4. Last time: classical AEP, Shannon entropy, q. ensemble.

Def [von Neumann entropy].

Let $\rho \geq 0$, $\operatorname{tr}\rho = 1$. $\rho \in \mathcal{L}(\mathbb{C}^d)$

Then $S(\rho) := -\operatorname{Tr}\rho\log\rho = -\sum_{i=1}^{d} \lambda_i \log \lambda_i$

where $\lambda_1, \dots, \lambda_d$ are eigvals of ρ.

eg. $\rho_0 = |0\rangle\langle 0|$, $f(0) = \frac{1}{2}$
$\rho_1 = |+\rangle\langle +|$, $f(1) = \frac{1}{2}$ $\Big\}$ BB92.

Ave state $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +|$

$$= \frac{1}{2}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + \frac{1}{2}\,\frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$$

$$= \begin{bmatrix}\frac{3}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4}\end{bmatrix} = \lambda_1 |e_1\rangle\langle e_1| + \lambda_2 |e_2\rangle\langle e_2|$$

where $\lambda_1 = 0.14645$, $|e_1\rangle = \begin{bmatrix}\sin\frac{\pi}{8}\\ -\cos\frac{\pi}{8}\end{bmatrix}$

$\lambda_2 = 0.85355$, $|e_2\rangle = \begin{bmatrix}\cos\frac{\pi}{8}\\ \sin\frac{\pi}{8}\end{bmatrix}$ $\Big\}$ Have little to do w/ states in the ensemble.

Quantum data compression (Schumacher compression):

Consider ensemble of pure states $\{ q(x), |\psi_x\rangle \}$

$$\rho = \sum_x q(x) \, |\psi_x\rangle\langle\psi_x| .$$

$\forall \delta > 0$, $\forall R > S(\rho)$

$\exists n_0$ s.t. $\forall n \geq n_0$, $\exists E_n, D_n$

s.t. ⓐ output dim of $E_n = 2^{nR}$

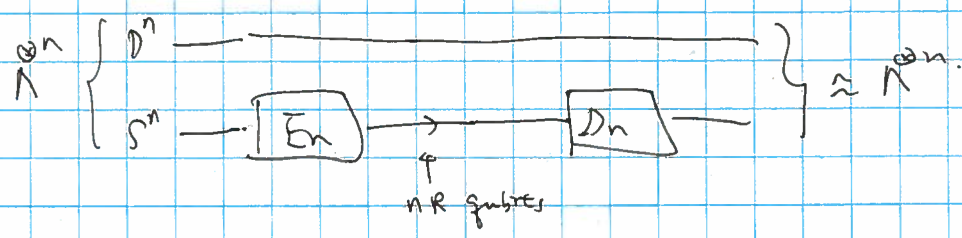
ⓑ Ave input-output fidelity $\geq 1 - \delta$.

Slightly more refined statements for ⓑ:

Let $$\Lambda^{\otimes n} = \sum_{x^n = x_1 \ldots x_n} q(x^n) \; \underbrace{|x^n\rangle\langle x^n|}_{D^n} \otimes \underbrace{|\psi_{x^n}\rangle\langle\psi_{x^n}|}_{S^n} .$$

$[\, |\psi_{x^n}\rangle = |\psi_{x_1}\rangle |\psi_{x_2}\rangle \cdots |\psi_{x_n}\rangle \,]$.

Then $$\| I_{D^n} \otimes (D_n \circ E_n)_{S^n} (\Lambda^{\otimes n}) - \Lambda^{\otimes n} \| \leq 2\sqrt{\delta} + \delta$$



Pf idea: transmit only the typical space of $\rho^{\otimes n}$, defined as follows.

Spectral decompose $\rho = \sum_{v=1}^{d} p(v)\, |e_v\rangle\langle e_v|$

Let V be rv with dist^n $p(v)$.

Let $T_{n,\varepsilon}$ be typical set for n iid draws of V.

For each $v^n = v_1 \dots v_n \in T_{n,\varepsilon}$

let $|e_{v^n}\rangle = |e_{v_1}\rangle |e_{v_2}\rangle \cdots |e_{v_n}\rangle$.

