11/4/10 QIC 710 Lecture 16 Quantum noiseless coding theorem, Entanglement concentration & distillation

Recall:

	(classical) Shannon entropy $H(\vec{p})$	(quantum) Von Neumann entropy $S(\rho)$
Def.	$H(\vec{p}) = -\sum_i p_i \log p_i = \mathbb{E}_{x\sim\vec{p}}[\log \frac{1}{p_x}]$	$S(\rho) = -\mathrm{Tr}(\rho \log \rho) = H(\{\text{eigenvalues } \lambda_i \text{ of } \rho\})$. $\Rightarrow S(U\rho U^\dagger) = S(\rho)$
for a d-dimensional system:	$0 \le H(\vec{p}) \le \log_2 d$; equality iff $\vec{p}$ is atomic; equality iff $\vec{p} = (\frac{1}{d}, \ldots, \frac{1}{d})$	$0 \le S(\rho) \le \log_2 d$; equal iff $\rho = \lvert\psi\rangle\langle\psi\rvert$; equal iff $\rho = \mathbb{1}/d$
subadditive:	for a joint distribution $\vec{p}_{AB}$: $H(\vec{p}_{AB}) \le H(\vec{p}_A) + H(\vec{p}_B)$, where $\vec{p}_A$ is the marginal dist $\vec{p}_A(i) = \sum_j p_{AB}(i,j)$; equality iff $\vec{p}_{AB} = \vec{p}_A \otimes \vec{p}_B$ independent	for a joint state $\rho_{AB} \in \mathcal{L}(H_A \otimes H_B)$, $S(\rho_{AB}) \le S(\rho_A) + S(\rho_B)$, where $\rho_A = \mathrm{Tr}_B \rho$, $\rho_B = \mathrm{Tr}_A \rho$. equality iff $\rho_{AB} = \rho_A \otimes \rho_B$ $\Rightarrow S(\rho \otimes \rho \otimes \cdots \otimes \rho) = n S(\rho)$.
however	while $H(\vec{p}_{AB}) \ge \max\{H(\vec{p}_A), H(\vec{p}_B)\}$	$S(\rho_{AB}) \ge \lvert S(\rho_A) - S(\rho_B)\rvert$ only.

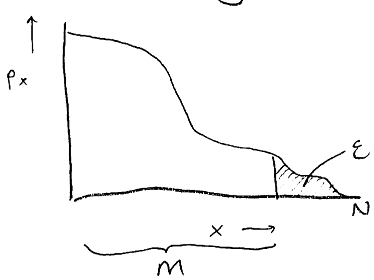
uncertainty measure
combinatorially: $\binom{n}{pn} \approx 2^{nH(p,1-p)}$

Recall: Classical data compression

Setting: Given a distribution $\vec{p}$ over N possible messages $\{1, 2, \ldots, N\}$.
Want a compression scheme into M messages (hopefully $M \ll N$)
so that $\mathbb{P}_{X\leftarrow\vec{p}}[\text{message } X \text{ can be recovered from its compression}] \approx 1$.

Optimal (information-theoretic) data compression:

1. Sort the messages according to their probabilities

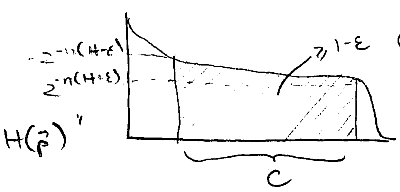


2. If we are willing to tolerate error ε, then throw away tail messages up to ε probability mass

Noiseless coding theorem: Fix a distribution $\vec{p}$. Let $\varepsilon > 0$. For n large enough, consider n independent copies of $\vec{p}$, ie. $\vec{p}^{\otimes n}$. Then

$$C \equiv \{x = x_1 x_2 \cdots x_n \mid \vec{p}^{\,n}(x) \in [2^{-n(H+\varepsilon)}, 2^{-n(H-\varepsilon)}]\}$$

satisfies $\mathbb{P}[X \in C] \ge 1-\varepsilon$.



