Quantum entropy

Entropy is a measure of the information content of a random variable. Classical entropy, measured in bits and called *Shannon entropy* after its inventor, is for a discrete random variable X with probabilities $\Pr[X = i] = p_i$, defined by the formula

$$H(X) = -\sum_i p(i) \log p_i .$$

This value can be interpreted informally as the average number of bits of the address of the random event that occurred. The basic property of Shannon entropy that shows that it actually expresses information content is the source coding theorem, also called the compression theorem or the noiseless channel capacity theorem. It shows that a sequence n of independent copies of a random variable X can be encoded by a sequence of bits of length $nH(X)$ with error probability asymptotically going to zero for large n. (The theorem is proved by noting that the vast majority of sequences have the expected distribution of the number of letters, and showing that there are almost exactly $2^{nH(X)}$ of such typical sequences of length n.)

The total information of a pair of random variables $H(X,Y)$ is at most the sum of $H(X) + H(Y)$, but can be smaller. For example, if Y is a function of X, then (X,Y) is completely determined by X and $H(X,Y) = H(X)$. The remaining information content of Y given knowledge of X is the conditional entropy $H(Y \mid X)$. (We should correctly say „average entropy“, since it is the average over the different values of X. The entropy $H(Y)$ is itself an average over different values of Y.) So the value $H(Y) - H(Y \mid X)$ expresses how much information we learn about Y if we know X. Similarly, $H(X) - H(X \mid Y)$ is a measure of the information Y reveals about X. The expected relation $H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$ holds. Value

$$I(X:Y) := H(X) + H(Y) - H(X,Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

is thus a measure of the dependence of the two quantities and is called *mutual information*. Note that this value is symmetric in X and Y.

The information content of a quantum system is given by the uncertainty about the measurement results. In other words, the result of a given measurement of a given system is a random variable with some entropy. However, the entropy of a quantum state must take into account all possible measurements. If the system is in a pure state, there is a measurement whose result is given uniquely (it is any measurement in the basis containing the measured state). Thus, the entropy of a quantum system is non-zero only in the case of mixed states and comes from the uncertainty about the prepared state. However, it is also affected by the nature of the ensemble. Let us illustrate this with the states that occur in the BB84 protocol. If we know what base Alice encodes in, but we don't know what bit she encodes, the system is in a state

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}E\,.$$

Measuring in the canonical basis is equivalent to accepting a random bit. The fact that the value was encoded using quantum states rather than classically plays no role here. The entropy of such a state should therefore be equal to one. If, on the other hand, we know that Alice encoded zero, but we do not know in what basis, we get a matrix

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +| = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix} \sim \begin{pmatrix} \frac{2+\sqrt{2}}{4} & 0 \\ 0 & \frac{2-\sqrt{2}}{4} \end{pmatrix} .$$

The diagonal form is with respect to the (normalized) eigenvectors

$$|v_1\rangle = \frac{1}{\sqrt{4-2\sqrt{2}}} \begin{pmatrix} -1 \\ \sqrt{2}+1 \end{pmatrix}, \qquad |v_2\rangle = \frac{1}{\sqrt{4+2\sqrt{2}}} \begin{pmatrix} 1 \\ \sqrt{2}-1 \end{pmatrix} .$$

Thus, we get the same density matrix if we choose v_1 or v_2 with probabilities $\frac{2+\sqrt{2}}{4}$ and $\frac{2-\sqrt{2}}{4}$. The classical entropy of such a random variable is

$$-\frac{2+\sqrt{2}}{4} \log \frac{2+\sqrt{2}}{4} - \frac{2-\sqrt{2}}{4} \log \frac{2-\sqrt{2}}{4} \doteq 0.6 .$$

The entropy is not equal to one because the selected bit was encoded into two states that are not completely distinguishable, thus some information was lost.

We can also use this example to illustrate the independence of the measurement result from the way the density matrix was created. Let us examine the probability of obtaining a result corresponding to $|0\rangle$, $|1\rangle$ when measuring in the $|0\rangle$ basis. By Postulate 3', this is $\operatorname{tr}(|0\rangle\langle 0|\rho) = 3/4$. This corresponds to simple reasoning: with probability one-half we have a state $|0\rangle$, and then the measurement result corresponds to $|0\rangle$ with certainty; with probability one-half we have a state $|+\rangle$, where the result corresponds to $|0\rangle$ with probability one-half. Similarly, we could verify that if $|v_1\rangle = \alpha_1|0\rangle + \beta_1|1\rangle$ and $|v_2\rangle = \alpha_2|0\rangle + \beta_2|1\rangle$, then

$$\frac{2+\sqrt{2}}{4}|\alpha_1|^2 + \frac{2-\sqrt{2}}{4}|\alpha_2|^2 = \frac{3}{4} .$$

These examples lead to the definition of the *von Neumann entropy* density matrix ρ. It is the entropy of the random variable corresponding to the choice of the eigenvectors ρ, which can be written concisely as

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

Let us list some properties of quantum entropy. For the entropy of the mixed state $\rho = \sum_i p_i \rho_i$, the following holds

$$S(\rho) \leq H(X) + \sum_i p_i S(\rho_i),$$

where X is the random variable of the state selection ρ_i, i.e. a discrete random variable with probability $\Pr[X = i] = p_i$. Equality holds if and only if the states ρ_i are distinguishable, i.e. if they are defined on mutually orthogonal spaces. The entropy of a state ρ is at most the entropy of the corresponding choice of ρ_i plus the average entropy contained in ρ_i itself. If the states ρ_i are pure, we get $S(\rho) \leq H(X)$. If they are pure and distinguishable, we have $S(\rho) = H(X)$. Then it is a classical random variable, it does not matter that we encode its values by quantum states.

For complex systems, *strong subadditivity* holds:

$$S(\rho^{ABC}) + S(\rho^B) \leq S(\rho^{AB}) + S(\rho^{BC}) .$$

Let us define the mutual information of two quantum states as

$$S(\rho^A : \rho^B) := S(\rho^A) + S(\rho^B) - S(\rho^{AB}) .$$