ε-typical space of $\rho^{\otimes n}$

Let $S = \text{span}\{|e_{v^n}\rangle : v^n \in T_{n,\varepsilon}\}$.

Define $\Pi_S := \sum_{v^n \in T_{n,\varepsilon}} |e_{v^n}\rangle\langle e_{v^n}|$

① $\dim S = |T_{n,\varepsilon}| \leq 2^{n(H(V)+\varepsilon)} = 2^{n(S(\rho)+\varepsilon)}$.

② $\text{Tr}(\rho^{\otimes n}\, \Pi_S) = \sum_{v'^n \in T_{n,\varepsilon}} p(v'^n) \geq 1-\delta$. $\leftarrow$ if $n \geq n_0 = \dots$

$\sum_{\text{all } v^n} |e_{v^n}\rangle\langle e_{v^n}|\, p(v^n) \quad \sum_{v'^n \in T_{n,\varepsilon}} |e_{v'^n}\rangle\langle e_{v'^n}|$ $\leftarrow$ diagonal in the same basis.

NB ~~last obs is a~~ Above: defs & math statement.

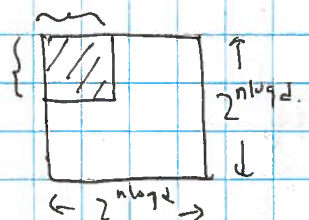
No where we project $\rho^{\otimes n}$ onto typical space.

Pf (ctd): What we actually do:

$\mathcal{E}_n$: ① performs binary measurement with projectors $\{\Pi_S, I-\Pi_S\}$.

② Rotate S to last $n(S(\rho)+\delta)$ qubits ← if outcome "Π_S"

$\uparrow$
U_n.

$U_n \Pi_S U_n^+$  $\approx |0\rangle\langle 0|^{\otimes n(\log d - S(\rho) - \delta)} \otimes$ 

$|0\rangle\langle 0|$: Q_2.   box: Q_1

③ Transmit Q_1

$\mathcal{D}_n$: Invert U_n. [attach $|0\rangle$'s to Q_2 first.]

unlike classical case, msg is not sent vs discard, always send $\Pi_S|\Psi_{x^n}\rangle$

Recall w.p. $p(x^n)$, input is $|\Psi_{x^n}\rangle$, output is $\Pi_S|\Psi_{x^n}\rangle$.

$$\text{Ave fidelity} = \sum_{x^n} p(x^n) \; |\underbrace{\langle\Psi_{x^n}|}_{\text{input}} \underbrace{\Pi_S|\Psi_{x^n}\rangle}_{\text{output}}|$$

Typical/not depends on S.

$$= \sum_{x^n} p(x^n) \; \langle\Psi_{x^n}|\Pi_S|\Psi_{x^n}\rangle$$

$$= \mathrm{tr}\left[\Pi_S \sum_{x^n} p(x^n) \; |\Psi_{x^n}\rangle\langle\Psi_{x^n}|\right]$$

Note: $\sum_{\text{any } x^n}$. no typical vs atypical here

$$= \mathrm{tr}\left[\rho^{\otimes n} |\Psi_{x^n}\rangle\langle\Psi_{x^n}|\right]$$

$$\geq 1-\delta.$$

For the more refined statement:

$$\Lambda^{\otimes n} = \sum_{x^n} q(x^n)\ |x^n\rangle\langle x^n| \otimes |\psi_{x^n}\rangle\langle\psi_{x^n}|.$$

$$\xrightarrow{E_n} \sum_{x^n} q(x^n)\ |x^n\rangle\langle x^n| \otimes U_n \Pi_s |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_s U_n^\dagger \otimes |0\rangle\langle 0|_C$$

$$+ \text{ 2nd term.}$$

$$\xrightarrow{D_n} \underbrace{\sum_{x^n} q(x^n)\ |x^n\rangle\langle x^n| \otimes \Pi_s |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_s \otimes |0\rangle\langle 0|_C}_{\text{1st term.}}$$

