

Entropy: From the Boltzmann equation to the Maxwell Boltzmann distribution

A formula to relate entropy to probability

Often it is a lot more useful to think about entropy in terms of the probability with which different states are occupied. Lets see if we can describe entropy as a function of the probability distribution between different states.

$$W_{N \text{ particles}} = \frac{N!}{n_1! n_2! \dots n_t!}$$

stirling

$$W_{N \text{ particles}} = \frac{(N/e)^N}{(n_1/e)^{n_1} (n_2/e)^{n_2} \dots (n_t/e)^{n_t}} = \frac{N^N}{n_1^{n_1} n_2^{n_2} \dots n_t^{n_t}}$$

with

$$p_i = \frac{N}{n_i}$$

$$W_{N \text{ particles}} = \frac{1}{p_1^{n_1} p_2^{n_2} \dots p_t^{n_t}}$$

take ln

$$\ln W_{N \text{ particles}} = - \sum_{i=1}^t n_i \ln p_i$$

divide $N_{\text{particles}}$

$$\ln W_{1 \text{ particle}} = - \sum_{i=1}^t p_i \ln p_i$$

times k

$$k \ln W_{1 \text{ particle}} = -k \sum_{i=1}^t p_i \ln p_i = S_{1 \text{ particle}}$$

and

$$S_{N_A} = -N_A k \sum_{i=1}^t p_i \ln p_i = -R \sum_{i=1}^t p_i \ln p_i$$

Think about how this equation behaves for a moment. If any one of the states has a probability of occurring equal to 1, then the $\ln p_i$ of that state is 0 and the probability of all the other states has to be 0 (the sum over all probabilities has to be 1).

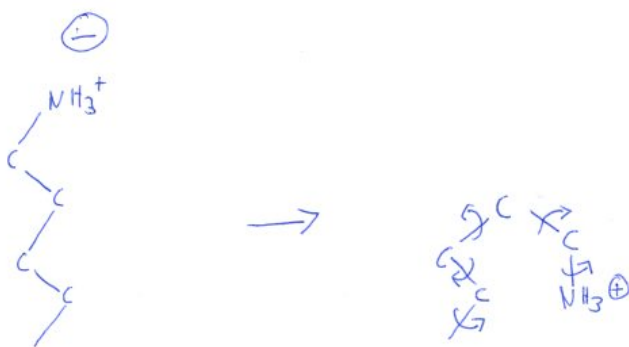
So the entropy of such a system is 0. This is exactly what we would have expected. In our coin flipping case there was only one way to be all heads and the W of that configuration was 1. Also, and this is not so obvious, the maximal entropy will result if all states are

equally populated (if you want to see a mathematical proof of this check out Ken Dill's book on page 85). In other words a completely flat distribution of probabilities, has the highest entropy.

Calculating with our formulas for entropy

We now have spent a lot of time talking about coin flips and tetrahedrons and derived two equations for entropy. It is now time to put some flesh on the bones of this theory and use it.

First, let's look at a simple molecular example. Let's say we have a lysine side-chain that is being pulled straight by an external electrostatic field. We now turn off the field instantaneously. Then we watch the reaction in which the stretched out lysine slowly starts to adopt a random conformation. What is the standard free energy for this reaction?



Our reaction is:

Lysine_{ordered} → Lysine_{disordered}

$$\text{And } \Delta G^0 = \Delta H^0 - T \Delta S^0$$

For now we can assume that ΔH^0 is zero for this reaction so

$$\Delta G^0 = - T \Delta S^0$$

How can we calculate ΔS^0 ? Simple, we just use the Boltzmann Equation.

$$S^0 = R \ln W$$

Of course there is only one way to be fully extended so the W for the stretched out form is 1 and $S^0 = 0$.

How many ways are there for the Lysine to be disordered? One might be tempted to think that this number is infinity, but common chemical knowledge tells us otherwise.

