Instructor's guide

This is the lecture notes for the first week of PH 441.

The only reference in the text is (Schroeder Problem 6.43).

There are two different approaches for deriving the results of statistical mechanics. These results differ in what fundamental postulates are taken, but agree in the resulting predictions. The textbook takes a traditional microcanonical Boltzmann approach.

This week, before using that approach, we will reach the same results using the Gibbs formulation of the entropy (sometimes referred to as the "information theoretic entropy"), as advocated by Jaynes. Note that Boltzmann also used the more general Gibbs entropy, even though it doesn't appear on his tombstone.

0.1 Microstates vs. macrostates

You can think of a **microstate** as a quantum mechanical energy eigenstate. As you know from quantum mechanics, once you have specified an energy eigenstate, you know all that you can about the state of a system. Note that before quantum mechanics, this was more challenging. You *can* define a microstate classically, and people did so, but it was harder. In particular, the number of microstates classically is generally both infinite and non-countable, since any real number for position and velocity is possible. Quantum mechanics makes this all easier, since any finite (in size) system will have an infinite but countable number of microstates.

When you have a non-quantum mechanical system (or one that you want to treat classically), a microstate represents one of the "primitive" states of the system, in which you have specified all possible variables. In practice, it is common when doing this to specify what we might call a "mesostate", but call it a microstate. e.g. you might hear someone describe a microstate of a system of marbles in urns as defined by how many marbles of each color are in each urn. Obviously there are many quantum mechanical microstates corresponding to each of those states.

Small White Boards Write down a description of one particular macrostate.

A **macrostate** is a state of a system in which we have specified all the properties of the system that will affect any measurements we may care about. For instance, when defining a macrostate of a given gas or liquid, we could specify the internal energy, the number of molecules (or equivalently mass), and the volume. We need to specify all three properties (if we want to ask, for instance, for the entropy), because otherwise we won't have a unique answer. For different sorts of systems there are different ways that we can specify a macrostate. In this way, macrostates have a flexibility that real microstates do not. e.g. I could argue that the macrostate of a system of marbles in urns would be defined by the number of marbles of each color in each urn. After all, each macrostate would still correspond to many different energy eigenstates.

0.2 Probabilities of microstates

The name of the game in statistical mechanics is determining the probabilities of the various microstates, which we call $\{P_i\}$, where i represents a microstate. I will note here the term **ensemble**, which refers to a set of microstates with their associated probabilities. We define ensembles according to what constraints

we place on the microstates, e.g. in this discussion we will constrain all microstates to have the same volume and number of particles, which defines the **canonical ensemble.** Next week/chapter we will discuss the microcanonical ensemble (which also constrains all microstates to have identical energy), and other ensembles will follow. Today's discussion, however, will be largely independent of which ensemble we choose to work with, which generally depends on what processes we wish to consider.

0.2.1 Normalization

The total probability of all microstates added up must be one.

$$\sum_{i}^{\text{all } \mu\text{states}} P_i = 1 \tag{1}$$

This may seem obvious, but it is very easy to forget when lost in algebra!

0.2.2 From probabilities to observables

If we want to find the value that will be measured for a given observable, we will use the weighted average. For instance, the internal energy of a system is given by:

$$U = \sum_{i}^{\text{all } \mu\text{states}} P_i E_i \tag{2}$$

$$= \langle E_i \rangle \tag{3}$$

where E_i is the energy eigenvalue of a given microstate. The $\langle E_i \rangle$ notation simply denotes a weighted average of E. The subscript in this notation is optional.

This may seem wrong to you. In quantum mechanics, you were taught that the outcome of a measurement was always an eigenvalue of the observable, *not* the expectation value (which is itself an average). The difference is in how we are imagining performing a measurement, and what the size of the system is thought to be.

In contrast, imagine measuring the mass of a liter of water, for instance using a balance. While you are measuring its mass, there are water molecules leaving the glass (evaporating), and other water molecules from the air are entering the glass and condensing. The total mass is fluctuating as this occurs, far more rapidly than the scale can tip up or down. It reaches balance when the weights on the other side balance the *average* weight of the glass of water.

The process of measuring pressure and energy are similar. There are continually fluctuations going on, as energy is going back and forth between your system and the environment, and the process of measurement (which is slow) will end up measuring the average.

In contrast, when you perform spectroscopy on a system, you do indeed see lines corresponding to discrete eigenvalues, even though you are using a macroscopic amount of light on what may be a macroscopic amount of gas. This is because each photon that is absorbed by the system will be absorbed by a single molecule (or perhaps by two that are in the process of colliding). Thus you don't measure averages in a direct way.

