S(\rho'_{AB}) = S(U\rho_{AB}U^\dagger) = \underbrace{S(\rho_{AB})}_{= S(\rho_A) + S(\rho_B)}$$

"Mutual information"

$$\begin{aligned} S(A':B') &= S(\rho'_A) + S(\rho'_B) - S(\rho'_{AB}) \\ &= [S(\rho'_A) - S(\rho_A)] + [S(\rho'_B) - S(\rho_B)] \\ &= \Delta S_A + \Delta S_B \geq 0 \end{aligned}$$

"Sum of entropy change is always positive"

③ Quantum Free energy.



① Bath state : $\gamma_B = \frac{1}{Z_B} e^{-\beta H_B}$

② Energy conservation law.

$$[U_{int}, H_S + H_B] = 0$$

$\Rightarrow$ Thermodynamic channel / operation

$$\Lambda(\rho_S) = \mathrm{Tr}_B[U_{int}(\rho_S \otimes \gamma_B)U_{int}^\dagger]$$

* If the system is at equilibrium, $\gamma_S = \frac{1}{Z_S} e^{-\beta H_S}$

$$\Lambda(\gamma_S) = \gamma_S \quad \left(\because \; \begin{aligned} \Lambda(\gamma_S) &= \frac{1}{Z_S Z_B} \mathrm{Tr}_B[U_{int} e^{-\beta(H_S + H_B)} U_{int}^\dagger] \\ &= \frac{1}{Z_S Z_B} \mathrm{Tr}_B[e^{-\beta U_{int}(H_S + H_B)U_{int}^\dagger}] \\ &= \frac{1}{Z_S Z_B} \mathrm{Tr}_B[e^{-\beta H_S} e^{-\beta H_B}] \\ &= \frac{1}{Z_S} e^{-\beta H_S}. \end{aligned}\right.$$

- We can again apply the monotonicity of the relative entropy.

$$S(\rho_s \| \gamma_s) \geq S(\Delta(\rho_s) \| \Delta(\gamma_s)) = S(\Delta(\rho_s) \| \gamma_s)$$

" What does $S(\rho_s \| \gamma_s)$ mean?

$$\begin{aligned} S(\rho_s \| \gamma_s) &= \mathrm{Tr}[\rho_s(\log \rho_s - \log \gamma_s)] \\ &= -S(\rho_s) - \mathrm{Tr}[\rho_s \log(e^{-\beta H_s}/Z_s)] \\ &= -S(\rho_s) - \mathrm{Tr}[\rho_s(-\beta H_s)] - \mathrm{Tr}[\rho_s \log(1/Z_s)] \\ &= -S(\rho_s) + \beta\langle H_s\rangle + \log Z_s \\ &= \beta \underbrace{F_Q(\rho_s)}_{\text{quantum free energy}} + \log Z_s \end{aligned}$$

$[F = U - TS \text{ (Free energy)}]$

$\Rightarrow$ Quantum Free energy

How far the state is from the equilibrium state.

$$F_Q(\rho_s) = k_B T \cdot (S(\rho_s \| \gamma_s) - \log Z_s)$$

$$\Rightarrow \quad \Delta F_Q(\rho_s) = F_Q(\Delta(\rho_s)) - F_Q(\rho_s) \geq 0$$

" Quantum free energy is non-increasing under thermal channels (operations) " (c.f.1) Strong subadditivity of quantum entropy

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$$

$\Longleftrightarrow$ Monotonicity of the quantum relative entropy

$$S(\rho \| \sigma) \geq S(\mathcal{N}(\rho) \| \mathcal{N}(\sigma))$$

$\Longleftrightarrow$ Joint convexity of the quantum relative entropy

$$S\left(\sum_i p_i \rho_i \,\Big\|\, \sum_i p_i \sigma_i\right) \leq \sum_i p_i \, S(\rho_i \| \sigma_i)$$

(c.f.2) Monotonicity Proof (Another file)