$$+ \text{ 2nd term.} = \tilde{\Lambda}^n$$

$$\text{tr (1st term)} = \text{tr} \sum_{x^n} q(x^n)\ |\psi_{x^n}\rangle\langle\psi_{x^n}| \cdot \Pi_s \overset{①}{\geq} 1-\delta.$$

$$\therefore \text{tr (2nd term)} \leq \delta.$$

$$\|\tilde{\Lambda}^n - \Lambda^{\otimes n}\|_1 \leq \|\Lambda^{\otimes n} - \text{1st term}\|_1 + \|\text{2nd term}\|_1$$

$$\leq \sum_{x^n} q(x^n)\ \|\Pi_s |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_s - |\psi_{x^n}\rangle\langle\psi_{x^n}|\|_1$$

$$+ \delta.$$

$$\left\| \begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array} - \begin{array}{c|c} C & 0 \\ \hline 0 & D \end{array} \right\|_1 = \|A-C\|_1 + \|B-D\|_1$$

$$= \sum_{x^n} q(x^n) \sqrt{1 - |\langle\psi_{x^n}| \Pi_s |\psi_{x^n}\rangle|^2} + \delta.$$

$$\leq 2 \sum_{x^n} q(x^n) \sqrt{1 - \langle\psi_{x^n}|\Pi_s|\psi_{x^n}\rangle} + \delta.$$

$\sqrt{\cdot}$ concave

$$\leq 2\sqrt{\sum_{x^n} q(x^n)\left(1 - \langle\psi_{x^n}|\Pi_s|\psi_{x^n}\rangle\right)} + \delta = 2\sqrt{\delta} + \delta$$

Remarks:

① Converse holds [proof later]. [follows from Holevo bdd].

② Mixed state ensembles:

Say $\mathcal{E} = \{ p(x), \rho_x \}$

where $\rho_x = \sum_j \mu_j^x \, |a_j^x\rangle\langle a_j^x|$

Consider $\mathcal{E}' = \{ p(x)\mu_j^x, \; |a_j^x\rangle \}$

Same ave state.

This ensemble can be transmitted

With $\approx S(\rho)$ per draw.

The info on "which j" is preserved necessarily.

In general: Schumacher compression works but unlikely optimal.

Conjectured: $S(\rho) - \sum_x p(x) S(\rho_x)$ sufficient.

[will see why necessary soon].

Main open problem in the field

③ Var length code: much harder.

Code is complicated

④ Universal compression also possible.

Recall $H(X)$ measures the ignorance on X.

Let X, Y be two rv's with dist^n $p(xy)$ [& Ω_X, Ω_Y as sample spaces]

① $H(XY)$ well defined for composite rv XY. ← X Y sym.

② $\forall y \in \Omega_Y$, let $p(y) = \sum_x p(xy)$.

$$q_y(x) = \frac{p(xy)}{p(y)}$$

ie $q_y(x)$ = condition dist^n of X knowing $Y = y$.

Def Conditional Entropy. ← X, Y not sym.

(Notations as above).

$$\underbrace{H(X|Y) := \sum_y p(y) \underbrace{H(q_y)}_{\text{uncertainty of } X \text{ given } Y=y}}_{\text{ave over } Y}$$

Thm: (Chain rule).

$H(XY) = H(X) + H(Y|X)$. [$= H(Y) + H(X|Y)$]

Pf: $H(XY) = -\sum_x \sum_y p(xy) \log \underbrace{p(xy)}_{p(x)p(y|x)}$

$[\log p(x) + \log p(y|x)]$

⋮

Algebra & defs.

Def [Relative entropy]. [Kullback Leibler distance] $\leftarrow p, q$ NOT sym

Let $p(x)$, $q(x)$ be two dist^{n}s on Ω [common].

$$D(p\|q) := \sum_{x\in\Omega} p(x) \log \frac{p(x)}{q(x)}.$$

$\leftarrow$ if $\mathrm{supp}(p) \not\subseteq \mathrm{supp}(q)$ then $D(p\|q) = \infty$.

[Where $0 \log \frac{0}{q} = 0$, $p \log \frac{p}{0} = \infty$.].

NB. $H(X) = \mathbb{E}_p[-\log p(x)]$, $D(p\|q) = \mathbb{E}_q\left[\log \frac{p(x)}{q(x)}\right]$.

