

Entropy and Probabilities

We've been using the idea that the entropy is the logarithm of the number of states accessible to the system. We've also said that each state is equally likely. At this point, I'd like to make the connection between entropy and probability. This allows one to construct an expression for the entropy of a system that isn't in equilibrium. It should also improve our intuition about the entropy and the partition function.

We might expect that an expression for entropy can be written in terms of the probabilities that a system is in a particular state. If there are g states, and the probabilities are p_1, p_2, \dots, p_g , then we would like to write

$$\sigma = \sigma(p_1, p_2, \dots, p_g) .$$

One of the things that will guide us in our selection of the function is that the entropy should be additive (i.e., an extensive parameter). If we have two non-interacting systems with total numbers of states g_1 and g_2 , entropies σ_1 and σ_2 , and probabilities p_{1i} and p_{2j} (the first index is the system, the second index is the state), we can also think of it as a single system with $g = g_1 g_2$ states, $\sigma = \sigma_1 + \sigma_2$ and, since any state in system 1 can be combined with any state in 2, the probability of a state in the combined system must be $p_{ij} = p_{1i} p_{2j}$.

Since the probabilities multiply, while the entropies add, we might expect that the entropy should involve the log of the probability. The first guess might be

$$\sigma_1 = - \sum_i \log p_{1i} \quad \text{wrong} .$$

Since $p_{1i} \leq 1$, the minus sign is inserted to make the entropy positive. Why doesn't this expression work? There are several reasons. First, suppose one has a totally isolated system. Then only states with the exact energy of the system are allowed. Disallowed states have $p_{1i} = 0$ and this will lead to problems with the logarithm. In addition, with the above expression, the entropy is not additive. To fix up the problem with $p_{1i} = 0$, we might try multiplying by p_{1i} since in the limit $x \rightarrow 0$, $x \log x \rightarrow 0$. Does this make the entropy additive? Consider

$$\begin{aligned} \sigma &= - \sum_{i,j} p_{1i} p_{2j} \log p_{1i} p_{2j} , \\ &= - \sum_{i,j} p_{1i} p_{2j} \log p_{1i} - \sum_{i,j} p_{1i} p_{2j} \log p_{2j} , \\ &= - \sum_i p_{1i} \log p_{1i} - \sum_j p_{2j} \log p_{2j} , \\ &= \sigma_1 + \sigma_2 . \end{aligned}$$

We used the fact that $\sum_i p_{1i} = \sum_j p_{2j} = 1$. We adopt the following expression for the entropy in terms of the probabilities.

$$\sigma = - \sum_i p_i \log p_i ,$$

where we can include or omit states with probability 0 without affecting the value of the entropy.

What set of probabilities maximizes the entropy? The answer depends on the conditions under which we seek a maximum. Suppose we are dealing with a completely isolated system. Then a state can have non-zero probability only if it has the required energy (and any other conserved quantities). So let's limit our sum to allowed states. (Here, we're doing this for convenience, not because our expression might blow up!) The other thing we know is that the probabilities of the allowed states sum to 1. The problem we want to solve is maximizing the entropy under the constraint that the probabilities sum to 1. How do we maximize with a constraint? Lagrange multipliers! So we seek to maximize

$$\begin{aligned} X(p) &= \sigma(p) + \lambda \left(1 - \sum_i p_i \right) , \\ &= - \sum_i p_i \log p_i + \lambda \left(1 - \sum_i p_i \right) . \end{aligned}$$

We set the derivative of X with respect to p_i to zero,

$$0 = \frac{\partial X}{\partial p_i} = - \log p_i - 1 - \lambda .$$

This gives

$$p_i = e^{-(\lambda + 1)} ,$$

so the probabilities of all allowed states are the same when the entropy is a maximum. We also set the derivative of X with respect to λ to 0 which recovers the condition that the probabilities sum to 1. Solving for λ , we find $\lambda = \log g - 1$. (g is the number of allowed states and the number of terms in the sum.) Finally,

$$\sigma = - \sum_i \frac{1}{g} \log \frac{1}{g} = - \log \frac{1}{g} = \log g ,$$

as we had before.

