We did know the concept of thermodynamics entropy: $dS = \frac{\delta Q}{T}$

It underlies the classical thermodynamics and was proposed from an engineering point of view about work extraction.

Also, we are familiar with the definition of Classical Boltzmann entropy. It is defined at equilibrium as the maximum's system entropy:

Considering that the probability of a macrostate of a gas could be identified with the number of compatible microstates. In the following way:

Suppose you have 3 true coins: with head(H) & tail(T) sides

All microstates are equally probable (as it should be in Boltzmann entropy)

We have 8 possible outcomes, microstates ($2^N = 2^3 = 8$)

and 4 possible macrostates

- 3H: 1 microstate (HHH)
- 2H,1T: 3 microstates (THH, HTH, HHT)
- 1T,2H: 3 microstates (TTH, THT, TTH)
- 3T: 1 microstate (TTT)

Microstates of a certain macrostate = multiplicity Ω

Being at a certain macrostate implies being in one of its microstates. The more microstates, the greater the uncertainty about being in which microstate (2^{-N})

$S = K \ln \Omega$ → measures uncertainty

↕

Boltzmann constant as scaling factor between macroscopic and microscopic.

Let's dig on the idea of information theoretic interpretation of entropy: It detaches from details of thermodynamics, phase space, Hilbert space, etc.; and focuses on the gain of information of an observer interacting with a system.

↳ Then, this concept of entropy can be used when the nature of the system is only partially known.

Shannon entropy

Shannon was interested in quantifying the information that can be transferred via a communication channel.

Suppose that one receives a message that consists of a string of symbols a and b (i.e. aaabababbb...)

a occurs with probability p

b occurs with probability $1-p$

Consider long messages with n letters ($n \gg 1$)

Then the number of distinct strings is of order the binomial coefficient $\binom{n}{np}$

$$N = \frac{n!}{(pn)![(1-p)n]!} \Rightarrow \log N = \log\left(\frac{n!}{(pn)![(1-p)n]!}\right) \sim$$

[Stirling approx. $\log n! = n\log n - n + O(\log n)$]

$$\sim n\log n - n - [np\log np - np + n(1-p)\log n(1-p) - n(1-p)] =$$

$$= \log n[n - np - n(1-np)] - np\log p - n(1-p)\log(1-p) = nS$$

where S is the Shannon entropy per letter:

$$S = -p\log p - [(1-p)\log(1-p)]$$

Total number of messages of that length: 2^{nS}

# of bits of information one gains observing a message: nS

In general we have not a binary alphabet (i letters, each one with different probability). Similar derivation gives:

$$S = \sum_i - p_i \log p_i$$

The amount of possible messages is at least one, so:

$$S \geq 0$$

- $S = 0 \Rightarrow$ just one message (no uncertainty!)
- Maximum entropy: In an alphabet with K letters, all probabilities p_i are $1/K$

$$S = -\sum_{i=1}^{K} \frac{1}{K} \log \frac{1}{K} = \log K$$

Same probability, less information about the possible message

Bit: Amount of information conveyed when we learn the outcome of flipping a coin ($\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2} = 1$). So $S = 1$

It's the maximum information we can have after seen the result. If we cheat and we know the result is going to be head: $p_H = 1$, $p_T = 0 \Rightarrow S = 0$

There is no uncertainty

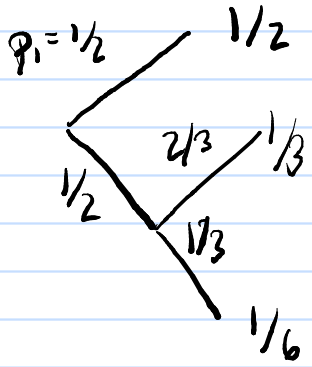
Ex. See previous 3 coins example and compute the entropy

See that result is: $S = 1'811$ bits

Basic properties

- S is continuous in p_i
- S is extensive
- If all p_i are equal to $p_i = \frac{1}{n}$, S is a monotonically increasing function of n (probe it as exercise)
- If breaking a choice into two successive choices, the original S should be the weighted sum of the individual values of S

Ex.



$$S\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = S\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} S\left(\frac{2}{3}, \frac{1}{3}\right)$$

Conditional entropy

Imagine Alice sends a message to Bob that consists in a string of many letters. Each letter is represented by a random variable X, whose possible values are $x_1, \dots, x_k$. As the communication channel is not perfect, there would be errors on it. Then Bob receives a set of letters of a random variable Y with possible values $y_1, \dots y_k$.