In thermal systems such as we are considering in this course, we will consider the kind of observable where the average value of that observable is what is measured. This is why statistics are relevant!

0.3 Energy as a constraint

Energy is one of the most fundamental concepts. When we describe a macrostate, we will (almost) always need to constrain the energy. For real systems, there are always an infinite number of microstates with no upper bound on energy. Since we never have infinite energy in our labs or kitchens, we know that there is a practical bound on the energy.

We can think of this as applying a mathematical constraint on the system: we specify a U, and this disallows any set of probabilities $\{P_i\}$ that have a different U.

Small Group Question Consider a system that has just three microstates, with energies $-\epsilon$, 0, and ϵ. Construct three sets of probabilities corresponding to $U = 0$.

I picked an easy U. Any "symmetric" distribution of probabilities will do. You probably chose something like:

E_i:	$-\epsilon$	0	ϵ
P_i:	0	1	0
P_i:	$\frac{1}{2}$	0	$\frac{1}{2}$
P_i:	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Question Given that each of these answers has the same U, how can we find the correct set of probabilities are for this U? Vote on which **you** think most likely!

The most "mixed up" would be ideal. But how do we define mixed-up-ness? The "mixed-up-ness" of a probability distribution can be quantified via the Gibbs formulation of entropy:

$$S = -k \sum_{i}^{\text{all } \mu\text{states}} P_i \ln P_i \tag{4}$$

$$= \sum_{i}^{\text{all } \mu\text{states}} P_i \left(-k \ln P_i\right) \tag{5}$$

$$= \langle -k \ln P_i \rangle \tag{6}$$

So entropy is a kind of weighted average of $-\ln P_i$ (which is also the average of $\ln(1/P_i)$).

The Gibbs entropy expression (sometime referred to as the information theory entropy, or Shannon entropy) can be shown to be the only possible entropy function (of $\{P_i\}$) that has a reasonable set of properties:

1. It must be extensive. If you subdivide your system into uncorrelated and noninteracting subsystems (or combine two noninteracting systems), the entropy must just add up. Solve problem 2 on the homework this week to show this. (Technically, it must be additive even if the systems interact, but that is more complicated.)

2. The entropy must be a continuous function of the probabilities $\{P_i\}$. Realistically, we want it to be analytic.

3. The entropy shouldn't change if we shuffle around the labels of our states, i.e. it should be symmetric.

4. When all microstates are equally likely, the entropy should be maximized.

5. All microstates have zero probability except one, the entropy should be minimized.

Note The constant k is called Boltzmann's constant, and is sometimes written as k_B. Kittel and Kroemer prefer to set $k_B = 1$ in effect, and defines σ as S/k_B to make this explicit. I will include k_B but you can and should keep in mind that it is just a unit conversion constant. Note also that changing the base of the logarithm in effect just changes this constant.

How is this mixed up?

Small Group Question Compute S for each of the above probability distributions.

Answer

E_i:	$-\epsilon$	0	ϵ	S/k_B
P_i:	0	1	0	0
P_i:	$\frac{1}{2}$	0	$\frac{1}{2}$	$\ln 2$
P_i:	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\ln 3$

You can see that if more states are probable, the entropy is higher. Or alternatively you could say that if the probability is more "spread out", the entropy is higher.

0.3.1 Maximize entropy

The correct distribution is that which maximizes the entropy of the system, its Gibbs entropy, subject to the appropriate constraints. Yesterday we tackled a pretty easy 3-level system with $U = 0$. If I had chosen a different energy, it would have been much harder to find the distribution that gave the highest entropy.

With only three microstates and two constraints (total probability = 1 and average energy = U), we could just maximize the entropy by eliminating two probabilities and then setting a derivative to zero. But this wouldn't work if we had more than three microstates.

Small Group Question Find the probability distribution for our 3-state system that maximizes the entropy, given that the total energy is U.

