Quantum Information Theory

Scope:

1. Transmission of classical information over quantum channels.

2. The tradeoff between acquisition of quantum state information and disturbance of the state.

3. Quantifying quantum entanglement.

4. Transmission of quantum information over quantum channels.

Mainly accomplished by the interpretation and application of the **Von Neumann entropy**.

Classical Information Theory

A message is a string of letters chosen from an alphabet of k letters

$$\{a_1, a_2, \ldots, a_k\}.$$

The letters are independent and occurs with probability $p(a_x)$, and $\sum_{x=1}^k p(a_x) = 1$.

A typical message of length n will contain $np(a_x)$ a_x's for each x. So the number of typical strings is

$$\frac{n!}{\prod_{x=1}^k (np(a_x))!}.$$

By the Stirling approximation $\log n! = n \log n - n + O(\log n)$ we have

$$\begin{aligned} & \log \frac{n!}{\prod_{x=1}^k (np(a_x))!} \\ = \ & \log n! - \sum_{x=1}^k \log(np(a_x))! \end{aligned}$$

$$\begin{aligned}
&= n \log n - n - \sum_{x=1}^{k} \left(np(a_x) \log np(a_x) - np(a_x) \right) \\
&= n \log n - \sum_{x=1}^{k} np(a_x)(\log n + \log p(a_x)) \\
&= -n \sum_{x=1}^{k} p(a_x) \log p(a_x) \\
&= nH(X),
\end{aligned}$$

where

$$H(X) = -\sum_{x=1}^{k} p(a_x) \log p(a_x)$$

is the **Shannon entropy** of the ensemble $X = \{a_x, p(a_x)\}$.

So there are approximately

$$2^{nH(X)}$$

typical strings of length n for the letter ensemble X. Hence if we consider the typical strings as the only strings that can appear, then a string of length n can be compressed to $nH(X)$ bits, that is, only $nH(X)$ bits are needed to store any length n string.

The noiseless coding theorem states that the optimal code compresses each letter to $H(X)$ bits asymptotically. It's the highest compression rate given the requirement that messages must be decoded without errors as $n \to \infty$.

Another Perspective

For a particular length n message

$$x_1x_2\ldots x_n,$$

its prior probability is

$$P(x_1x_2\ldots x_n) = p(x_1)p(x_2)\ldots p(x_n)$$

and

$$\log P(x_1x_2\ldots x_n) = \sum_{i=1}^{n} \log p(x_i).$$

By the central limit theorem, when n is large enough most messages has probability P satisfying

$$\begin{aligned} -\frac{1}{n}\log P(x_1x_2\ldots x_n) &= -\frac{1}{n}\sum_{i=1}^{n}\log p(x_i) \\ &\approx \langle -\log p(x)\rangle \\ &\equiv H(X), \end{aligned}$$

where the random varaiable x represent a letter chosen from X.

So for the typical sequences its probability P satisfies

$$H(X) - \delta < -\frac{1}{n} \log P(x_1 x_2 \dots x_n) < H(X) + \delta,$$

or

$$2^{-n(H(X)-\delta)} > P(x_1 x_2 \dots x_n) > 2^{-n(H(X)+\delta)},$$

where $\delta > 0$ is small.

Interpretation of Shannon Entropy

The Shannon entropy for a specific source X can be seen as the amount of our ignorance about the value of the next letter, or the amount of indeterminancy of the unknownm letter. It can also be seen as the amount of information we gain after receiving one letter, in the usual case where the logarithm in H is with base 2, the unit of H is bits.

Binary Entropy

Suppose that the alphabet is bits, that is, $X = \{0, 1\}$, with probability $p_0 = p$ and $p_1 = 1 - p$. The entropy for this case is

$$H(X) = H(p) = -p \log p - (1 - p) \log(1 - p).$$

When $p_0 = \frac{1}{2}$, the bit is completely random, hence

$$H(\frac{1}{2}) = 1$$

is the maximum attainable value for the entropy, that is, we are maximally ignorant about the value of the next letter, or that we gain the most information (one bit) by receiving one letter. When $p_0 = 1$ or $p_1 = 1$, the next bit is completely predictable, the entropy in this case is

$$H(0) = H(1) = 0,$$

so we are not ignorant about the value of next bit at all, it also means that we get no information by receiving one letter. All other cases have entropy between these two extremes.