Thm $D(p\|q) \geq 0$, "=" iff $p = q$.

Pf: algebra +

concavity of log i.e. $\mathbb{E}_x \log f(x) \leq \log \mathbb{E}_x f(x)$. ☆.

Aside: Hypothesis testing.

Dist^n is either p or q, n iid draws available.

Constrain: prob (conclude p | p) $\geq 1-\varepsilon$.

[When test deterministic, test given by $A \subseteq \Omega^n$

s.t. $x^n \in A \Rightarrow$ conclude p.

$x^n \notin A \Rightarrow$ " " q]

$\boxed{\text{NB } D(p\|q)=\infty \text{ If } q^n, \text{ will not conclude "}p\text{" at all.}}$

Let $\beta(n,\varepsilon) := \min_{\substack{A\subseteq\Omega^n \\ \underbrace{\text{under } p^n \to p(A) \geq 1-\varepsilon}_{\text{contrain}}}} q^n(A)$ = prob dist is q & we conclude p under test A.

Then: $\beta(n,\varepsilon) \sim 2^{-n D(p\|q)}$

Pf: Similar to AEP !

Def [Mutual info] :

← A derived concept based on $D(\cdot \| \cdot)$.

$$I(X:Y) := D(\,p(xy) \,\|\, p(x)\cdot p(y)\,).$$

$$= \sum_x \sum_y p(xy) \log \frac{p(xy)}{p(x)\cdot p(y)}$$

NB: $I(X:Y) \geq 0$ "=" iff X, Y indep. [$\because$ $p(xy) = p(x)\cdot p(y)$].

$I(X:Y) = I(Y:X)$.

Properties of H, D, & I:

(H1) Range:

$$0 \leq H(X) \leq \log|\Omega|$$

"=" iff $\exists a$ s.t. $p(x)=0 \ \forall x \neq a$.

"=" iff $p(x)=\frac{1}{|\Omega|} \ \forall x$

Ω: Sample space for H. WLOG, restrict to supp(p).

(H2) $H(X|Y) \geq 0$ (from (H1)) (& $H(X|Y)$ = conv combination of entropies).

(H3) $H(XY) \geq H(X)$. (from (H2), chain rule).

the more the merrier / messier.

(H4). $I(X:Y) = H(X) - H(X|Y)$ "How mutual info got its name"

$$\left[\begin{array}{l} = H(Y) - H(Y|X) \\ = H(X) + H(Y) - H(XY) \end{array}\right]$$

chain rule

Pf: replace $\frac{p(xy)}{p(y)}$ by $p(x|y)$ in "log" & alg.

Interpretation: $I(X:Y)$ = decrease in ignorance of X given Y = any info of X carried by Y. Sym "mutual!"

(H5) Subadditivity. (SA)

$H(XY) \le H(X) + H(Y)$ "=" iff X, Y indep.

(By non-neg of $I(X:Y)$.)

(H6) Conditioning reduces entropy.

$H(X|Y) \le H(X)$. "=" iff X, Y indep

(By Chain rule + Sub add.).

(H7) Concavity : mixing ↑ entropy.

Let $\alpha_k \ge 0$, $\sum_{k=1}^{t} \alpha_k = 1$

& $p_1, p_2, \ldots p_t$ dist^n's on Ω.

Then $\sum_{k=1}^{t} \alpha_k H(p_k) \le H\left(\sum_{k=1}^{t} \alpha_k p_k\right)$

Pf: Let K, X be r.v's with sample spaces $[t]$ & Ω.

$\text{prob}(K\,x) = \alpha_k \cdot p_k(x)$.

K chosen w.p. α_k, condition on K, x drawn according to p_k.

$\sum_{k=1}^{t} \alpha_k H(p_k) = H(X|K)$

$H\left(\sum_{k=1}^{t} \alpha_k p_k\right) = H(X)$. $\left(\text{prob}(x) = \sum_k \alpha_k p_k(x)\right)$

$\therefore$ (H7) follows from (H6).

H8 Strong sub add SSA or Data processing ineq (DPI)