Classical Information Theory

- Definition [Shannon 1949] Let $p \in \mathbb{R}^d$ be a probability distribution. Its *entropy*, $H(p)$, is $\sum_{i=1}^{d} p_i \log_2(1/p_i)$
  - Convention: $0 \log(1/0) = 0$
- One intuition: If you had to write code to simulate a draw from p, $H(p)$ is the least number of truly random coin flips you'd need on average.
- Example: $d = 3$, $p = (1/2, 1/4, 1/4)$. How to generate p?
- Facts: $0 \leq H(p) \leq \log d$, first equality holds if and only if $p_i = 1$ for some i, and second equality holds if and only if p is uniform

Quantum Information Theory

- Definition: The (von Neumann) entropy of a mixed state $\rho \in \mathbb{C}^{d \times d}$ is

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

  where ρ has eigenvalues $\lambda_1, \ldots, \lambda_d$

- Example: 100% the qubit $| + \rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$

- Interpretation $H(\rho)$ is the least number of coin flips needed to simulate the d measurement outcomes of your favorite orthonormal basis measurement
- Example 1: 100% the qubit $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$. Measure in … basis?
- Example 2: 2-dimensional maximally mixed state $\rho = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$. For any basis, measurement is 50%-50%.
- Fact: $0 \leq H(\rho) \leq \log d$, the first equality holds if and only if ρ is pure, and the second equality if and only if ρ is maximally mixed.

Probability 101

- Say X, Y are random variables each taking values in $[d] = \{1,2,\ldots,d\}$
- They have joint distribution p on $[d] \times [d]$
- Example: $d = 4$, (X, Y) uniform such that $X + Y$ even
- Say Alice holds X, Bob holds Y. Distribution of just X is $p_A \in \mathbb{R}^d$, called Alice's marginal distribution (similar for p_B)
- In example, p_A, p_B both uniform on $[d]$
- Formula: $p_A(x) = \sum_y p(x, y)$

Quantum case

- Alice has a qubit particle, Bob does too, and they're potentially entangled
- Joint state is some $\rho \in \mathbb{C}^{d^2 \times d^2}$
- Example: $d = 2$, Alice and Bor share an EPR pair. Then ρ = ?
- What is "state" of Alice's qubit alone?
  - It is mixed: whatever you'd get if Bob measured, $\rho_A = ?$
  - Also, ρ_B = ? But $\rho \neq \rho_A \otimes \rho_B$
- The operation $\rho \mapsto \rho_A$ is called "partial trace" over Bob's register, denoted by $\rho_A = \mathrm{tr}_B(\rho)$

Classical Information Theory 101

- Say p is joint probability distribution on $[d] \times [d]$
- Example: (X, Y) uniform on $\{(1,1), (1,3), (2,2), (2,4), (3,1), (3,3), (4,2), (4,4)\}$
- $H(p_A)$ = ?, $H(p_B)$ = ?, $H(p)$ = ?
- Obvious fact: $H(p_A), H(p_B) \leq H(p)$ always
- Definition: 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 generate jointly = savings when generating jointly

- $I(X;Y)$ = ?
- Another interpretation: number of bits of information about X that Bob learns upon seeing Y, and also, conversely, that Alice learns about Y upon seeing X
- Properties:
  - $I(X;Y) \geq 0$ with equality if and only if X, Y independent
  - $I(X;Y) \leq H(p_B), I(X;Y) \leq H(p_A)$
  - Say Alice and Bob separated, Bob takes Y and somehow locally produces new random variable Z. Then $I(X;Z) \leq I(X;Y)$. That is, Bob cannot create more mutual information by local actions.

The Quantum Case

- Alice and Bob share an EPR pair
  - Let ρ be associated density matrix
  - $H(\rho) = 0$, since EPR pair is pure
  - What is ρ_A and $H(\rho_A)$? $H(\rho_A) \leq H(\rho)$??? Disturbing…
- 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 $|\Psi\rangle$ is a product state, that is, $|\Psi\rangle = |\Psi_A\rangle \otimes |\Psi_B\rangle$
  - Fact 2: ρ_A and ρ_B have same eigenvalues, hence $H(\rho_A) = H(\rho_B)$. This is called "measure of entanglement of $|\Psi\rangle$

- Facts: $H(\rho) \leq H(\rho_A) + H(\rho_B)$
- If we define quantum mutual information $I(\rho) = I(\rho_A; \rho_B) = H(\rho_A) + H(\rho_B) - H(\rho)$, then this is ≥ 0, and equality holds if and only if $\rho = \rho_A \otimes \rho_B$
- $I(\rho_A; \rho_B) \leq H(\rho_A), H(\rho_B)$?
- Example: Alice and Bob have entangled qubits. Bob now operates locally on his, and obtains $\Phi(\rho_B)$.
- $I(\rho_A; \Phi(\rho_B) \leq I(\rho_A; \rho_B)$?
- This fact is known as “strong subadditivity of von Neumann entropy”

Holevo's bound

- Suppose p is a classical probability distribution on $\{0,1\}^n$
- Alice gets $X \sim p$ and forms string $Y = Y_X$, and sends y to Bob. Bob wants to learn about X and Bob knowns p.
- Classically: Bob learns $I(X;Y)$ bits about X. If Y limited to b bits, then $I(X;Y) \leq H(Y) \leq b$
- Question: What if Alice can send quantum states $\sigma = \sigma_X$? X is still classical. Still interested in how much classical information Bob can learn from σ_X about X

- Alice: $X \sim p$. Her state is $\rho_A = \sum_{x \in \{0,1\}^n} p_x |x\rangle\langle x|$
- She attaches σ_x on getting x. Now joint sate is $\rho = \sum_x p_x |x\rangle\langle x| \otimes \sigma_x$
- Bob's half of ρ is $\rho_B = \sum_x p_x \sigma_x$, and he can now derive classical information from ρ_B. Say $Y = \Phi(\rho_B)$
- "Strong subadditivity" implies that $I(X; Y) \leq I(\rho_A, \rho_B)$
- Say σ_x is restricted to b qubits. Then

$$I(\rho_A; \rho_B) = \chi(\rho, \sigma) := H(\rho_B) - \sum_x p_x H(\sigma_x) \leq H(\rho_B)$$