Because of steric clashes, each of the carbon bonds will adopt only one of three different rotamers. So if we assume that all combinations of rotamers would be possible we would get a W for our completely disordered form of:

$$W = 3 \cdot 3 \cdot 3 \cdot 3 = 81$$

so

$$\begin{aligned} S^0_{\text{disordered}} &= R \ln 81 \\ &= R \cdot 4.4 \end{aligned}$$

then

$$\begin{aligned} \Delta G^0_{\text{ordered} \rightarrow \text{disordered}} &= -T (\Delta S^0_{\text{disordered}} - \Delta S^0_{\text{ordered}}) \\ &= -T (\Delta S^0_{\text{disordered}} - 0) \\ &= -T R \cdot 4.4 \end{aligned}$$

and with $RT = 0.592 \approx 0.6$ kcal/mol

$$G^0_{\text{ordered} \rightarrow \text{disordered}} = 2.64 \text{ kcal/mol}$$

A common question / objection.

Just when I turn around from finishing this calculation on the blackboard there are usually a number of students who will object to this result, because the lysine side chain is never quite rigid in either the “fully-extended” conformation or in any of the other rotamers. Instead, there are clearly many conformational substates and vibrations that contribute to entropy in each one of these states and because I neglect all these substates my cartoonish representation of lysine will lead to completely meaningless number for the entropy change. After all there must be thousands of substates for each one of the rotamers. How can we know exactly how many substates there are and how do you define what a substate is anyway? **These are all very good points!** (And it is always good to use your common sense and experience with other stuff to double check what you learn in this class – or in any other class.)

But before you all throw up your hands and conclude that doing all the math that got us to this point was a complete waste of time, I would encourage you to use this very math to see, if the presence of this large number of substates or the choice when we call something a separate substate actually impacts our results.

So for the sake of the argument, lets assume that the extended conformation actually consists of 10,000 conformational substates and that the same is true for each one of our rotamers.

$$W_{ordered} = 1 \cdot 10,000$$

so

$$S^0_{ordered} = R \ln(1 \cdot 10,000)$$

and

$$W_{disordered} = 81 \cdot 10,000$$

$$S^0_{disordered} = R \ln(81 \cdot 10,000)$$

$$\Delta S^0_{ordered \rightarrow disordered} = R \ln(81 \cdot 10,000) - R \ln(1 \cdot 10,000)$$

$$\text{with } \ln(a \cdot b) = \ln a + \ln b$$

$$\Delta S^0_{ordered \rightarrow disordered} = R(\ln 81 + \ln 10,000) - R(\ln 1 + \ln 10,000)$$

$$\Delta S^0_{ordered \rightarrow disordered} = R \cdot \ln 81 - R \cdot \ln 1$$

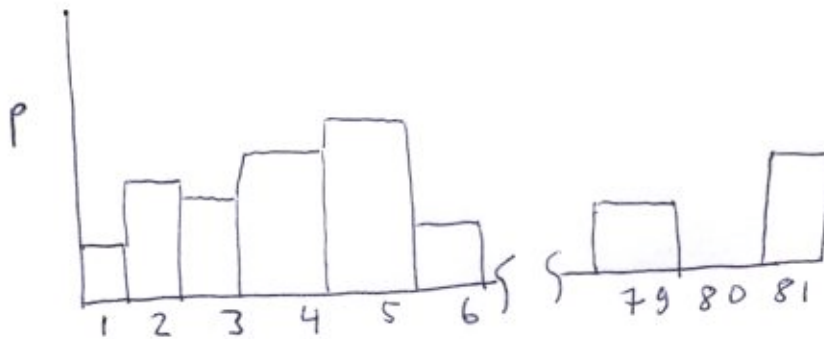
$$\Delta S^0_{ordered \rightarrow disordered} = R \cdot 4.4$$

As you see, the fact that each of our rotamers may actually consist of many, many substates does not effect the entropy from going to the ordered state to the disordered state, nor does the way we define what makes a substate a substate. We could say that each substate actually consists of another 1000 sub-substates and we would get the very same result.

