March 2024

Easier

In the parlour game Twenty Questions one person choses an object at random and another asks yes/no questions to determine what it is. A useless question does not cut down the options at all. But a good question, e.g. one that cuts down the number of options by a factor of eight, is worth three average questions that give a factor of two. We could measure the effectiveness of a question, the information in the answer, in terms of how many average (50/50) questions it would take to produce the same reduction in the number of options remaining. The useless question produced no information, whereas the good question produced 3 units of information, 3 *bits*.

1) Provide a motivation for the statement that the information we gain when we perform an experiment is proportional to the logarithm of the probability of the result:

$$\text{Information} = -\log_2(\text{Probability})$$

2) If we perform an experiment which has outcomes with probabilities P_i , show that the average amount of information we expect to gain by performing the experiment is the *entropy*:

$$H = -\sum_i P_i \log_2 P_i$$

1) In the game Twenty Questions we saw that the information gained when asking a question was related to how many possible objects remained. A response that reduced the options available by a factor of eight would be equivalent to three purely binary responses. If we take as our basic unit of information to be that of a binary response, then the information in a response is proportional to the logarithm of the factor by which the options were reduced. Clearly this need not only apply to powers of two. If our questions were poor and it took two questions to halve the number of options we could argue that each one was only equivalent to part of a question.

In order to specify an object we can expect to ask many questions. We could write down the responses on a piece of paper; e.g. "It is a car" or "It is a car. It is red. It has black seats. It is left hand drive." As broader categories tend to have shorter descriptions than narrow ones it would make sense if our concept of information was proportional to the length of the description. This would indicate that our notion of information is additive.

These two properties – information as logarithm of the reduction factor, and information being additive – are not exclusive and can be combined. We could make the length of the text describing a response proportional to the number of equivalent binary questions, and if we do that the total length of the description (the total information) would be proportional to the logarithm of the number of objects remaining as a fraction of the total:

$$\text{Information} \propto \log_2(\text{Probability})$$

To provide a standard for this information measure we need to define a scale for measuring the length of text, and by convention we use positive 'bits' in which case

$$\text{Information} = -\log_2(\text{Probability})$$

Lastly, in order to relate this to an experiment we need only note that our objects could be the outcomes of the experiment.

2) The second part of the question is much easier. To find the average we find the sum of the values multiplied by their probability they occur. If the entropy is the average information we have

$$H = \sum_i P_i \,\text{Information}(i) = -\sum_i P_i \log_2(P_i)$$

Harder

If two events are statistically independent the probability of them both happening is the product of the probabilities of each one separately:

$$P_{ij} = P_i P_j$$

1) If they are correlated (i.e. $P_{ij} \neq P_i P_j$) the average information content will be reduced by the amount of *mutual information*. Find a formula for the mutual information in terms of the P_{ij}.

Entanglement is a property that leads to correlations in quantum experiments. A quantum system consisting of two *independent* sub-systems will be in a product state:

$$|\Psi\rangle = |\Theta\rangle \otimes |\Phi\rangle$$
$$= \sum_{i,j} \theta_i \varphi_j |i\rangle \otimes |j\rangle$$

where i and j denote eigenstates of the first and second sub-systems. In general, however, a system will be in an entangled state

$$|\Psi\rangle = \sum_{i,j} \psi_{ij} |i\rangle \otimes |j\rangle$$

2) By forming the obvious analogy with the first part of the question, find a maximally entangled 2 qubit state; i.e. a superposition of eigenstates where i and j take the values 0 or 1, and the mutual information is as large as possible. These states are called Bell States, and they are not unique.

1) The average information (i.e. the entropy) in the probability distribution is

$$H_{IJ} = -\sum_{i,j} P_{ij} \log_2(P_{ij})$$

The probability that the first event (the one whose outcomes is labelled by i) is found by summing P_{ij} over all the possible outcome for the second even, i.e. by summing over j:

$$P_i = \sum_j P_{ij}$$

And by analogy we also have

$$P_j = \sum_i P_{ij}$$

The average information content in these two separately is

$$H_I = -\sum_i P_i \log_2(P_i)$$

and

$$H_J = -\sum_j P_j \log_2(P_j)$$

Hence the mutual information, which is usually given the symbol I, is:

$$\begin{aligned} I &= H_I + H_J - H_{IJ} \\ &= -\sum_i P_i \log_2(P_i) - \sum_j P_j \log_2(P_j) + \sum_{i,j} P_{ij} \log_2(P_{ij}) \\ &= \sum_{i,j} P_{ij} \log_2\left(\frac{P_{ij}}{P_i P_j}\right) \end{aligned}$$

2) Writing $|i,j\rangle$ for $|i\rangle \otimes |j\rangle$, a general 2 qubit state is of the form

$$|\Psi\rangle = \psi_{00}|0,0\rangle + \psi_{01}|0,1\rangle + \psi_{10}|1,0\rangle + \psi_{11}|1,1\rangle$$

By the Born rule, the probabilities of these states are:

$$P_{ij} = |\psi_{ij}|^2$$

And these probabilities can be used to calculate the mutual information.

It is not immediately obvious what the state of greatest mutual information ought to be so it is worth re-examining the formula for the mutual information:

$$
\begin{aligned}
I &= H_I + H_J - H_{IJ} \\
&= -\sum_i P_i \log_2(P_i) - \sum_j P_j \log_2\left(P_j\right) + \sum_{i,j} P_{ij} \log_2(P_{ij})
\end{aligned}
$$

Instead of combining all of the terms immediately into a single sum we can write:

$$
\begin{aligned}
I &= H_I - \sum_j P_j \log_2\left(P_j\right) + \sum_{i,j} P_{ij} \log_2(P_{ij}) \\
&= H_I + \sum_{i,j} P_{ij} \log_2\left(\frac{P_{ij}}{P_j}\right) \\
&= H_I - \sum_j P_j \left(-\sum_j \frac{P_{ij}}{P_j} \log_2\left(\frac{P_{ij}}{P_j}\right)\right)
\end{aligned}
$$

The combination P_{ij}/P_j is a probability in its own right: it is the probability of i given j. Given that probabilities are positive and less than 1, the entropy (the sum in brackets) is positive (all entropies are positive), so the mutual information is less than or equal to the entropy H_I. So the mutual information shared between with a subsystem cannot be greater than the entropy of that subsystem.

In our case it is easy to arrange equality simply by setting the probabilities to have one of the following forms:

$$
\begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 1-a \end{bmatrix} \text{ or } \begin{bmatrix} 0 & a \\ 1-a & 0 \end{bmatrix}
$$

If we do this we find

$$
H_I = H_J = H_{IJ} = a \log_2(a) + (1-a) \log_2(1-a)
$$

Unsurprisingly given the symmetry, the maximum occurs at $a = 1/2$. Tracking back first to the P_{ij} and thence to the ψ_{ij} we find the maximally entangled states to be either

$$
|\Psi\rangle = \frac{1}{\sqrt{2}}\left(e^{i\theta}|0,0\rangle + e^{i\varphi}|1,1\rangle\right)
$$

or

$$
|\Psi\rangle = \frac{1}{\sqrt{2}}\left(e^{i\theta}|0,1\rangle + e^{i\varphi}|1,0\rangle\right)
$$

where $e^{i\theta}$ and $e^{i\varphi}$ are arbitrary phase factors.