Generalize the result in the previous section to general sources X, the entropy H is zero whenever any one of the letters occurs with certainty. That is,

$$H_{min}(X) = -\log 1 = 0.$$

And maximum entropy is achieved when all letters occur with equal probability, that is, with a uniform probability distribution. For a X with d letters, the entropy of uniform probability is

$$H_{max}(X) = -\sum_i \frac{1}{d} \log \frac{1}{d} = -\log \frac{1}{d} = \log d.$$

In the general case of source X with d letters, its information per letter is

$$0 \leq H(X) \leq \log d.$$

Relative Entropy

If $p(x)$ and $q(x)$ are two probability distributions over the same index set x (or a given set of letters), then the relative entropy of $p(x)$ to $q(x)$ is defined as

$$\begin{aligned} H(p(x)||q(x)) &\equiv \sum_x p(x) \log \frac{p(x)}{q(x)} \\ &= -H(X) - \sum_x p(x) \log q(x). \end{aligned}$$

The relative entropy is a measure of the closeness of these two probability distributions. Since $\ln y \leq y - 1$, we have

$$\begin{aligned} H(p(x)||q(x)) &= -\sum_x p(x) \log \frac{q(x)}{p(x)} \\ &\geq \frac{1}{\ln 2} \sum_x p(x) \left(1 - \frac{q(x)}{p(x)}\right) \\ &= \frac{1}{\ln 2} \sum_x (p(x) - q(x)) \\ &= 0. \end{aligned}$$

With equality when $p(x) = q(x)$ for all x.

Mutual Information

Suppose a message composed from X are transmitted through a noisy channel, and a message composed from Y is received, that is, the channel distorts a letter $x \in X$ into $y \in Y$ with conditional probability $p(y|x)$. When the message is received, the probability distribution for x can be updated to

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)},$$

where $p(y|x)$ represent properties of the channel, $p(x)$ the a priori probabilities of ensemble X, and $p(y) = \sum_x p(y|x)p(x)$. So the message composed from Y contains some information about the original message from X. Using the $p(x|y)$'s we can defined the conditional entropy as

$$H(X|Y) = \langle -\log p(x|y) \rangle = -\sum_{xy} p(x,y) \log p(x|y).$$

Note that

$$\begin{aligned} H(X|Y) &= \langle -\log p(x,y) + \log p(y)\rangle \\ &= \langle -\log p(x,y)\rangle - \langle -\log p(y)\rangle \\ &= H(X,Y) - H(Y), \end{aligned}$$

where $H(X,Y) \equiv -\sum_{xy} p(x,y) \log p(x,y)$, similarly

$$H(Y|X) = H(X,Y) - H(X).$$

We need $H(X)$ bits per letter to decode messages from X, after receiving via the noisy channel a message from Y, we need $H(X|Y)$ more bits per letter to decode the message. In other words

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) \\ &= H(Y) - H(Y|X). \end{aligned}$$

bits of information per letters is gained by receiving the distorted message. $I(X;Y)$ is the **mutual information**, which is symmetric.

From the properties of the logarithm we have

$$H(X) \geq H(X|Y) \geq 0,$$

$$H(Y) \geq H(Y|X) \geq 0,$$

so

$$I(X;Y) \geq 0,$$

$$H(X) + H(Y) \geq H(X,Y).$$

That is, we will not lose any knowledge of a message from X by receiving a message from Y.
Equality occurs when X and Y is independent, then

$$\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= H(X) - \langle - \log p(x|y) \rangle \\
&= H(X) - \left\langle - \log \frac{p(x,y)}{p(y)} \right\rangle \\
&= H(X) - \left\langle - \log \frac{p(x)p(y)}{p(y)} \right\rangle \\
&= H(X) - \langle - \log p(x) \rangle \\
&= 0
\end{aligned}$$

The Noisy Coding Theorem

With $X = \{x, p(x)\}$ for the input letters, we send a length n message through a memoryless noisy channel specified by $p(y|x)$'s. The output letters $Y = \{y, p(y)\}$ can be found by knowledge of X and the channel.

Intuitively it seems we can send no more than $I(X;Y)$ bits per letter over the noisy channel, the value of which depends on the $p(y|x)$'s (channel) and $p(x)$'s (input ensemble). This is the noisy coding theorem.

Coding and Transmission of Messages Using Quantum States

The quantum equivalent of the previous situation is to replace message letters with quantum states. Suppose for a particular physical system we have the states $|\psi_x\rangle$ each occuring with probability $p(x)$, where $\sum_x p(x) = 1$. Then the density operator for a particular state (letter) is

$$\rho = \sum_x p(x) |\psi_x\rangle\langle\psi_x|.$$

Since the states $|\psi_x\rangle$ may not be mutually orthogonal, different states are not completely distinguishable, that is, they overlap in the state space, hence the entropy for this case is not

$$H(X) = -\sum_x p(x) \log p(x).$$

Two overlapping letters are not exactly two letters, they are effectively less than two letters, although always more or same as one letter.

