PHY 4105: Quantum Information Theory
Lecture 5

Anil Shaji
School of Physics, IISER Thiruvananthapuram
(Dated: August 16, 2013)

Two random variables

If we have two random variables X and Y distributed according to the joint probability function $p(x, y)$ we can compute a joint entropy,

$$H(X, Y) = \sum_{x,y} p(x, y) \log p(x, y),$$

which quantifies our ignorance of both the variables taken together. We can also find the information/ignorance about one variable given that the distribution of the other is known. This is quantified by the conditional entropy,

$$H(X|Y) = \sum_y p(y) \Big(- \sum_x p(x|y) \log p(x|y) \Big).$$

The term within brackets is the information in X given $Y = y$ and that is, in turn, averaged over Y. Using Bayes' theorem, $p(x|y) = p(x, y)/p(y)$, we have

$$\begin{aligned} H(X|Y) &= -\sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(y)} \\ &= -\sum_{x,y} p(x, y) \log p(x, y) + \sum_y p(y) \log p(y) \\ &= H(X, Y) - H(Y). \end{aligned}$$

Rearranging the terms, we have the analogue of Bayes theorem in the language of Shannon entropies as

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

Finally we can define the mutual information $H(X : Y)$ which quantifies the reduction in information/ignorance about X on knowing Y (or vice-versa) as

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

The mutual information is symmetric between X and Y. The mutual information is positive, as

proven below:

$$\begin{aligned} H(X:Y) &= -\sum_x p(x)\log p(x) + \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(y)} \\ &= -\sum_{x,y} p(x,y)\log p(x) + \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(y)} \\ &= \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= H(p(x,y)\|p(x)p(y)) \geq 0. \end{aligned}$$

The properties of Shannon entropy for two variables is described best by the Venn diagram below (Fig. 1):

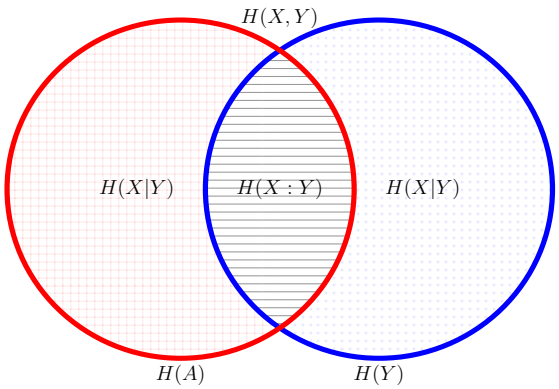

FIG. 1: The (red) circle on the left denotes the entropy associated with a state of system X while the (blue) circle on the right denotes the entropy associated with a state of system Y. The area on the left filled in with the (red) grid is the information missing about X given that all information about Y is available ($H(Y) = 0$) and so this area denotes the conditional entropy $H(X|Y)$. Similarly the area on the right filled in with (blue) dots denotes $H(Y|X)$. The overlap between the two circles filled with horizontal lines denotes the mutual information $H(X:Y)$ which is the information contained in X about Y and vice versa. The combined envelop of the two circles is the (classical) joint entropy $H(X,Y)$. From the diagram clearly $H(X:Y) = H(X) + H(Y) - H(X,Y)$ and also $H(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.

We can enumerate the properties of the Shannon information for two variables:

1.

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \Rightarrow H(X), H(Y) \leq H(X,Y).$$

2.

$$\begin{aligned} H(X:Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) = H(Y:X) \geq 0 \\ &\Rightarrow H(X) \geq H(X|Y) \\ &\Rightarrow H(Y) \geq H(Y|X) \\ H(X:Y) &\leq H(X),\ H(Y). \end{aligned}$$

3.

$$\begin{aligned} H(X,Y) &= H(X) + H(Y) - H(X:Y) \leq H(X) + H(Y) \\ &= H(X|Y) + H(Y|X) + H(X:Y) \end{aligned}$$

The statement $H(X,Y) \geq H(X) + H(Y)$ is called the strong sub-additivity of Shannon entropy and sub-additivity is something that we will come back to many times in the context of quantum information as well. Equality holds when X and Y are independent random variables. It is also intuitively clear that conditioning will reduce the entropy. That is,

$$H(X|Y,Z) \leq H(X|Y).$$

Strong Subadditivity: $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$ with equality holding when $Z \to Y \to X$ forms a Markov chain. Markov chains capture very well, and very often what is meant by information processing. A Markov chain is a sequence $X_1 \to X_2 \to \cdots$ of random variables such that X_{n+1} is independent of $X_1, \ldots X_{n-1}$ given X_n. In other words,

$$p(X_{n+1} = x_{n+1}|x_n, \ldots x_1) = p(x_{n+1}|x_n).$$

The **data processing inequality** looks at how a Markov chain loses information about its early values as time progresses. Suppose $X \to Y \to Z$ is a Markov chain, then

$$H(X) \geq H(X:Y) \geq H(X:Z).$$

The firs inequality is saturated if and only if given Y one can completely reconstruct X. The inequality tells us that if the random variable X is subject to noise, producing the new random variable Y then further data processing on our part cannot be used to increase the amount of mutual information between the output of the process and the original information X.
Proof:: We have already proved the first inequality, $H(X) \geq H(X:Y)$.

$$\begin{aligned} H(X:Y) &\geq H(X:Z) \\ \Rightarrow H(X) - H(X|Y) &\geq H(X) - H(X|Z) \\ \Rightarrow H(X|Y) &\leq H(X|Z). \end{aligned}$$

We can show that if $X \to Y \to Z$ is a Markov chain then $Z \to Y \to X$ is also a Markov chain (use Bayes theorem). This means that

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

So the problem is reduced to showing

$$H(X,Y,Z) - H(Y,Z) = H(X|Y,Z) \leq H(X|Z) = H(X,Z) - H(Z).$$

This is just the strong subadditivity property.

The probabilistic description of states and the entropic measures of information lets us formulate and understand the dynamics of information when when we manipulate realistic alternatives which encode the information. In the quantum case there are no realistic alternatives and there are no joint probability distributions for non-commuting variables. Probabilities enter the picture in a intrinsic way when quantum states and dynamics are coupled to readouts and measurements.