Entropy as Ignorance: Information

Recall if we had N atoms and place one of them into one of two boxes without looking, the system gains $k_B \ln 2$ worth of entropy. This suggests entropy has something to do with knowledge about the system.

The most general interpretation of entropy is as a measure of our ignorance of a system.

For example, the equilibrium state of a system maximizes the entropy because we have lost all information about the initial conditions except for the conserved quantities.

Entropy, then, represents our knowledge about the system.

Equilibrium systems are a subset of the more general probability distributions, for which we can give a more general definition of entropy.

Non-equilibrium entropy

Out of equilibrium, we may describe our partial knowledge about the system as a probability distribution ρ, defining an ensemble of states.

In the discrete case, we know from counting" entropy that

$$S_{counting} = k_B \ln [\#\text{ of configurations}]$$

For M equally likely states, the i^{th} state occurs with probability

$$p_i = \frac{1}{M} \quad \text{(independent events)}$$

Now,

$$S_{\text{counting}}(M) = k_B \ln M = -k_B \ln\left(\frac{1}{M}\right)$$

$$= -k_B \langle \ln p_i \rangle$$

So in the case of a discrete distribution of states,

$$\boxed{S_{\text{Discrete}} = -k_B \langle \ln p_i \rangle = -k_B \sum_i p_i \ln p_i}$$

Out of equilibrium, S_{Discrete} describes the entropy given the probability distribution. This is <u>statistically general</u>.

In the continuous case, any non-equilibrium state of a classical Hamiltonian system is described with a probability density in phase space $\rho(\mathbb{P},\mathbb{Q})$.

Then, in general,

$$S_{\text{non-equilibrium}} = -k_B \langle \ln \rho \rangle = -k_B \int_{E < H(\mathbb{P},\mathbb{Q}) < E+\delta E} \frac{d\mathbb{P}\, d\mathbb{Q}}{h^{3N}} \rho(\mathbb{P},\mathbb{Q}) \ln \rho(\mathbb{P},\mathbb{Q})$$

In the microcanonical ensemble

$$\rho_{\text{equilibrium}} = \frac{1}{\Omega(E)\delta E}$$

Then the nonequilibrium definition of entropy is shifted from our equilibrium definition $S = k_B \ln \Omega$ by a tiny, negligible amount

$$\frac{k_B \ln \delta E}{N} \quad \text{per particle}$$

$$S_{micro} = -k_B \ln \rho_{equilibrium} = k_B \ln (\Omega(E)\delta E)$$

$$= k_B \ln \Omega + k_B \ln \delta E$$

The arbitrary choice of δE in the microcanonical ensemble then corresponds to any arbitrary choice of the zero of classical entropy.

$$S_{non\,equilibrium} = -k_B \int_{E < \mathcal{H}(\mathbb{P},\mathbb{Q}) < E+\delta E} \frac{d\mathbb{P}\, d\mathbb{Q}}{h^{3N}} \rho(\mathbb{P},\mathbb{Q}) \ln \rho(\mathbb{P},\mathbb{Q})$$

is defined for $H(\mathbb{P},\mathbb{Q}) \Rightarrow$ the microscopic, time-reversal invariant laws of motion.

Therefore, we can guess that these microscopic entropies will be time independent, since microscopically, the system does not know in which direction of time entropy should increase (See prob 5.7)

Information Entropy

In information theory, entropy is a measure of the uncertainty in a random variable.

For example, we can define a bit as the basic unit of information. Each bit occurs in a string and can have the value 0 or 1 with equal probability.

By definition, let's say that a single toss of a fair coin (which acts like a bit) has an entropy of one bit. A series of two tosses has an entropy of two bits, etc. Therefore, the number of fair coin tosses is its entropy in bits.

Mathematically:

$$S_{Bit} = \log_2 (2^N) = N$$

This fixes the entropy rate in a string of N bits to be one bit per toss.

We can generalize this using Shannon entropy, which takes the non equilibrium form

$$S_S = -\sum_i p_i \log_2 p_i \qquad \text{Entropy measured in bits}$$

Note each bit doubles the number of possible states Ω, so $\log_2 \Omega$ is the number of bits

When applied to an ensemble of possible messages, or images, shannon entropy can be used to put a fundamental limit on the amount they can be compressed.

For example, in image transmission if the last six pixels were white, the region being depicted is likely to be a white background, AND the next pixel is also likely to be white.

Six pixels has a low shannon entropy, and is therefore "predictable"

With each stage of image compression information is lost, but the dataset reduces its Shannon entropy.

As you will see in the homework, messages passed along a network communicate information, reducing the information entropy for the receiver

This may be the most general form of entropy (independent of any well defined system) and has many deep implications.

Final Exam

Assigned Friday, due June 12 @ 7pm (Time of final)

5 problems from Sethna, some challenging. This is like a typical grad school final (except w/ way more time!)

Work alone. You may ask me any questions or clarifications. To be fair, you must EMAIL questions, so that Q/As can be seen by everybody