Quantum Information Theory

In our tour of classical information theory, we considered several definitions and (more importantly) began to see some of the questions these definitions can be used to answer:

What are the ultimate limits to the compressibility of information?

How do we do hypothesis testing and distinguish distributions using observations?

How do we describe correlations between subsystems, how are these correlations affected by the decomposition of the system, and how are they affected by time evolution?

Quantum Information Theory

In our tour of classical information theory, we considered several definitions and (more importantly) began to see some of the questions these definitions can be used to answer:

What are the ultimate limits to the compressibility of information?

Quantum states are capable of encoding quantum information, and can also be measured to yield classical information in a large number of different bases.

How do we do hypothesis testing and distinguish distributions using observations?

If I have a hypothesis about an unknown quantum state, then how many measurements of which type should I use to test my hypothesis?

How do we describe correlations between subsystems, how are these correlations affected by the decomposition of the system, and how are they affected by time evolution?

General quantum states contain both quantum correlations (entanglement) as well as classical correlations. Quantifying the effect of these correlations is important in many settings, ranging from input/output correlations over a noisy channel, to understanding the physics of entangled phases of quantum matter.

Quantum Information Theory

Since classical information theory answers to these questions for probability distributions, our first approach to quantifying quantum information might naturally consider the entropy of measurement distributions.

Consider an n-qubit quantum state $|+^n\rangle = |+\rangle \otimes |+\rangle \otimes ... \otimes |+\rangle$. We can measure this state in any basis of our choice to obtain a probability distribution over measurement outcomes.

Suppose we measure the state $|+^n\rangle$ is the computational basis, obtaining the distribution:

$$p_Z(a) = |\langle a|+^n\rangle|^2 \quad , \quad a \in \{0,1\}^n$$

(where the subscript Z indicates the Z basis). What is the Shannon entropy of this distribution?

Suppose instead we consider the measurement distribution of the state above in +/- basis,

$$p_X(a) = |\langle a|+^n\rangle|^2 \quad , \quad a \in \{+,-\}^n$$

What is the Shannon entropy of this distribution? Compare $S(p_Z), S(p_X)$.

Quantum Information Theory

Following the connection between Shannon entropy and compressibility, we want our notion of the entropy of a quantum state to capture the most compact possible description of the state.

For the n-qubit quantum state $|+^n\rangle = |+\rangle \otimes |+\rangle \otimes ... \otimes |+\rangle$, characterizing the information content in terms of the computational basis would be a gross overestimate. From the point of view of compressibility, the quantum entropy should be no larger than the Shannon entropy in any choice of measurement distribution.

Furthermore, this quantum entropy should be basis independent and hence invariant under unitaries:

$$S(\rho) = S(U\rho U^\dagger)$$

The definition which turns out to capture these properties is the **von Neumann entropy**:

$$S(\rho) = -\mathrm{tr}(\rho \log \rho)$$

In the basis that diagonalizes the density matrix, this is the Shannon entropy of the distribution over quantum states that is represented by that density matrix:

$$\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i| \implies S(\rho) = -\sum_i \lambda_i \log \lambda_i$$

Quantum Information Theory

Because the von Neumann entropy is the Shannon entropy of a probability distribution associated with ρ, it inherits the same extremes we saw in the classical case.

The minimum von Neumman entropy is 0, and this corresponds to a density matrix that is also a pure state:

$$\rho = |\psi\rangle\langle\psi| \implies S(\rho) = 0$$

The maximum value of the von Neumann entropy is $\log D$, where D is the dimension of the quantum system. For an n-qubit state ρ this implies

$$S(\rho) \leq n$$

A density matrix that saturates this bound must have eigenvalues that correspond to the uniform distribution:

$$\rho = \frac{1}{2^n} \sum_{i=1}^{2^n} |\psi_i\rangle\langle\psi_i| = \frac{1}{2^n} I$$

Where the last step follows because the eigenvectors form an orthonormal basis. This state of maximum entropy is called the maximally mixed state, and evidently it has the same form in every basis.

