Lecture 23 - Quantum Information Theory

[Last time we discussed "Quantum Probability 101" in analogy with "Classical Probability 101". Today we'll do the same with Information Theory. We'll just barely scratch the surface of this very exciting subject, on which whole books have been written (E.g.: [Watrous '15]) !

Classical Info Theory 101 = [Shannon '49]

def#1: Let $p \in \mathbb{R}^d$ be a prob. distribution. Its <u>entropy</u>, $H(p)$, is $\sum_{i=1}^{d} p_i \log_2\left(\frac{1}{p_i}\right)$.

$(0 \cdot \log \frac{1}{0} := 0)$

[I hope you have some familiarity w/ this def. already.]

One intuition: [of many] If you had to write code to simulate a draw from p, $H(p)$ is least # of truly random coin flips (bits) you'd need (on average).

[This is exactly correct if you amortize over generation of many draws from p.]

[Another intuition is: # of bits needed (on avg.) to store a draw from p, with best compression scheme. Should be the same: given a draw, you can store the truly random bits that generate that draw.] //

E.g.: $d=3$, $p=(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

To generate p: Flip coin: H → output 1; T → flip coin for 2/3

$$E[\#\text{coins}] = 1 + \tfrac{1}{2}\cdot 1 = 1.5 = H(p) \checkmark$$

Facts: $0 \le H(p) \le \log d$

$0 \le$: equal if $p_i = 1$ for some i (p is deterministic)

$\le \log d$: equal if p is uniform, $(\frac{1}{d}, \ldots, \frac{1}{d})$

[Nothing to do but flip $\log d$ coins.]

Quantum Info Theory def #1:

The (von Neumann) entropy of a mixed state $\rho \in \mathbb{C}^{d \times d}$ is $H(\rho)$.... [[more usual notation is $S(\rho)$]]

[[There are many prob. dists. on pure states that have the same density matrix. But for this def., you should take a "canonical" one.]]

... if ρ has eigenvalues $\lambda_1, \ldots, \lambda_d$

[[recall, these form a prob. distribution — on an orthonormal basis of eigenvectors]]

$$H(\rho) := \sum_{i=1}^{d} \lambda_i \log_2 \frac{1}{\lambda_i}.$$

[[Represents the # of bits of "uncertainty" about which pure state — among d orthonorm ones — mixed state is in.]]

Quiz: The qubit $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ has density mtx

$\rho = |+\rangle\langle+| = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$. What is $H(\rho)$?

Answer: 0. Pure states have 0 entropy:

ρ's eigs: $\lambda_1 = 1, \lambda_2 = 0$, on $|+\rangle, |-\rangle$

Interp.: $H(\rho)$ = least # bits (coin flips) needed to simulate the d measurement outcomes of your favorite (i.e., cheapest) o.n. basis measurement.

E.g.: $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$, $\rho = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$

Measure in std basis? 50% "0" ⇝ 1 flip to simul.
50% "1"

Measure in $|+\rangle, |-\rangle$ basis? 100% "+" ⇝ 0 flips to simul
0% "−"

Vs.: 2-dim. max. mixed state, $\rho = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$

For any basis, measurement is 50%-50%.

Indeed, eigs are $\frac{1}{2}, \frac{1}{2}$, $H(\rho) = 1$.

fact: For $\rho \in \mathbb{C}^{d \times d}$, $0 \leq H(\rho) \leq \log d$

equal iff ρ pure

equal iff ρ max mixed.

Interp.: $H(\rho)$ = least # bits (coin flips) needed to simulate the d measurement outcomes of your favorite (i.e., cheapest) o.n. basis measurement.

E.g.: $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$, $\rho = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$

Measure in std basis? 50% "0" ⇝ 1 flip to simul.
50% "1"

Measure in $|+\rangle, |-\rangle$ basis? 100% "+" ⇝ 0 flips to simul
0% "−"

Vs.: 2-dim. max. mixed state, $\rho = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$

For any basis, measurement is 50%-50%.

Indeed, eigs are $\frac{1}{2}, \frac{1}{2}$, $H(\rho) = 1$.

fact: For $\rho \in \mathbb{C}^{d \times d}$, $0 \leq H(\rho) \leq \log d$

equal iff ρ pure

equal iff ρ max mixed.

[Back to classical info theory. Life only really starts to get fun when are two parties. Communication, mutual information, etc....]

Prob. 101: Say X, Y are random variables each taking values in $[d] = \{1, 2, \ldots, d\}$.

(Not necessarily independent, so....)

They have a joint distribution p on $[d] \times [d]$.

Example: $d = 4$, X, Y uniform s.t. $X+Y$ even.

I.e. unif. on $\{(1,1), (1,3), (2,2), (2,4), \ldots, (4,4)\}$

$p(1,1) = 1/8$, $p(1,2) = 0$, $p(1,3) = 1/8$, etc.

Say Alice "holds" X, Bob "holds" Y.