Now suppose we consider a system which is not isolated, but is in equilibrium thermal contact with a heat bath so that the average value of its internal energy is U . Again, we sum only over allowed states. This time states with energies other than U are allowed,

provided the average turns out to be U . We want to find the probabilities that maximize the entropy under the constraints that the probabilities sum to 1 and average energy is U . We find the maximum of

$$X(p) = - \sum_i p_i \log p_i + \lambda_1 \left(1 - \sum_i p_i \right) + \lambda_2 \left(U - \sum_i p_i E_i \right) ,$$

where E_i is the energy of state i . We want

$$0 = \frac{\partial X}{\partial p_i} = -\log p_i - 1 - \lambda_1 - \lambda_2 E_i .$$

It follows that

$$p_i = e^{-1 - \lambda_1 - \lambda_2 E_i} .$$

You are asked to show in the homework that $\lambda_2 = 1/\tau$. So, the probabilities wind up with a Boltzmann factor!

Consider an ensemble of systems. The case in which the energy of each system is identical and equal probabilities are assigned is known as the *micro-canonical ensemble*. The case in which the energies vary and the probabilities are assigned with Boltzmann factors is known as the *canonical ensemble*.

Heat Capacity

In general, the amount of energy added to a system in the form of heat, dQ , and the rise in temperature $d\tau$ resulting from this addition of heat are proportional,

$$dQ = C d\tau ,$$

where C is “constant” of proportionality. Why is constant in quotes? Answer: the bad news is that it can depend on just about everything. The good news is that over a small range of temperature it doesn't vary too much, so it can be treated as a constant.

One of the things it obviously depends on is the amount of material in the system. To remove this dependence, one often divides by something related to the amount of material and then speaks of the specific heat. For example, dividing by the mass of the system gives the heat capacity per unit mass, $c = C/m$. Of course, this is only useful if one is dealing with a homogeneous material. That is you might speak of the specific heat of aluminum and the specific heat of water, but for boiling water in an aluminum pan you would be concerned with the heat capacity (which you could calculate from the masses of the pan and the water and the specific heats from the *Handbook of Chemistry and Physics*). In the case of gasses, the amount of material is usually measured in moles and the heat capacity

is divided by the number of moles to give molar specific heat or molar heat capacity. This is usually a number of order the gas constant. In statistical physics, we often speak of the heat capacity per molecule. This is usually a number of order Boltzmann's constant.

All the above is mainly bookkeeping. Of more significance is the fact that heat capacities can depend on the manner in which the heat is added to a system. Heat added to a system whose volume is kept constant will result in a different temperature rise than the same amount of heat added to an identical system whose pressure is kept constant. The heat capacities in these cases are often written C_V , constant volume, or C_p , constant pressure. Of course, one can think of more complicated way in which heat may be added, and these can lead to additional definitions of heat capacity.

After a while, you might get the idea that we are overly fixated on heat capacities. Measurements of heat capacities, aside from their engineering utility, are also a good way to test theories of atomic and molecular structure and interactions, A theory will predict the energy levels of a molecular system, and statistical mechanics relates the microscopic energy levels to macroscopic quantities. Heat added and temperature rise are relatively easy macroscopic quantities to measure, so these can provide tests of microscopic theories.

For the moment, we will consider the addition of heat at constant volume and particle number. Then the internal energy changes only because of the addition of the heat. (No change due to work done, no change due to adding particles.) Then,

$$C_V = \left. \frac{dQ}{d\tau} = \frac{\tau d\sigma}{d\tau} = \tau \frac{\partial \sigma}{\partial \tau} \right)_V = \left. \frac{\partial U}{\partial \tau} \right)_V ,$$

where we can use dU for $\tau d\sigma$ since we are considering a constant volume (as indicated by the subscript) and constant particle number process. Changing to partial derivatives is just a clean up of the mathematical notation.