Quantum Information Theory

Since the von Neumann entropy of pure state is zero, we may wonder whether this measure is suitable for describing pure state entanglement.

The **entanglement entropy** of a bipartite state ρ_{AB} (which may be pure) is the von Neumann entropy of the reduced state on either subsystem:

$$\rho_{AB} = \sum_{i=1}^{m} \lambda_i |\psi_i^A\rangle |\psi_i^B\rangle \implies \rho_A = \sum_{i=1}^{m} |\lambda_i|^2 |\psi_i^A\rangle\langle\psi_A^i| \implies S(\rho_A) = -\sum_{i=1}^{m} |\lambda_i|^2 \log |\lambda_i|^2$$

For any state ρ_{AB} we define the quantum mutual information I(A:B) in terms of the von Neumann entropies:

$$I(A:B) = S(A) + S(B) - S(AB)$$

For pure states $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ we have S(AB) = 0 so I(A:B) is twice the entanglement entropy. Therefore the quantum mutual information is suitable for detecting entanglement in pure states.

Quantum Information Theory

In the previous example we saw that the von Neumann entropy of a subsystem can be larger than that of the full system. This is a uniquely quantum phenomenon. In particular the quantum relative entropy

$$S(A|B) = S(AB) - S(B)$$

can be negative (in particular for entangled pure states). Classically this does not occur (because the conditional entropy is the expected entropy in the conditional distribution).

A negative conditional entropy is an indicator of entanglement, and later we will see it has an operational meaning in terms of the communication cost of "quantum state merging." But for now it is also a cautionary tale about what may change as we generalize classical definitions to the quantum setting!

Quantum Information Theory

Is the quantum mutual information nonnegative? Just as in the classical case, we will investigate this question using the relative entropy. The quantum relative entropy of two states ρ, σ is defined to be

$$S(\rho\|\sigma) = \mathrm{tr}\left(\rho \log \rho - \rho \log \sigma\right)$$

As in the classical case, our strategy is to prove this quantum relative entropy is nonnegative, and then relate a particular case of the quantum relative entropy to the quantum mutual information.

The first step is to let $\rho = \sum_i p_i |i\rangle\langle i|$ and let $\sigma = \sum_j q_j |j\rangle\langle j|$, so that

$$S(\rho\|\sigma) = \mathrm{tr}\left(\rho \log \rho - \rho \log \sigma\right) = \sum_i p_i \log p_i - \sum_{ij} p_i \langle i|j\rangle \log q_j \langle j|i\rangle$$

$$= \sum_i p_i \log p_i - \sum_{ij} p_i P_{ij} \log q_j \geq \sum_i p_i \log p_i - \sum_i p_i \log \left(\sum_j P_{ij} q_j \right) = \sum_i p_i \log(p_i / r_i)$$

Where $r_i = \sum_j P_{ij} q_j$ is a distribution because P is a stochastic, and since this has the form of a classical relative entropy we have shown that quantum relative entropy is nonnegative.

Quantum Information Theory

Now we can use the nonnegativity of quantum relative entropy to show the nonnegativity of quantum mutual information. Given a density matrix ρ_{AB} we imitate the classical proof and define:

$$\sigma_{AB} = \rho_A \otimes \rho_B$$

The following identity will be useful:

$$\log \sigma_{AB} = \log \rho_A \otimes I_B + I_A \otimes \log \rho_B$$

Using this we have

$$\begin{aligned} S(\rho_{AB} \| \sigma_{AB}) &= \mathrm{tr}\left(\rho_{AB} \log \rho_{AB} - \rho_{AB} \log \rho_A \otimes I - \rho_{AB} \log I \otimes \rho_B\right) \\ &= S(A) + S(B) - S(AB) = I(A:B) \end{aligned}$$

which shows that quantum mutual information is nonnegative (**subadditivity of von Neuman entropy**).