For a particular letter, the probability that Bob hears $Y = y_j$ has to be summed over all possible letters Alice could have sent

$$P_Y(y_j) = \sum_i \underbrace{P_{X,Y}(x_i, y_j)}_{\text{Probability that when Alice sends } x_i \text{, Bob hears } y_j}$$

Once Bob hears $Y = y_j$, he estimates the probability Alice sent x_i

$$P_{X|Y}(x_i|y_j) = \frac{P_{X,Y}(x_i, y_j)}{P_Y(y_j)} \Rightarrow \text{Conditional probability}$$

Then you calculate the entropy associated with the uncertainty on the letter Alice have really sent.

$$S_{X|Y=y_i} = -\sum_i P_{X|Y}(x_i|y_j) \log P_{X|Y}(x_i, y_j)$$

If you want to do this for all possible values of Y, giving a total entropy, you need to average previous entropy with the probability distribution of Y

$$S_{X|Y} = \sum_j P_Y(y_j) S_{X|Y=y_j} = S_{XY} - S_Y$$

It is the uncertainty that Bob still has after observing the message Y.

In other words, it quantifies the amount of information needed to describe the outcome of a random variable Y given that the value of another random variable X is known.

Mutual information

Information about **X** that Bob gains when he receives **Y**

$$I(X;Y) = S_X + S_Y - S_{XY} = S_X - S(X|Y)$$

It measures how much we learn about **X** by measuring **Y**. It quantifies the level of correlation between sources **X** and **Y**. If the sources are independant

$$S_X = S(X|Y) \Rightarrow I(X;Y) = 0$$

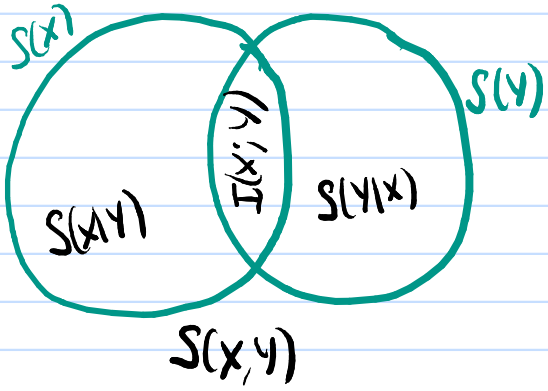
Properties

- Symmetric (Ex: probe it)
- $I(X;Y) \geq 0$
- Subadditivity of the entropy: $S_X + S_Y - S_{XY} \geq 0$
- Strong subadditivity: $S_{XY} + S_{YZ} \geq S_Y + S_{XYZ}$

  $$I(X;YZ) \overset{!!}{\geq} I(X;Y)$$

  ↳ or monotonicity of mutual information.

Graphic idea of these concepts:



Relative entropy

Suppose we are observing a random variable X. We have a theory predicting a probability distribution Q_X, saying that the prediction for final state $X = x_i$ is given by $q_i = Q_X(x_i)$.

But, maybe, our theory is wrong, and the process is described by another theory P_X, predicting $p_i = P_X(x_i)$

After observing the process N times, how sure could we be about the initial hypothesis of the theory?

The probability of what we have seen:

$$P = \underbrace{\prod_{i=1}^{S} q_i^{p_i N}}_{\text{Probability any specific such sequence (assuming } Q_X \text{ as correct)}} \underbrace{\frac{N!}{\prod_{j=1}^{S} (p_j N)!}}_{\text{number of sequences in which the outcome } x_i \text{ occurs } p_i N \text{ times}}$$

For large N

$$P \sim 2^{N \underbrace{\sum_i (\log p_i - \log q_i)}_{S(P_X \| Q_X) \text{ relative entropy}}}$$

(or Kullback-Liebler divergence)

It is non-negative, and zero only if $P_X = Q_X$, meaning the initial hypothesis was correct.

See that is also monotonous and asymmetric.

Von Neumann entropy

Consider a bipartite system with subsystems called A and B. The total Hilbert space is the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$.

In the simplest case the state of the total system ψ_{AB} is separable

$$\psi_{AB} = \psi_A \otimes \psi_B$$

But, generally a pure state ψ_{AB} is entangled. Any pure state can be written

$$\psi_{AB} = \sum_i p_i \, \psi_A^i \otimes \psi_B^i$$

Suppose we have a single system A, and a mixed state on it. We can always construct a larger system AB and a pure state on it ψ_{AB}. This is called purification. This process is not unique and implies adding a Hilbert space $\mathcal{H}_B$ with orthonormal states ψ_B. However, any other set of states (purifying the system) can be found by unitary transformation.