It should be noted that in our fundamental units, heat capacities are dimensionless. However, in conventional units the dimensions are energy per Kelvin.

As an example, we can apply the formalism of the partition function to the paramagnetic spin system we discussed in lecture 4. Recall, we worked that system out starting from the entropy. The Boltzmann factor showed up when we considered the numbers of aligned and anti-aligned magnets. Now we are going to use the partition function (the sum of Boltzmann factors) to work out the energy from the temperature. An individual magnetic dipole has energy $-E$ when aligned with the field and $+E$ when anti-aligned. (This is a slight difference in the definition of E used in the discussion of the partition function, but it's the same as we used in the discussion of the paramagnetic system.) Then

$$Z(\tau) = e^{+E/\tau} + e^{-E/\tau} = 2 \cosh(E/\tau) .$$

We use the earlier expression to get the average energy per magnet and multiply by N to get the total energy in the system,

$$\begin{aligned}
 U &= N\tau^2 \frac{\partial \log Z}{\partial \tau}, \\
 &= N\tau^2 \frac{\sinh E/\tau}{\cosh E/\tau} \frac{\partial}{\partial \tau} \left(\frac{E}{\tau} \right), \\
 &= N\tau^2 \tanh E/\tau \frac{-E}{\tau^2}, \\
 &= -NE \tanh E/\tau,
 \end{aligned}$$

as before! The heat capacity is $\partial U/\partial \tau$, so

$$\begin{aligned}
 C_V &= -NE \frac{\partial}{\partial \tau} \tanh E/\tau, \\
 &= -NE \operatorname{sech}^2 E/\tau \frac{\partial}{\partial \tau} \frac{E}{\tau}, \\
 &= +N \left(\frac{E}{\tau} \operatorname{sech} \frac{E}{\tau} \right)^2.
 \end{aligned}$$

At high and low temperatures the heat capacity goes to zero because of the τ^{-2} dependence and the exponential dependence, respectively. When the temperature is low, it's very hard for the bath to get together enough energy to flip a magnet, so increasing the temperature of the bath has little effect on the total energy of the system. When the temperature is high, the magnets are 50% aligned and 50% anti-aligned and the system cannot be made more random by adding energy.

Reversible Processes

We've glossed over the fact that adding heat is a process, not a state of the system. So exactly what happens when we add heat depends on just how we do it. In particular, we consider reversible processes. For such a process, the "driving force" is infinitesimally small, the process occurs infinitely slowly, and an infinitesimal change in the driving force can make the process run equally well in the reverse direction. While the process is underway, the state of the system (and heat bath) is not an equilibrium state. Because things are happening so slowly, we can imagine that the system is arbitrarily close to an equilibrium state. A reversible process can be thought of as a continuous sequence of equilibrium states carrying the initial state of the system to its final state.

In the case of heat transfer, the driving force is the temperature difference. To obtain an approximation to a reversible process the temperature difference between the system and the source of heat must be very small. The process must be carried out slowly so the system has time to get close to equilibrium (for the heat energy to be carried to all parts of the system). Furthermore, the system must not lose any energy by unwanted thermal conduction through the boundaries. (In actual practice, this means you want to perform the heat transfer as quickly as possible, before an appreciable amount can leak away!) In any case, a reversible process is a convenient idealization much like (and related to) the idealization of frictionless mechanical systems.

Food for thought: Do you suppose there's a connection between the "direction of the flow of time" and the increase of entropy in irreversible processes?

