Classical - Quantum Metrics:

Shannon entropy of POVM: Let $p(x) = \text{Tr}(\Lambda^y \rho)$
Entropy of POVM is $H(X)$.

Accessible Information of Ensemble:

Given $E = E(p(x), p_x)^Y$

$$I_{acc}(E) = \max \left\{ \Lambda, \gamma \right\} I(X; Y)$$

Classical Mutual Information:

$$I_c(p^{AB}) = \max \left( \Lambda^A, \gamma^{AB} \right) I(X; Y)$$

always maximized by rank 1 POVM.

How do we define quantum conditional entropy?

One idea: Condition on measurement outcome on $A$.

$$H(B|A)_\Pi = \sum_x p(x) H(\rho^B_x)$$

$x$ indicates which measurement outcome was obtained by measuring $A$.

Which measurement should be used? The best one?

Instead let $H(B|A) = H(A, B) - H(A)$

In the classical case, both definitions are the same.
In fact, in a classical-quantum system, all quantities are similar to classical quantities.
Joint Entropy:

\[ H(A, B) = -\text{Tr} \left( \rho^{AB} \log \rho^{AB} \right) \]

Something funny happens: \( H(A, B) \neq H(A) \)

Why? Think of von Neumann entropy as the minimum Shannon entropy over measurements.

Consider an entangled pure state: \( H(A, B) = 0 \)

\[ H(A) > 0 \leftarrow \text{Why?} \]

1) In fact, \( H(A) = H(B) \) for a pure state.

Proof: Schmidt Decomposition

\[ |\psi^{AB}⟩ = \sum_i \lambda_i |i⟩^A ⊗ |i⟩^B \]

\[ ⇒ \rho^A = \sum_i |i⟩^A ⟨i|^A, \quad \rho^B = \sum_i |i⟩^B ⟨i|^B \]

\[ H(A) = H(\{\lambda_i\}) = H(B) \]

Consider \( \rho^A \) with purification \( |\psi^{RA}⟩ \): \( H(R) = H(A) \)

In fact, they have the same spectrum.

Corollary to 1) For a pure state on \( (A, B, C) \)

\[ H(A) = H(B, C) \]

\[ H(B) = H(A, C) \]

\[ H(C) = H(A, B) \]

Joint entropy of product state is additive:

\[ H(\rho \otimes \sigma) = H(\rho) + H(\sigma) \]

Classical-Quantum State:
\[ H(X, A) = H(X) + \sum_x p(x) H(p_x^A) \]

**Proof:** Consider the spectral decomposition
\[ p_x^A = \sum_a p(a|x) |\phi_{x,a}^A\rangle \langle \phi_{x,a}^A| \]

First calculate \( H(X) \) and \( H(p_x^A) \)

Notice: \( p_x^A = \sum_a p(a|x) |x\rangle \langle x| \)

\[ \Rightarrow H(X) = \sum_x p(x) \log \frac{1}{p(x)} \]

Also, \( H(p_x^A) = \sum_x p(a|x) \log \frac{1}{p(a|x)} \)

\[ p_x^A = \sum_{x,a} p(x) p(a|x) |x\rangle \langle x| \otimes |\phi_{x,a}^A\rangle \langle \phi_{x,a}^A| \]

Notice that \( |x\rangle |\phi_{x,a}^A\rangle \) is an orthonormal basis

\[ \Rightarrow H(p_x^A) = H(p(x) p(a|x)) = H(X) + \sum_x p(x) H(p_x^A) \]

---

**Back to Conditional Entropy:**

**Define** \( H(A|B) \equiv H(A, B) - H(B) \)

We've already seen that this can be negative.

**Operational significance:**

**Classical result:** \( H(X|Y) \) is minimum lossless compression rate for \( X \)
if \( Y \) is known to the decoder.

**Quantum result:** \( H(A|B) \) is minimum quantum compression rate for transferring
\( A \)'s state if \( B \) is available to the decoder, assuming unlimited classical communication.

A negative value means you end up with a net gain in quantum communication.
Specifically, you end up with entanglement. Recall \([\|g\|]=|g\rangle\rangle\) with unit \([e\rightarrow\) 

**Simple example:** \(|\Phi^{+}\rangle_{AB}\) is an ebit (Bell state)  
\[H(A|B) = -\log_2 1\] bits  

Because \(|\Phi^{+}\rangle_{AB}\) is a known pure state,  
the decoder can produce a copy without communication.  
Furthermore, \(|\Phi^{+}\rangle_{AB}\) can be used for one \([e\rightarrow\) in the future (teleportation).

---

**Coherent Information:**

\[I(A > B) \equiv -H(A|B)\]

\[\Rightarrow I(A > B) = H(B) - H(A, B) = I(A; B) - H(A)\]

---

**Bounds on conditional entropy.**

\[H(A|B) \leq H(A)\] \(\leftarrow\) Will use relative entropy to show this.

\[H(A|B) \geq -H(A)\] \(\iff I(A > B) \leq H(A)\)

\[I(A > B) = H(A|E) \leq H(A)\]

shown in problem set.