Of course, the calculation above assumes, that each one of the rotamers has the same number of substates and this may not be true. However, we also do not have any reason to believe that any one of the rotamers would have a much larger number of substates than all the other ones. Feel free to do the math to see how our results change if you have only half as many substates in the extended form than in all the other forms. You will see that even such a drastic difference in the number of substates only results in a 10% change in the entropy. So in the absence of any information to the contrary, assuming that all of them have the same number of substates is a very good assumption and, even if the number of substates differs substantially, the error resulting from violating that assumption is not very dramatic.

Another point

One point I have to concede though is that not all of the 81 states will be occupied equally. For example, we did not account for the fact that there will be some combinations of rotamers that will not be occupied because of steric collisions. So if we really wanted to know the entropy, we could perform a molecular dynamics simulation in a computer of the free lysine sidechain.



We could then plot out the probability, with which each of the rotamers gets adopted. And use the formula

$$S^0 = -R \sum_{i=1}^I p_i \ln p_i$$

to calculate the entropy of the disordered form that way.

If restraints are present a system will tend towards the maximal entropy configuration that is consistent with these restraints.

But what is going on here, didn't we just say that the entropy of a system is maximal if the distribution is completely flat? Should a system not always tend to maximal entropy (principle of maximum multiplicity a.k.a. the principle of maximum entropy)?

Well yes, a system is always going to go towards maximal entropy, but often there are restraints imposed on a system. For example the number of different states that are accessible is limited. If this is the only restraint, the system will simply populate each accessible state with equal probability and thus achieve the maximal entropy possible.

In general, if restraints are present, the system adopts the maximal entropy configuration that is compatible with those restraints.

To understand this principle, let's look at dice. Because we only have 6 sides the maximal entropy state of a system of dice will be the case where each side faces up with a probability of 1/6.

Restraints on the average properties of a system

Let's see what happens if we introduce a restraint in our dice example.

If we apply our principle of maximal entropy, we will get a distribution that has equal numbers of 1 2 3 4 5 and 6.

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$$

but this maximal entropy configuration has a defined average score.

$$\langle \epsilon \rangle = \sum_i p_i \cdot \epsilon_i = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{21}{6} = 3.5$$

So the only way to change our average score is to change the values of the different p_i which -as we know- will decrease our entropy, because

$$S = -k \sum_i p_i \ln p_i$$

and S is minimal if all p_i are exactly equal ($p_1=p_2=p_3=\dots=p_i$). So our population, if left alone, will tend towards adopting that even distribution with its associated average $\langle \epsilon \rangle$.

However in many cases in real life there are environmental influences that dictate the average property of a system. The example that springs to mind right away is the average kinetic energy of a substance, which is directly controlled by the temperature. If we heat a system, the system will change its average kinetic energy and this will change its entropy. We are now asking what is the probability distribution that gives the maximal entropy while satisfying the imposed restraints.

Lets say we want to have an average value of 5 instead of 3.5. How could we get a distribution with that value?

- 1) we could make all dice show 5. That would obviously be a low entropy. As a matter of fact the entropy would be zero.
- 2) We could make half of the dice show 4 and half 6. Then our entropy would be $-k \cdot 2 \cdot 0.5 \cdot \ln 0.5 = -k \ln 0.5 = k \cdot 0.69$
- 3) Or we could get 1/3 of all dice to show 4, 5 and 6. Then our entropy would be $-k \cdot 3 \cdot 0.333 \cdot \ln 0.333 = -k \ln 0.333 = k \cdot 1.098$

So we see that we can maintain an average score of 5 and still get quite a boost in entropy by spreading out our probability distribution over a few values for the dice, but what is the absolutely highest entropy we can achieve while maintaining that average of 5.

Imposing restraints on systems leads to exponential distribution functions.

The answer is an exponential distribution function. Ken Dill nicely derives this very central result in his Book (see page 86). I am not going to show this here, so either take my word for it or look it up.