Pressure

Recall in lectures 2 and 3, we maximized the entropy with respect to changes in energy, volume, and particle number and made some definitions and came up with

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

with

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

Observe that if σ and N are constant, then the pressure is

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

What does it mean to keep σ and N constant? N is easy, we just keep the same number of particles. σ requires a bit more thought. If the entropy doesn't change, the number of states doesn't change. If we perform a volume change at constant entropy, we are changing the volume without changing the microstates of the system. In other words, the energy change produced by a change in volume at constant entropy is purely a mechanical energy change, not a thermal energy change. We know mechanics: the work done (and energy supplied) when the volume is increased by dV is just $-p dV$, where p is the ordinary mechanical pressure, force per unit area. This argument supports our identification of

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

as the conventional mechanical pressure.

Pressure in a Low Density Gas, I

Suppose we have some gas in equilibrium at low density. By low density, we mean that the molecules spend most of the time well separated from each other so that they are weakly interacting. Occasionally, there are collisions between molecules and these serve to randomize the velocities and maintain thermal equilibrium. To the extent that we can treat molecules as point masses and ignore their interactions, we have a model for an ideal gas.

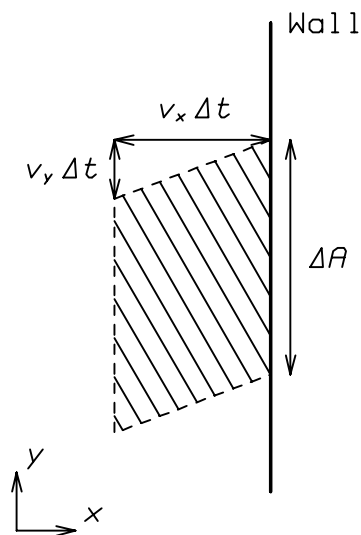
We are going to relate the pressure to the motions of the gas molecules. To start with, consider a wall of the container holding the gas. The wall is perpendicular to the x direction. There is a force on this wall because gas molecules are bouncing off it. If the mass of a molecule is m , and the x component of its velocity is v_x , then the change in momentum experienced by this molecule if it makes a perfectly elastic, perfectly reflecting collision from the wall is $\Delta P_x = -2mv_x$, and of course, the wall experiences an equal and opposite change in momentum.

A perfectly elastic and perfectly reflecting collision is one in which v_x changes sign and v_y and v_z are unchanged. In other words, the angle of reflection equals the angle of incidence and the kinetic energy is unchanged by the collision. Do we really think that all collisions between molecules and walls are like this? Of course not. A gas molecule doesn't collide with a "wall," it collides with one, or a small number, of molecules that are part of the wall, and this collision is just as "randomizing" as collisions between gas molecules. On the average, there can be no net change in v_y or v_z or the gas would start moving in the y or z directions. In the same way, there can be no net change in the energy caused by collisions with the wall or the gas would heat up or cool down, contrary to the assumption of thermal equilibrium. So as a convenience, which is consistent with the average behavior, we treat the collisions as perfectly elastic and reflecting.

The change in momentum of the wall in one collision is $2mv_x$. Consider a time interval Δt and an area of the wall ΔA , and consider molecules with velocities in the range $v_x \rightarrow v_x + dv_x$, $v_y \rightarrow v_y + dv_y$, $v_z \rightarrow v_z + dv_z$. If, at the beginning of the time interval, such a molecule is headed in the positive x direction and contained within the parallelepiped indicated schematically in the figure, it will collide with the wall during the time interval. The change in momentum caused by such molecules is

$$\delta P_x = \left[\frac{N}{V} \right] [v_x \Delta t \Delta A] [p(v_x, v_y, v_z) dv_x dv_y dv_z] [2mv_x] ,$$

where the first factor is the number of molecules per unit volume, the second factor is the volume, the third gives the fraction of molecules which have the specified velocity, and the last



is the change in momentum per collision. Note that $p(v_x, v_y, v_z)$ is the probability density for the velocities and is most likely the Maxwell velocity distribution, but all we require is that it be independent of direction. We find the total change in momentum during the time interval by adding up the contribution of all molecules