---

**Quantum Mutual Information:**

\[I(A; B)_{p} = H(A)_{p} + H(B)_{p} - H(A, B)_{p}\]
Conditional Mutual Information:

$$I(A;B|C) = H(A|C) + H(B|C) - H(A,B|C)$$

All classical chain rules still apply (by definition)

**Theorem:** $$I(A;B|C) \geq 0$$

Much of information theory rests on this.

**Corollary:** $$I(A;B) \geq 0$$

Already claimed above (i.e. $$H(A|B) \leq H(A)$$)

Quantum Relative Entropy:

$$D(\rho || \sigma) = Tr(\rho \log \rho - \rho \log \sigma)$$

Properties:

1. $$D(\rho || \sigma) \geq 0$$

2. If $$\text{supp}(\rho) \cap \text{supp}(\sigma^\perp) \neq \emptyset$$
   $$D(\rho || \sigma) = \infty$$

**Proof of 1:**

Let $$\rho = \sum_x p(x) |\psi_x\rangle \langle \psi_x|$$
$$\sigma = \sum_{x'} q(x') |\psi_{x'}\rangle \langle \psi_{x'}|$$

$$D(\rho || \sigma) = Tr(\rho \log \rho - \rho \log \sigma)$$
= $$\sum_x \langle \psi_x | (\rho \log \rho - \rho \log \sigma) | \psi_x \rangle$$
= $$\sum_x p(x) \log p(x) - \sum_x p(x) \sum_{x'} |\langle \psi_x | \psi_{x'} \rangle|^2 \log q(x')$$

This is because the bases are different.
Notice that \( \left| \langle \psi_x | \psi_x \rangle \right|^2 \) is doubly-stochastic.
That is \( \sum_x \left| \langle \psi_x | \psi_x \rangle \right|^2 = 1 \) \( \forall x \)
and \( \sum_x \left| \langle \psi_x | \psi_x \rangle \right|^2 = 1 \) \( \forall x' \).

\[ p(x'|x) = \left| \langle \psi_x | \psi_x \rangle \right|^2 \] is a valid conditional distribution.
and \( r(x) = \sum_x p(x) \left| \langle \psi_x | \psi_x \rangle \right|^2 \) is a valid pdf.

\[ \mu \sum_x p(x) \sum_x \left| \langle \psi_x | \psi_x \rangle \right|^2 \log q(x) = -\sum_x p(x) \sum_x p(x'|x) \log q(x') \]
\[ \geq -\sum_x p(x) \log \left( \sum_x p(x'|x) \log q(x') \right) \]
\[ = -\sum_x p(x) \log \left( \sum_x \left| \langle \psi_x | \psi_x \rangle \right|^2 \log q(x') \right) \]
\[ = -\sum_x p(x) \log r(x) \]

\[ D(p||q) \geq \sum_x p(x) \log p(x) - \sum_x p(x) \log r(x) \]
\[ = \sum_x p(x) \log \frac{p(x)}{r(x)} = D(p(x) || r(x)) \geq 0 \]

**Consequences:** \( I(A;B) = D(p^{AB} || p^A \otimes p^B) \)

Show in problem set.

\( I(A;B) \geq 0 \)

---

**Monotonicity of Relative Entropy:** (Data-Processing Ineq.)

\( D(p||q) \geq D(N(p)||N(q)) \)

We encountered this for trace distance.

**Proof:** Trick is to reduce to special case: \( D(p^{AB} || (\sigma^{AB})) \geq D(p^A || (\sigma^A)) \)

Believe but tedious to prove. (Appendix)

\( D(p||q) \geq D(p \otimes |0><0| \otimes I || \sigma \otimes |0><0| \otimes I) \)
\[ D(\rho \| \sigma) = D(\rho \otimes |0\rangle \langle 0|^E \| \sigma \otimes |0\rangle \langle 0|^E) \]
\[ = D(U(\rho \otimes |0\rangle \langle 0|^E)U^\dagger \| U(\rho \otimes |0\rangle \langle 0|^E)U^\dagger) \]
\[ \text{Relative Entropy is unitarily invariant} \]
\[ \geq D(N(\rho) \| N(\sigma)) \quad \text{under assumption 1.} \]

Consequences:
\[ D(\rho^{ABC} \| \rho^A \otimes \rho^{BC}) = D(\rho^{AB} \| \rho^A \otimes \rho^B) \]
\[ \Rightarrow \quad I(A;B) \geq 0 \]

Quantum Data Processing Ineq.:
\[ \text{Let } \rho^{A'B'} = (I^A \otimes N^{B \to B'}) (\rho^{AB}) \]
\[ I(A;B') \geq I(A;B) \quad \text{(Equivalently: } I(A \geq B) \geq I(A \geq B')) \]