$\Rightarrow$ "for product dists $\vec{p}^{\otimes n}$, typical msgs fall in a small set, determined by $H(\vec{p})$"

$|C| \le 2^{n(H+\varepsilon)}$

Quantum noiseless coding theorem

Fix a $d \times d$ density matrix ρ.

Let $\varepsilon > 0$ and n be large enough.

Consider n independent copies of ρ, $\rho^{\otimes n}$.

Then there exists a subspace $V \subset (\mathbb{C}^d)^{\otimes n}$, with orthogonal projection Π_V, such that

① $\dim(V) = \mathrm{Tr}(\Pi_V) \leq 2^{n(S(\rho)+\varepsilon)}$.

② $\mathrm{Tr}[\Pi_V \rho^{\otimes n}] \geq 1-\varepsilon$.

Observe: ② $\Rightarrow$ all but ε fraction of $\rho^{\otimes n}$ is supported on V (measuring $\Pi, \mathbb{1}-\Pi$ gives Π w/ prob. $\geq 1-\varepsilon$)

① $n(S(\rho)+\varepsilon)$ qubits suffice to hold an arbitrary state in V.

Proof:

Diagonalize ρ, $\rho = \sum_i p_i |i\rangle\langle i|$ for an orthonormal basis $\{|i\rangle\}$. In this basis, ρ is a probability distribution, so apply the classical noiseless coding theorem.

$\Rightarrow C = \{x = i_1 i_2 \cdots i_n \mid p^n(x) \in [2^{-n(S+\varepsilon)}, 2^{-n(S-\varepsilon)}]\}$ has $\mathbb{P}_{X \leftarrow \rho^{\otimes n}}[X \in C] \geq 1-\varepsilon$.

Let V be the span of C,

ie. $\Pi_V = \sum_{x \in C} |x\rangle\langle x|$.

① $\mathrm{Tr}\,\Pi_V = |C| \leq 2^{n(S+\varepsilon)}$ ✓

② $\mathrm{Tr}\,\Pi_V \rho^{\otimes n} = \sum_{x \in C} \mathrm{Tr}\,|x\rangle\langle x|\rho^{\otimes n}$

$= \sum_{x \in C} \langle x|\rho^{\otimes n}|x\rangle$

$= \sum_{x \in C} \mathbb{P}[X = x] = \mathbb{P}[X \in C] \geq 1-\varepsilon$ ✓ □

Typical application: Take $\rho^{\otimes n}$, measure $\Pi, \mathbb{1}-\Pi$. If get Π (usually you do), then rotate the space into $n(S+\varepsilon)$ qubits. Later can undo the rotation.

Moral: Just as holds classically, many copies of a quantum state are mostly concentrated in a small "typical subspace" (if $S(\rho)$ is small).

(This is also optimal)

entanglement is a fundamental new feature of quantum mechanics

Application: Entanglement measures. (for pure, bipartite states)

A. $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$

D. $\frac{1}{2}(|00\rangle + |11\rangle + |22\rangle + |33\rangle)$

B. $|0\rangle \otimes |0\rangle$

E. $\sqrt{1-3\varepsilon^2}\,|00\rangle + \varepsilon|11\rangle + \varepsilon|22\rangle + \varepsilon|33\rangle$

C. $\sqrt{\frac{2}{3}}\,|00\rangle + \sqrt{\frac{1}{3}}\,|11\rangle$

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

Order these states by "entanglement"

- possible answer: Schmidt rank

$B = 1 < A = C = 2 < D = E = 4$

but E is barely entangled!

- possible answer: "ε-approx. Schmidt rank" (which I just made up)

= min Schmidt rank of a state within ε

$B = E = 1 < A = C = 2 < D = 4$

still imprecise

Best answer: $E(|\psi\rangle_{AB}) = S(\rho_A) = S(\rho_B)$.