$$\Delta P_x = \int_0^{+\infty} dv_x \int_{-\infty}^{+\infty} dv_y \int_{-\infty}^{+\infty} dv_z p(v_x, v_y, v_z) \frac{N}{V} 2mv_x^2 \Delta t \Delta A ,$$

where the integral over v_x includes only $v_x > 0$ since we want the molecules that are about to collide with the wall, not those which have just collided. Pressure is the rate of change of momentum per unit area, so

$$p = \frac{\Delta P_x}{\Delta t \Delta A} = \frac{N}{V} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(v_x, v_y, v_z) mv_x^2 dv_x dv_y dv_z = \frac{N}{V} m \langle v_x^2 \rangle .$$

Since the distribution is independent of direction, we dropped the factor of two and extended the range of integration to $v_x < 0$. Also since the distribution is isotropic, we have $m \langle v_x^2 \rangle = m \langle v_y^2 \rangle = m \langle v_z^2 \rangle = \frac{1}{3} m \langle v^2 \rangle = \frac{2}{3} \langle E_{\text{tran}} \rangle$, where $\langle E_{\text{tran}} \rangle$ is the average translational kinetic energy per molecule. Finally,

$$pV = \frac{2}{3} N \langle E_{\text{tran}} \rangle = \frac{2}{3} U_{\text{tran}} ,$$

where U_{tran} is the translational kinetic energy of all the gas molecules. If, in fact, the probability density for the velocities is the Maxwell density, then $\langle E_{\text{tran}} \rangle = 3\tau/2$ (homework!) and

$$pV = N\tau = NkT = nRT ,$$

where n is the number of moles and R is the gas constant.

Pressure in a Low Density Gas, II

Look inside a gas, and consider molecules with velocity components in $dv_x dv_y dv_z \equiv d^3v$ at $(v_x, v_y, v_z) = \mathbf{v}$, these molecules have an x momentum density (momentum per unit volume) of

$$\frac{\delta P_x}{\delta V} = \frac{N}{V} m v_x p(\mathbf{v}) d^3v .$$

All this momentum is carried in the x direction at speed v_x . Note that positive momentum is carried in the positive direction while negative momentum is carried in the negative direction; both contribute to a momentum flux in the x direction. In fact, the flux of x momentum in the x direction (momentum per unit area perpendicular to x per unit time) is

$$\frac{\delta P_x}{\delta A \delta t} = \frac{N}{V} m v_x^2 p(\mathbf{v}) d^3v .$$

To get the total flux of momentum, we integrate over all velocities and come up with the same thing we had before. Momentum per unit area per second which is force per unit area which is pressure is

$$p = \frac{N}{V} \int d^3v m v_x^2 p(\mathbf{v}) = \frac{N}{V} m \langle v_x^2 \rangle .$$

So, why did we bother with this? For one thing, we don't have to introduce a wall to talk about pressure. Pressure exists throughout a fluid. Secondly, it's a first introduction to calculation of transport phenomena.

In the preceding we considered the transport of x momentum in the x direction. Of course, y momentum is transferred in the y direction and z momentum in the z direction. These are usually numerically equal to the flux of x momentum in the x direction and we have an isotropic pressure. One can also transport x momentum in the y and z directions, y momentum in the x and z directions and z momentum in the x and y directions. For the simple gas we've been considering, these fluxes are zero (can you see why?). However, in more complicated situations, they might not be zero; they correspond to viscous forces. In general, we need a nine component object to specify the transport of momentum (a vector) in any of three directions. This is a second rank tensor, usually called the stress tensor.

Summary: For an ideal gas, we've found an expression relating pressure, volume and translational kinetic energy. We related the energy to the temperature using the Maxwell velocity distribution, which was motivated by the Boltzmann factor. However, in writing down the Maxwell distribution, we "finessed" the issue of counting the states, so we haven't really derived the ideal gas law.