

Pressure

Last lecture, we considered two systems with entropy as a function of internal energy, volume and number of particles,

$$\sigma(U, V, N, U_1, V_1, N_1) = \sigma_1(U_1, V_1, N_1) + \sigma_2(U_2, V_2, N_2) .$$

We allowed them to exchange internal energy (that is, they were placed in thermal contact), and by requiring that the entropy be a maximum, we were able to show that the temperature is

$$\frac{1}{\tau} = \left(\frac{\partial \sigma}{\partial U} \right)_{V, N} .$$

Suppose we continue to consider our two systems, and ask what happens if we allow them to exchange volume as well as energy? (We're placing them in mechanical as well as thermal contact.) Again, the total entropy must be a maximum with respect to exchanges of energy and exchanges of volume. Working through similar mathematics, we find an expression for the change in total entropy and insist that it be zero (so the entropy is maximum) at equilibrium,

$$\begin{aligned} 0 &= \delta \sigma \\ &= \frac{\partial \sigma_1}{\partial U_1} \delta U_1 + \frac{\partial \sigma_2}{\partial U_2} \delta U_2 + \frac{\partial \sigma_1}{\partial V_1} \delta V_1 + \frac{\partial \sigma_2}{\partial V_2} \delta V_2 \\ &= \left(\frac{\partial \sigma_1}{\partial U_1} - \frac{\partial \sigma_2}{\partial U_2} \right) \delta U_1 + \left(\frac{\partial \sigma_1}{\partial V_1} - \frac{\partial \sigma_2}{\partial V_2} \right) \delta V_1 , \end{aligned}$$

from which we infer that at equilibrium,

$$\frac{\partial \sigma_1}{\partial U_1} = \frac{\partial \sigma_2}{\partial U_2} ,$$

which we already knew, and

$$\frac{\partial \sigma_1}{\partial V_1} = \frac{\partial \sigma_2}{\partial V_2} .$$

This last equation is new, and it must have something to do with the pressure. Why? Because, once the temperatures are the same, two systems exchange volume only if one system can “push harder” and expand while the other contracts. We define the pressure:

$$p = \tau \left(\frac{\partial \sigma}{\partial V} \right)_{U, N} .$$

We will see later that this definition agrees with the conventional definition of pressure as force per unit area.

Chemical Potential

Well, there's one variable left, guess what we're going to do now! Suppose we allow the two systems to exchange particles as well as energy and volume. Again, we want to maximize the entropy with respect to changes in all the independent variables and this leads to,

$$\begin{aligned} 0 &= \delta\sigma \\ &= \frac{\partial\sigma_1}{\partial U_1}\delta U_1 + \frac{\partial\sigma_2}{\partial U_2}\delta U_2 + \frac{\partial\sigma_1}{\partial V_1}\delta V_1 + \frac{\partial\sigma_2}{\partial V_2}\delta V_2 + \frac{\partial\sigma_1}{\partial N_1}\delta N_1 + \frac{\partial\sigma_2}{\partial N_2}\delta N_2 \\ &= \left(\frac{\partial\sigma_1}{\partial U_1} - \frac{\partial\sigma_2}{\partial U_2}\right)\delta U_1 + \left(\frac{\partial\sigma_1}{\partial V_1} - \frac{\partial\sigma_2}{\partial V_2}\right)\delta V_1 + \left(\frac{\partial\sigma_1}{\partial N_1} - \frac{\partial\sigma_2}{\partial N_2}\right)\delta N_1 . \end{aligned}$$

So, when the systems can exchange particles as well as energy and volume,

$$\frac{\partial\sigma_1}{\partial N_1} = \frac{\partial\sigma_2}{\partial N_2} .$$

The fact that these derivatives must be equal in equilibrium allows us to define yet another quantity, μ , the chemical potential

$$\mu = -\tau \left(\frac{\partial\sigma}{\partial N} \right)_{U,V} .$$

If two systems are allowed to exchange particles and the chemical potentials are unequal, there will be a net flow of particles until the chemical potentials are equal. Like temperature and pressure, chemical potential is an intensive quantity. Unlike temperature and pressure, you probably have not come across chemical potential in your elementary thermodynamics. You can think of it very much like a potential energy per particle. Systems with high chemical potential want to send particles to a system with low potential energy per particle. Note that we can write a change in the entropy of a system, specified in terms of U , V , and N as

$$d\sigma = \frac{1}{\tau} dU + \frac{p}{\tau} dV - \frac{\mu}{\tau} dN ,$$

or rearranging,

$$dU = \tau d\sigma - p dV + \mu dN .$$

Which is the conservation of energy (first law of thermodynamics) written for a system which can absorb energy in the form of heat, which can do mechanical pV work, and which can change its energy by changing the number of particles.