Distribution of just X is $p_A \in \mathbb{R}^d$, called Alice's marginal dist. (Sim for p_B)

In example, p_A, p_B both uniform on $\{1,2,3,4\}$.

[But not independent; p not unif. on $[4] \times [4]$.]

Formula: $p_A(x) = \sum_y p(x,y)$.

Quantum case

Alice has a qudit particle, Bob does too, and they're (potentially) entangled.

Joint state is some $\rho \in \mathbb{C}^{d^2 \times d^2}$.

E.g. $d=2$, A & B share (pure) EPR pair,

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{bmatrix}. \quad \rho = \begin{bmatrix} | \\ | \\ | \\ | \end{bmatrix} \begin{bmatrix} — & — & — & — \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}$$

What is "state" of Alice's qudit alone?

It's mixed: whatever you'd get if Bob measured.

In example: ρ_A = "50% $|0\rangle$, 50% $|1\rangle$" $= \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$

= max. mixed qubit [[not pure]]

Also $\rho_B = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$. But $\rho \neq \rho_A \otimes \rho_B$.

The operation $\rho \longmapsto \rho_A$ could be called "quantum marginalization". Q.M. nerds call it "partial trace (over Bob's register)", denoted "$\rho_A = tr_B(\rho)$." (Sim., $\rho_B = tr_A(\rho)$.)

[[Can write a formula that sorta justifies name "partial trace", but not really worth it, IMHO.]]

[Back to Classical Info. Theory 101]

Say p is joint prob. distrib on $[d] \times [d]$

(e.g. of (X,Y) unif on $\{(1,1),(1,3),(2,2),\ldots\}$)

$H(p_A)$ = # bits needed to gen X, = 2 in example.

(on average, amortized)

$H(p_B)$ = ~~~~~~~~~~ Y, = 2 in example

$H(p)$ = ~~~~~~~~~~ (X,Y) = 3 in example

Obvious fact:
$H(p_A), H(p_B) \leq H(p)$
always ⊛

flip 1 coin for (odd, odd)/(even, even)
then 1 coin to finish X,
1 coin to finish Y.

def: Mutual information is "$I(p)$" or "$I(X;Y)$":

$$H(p_A) + H(p_B) - H(p).$$

cost to generate X,Y separately — cost to gen. jointly

$\Rightarrow I(X;Y)$ = savings when generating jointly.

$I(X;Y) = 2+2-3 = 1$ in running example.

[Intuition: the 1 bit is odds/evens; the "mutual rand bit"]

Another interp.: "# of bits of info about X that Bob learns upon seeing Y" (on average)

[[And also, conversely, that Alice learns about Y upon seeing X, $\because I(X;Y)$ obviously symm.]]

Properties:

- $I(X;Y) \geq 0$, $\because H(\rho) \leq H(\rho_A) + H(\rho_B)$

  [[Least you can save is... nothing.]]

  With equality iff X, Y independent.

- $I(X;Y) \leq H(\rho_B)$ $\because H(\rho_A) \leq H(\rho)$

  (& $\leq H(\rho_A)$)

  [["Obvious fact" ☺]]

  [[Most you can save is all of $H(\rho_B)$ or $H(\rho_A)$.

  Equiv: most Alice learns about Y from X is $H(Y) = H(\rho_B)$.]]

- Say Alice & Bob separated. Bob takes Y & somehow locally produces new rand. var Z. [["Lumps", "adds randomness", whatever.]] Then $I(X;Z) \leq I(X;Y)$. I.e., Bob cannot create more mutual info by local actions. (E.g., if X, Y indep., $I(X;Y)=0$, Bob can't make a Z dependent on X.)

The Quantum Case ... a surprise ....

Say Alice & Bob share an EPR pair.

Let ρ be assoc. density matrix.

$H(\rho) =$ 0, since EPR pair is pure.

What is ρ_A? Max. mixed qubit state.

$H(\rho_A) =$ 1 !! $H(\rho_A) \neq H(\rho)$. Disturbing!

[Entropy created by discarding part of ρ.]

Spooky, & due to entanglement. [Purely quantum phenomenon.]

Aside: Let $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ be a pure "bipartite" state.

Write $\rho = |\psi\rangle\langle\psi|$

Fact 1: $H(\rho_A) = 0$ if and only if $|\psi\rangle$ a product state, $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$.

Fact 2: [Requires 5 mins of linear algebra; namely, S.V.D.]

ρ_A, ρ_B have same eigenvalues, hence $H(\rho_A) = H(\rho_B)$.

This # is called "measure of entanglement" of $|\psi\rangle$.

[Wide open area of research: define & understand measures of entanglement for non-pure bipartite mixed states.]

‖Disturbing that $H(\rho_A) \not\leq H(\rho)$! Fortunately, most other properties are okay.... ‖

Facts: $H(\rho) \leq H(\rho_A) + H(\rho_B)$? Yes!