Note that for the following I will also use sometimes the notation of density matrix, that in the general case for a system A

$$\rho_A = \sum_i p_i \, |\psi_A^i\rangle\langle\psi_A^i|$$

Having a total system AB, we can always integrate out an unobserved subsystem and obtain the reduced density matrix for the other subsystem, via taking the partial trace

$$\rho_A = Tr_B \rho_{AB}$$

leaving a density operator on $\mathcal{H}_A$

We can define the von Neumann entropy in this bipartite system as

$$S_A = - Tr_B \rho_A \log \rho_A$$

[We also can see it as the quantum version of Shannon. If we choose an orthonormal basis where ρ is diagonal and using Taylor expansion for the logarithm, we get the Shannon entropy].

Let's see a couple of examples.

Ex 1: Each subsystem A and B consists in a single qubit. So Hilbert space of the system is spanned by $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ (first bit refers to A, second to B)

Suppose the system in an entangled pure state:

$$|\psi\rangle = \frac{1}{2}(|00\rangle + |11\rangle)$$

$\rho = |\psi\rangle\langle\psi|$, so the reduced matrix of subsystem A

$$\rho_A = Tr_B \rho = \frac{1}{2} {}_B\langle 0| (|000\rangle + |111\rangle)(\langle 111| + \langle 000|)|0\rangle_B$$

$$+ \frac{1}{2} {}_B\langle 1| (|000\rangle + |111\rangle)(\langle 111| + \langle 000|)|1\rangle_B =$$

$$= \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \propto \mathbb{1}$$

ρ_A is maximally mixed

The entanglement entropy of A:

$$S_A = -Tr_B \rho_A \log \rho_A = 2 \times \frac{1}{2} \log \frac{1}{2} = \log 2$$

This saturates the maximum possible entropy (because it is a maximally entangled state)

Ex 2: Now we have a separable pure state:

$$|\psi\rangle = \frac{1}{2}\left((|1\rangle_A + |0\rangle_A) \otimes (|1\rangle_B + |0\rangle_B)\right)$$

See that $\rho_A = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

This is a pure state for A, so $S_A = 0$

Properties of von Neumann entropy

- Continuous
- Extensive
- Subadditive : $S(AB) \leq S(A) + S(B) \Rightarrow I(X:Y) \geq 0$
- Upper bounded $S(\rho) \leq \log K$
  for maximally mixed states
- $S \geq 0$ and zero for a pure state
- In a bipartite system in a pure state $S(AB) = 0$
  $$S(A) = S(B)$$
  I see it also that system B is purification of A
- Strong subadditivity $S(ABC) + S(B) \leq S(AB) + S(BC)$
- Triangle inequality $S(AB) \leq |S(A) - S(B)|$
- Invariant under unitary transformation
- It's concave

Quantum conditional entropy

$$S(A|B) = S_{AB} - S_B$$

Unlike the classical case : it can be negative !

Supose AB in an entangled pure state $S_{AB} = 0$ but B is in a mixed state $S_B > 0$ so $S(A|B) < 0$

↓

for entangled states

Quantum mutual information

$$I(A;B) = S_A + S_B - S_{AB}$$

$I(A;B) \geq 0$ and $I(A;B) = 0$ only if density matrix factorizes $\rho_{AB} = \rho_A \otimes \rho_B$

Monotonicity $I(A;BC) \geq I(A;B)$

↳ or strong subadditivity $S_{AB} + S_{BC} \geq S_B + S_{ABC}$

⚠ Only measure of entanglement in bipartite systems!

Relative quantum entropy

$$S(A\|B) = \mathrm{Tr}_A(\log A - \log B) \geq 0$$

Monogamy of entanglement

Consider a system ABCD in a pure state, so

$S_{AB} = S_{CD}$, $S_{ABC} = S_D$

So taking strong subadditivity

$$S_{CD} + S_{BC} \geq S_B + S_D$$

For instance $S(C|D) = S_{CD} - S_D$ can be negative, or

$S(C|B) = S_{BC} - S_B$ can be negative, but

$$S(C|D) + S(C|B) \geq 0$$

So C can be entangled with D reducing $S(C|D)$ or with B reducing $S(C|B)$, but not both!

This is related to the monogamy of entanglement. If 2 subsystems **A** and **B** are maximally entangled, they cannot be entangled with a third subsystem **C**.

Remember also that von Neumann entropy for a closed system under Hamiltonian evolution stays strictly constant; this expresses the preservation of information in the evolution of the density matrix.

$$\Delta S = 0 \Rightarrow \text{unitarity}$$