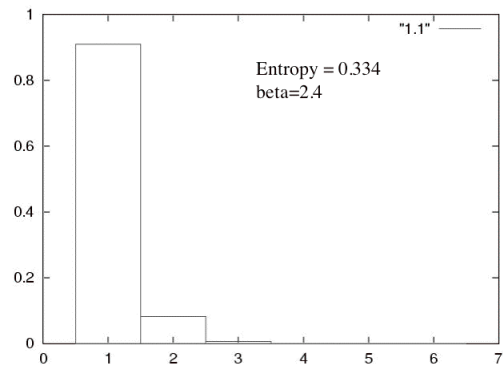
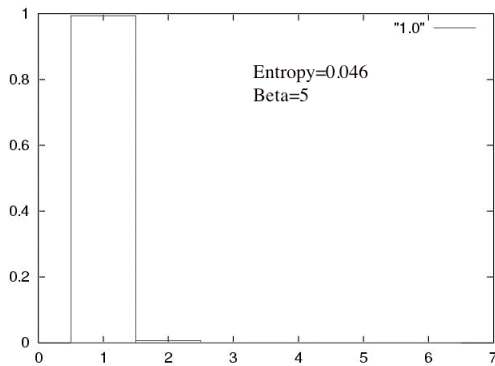
Specifically for a system with t states the exponential function that will give us the maximal entropy will look like this

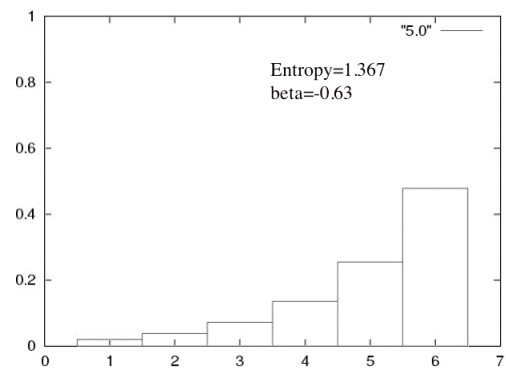
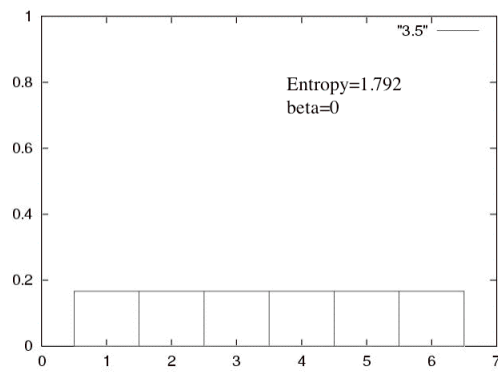
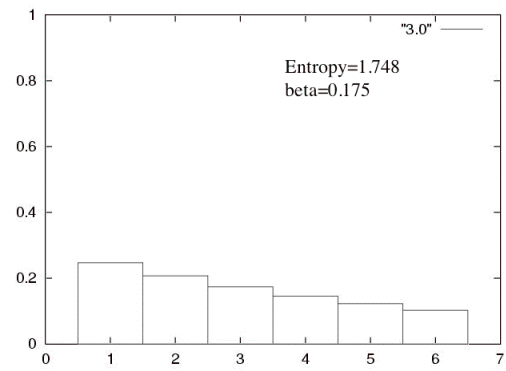
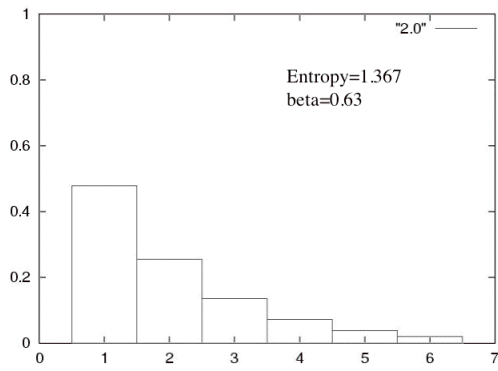
$$p_i = \frac{e^{-\beta \varepsilon_{p_i}}}{\sum_{i=1}^t e^{-\beta \varepsilon_{p_i}}}$$