‖Not hard to show. Proof in words: Recall interp. of $H(\rho)$: least # of coins needed to simulate measurement outcomes in my fave o.n. basis of $\mathbb{C}^{d^2}$. Well, can just measure A-half in your fave basis there, doesn't change ρ_B, so can measure that in fave B-way. You've paid $H(\rho_A) + H(\rho_B)$ now, so that upper-bounds $H(\rho)$. ‖

∴ if we define Quantum Mutual Info,

$$\text{“}I(\rho)\text{”} = \text{“}I(\rho_A;\rho_B)\text{”} = H(\rho_A) + H(\rho_B) - H(\rho),$$

then this is ≥ 0. ☺

(Fact: Equality iff $\rho = \rho_A \otimes \rho_B$.)

[What about upper bounds?]

$I(\rho_A;\rho_B) \leq H(\rho_A), H(\rho_B)$? X

[precisely b/c of disturbing fact]

e.g. for EPR pair, $I=2$, H's $=1$.

[What about Bob's inability to locally increase mutual info?]

Say Alice & Bob have entangled qubits.

Bob now operates locally on his.

[E.g. adds some more qubits, does unitaries, does partial measurements....]

[some (quantum) operations]

$I(\rho_A;\Phi(\rho_B)) \leq I(\rho_A;\rho_B)$? Yes!!

This fact is (equiv. to) "Strong Subadditivity of von Neumann Entropy"

[Elliott Lieb, Mary Beth Ruskai; '73]

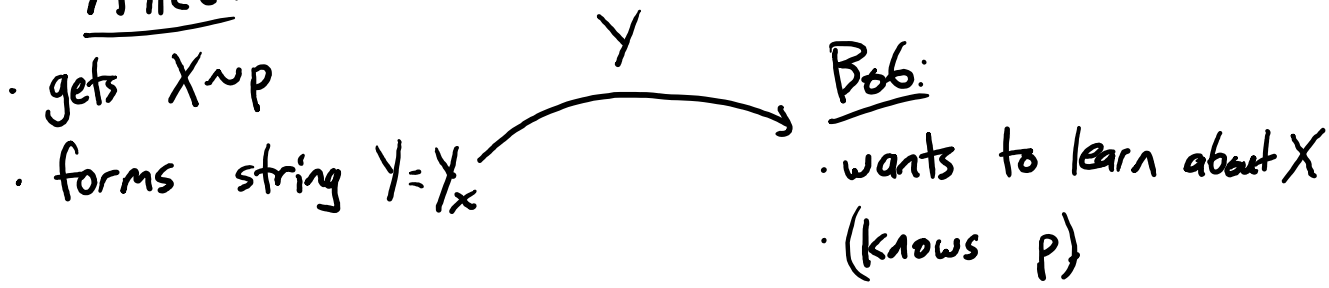
[Conjectured multiple times in '60s]

[Surprisingly difficult math thm. Like, a 2-lecture proof.

Now a foundational basic theorem in Quantum Information Theory.]

Implies "Holevo's Bound":

Say p is a classical prob. dist on $\{0,1\}^n$.



Classically: "Bob learns $I(X;Y)$ bits about X".

If Y limited to b bits, $\hookrightarrow \leq H(Y) \leq \log(2^b) = b$.

Q: What if Alice can send quantum states, $\sigma = \sigma_x$?

X: still classical. Still interested in how much classical info Bob can learn from σ_x about X.

[We have a half-classical, half-quantum scenario here.]

[Think of Alice's situation in quantum notation...]

Alice: $x \sim p$

$\Rightarrow$ her state is

$$\rho_A = \sum_{x \in \{0,1\}^n} p_x |x\rangle\langle x|.$$

She attaches σ_x on getting x.

$\Rightarrow$ Joint state is

$$\rho = \sum_x p_x |x\rangle\langle x| \otimes \sigma_x.$$

Bob's half of ρ is

$$\rho_B = \sum_{x \in \{0,1\}^n} p_x \cdot \sigma_x$$

Bob can now derive classical info from ρ_B. [Add qubits, unitary, measurement.]

Say $Y = \Phi(\rho_B)$

"Strong subadditivity" $\Rightarrow I(X;Y) \leq I(\rho_A;\rho_B)$

classical → $I(X;Y)$; $I(\rho_A;\rho_B)$ ← quantum

Say σ_x restricted to 6 qubits.

$$I(\rho_A;\rho_B) \overset{?}{\leq} H(\rho_B) \overset{\checkmark}{\leq} \log(2^6) = 6?$$

Not true in general. (by disturbing fact) →

But: true in this "semiclassical case!"

☺ $\Rightarrow$ 6 qubits can only convey ≤ 6 classical bits of information!

ex: [[Honestly, it's not too hard. The hard part of Holevo's Thm. is Strong Subadditivity.]]

$$I(\rho_A;\rho_B) = \text{“}\chi(p,\sigma)\text{”} := H(\rho_B) - \sum_x p_x H(\sigma_x)$$

(in this case)

(“Holevo's chi”)

$$\downarrow$$

evidently $\leq H(\rho_B)$.