Von Neumann Entropy

The **Von Neumann entropy** for the density operator defined previously is defined as

$$S(\rho) \equiv -\mathrm{tr}\,(\rho \log \rho)\,.$$

The logarithm of a matrix is defined as the inverse of the exponential of a matrix. For matrices A and B if

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = B,$$

then

$$\log B = A.$$

The logarithm of a matrix is normally very hard to calculate, but for diagonal matrix A where $A_{ij} = \delta_{ij} a_i$, its exponential is

$$(e^A)_{ij} = \delta_{ij} e^{a_i} = B_{ij},$$

so B is diagonal, with $B_{ij} = \delta_{ij} b_i$ we have

$$(\log B)_{ij} = \delta_{ij} \log b_i.$$

Since any density operator can be diagonalized, suppose the eigenvalues of ρ is λ_i, that is,

$$\rho = \sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|,$$

where the $|\varphi_i\rangle$'s are mutually orthonormal, then

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i.$$

This is the same as the entropy of an ensemble of letters each with probabilities λ_i, since all density operators have unit trace. This equality is not surprising since orthogonal states are completely distinguishable, hence can be treated as classical letters.

Density operators can also be treated as letters, for the ensemble $X = \{\rho_x, p(x)\}$, the density operator for each letter is

$$\rho = \sum_x p(x)\rho_x.$$

This is the most general case in which even individual letters are in a mixed state, but how can such a message be sent?

The Von Neumann entropy represents three physical quantities:

1. The quantum information per letter.

2. The classical information per letter.

3. The amount of entanglement.

Yet the theories and methods developed by use of the Von Neumann entropy may somehow be limited due to large correspondence with classical information theory. For example, letters are generally represented physically as mixed states rather than pure states, that is, without relative phase information. The Von Neumann entropy may be a special case of a more general complex entropy?

Quantum (Von Neumann) Relative entropy

This is the quantum version of relative entropy. For density matrices ρ_1 and ρ_2, the relative entropy of ρ_1 to ρ_2 is defined as

$$S(\rho_1||\rho_2) \equiv \mathrm{tr}\left(\rho_1 \log \rho_1\right) - \mathrm{tr}\left(\rho_1 \log \rho_2\right).$$

The relative entropy is likewise non-negative, and equals zero when $\rho_1 = \rho_2$.

Diagonalize both ρ_1 and ρ_2:

$$\rho_1 = \sum_i p_i |\psi_i\rangle\langle\psi_i|,\, \rho_2 = \sum_i q_i |\varphi_i\rangle\langle\varphi_i|,$$

then

$$\begin{aligned} & S(\rho_1||\rho_2) \\ = \; & S(\rho_1) - \mathrm{tr}\left(\rho_1 \log \rho_2\right) \\ = \; & \sum_i p_i \log p_i - \sum_i \langle\psi_i|\rho_1 \log \rho_2|\psi_i\rangle \\ = \; & \sum_i p_i \log p_i - \sum_i p_i \langle\psi_i| \log \rho_2|\psi_i\rangle \end{aligned}$$

$$\begin{aligned}
&= \sum_i p_i \log p_i - \sum_i p_i \langle\psi_i| \left(\sum_j (\log q_j) |\varphi_j\rangle\langle\varphi_j| \right) |\psi_i\rangle \\
&= \sum_i p_i \log p_i - \sum_i p_i \sum_j \left|\langle\psi_i|\varphi_j\rangle\right|^2 \log q_j \\
&= \sum_i p_i \left(\log p_i - \sum_j \left|\langle\psi_i|\varphi_j\rangle\right|^2 \log q_j \right)
\end{aligned}$$

Since the logarithm is strictly concave, we have

$$\sum_j \left|\langle\psi_i|\varphi_j\rangle\right|^2 \log q_j \leq \log \left(\sum_j \left|\langle\psi_i|\varphi_j\rangle\right|^2 q_j \right),$$

with equality if and only if $\forall i \exists j |\varphi_j\rangle = |\psi_i\rangle$, so

$$\begin{aligned}
S(\rho_1||\rho_2) &= \sum_i p_i \left(\log p_i - \sum_j \left|\langle\psi_i|\varphi_j\rangle\right|^2 \log q_j \right) \\
&\geq \sum_i p_i \left(\log p_i - \log \left(\sum_j \left|\langle\psi_i|\varphi_j\rangle\right|^2 q_j \right) \right)
\end{aligned}$$

in the case of equality,

$$S(\rho_1||\rho_2) = \sum_i p_i \left(\log p_i - \log q_i\right) \geq 0,$$

since S becomes a (classical) relative entropy.