Answer We have two constraints,

$$\sum_i P_i = 1 \tag{7}$$

$$\sum_i P_i E_i = U \tag{8}$$

and we want to maximize

$$S = -k \sum_i P_i \ln P_i. \tag{9}$$

Fortunately, we've only got three states, so we write down each sum explicitly, which will make things easier.

$$P_- + P_0 + P_+ = 1 \tag{10}$$
$$-\epsilon P_- + \epsilon P_+ = U \tag{11}$$
$$P_+ = P_- + \frac{U}{\epsilon} \tag{12}$$
$$P_- + P_0 + P_- + \frac{U}{\epsilon} = 1 \tag{13}$$
$$P_0 = 1 - 2P_- - \frac{U}{\epsilon} \tag{14}$$
$$\tag{15}$$

Now that we have all our probabilities in terms of P_- we can simplify our entropy:

$$-\frac{S}{k} = P_- \ln P_- + P_0 \ln P_0 + P_+ \ln P_+ \tag{16}$$
$$= P_- \ln P_- + \left(P_- + \tfrac{U}{\epsilon}\right) \ln \left(P_- + \tfrac{U}{\epsilon}\right) + \left(1 - 2P_- - \tfrac{U}{\epsilon}\right) \ln \left(1 - 2P_- - \tfrac{U}{\epsilon}\right) \tag{17}$$

Now we can maximize this entropy by setting its derivative to zero!

$$-\frac{\frac{dS}{dP_-}}{k} = 0 \tag{18}$$
$$= \ln P_- + 1 - 2\ln\left(1 - 2P_- - \tfrac{U}{\epsilon}\right) - 2 + \ln\left(P_- + \tfrac{U}{\epsilon}\right) + 1 \tag{19}$$
$$= \ln P_- - 2\ln\left(1 - 2P_- - \tfrac{U}{\epsilon}\right) + \ln\left(P_- + \tfrac{U}{\epsilon}\right) \tag{20}$$
$$= \ln\left(\frac{P_-\left(P_- + \frac{U}{\epsilon}\right)}{\left(1 - 2P_- - \frac{U}{\epsilon}\right)^2}\right) \tag{21}$$
$$1 = \frac{P_-\left(P_- + \frac{U}{\epsilon}\right)}{\left(1 - 2P_- - \frac{U}{\epsilon}\right)^2} \tag{22}$$

And now it is just a polynomial equation…

$$P_-\left(P_- + \tfrac{U}{\epsilon}\right) = \left(1 - 2P_- - \tfrac{U}{\epsilon}\right)^2 \tag{23}$$
$$P_-^2 + \tfrac{U}{\epsilon}P_- = 1 - 4P_- - 2\tfrac{U}{\epsilon} + 4P_-^2 + 4P_-\tfrac{U}{\epsilon} + \tfrac{U^2}{\epsilon^2} \tag{24}$$

At this stage I'm going to stop. Clearly you could keep going and solve for P_- using the quadratic equation, but we wouldn't learn much from doing so. The point here is that we *can* solve for the three probabilities given the internal energy constraint. However, doing so is a major pain, and the result is not looking promising in terms of simplicity. There is a better way!

0.4 Lagrange multipliers (for those who are curious)

If you have a function of N variables, and want to apply a single constraint, one approach is to use the constraint to algebraically eliminate one of your variables. Then you can set the derivatives with respect to all remaining variables to zero to maximize. A nicer approach for maximization under constraints is the method of Lagrange multipliers.

The idea of Lagrange multipliers is to introduce an additional variable (called the Lagrange multiplier) rather than eliminating one. This may seem counterintuitive, but it allows you to create a new function that can be maximized by setting its derivative with to all N variables to zero, while still satisfying the constraint.

Suppose we have a situation where we want to maximize F under some constraints:

$$F = F(w, x, y, z) \tag{25}$$
$$f_1(w, x, y, z) = C_1 \tag{26}$$
$$f_2(w, x, y, z) = C_2 \tag{27}$$

We define a new variable L as follows:

$$L \equiv F + \lambda_1(C_1 - f_1(w, x, y, z)) + \lambda_2(C_2 - f_2(w, x, y, z)) \tag{28}$$

Note that $L = F$ provided the constraints are satisfied, since the constraint means that $f_1(x, y, z) - C_1 = 0$. We then maximize L by setting its derivatives to zero:

$$\left(\frac{\partial L}{\partial w}\right)_{x,y,z} = 0 \tag{29}$$
$$= \left(\frac{\partial F}{\partial w}\right)_{x,y,z} - \lambda_1 \frac{\partial f_1}{\partial w} - \lambda_2 \frac{\partial f_2}{\partial w} \tag{30}$$
$$\left(\frac{\partial L}{\partial x}\right)_{w,y,z} = 0 \tag{31}$$
$$= \left(\frac{\partial F}{\partial x}\right)_{w,y,z} - \lambda_1 \frac{\partial f_1}{\partial x} - \lambda_2 \frac{\partial f_2}{\partial x} \tag{32}$$

$$\left(\frac{\partial L}{\partial y}\right)_{w,x,z} = 0 \tag{33}$$
$$= \left(\frac{\partial F}{\partial y}\right)_{w,x,z} - \lambda_1 \frac{\partial f_1}{\partial y} - \lambda_2 \frac{\partial f_2}{\partial y} \tag{34}$$
$$\left(\frac{\partial L}{\partial z}\right)_{w,x,y} = 0 \tag{35}$$
$$= \left(\frac{\partial F}{\partial z}\right)_{w,x,y} - \lambda_1 \frac{\partial f_1}{\partial z} - \lambda_2 \frac{\partial f_2}{\partial z} \tag{36}$$