Quantum Information Theory

Now suppose we consider a quantum channel that maps a density matrix ρ to a mixture of density matrices $\sum_i p_i \rho_i$, where the ρ_i have support on orthogonal subspaces (e.g. different bit strings).

This operation increases entropy, because if $\rho_i = \sum_j \lambda_i^j |\psi_i^j\rangle\langle\psi_i^j|$ then

$$S\left(\sum_i p_i \rho_i\right) = \sum_{ij} -p_i \lambda_i^j \log p_i \lambda_i^j = -\sum_i p_i \log p_i - \sum_i p_i \sum_j \lambda_i^j \log \lambda_i^j$$

$$= S(p_i) + \sum_i p_i S(\rho_i)$$

So this channel, which could represent a projective measurement, takes our initial state to a new mixture. The new state has two sources of entropy: one is the entropy from being a weighted combination of possibilities, and the other is the weighted average of the entropy of each of the possibilities.

Quantum Information Theory

A projective measurement in the basis $\sum_k \Pi_k = I$ always increases the von Neumann entropy. To see this we can compute the relative entropy between the state before and after the measurement channel is applied.

$$\rho' = \sum_k \Pi_k \rho \Pi_k \implies S(\rho \| \rho') = -S(\rho) - \mathrm{tr} \rho \log \rho'$$

$$= -S(\rho) - \mathrm{tr} \sum_k \Pi_k \rho \log \rho'$$

$$= -S(\rho) - \sum_k \mathrm{tr}\ \Pi_k \rho (\log \rho') \Pi_k$$

$$= -S(\rho) + S(\rho')$$

Which implies that the measurements necessarily increase the entropy:

$$S(\rho') \geq S(\rho)$$

How much of an increase can there be in an n qubit system?

Quantum Information Theory

Previously we've shown that a quantum channel that maps a density matrix ρ to a mixture of density matrices $\sum_i p_i \rho_i$, where the ρ_i have support on orthogonal subspaces satisfies:

$$S\left(\sum_i p_i \rho_i\right) = S(p_i) + \sum_i p_i S(\rho_i)$$

Therefore if the ρ_i have orthogonal support, we have an inequality reminiscent of concavity:

$$S\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i S(\rho_i)$$

However, it also turns out that concavity holds for all linear combinations of density matrices $\sum_i p_i \rho_i$, which we will now prove by using nonnegativity of quantum mutual information.

Quantum Information Theory

This method for showing concavity of the von Neumann entropy is based on a trick which adds an auxillary subsystem B, while labeling the original system A:

$$\rho_{AB} = \sum_i p_i \rho_i \otimes |i\rangle\langle i|$$

Since ρ_{AB} is a linear combination of density matrices with orthogonal support, the joint entropy is:

$$S(\rho_{AB}) = S(p_i) + \sum_i p_i S(\rho_i)$$

And we can also compute the entropy of the subsystems: $S(A) = S\left(\sum_i p_i \rho_i\right), S(B) = S(p_i)$

By subadditivity, $S(A,B) \leq S(A) + S(B) \implies \sum_i p_i S(\rho_i) \leq S\left(\sum_i p_i \rho_i\right)$

Which is the concavity of von Neumann entropy.

Quantum Information Theory

In the classical setting, many of the most powerful results were obtained from strong subadditivity:

$$S_{ABC} + S_B \leq S_{AB} + S_{BC}$$

It turns out that this strong subadditivity also holds for the von Neumann entropy. The proof is nontrivial: the result was conjectured in 1968, but not proven until 1973 by Lieb and Ruskai.