First Derivatives versus Second Derivatives

You will notice that the quantities defined by first derivatives are not material specific. For example, whether it's nitrogen gas or a block of steel, the rate of change of energy with entropy (at constant volume and particle number) is the temperature.

We'll eventually define Helmholtz and Gibbs free energies and enthalpy (different independent variables) and it will always be the case that the first derivatives of these quantities produce other quantities that are not material specific.

To get to material specific quantities, one must go to second derivatives. For example, suppose we have a block of steel and nitrogen gas inside a container that is thermally insulating, fixed in volume, and impermeable. Then at equilibrium,

$$\left(\frac{\partial U_{\text{steel}}}{\partial V_{\text{steel}}}\right)_{S,N} = \left(\frac{\partial U_{\text{nitrogen}}}{\partial V_{\text{nitrogen}}}\right)_{S,N} = -p.$$

If we make a change to the volume of the container, we might be interested in

$$\frac{\partial p}{\partial V} = -\frac{\partial^2 U}{\partial V^2}.$$

This quantity is related to the compressibility of the material. Nitrogen gas is much more compressible than steel and most of the volume change will be taken up by the gas, not the steel. In other words, the material specific quantity (second derivative) is different for the two materials.

Probability

Here, we will introduce some basic concepts of probability. To start with, one imagines some experiment or other process in which several possible outcomes may occur. The possible outcomes are known, but not definite. For example, tossing a die leads to one of the 6 numbers 1, 2, 3, 4, 5, 6 turning up, but which number will occur is not known in advance. Presumably, a set of elementary outcomes can be defined and all possible outcomes can be specified by saying which elementary outcomes must occur. For example, the tossing of the die resulting in an even number would be made up of the elementary events: the toss is 2 or the toss is 4 or the toss is 6. A set of elementary events is such that one and only one event can occur in any repetition of the experiment. For example, the events (1) the toss results in a prime number and (2) the toss gives an even number could not both be part of a set of elementary events, because if the number 2 comes up, both events have occurred!

One imagines that a very large number of tosses of the die take place. Furthermore, in each toss, an attempt is made to ensure that there is no memory of the previous toss. (This is another way of saying successive tosses are *independent*.) Then the probability of an event is just the fraction of times it occurs in this large set of experiments, that is, n_e/N , where n_e is the number of times event e occurs and N is the total number of experiments. In principle, we should take the limit as the number of trials goes to ∞ . From this definition it is easy to see that the probabilities of a set of elementary events must satisfy

$$p_i \geq 0,$$

and

$$\sum_i p_i = 1,$$

where p_i is the probability of event i and i is an index that ranges over the possible elementary events.

The above definition is intuitive, but gives the sense of a process occurring in time. That is, we throw the same die over and over again and keep track of what happens. Instead, we can imagine a very large number of dice. Each die has been prepared, as nearly as possible, to be identical. Each die is shaken (randomized) and tossed independently. Again, the probability of an event is the fraction of the total number of trials in which the event occurs. This collection of identically prepared systems and identically performed trials is called an *ensemble* and averages that we calculate with this construction are called ensemble averages.

You are probably thinking that for the die, the probabilities of each of the six elementary events 1, 2, 3, 4, 5, 6 must be 1/6. Well, they could be, but it's not necessary! You've heard of loaded dice, right? All that's really necessary is that each p_i be non-negative and that their sum be 1. Probability theory itself makes no statement about the values of

the probabilities. The values must come from somewhere else. In general, we just assign probabilities to the elementary events. Often we will appeal to symmetry or other arguments to assign the probabilities. For example, since a die is a symmetric cube, no face can be distinguished (mechanically) from any other face and we can plausibly argue that the probabilities should be equal.

Aside: well, the dots have to be painted on, so the die isn't perfectly symmetric. Presumably, differences in the amount and pattern of paint on each face make a negligible difference in the mechanical properties of the die (such as cm, moment of inertia, etc.) so it's a very good approximation to regard the die as symmetric. However, some dice have rather large indentations for each dot. I've occasionally wondered if this might make a detectable difference in the probabilities.

In our discussion of the entropy, we postulated that a system is equally likely to be in any microscopic state consistent with the constraints. This amounts to assigning the probabilities and is basically an appeal to symmetry in the same way that assigning equal probabilities to each face of a die is an appeal to symmetry!

Averages

Assuming there is some numeric value associated with each elementary event, we can calculate its average value just by adding up all the values and dividing by the total number of trials—exactly what you think of as an average. So, if event i produces the value y_i , then its average value is

$$\begin{aligned}\langle y \rangle &= \frac{1}{N} (\underbrace{y_1 + y_1 + \cdots y_1}_{n_1 \text{ times}} + \underbrace{y_2 + y_2 + \cdots y_2}_{n_2 \text{ times}} + \cdots) \\ &= \frac{1}{N} (n_1 y_1 + n_2 y_2 + \cdots) \\ &= \sum_i y_i p_i.\end{aligned}$$

Quantities like y , whose value varies across an ensemble, are called random variables.