$B = 0 < E = H(1-3\varepsilon^2, \varepsilon^2, \varepsilon^2, \varepsilon^2) \ll \underset{\overset{\shortparallel}{C}}{H(\tfrac{2}{3}, \tfrac{1}{3})} < \underset{\overset{\shortparallel}{A}}{H(\tfrac{1}{2}, \tfrac{1}{2})} = 1 < H(\tfrac{1}{4}, \dots, \tfrac{1}{4}) = 2 = D$

Why?

Data compression $\Rightarrow$

$n(S(\rho_A) + \varepsilon)$ copies of $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)_{AB}$ suffice to create $|\psi\rangle_{AB}^{\otimes n}$ using classical communication

Conversely, n copies of $|\psi\rangle_{AB}$ can be locally converted into $n(S(\rho_A) - \varepsilon)$ copies of $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$

** (Bipartite, pure) entanglement is asymptotically interconvertible, with Bell pairs $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)_{AB}$ as a basic unit.

e.g. if $\rho_{AB} = |\psi\rangle\langle\psi|$, $\sigma_{AB} = |\varphi\rangle\langle\varphi|$

$S(\rho_A) = 5$ $\qquad S(\sigma_A) = 4$

then asymptotically $5n$ copies of $|\psi\rangle$ can be locally converted into $\sim 4n$ copies of $|\varphi\rangle$, and vice versa!

(this is not true for single states)

Proof:

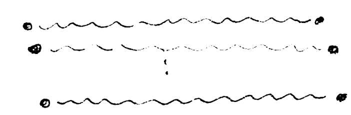
1. Use $\approx nS(\rho_A)$ Bell pairs $|00\rangle + |11\rangle$ to generate n copies of $|\psi\rangle$ between Alice & Bob.
2. Use n copies of $|\psi\rangle$ to generate $\approx nS(\rho_A)$ Bell pairs.

<u>Setting</u>

Alice　　　　　　　　Bob

classical telephone line
(can't create quantum entanglement!)

n(S+ε) Bell pairs



①: Alice prepares $|\psi\rangle^{\otimes n}$ <u>on her side</u>
Using data compression, she compresses half the states down to a $2^{n(S+\varepsilon)}$-dimensional space, which she rotates into $n(S+\varepsilon)$ qubits. (Fails w/ prob. ε)
She <u>teleports</u> these qubits to Bob, and he rotates them back to V.

② Converting $|\psi\rangle_{AB}^{\otimes n} \longrightarrow \left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right)_{AB}^{\otimes(n(S-\varepsilon))}$:

<u>Intuitive</u> idea (see [Preskill, 5.5] for more):

$$\frac{1}{\sqrt{2^m}}(|00\rangle + |11\rangle)_{AB}^{\otimes m} = \frac{1}{\sqrt{2^m}} \sum_{b_1=0}^{1} \sum_{b_2=0}^{1} \cdots \sum_{b_m=0}^{1} |b_1 b_1\rangle_{AB} \otimes |b_2 b_2\rangle_{AB} \otimes \cdots \otimes |b_m b_m\rangle_{AB}$$

$$= \frac{1}{\sqrt{2^m}} \sum_{b \in \{0,1\}^m} |b, b\rangle_{AB}$$

has uniform Schmidt coefficients $\frac{1}{\sqrt{2^m}}$.

- On the typical subspace V, eigenvalues of ρ_A lie in $[2^{-n(S+\varepsilon)}, 2^{-n(S-\varepsilon)}]$, i.e. are also roughly flat, & these are the Schmidt coefficients of $(\Pi_V \otimes \Pi_V)_{AB} |\psi\rangle_{AB}^{\otimes n}$! ∴ If Alice & Bob both apply data compression, they are close to getting nS Bell pairs (or a state w/ same Schmidt coeffs).
Technical issues: eigenvalues not flat enough, rank not a power of 2...