Just as in the classical case, strong subadditivity is the key to proving the monotonicity of the relative entropy under partial trace, which is the quantum data processing inequality:

$$S(\rho_{AB} \| \sigma_{AB}) \geq S(\rho_A \| \rho_B)$$

Since the quantum mutual information is a particular quantum relative entropy, we also have the monotonicity of the quantum mutual information under partial trace:

$$I(A : BC) \geq I(A : B)$$

Quantum Information Theory

Since the quantum mutual information is a particular quantum relative entropy, we can imitate the steps from the classical case to show the monotonicity of the quantum mutual information under partial trace:

$$I(A:BC) \geq I(A:B)$$

Finally, since every quantum channel $\mathcal{E}$ arises by acting with a unitary on a joint system and performing a partial trace, and since the von Neumann entropy is invariant under conjugation by unitaries, the monotonicity of quantum relative entropy under partial trace also implies

$$S(\rho\|\sigma) \geq S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))$$

For every quantum channel $\mathcal{E}$. Just as in the classical case, this data processing inequality implies that no quantum channel can increase the distinguishability of two quantum states ρ, σ.

Quantum Information Theory

In summary, the von Neumann entropy of a quantum state ρ is the Shannon entropy of the eigenvalues of ρ,

$$S = -\mathrm{tr}(\rho \log \rho)$$

For a bipartite state ρ_{AB}, a suitable measure of entanglement called the entanglement entropy is defined to be von Neumann entropy of the reduced state on either subsystem.

Using the von Neumann entropy we can imitate the classical definitions and define the quantum conditional entropy, the quantum relative entropy, and the quantum mutual information:

$$S(A|B) = S_{AB} - S_B \quad , \quad S(\rho\|\sigma) = \mathrm{tr}(\rho \log \rho) - \mathrm{tr}(\rho \log \sigma) \quad , \quad I(A:B) = S_A + S_B - S_{AB}$$

We noted that a negative quantum conditional entropy is an indicator of entanglement. We proved the nonnegativity of quantum relative entropy, which also gives subadditivity $I(A:B) \geq 0$.

We stated the celebrated result known as the strong subadditivity of von Neumann entropy, and just as in the classical case this leads to a data processing channel for all states ρ, σ and all quantum channels $\mathcal{E}$,

$$S(\rho\|\sigma) \geq S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))$$

Copying Quantum States

An important feature of quantum theory is the ability to measure a state in an uncountable number of different bases, so a single quantum state can yield samples from an uncountable number of probability distributions.

However, we've also seen that all these different bases of measurement events correspond to incompatible observables which do not commute. Therefore measurement in one basis disturbs the state in another.

Suppose we have a qubit state $|\psi\rangle$ and we would like to obtain a sample from both the Z basis and X basis measurement distributions of $|\psi\rangle$. We could do this if we could make a copy of the state in a tensor product:

$$|\psi\rangle \rightarrow |\psi\rangle \otimes |\psi\rangle$$

This would suffice since one system could be measured in the Z basis, and the other in the X basis. If we make more copies we can measure in as many different incompatible bases as we want.

It turns out that nature doesn't allow us to do this. In its simplest form, the quantum no-cloning theorem states that there is no quantum operation that duplicates arbitrary quantum states

Copying Quantum States

Theorem (quantum no-cloning): there does not exist any linear operator U that can clone arbitrary states

$$(\nexists U)(\forall|\psi\rangle) : U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle$$

Proof: suppose U is an operator that can clone arbitrary quantum states, so that for all states $|\psi\rangle, |\phi\rangle$

$$U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle \quad , \quad U(|\phi\rangle \otimes |0\rangle) = |\phi\rangle \otimes |\phi\rangle$$

Then by assumption U can also clone the superposition state, $|\psi\rangle + |\phi\rangle$,

$$U\left((|\psi\rangle + |\phi\rangle) \otimes |0\rangle\right) = (|\psi\rangle + |\phi\rangle) \otimes (|\psi\rangle + |\phi\rangle)$$

Therefore U cannot be a linear operator, since $U|\psi\rangle|0\rangle + U|\phi\rangle|0\rangle = |\psi\rangle|\psi\rangle + |\phi\rangle|\phi\rangle$, and this is not equal to the above,

$$U\left((|\psi\rangle + |\phi\rangle) \otimes |0\rangle\right) \neq U|\psi\rangle|0\rangle + U|\phi\rangle|0\rangle$$

Copying Quantum States

There is an additional subtlety of this argument that is sometimes glossed over. To understand it, is also useful to see that no linear stochastic map M can clone arbitrary classical probability distributions in tensor product.