Notice how the probability of all p_i depends only on that property of state i that contributes to the average property of the population and on the parameter beta. So if we want to find the maximal entropy configuration of a system with a certain average property, all we need to do is to turn up or down beta. And use

$$\langle \varepsilon \rangle = \sum_i p_i \cdot \varepsilon_i \quad \text{to calculate our average property.}$$

On the next page I show you what our distributions for the dice example look like for different betas and what the entropy of these different states are. Note that beta =0 gives you the flat distribution, and that this distribution has the highest entropy as well as the average value 3.5 (shown in quotation marks). Also note that the exponential distribution with an average value of 5 has an entropy of 1.367, significantly better than the entropy of the 1/3 4's, 1/3 5's, 1/3 6's distribution.





The equation that describes the exponential distribution giving us the maximal entropy under restrained conditions

$$p_i = \frac{e^{-\beta \epsilon_{p_i}}}{\sum_{i=1}^t e^{-\beta \epsilon_{p_i}}}$$

is called the Maxwell-Boltzmann distribution.

The form in which we will see it most of the time is

$$p_i = \frac{e^{-\frac{G_i}{RT}}}{\sum_{i=1}^I e^{-\frac{G_i}{RT}}}$$

where the property of state i is the free energy of this state and the factor beta is $1/RT$. This last equation describes a very fundamental property of systems with large numbers of particles. In one form or another the Maxwell-Boltzmann distribution shows up in all sorts of different places in the natural sciences.

For example, if we want to know the relative probability of two different states and we know the state's energies we can use this formula to calculate the former from the latter. Conversely, if we know the relative probability of two different states at a given temperature we can get their relative energies.

$$\frac{p_a}{p_b} = \frac{\frac{e^{-\frac{G_a}{RT}}}{\sum_{i=1}^I e^{-\frac{G_i}{RT}}}}{\frac{e^{-\frac{G_b}{RT}}}{\sum_{i=1}^I e^{-\frac{G_i}{RT}}}} = \frac{e^{-\frac{G_a}{RT}}}{e^{-\frac{G_b}{RT}}} = e^{-\frac{\Delta G}{RT}}$$

Summary

I hope I have convinced you that entropy is nothing mysterious or that it is particularly difficult to understand, when we approach it from the molecular level. Entropy is proportional to the multiplicity W of microscopic states that contribute to a macroscopic state. This is to say, a system tends towards a state of maximal entropy, simply because there are more microscopic ways to be in a macroscopic state with high multiplicity/entropy. The Boltzmann equation allows us to calculate the entropy from the multiplicity.

$$S = N_A k \ln W$$

We can convert the Boltzmann equation into a different form that relates probability distributions to entropy:

$$S = -N_A k \sum_{i=1} p_i \ln p_i \quad \text{with} \quad \sum_{i=1} p_i = 1$$

A different way of looking at this is to look at the distribution of probabilities, with which a molecule adopts a number of different states. Given a fixed number of microscopic

states, the state of maximal entropy is the state that has the flattest distribution – i.e. the state in which each microscopic state has exactly the same probability. So if we place no restraints on our system – other than the number of states that are accessible – we will always tend towards a state with a completely flat distribution of probabilities.

Finally we realized that the maximum entropy state of a system will determine many of the macroscopic properties of a system. If we now force our system to change those average properties, we will automatically force the system into a state of lower entropy. The question now is, “What distribution across the microscopic states retains the maximal entropy, while achieving that average property”. The answer is “An exponential distribution”. Specifically it is the Maxwell-Boltzmann distribution, which has the following equation.

$$P_i = \frac{e^{-\frac{G_i}{RT}}}{\sum_{i=1}^t e^{-\frac{G_i}{RT}}}$$

This equation relates the free energy of a specific state G_i to the probability of populating state i via the temperature of our system and the energies of all other states accessible to our system.