After the average, we will often be most interested in the variance (often called the square of the standard deviation.) This is just the average value of the square of the deviation from the average.

$$\text{var}(y) = \sigma_y^2 = \langle (y - \langle y \rangle)^2 \rangle,$$

where σ_y is the standard deviation in y , not the entropy! The standard deviation is a measure of the spread about the average.

Probabilities for Continuous Variables.

Rather than giving one of a finite (or infinite) number of discrete outcomes, an experiment might result in the measurement of a random variable which is continuously distributed over some finite (or infinite) range. In this case we deal with a probability density rather than discrete probabilities. For example, we might make a measurement of a continuous variable x . Then the probability that the measurement falls in a small range dx around the value x is

$$\text{Prob}(x < \text{result} < x + dx) = p(x) dx ,$$

where $p(x)$ is the probability density. Just as for discrete probabilities, the probability density must satisfy

$$p(x) \geq 0 ,$$

and

$$\int_{\text{allowed range of } x} p(x) dx = 1 .$$

We can simply define $p(x) = 0$ when x is outside the allowed range, so the normalization becomes

$$\int_{-\infty}^{+\infty} p(x) dx = 1 .$$

The average of any function of x , $y(x)$ is defined by

$$\langle y \rangle = \int_{-\infty}^{+\infty} y(x)p(x) dx .$$

The Binomial Distribution

As an example of working with probabilities, we consider the binomial distribution. We have N trials or N copies of similar systems. Each trial or system has two possible outcomes or states. We can call these heads or tails (if the experiment is tossing a coin), spin up or spin down (for spin 1/2 systems), etc. We suppose that each trial or system is independent and we suppose the probability of heads in one trial or spin up in one system is p and the probability of tails or spin down is $1 - p = q$. (Let's just call these up and down, I'm getting tired of all these words!)

To completely specify the state of the system, we would have to say which of the N systems are up and which are down. Since there are 2 states for each of the N systems, the total number of states is 2^N . The probability that a particular state occurs depends on the number of ups and downs in that state. In particular, the probability of a particular state with n up spins and $N - n$ down spins is

$$\text{Prob}(\text{single state with } n \text{ up spins}) = p^n q^{N-n} .$$

Usually, we are not interested in a single state with n up spins, but we are interested in all the states that have n up spins. We need to know how many there are. There is 1 state with no up spins. There are N different ways we have exactly one of the N spins up and $N - 1$ down. There are $N(N - 1)/2$ ways to have two spins up. In general, there are $\binom{N}{n}$ different states with n up spins. These states are distinct, so the probability of getting any state with n up spins is just the sum of the probabilities of the individual states. So

$$\text{Prob}(\text{any state with } n \text{ up spins}) = \binom{N}{n} p^n q^{N-n} .$$

Note that

$$1 = (p + q)^N = \sum_{n=0}^N \binom{N}{n} p^n q^{N-n} ,$$

and the probabilities are properly normalized.

To illustrate a trick for computing average values, suppose that when there are n up spins, a measurement of the variable y produces n . What are the mean and variance of y ? To calculate the mean, we want to perform the sum,

$$\langle y \rangle = \sum_{n=0}^N n \binom{N}{n} p^n q^{N-n} .$$

Consider the binomial expansion

$$(p + q)^N = \sum_{n=0}^N \binom{N}{n} p^n q^{N-n} ,$$

and observe that if we treat (for the moment) p and q as independent mathematical variables and we differentiate both sides of this expression with respect to p (keeping q fixed), we get

$$N(p+q)^{N-1} = \sum_{n=0}^N n \binom{N}{n} p^{n-1} q^{N-n} .$$

The RHS is almost what we want—it's missing one power of p . No problem, just multiply by p ,

$$Np(p+q)^{N-1} = \sum_{n=0}^N n \binom{N}{n} p^n q^{N-n} .$$

This is true for any (positive) values of p and q . Now specialize to the case where $p+q = 1$. Then

$$Np = \sum_{n=0}^N n \binom{N}{n} p^n q^{N-n} = \langle y \rangle .$$

A similar calculation gives

$$\text{var}(y) = Npq .$$

The fractional spread about the mean is proportional to $N^{-1/2}$. This is typical; as the number of particles grows, the fractional deviations from the mean of physical quantities decreases in proportion to $N^{-1/2}$. So with $\sim N_0$ numbers of particles, fractional fluctuations in physical quantities are $\sim 10^{-12}$. This is extremely small. Even though the macroscopic parameters in statistical mechanics are random variables, their fluctuations are so small that they can usually be ignored. We speak of *the* energy of a system and write down a single value, even though the energy of a system in thermal contact with a heat bath is properly a random variable which fluctuates continuously.