To see this, let μ, ν be distributions (unit vectors in the 1-norm), and suppose M copies arbitrary distributions:

$$M(\mu \otimes \mathbf{0}) = \mu \otimes \mu \quad , \quad M(\nu \otimes \mathbf{0}) = \nu \otimes \nu$$

Then by assumption M can also clone the mixture, $\alpha\mu + \beta\nu$, where $\alpha, \beta \geq 0, \alpha + \beta = 1,$

$$M\left((\alpha\mu + \beta\nu) \otimes \mathbf{0}\right) = (\alpha\mu + \beta\nu) \otimes (\alpha\mu + \beta\nu)$$

Therefore M cannot be linear, by the same argument we just gave for quantum states. So there is no stochastic map that can clone arbitrary probability distributions in a tensor product. What's going on here?

Copying Quantum States

The subtlety is in the phrase "copy in tensor product." This means our demand is to produce an **uncorrelated** copy:

$$\mu \to \mu \otimes \mu$$

Suppose the distribution is a biased coin, $\mu = (1-\epsilon, \epsilon)$. It is somewhat intuitive that without knowing ϵ, we cannot produce another biased coin of the same type that is uncorrelated with the first.

But we could copy the coin in a completed correlated way. We just say that coin #2 is heads if the coin #1 is heads, and similarly for tails.

The key point is that if we do this, the marginal distribution on coin #1 is still $\mu = (1-\epsilon, \epsilon)$, and the marginal distribution on coin #2 is also $\mu = (1-\epsilon, \epsilon)$, so we have effectively copied the state. (and this is how we really copy classical information). The point is just that the joint distribution is highly correlated.

Now we are prepared to understand the subtlety. This extra correlation introduced by copying does not affect the correctness of the reduced state of coin #1 and coin #2. But in the quantum case, the extra correlations do affect the reduced state of the system. Therefore the only way to avoid this would be to copy the state in tensor product, but this is impossible.

Copying Quantum States

Another way of describing our classical cloning procedure is to say that we copied the information in a particular choice of basis (the heads / tails basis). Quantumly, we can define a unitary U with the property:

$$U|0\rangle|0\rangle = |0\rangle|0\rangle \quad , \quad U|1\rangle|0\rangle = |1\rangle|1\rangle$$

Which copies the computational basis state of qubit 1 into a computational basis state of qubit 2. Now suppose we apply the unitary with the above properties to a superposition:

$$U(\alpha|0\rangle + \beta|1\rangle) \otimes |0\rangle = \alpha|00\rangle + \beta|11\rangle$$

Was this distribution successful in copying our quantum superposition? No, because if we look at the reduced state on either of the qubits by themselves we just see an incoherent mixture:

$$\rho_A = |\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1|$$

Therefore an understanding of the no-cloning theorem begins with understanding why we would need to impose the tensor product form on the cloned state. Then we show no linear operation exists to do this.

Copying Quantum States

Another way of describing our classical cloning procedure is to say that we copied the information in a particular choice of basis (the heads / tails basis). Quantumly, we can define a unitary U with the property:

$$U|0\rangle|0\rangle = |0\rangle|0\rangle \quad , \quad U|1\rangle|0\rangle = |1\rangle|1\rangle$$

Which copies the computational basis state of qubit 1 into a computational basis state of qubit 2. Now suppose we apply the unitary with the above properties to a superposition:

$$U(\alpha|0\rangle + \beta|1\rangle) \otimes |0\rangle = \alpha|00\rangle + \beta|11\rangle$$

Was this distribution successful in copying our quantum superposition? No, because if we look at the reduced state on either of the qubits by themselves we just see an incoherent mixture:

$$\rho_A = |\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1|$$

Therefore an understanding of the no-cloning theorem begins with understanding why we would need to impose the tensor product form on the cloned state. Then we show no linear operation exists to do this.