This gives us four equations. But we need to keep in mind that we also have the two constraint equations:

$$f_1(x, y, z) = C_1 \tag{37}$$
$$f_2(x, y, z) = C_2 \tag{38}$$

We now have six equations and *six* unknowns, since λ_1 and λ_2 have also been added as unknowns, and thus we can solve all these equations simultaneously, which will give us the maximum under the constraint. We also get the λ values for free.

0.4.1 The meaning of the Lagrange multiplier

So far, this approach probably seems pretty abstract, and the Lagrange multiplier λ_i seems like a strange number that we just arbitrarily added in. Even were there no more meaning in the multipliers, this method would be a powerful tool for maximization (or minimization). However, as it turns out, the multiplier often (but not always) has deep physical meaning. Examining the Lagrangian L, we can see that

$$\left(\frac{\partial L}{\partial C_1}\right)_{w,x,y,z,C_2} = \lambda_1 \tag{39}$$

so the multiplier is a derivative of the lagrangian with respect to the corresponding constraint value. This doesn't seem to useful.

More importantly (and less obviously), we can now think about the original function we maximized F as a function (after maximization) of just C_1 and C_2. If we do this, then we find that

$$\left(\frac{\partial F}{\partial C_1}\right)_{C_2} = \lambda_1 \tag{40}$$

I think this is incredibly cool! And it is a hint that Lagrange multipliers may be related to Legendre transforms.

0.5 Maximizing entropy

When maximizing the entropy, we need to apply *two* constraints. We must hold the total probability to 1, and we must fix the mean energy to be U:

$$\sum_i P_i = 1 \tag{41}$$

$$\sum_i P_i E_i = U \tag{42}$$

I'm going to call my lagrange multipliers αk_B and βk_B so as to make all the Boltzmann constants go away.

$$L = S + \alpha k_B \left(1 - \sum_i P_i\right) + \beta k_B \left(U - \sum_i P_i E_i\right) \tag{43}$$

$$= -k_B \sum_i P_i \ln P_i + \alpha k_B \left(1 - \sum_i P_i\right) + \beta k_B \left(U - \sum_i P_i E_i\right) \tag{44}$$

where α and β are the two Lagrange multipliers. I've added here a couple of factors of k_B mostly to make the k_B in the entropy disappear. We want to maximize this, so we set its derivatives to zero:

$$\frac{\partial L}{\partial P_i} = 0 \tag{45}$$

$$= -k_B \left(\ln P_i + 1\right) - k_B\alpha - \beta k_B E_i \tag{46}$$

$$\ln P_i = -1 - \alpha - \beta E_i \tag{47}$$

$$P_i = e^{-1-\alpha-\beta E_i} \tag{48}$$

So now we know the probabilities in terms of the two Lagrange multipliers, which already tells us that the probability of a given microstate is exponentially related to its energy. At this point, it is convenient to invoke the normalization constraint…

$$1 = \sum_i P_i \tag{49}$$

$$= \sum_i e^{-1-\alpha-\beta E_i} \tag{50}$$

$$= e^{-1-\alpha} \sum_i e^{-\beta E_i} \tag{51}$$

$$e^{1+\alpha} = \sum_i e^{-\beta E_i} \tag{52}$$

$$\tag{53}$$

Where we define the normalization factor as

$$Z \equiv \sum_i^{\text{all states}} e^{-\beta E_i} \tag{54}$$

which is called the **partition function**. Putting this together, the probability is

$$P_i = \frac{e^{-\beta E_i}}{Z} \tag{55}$$

$$= \frac{\textit{Boltzmann factor}}{\textit{partition function}} \tag{56}$$

At this point, we haven't yet solved for β, and to do so, we'd need to invoke the internal energy constraint:

$$U = \sum_i E_i P_i \tag{57}$$

$$U = \frac{\sum_i E_i e^{-\beta E_i}}{Z} \tag{58}$$

As it turns out, $\beta = \frac{1}{k_B T}$. This follows from my claim that the Lagrange multiplier is the partial derivative with respect to the constaint value

$$k_B\beta = \left(\frac{\partial S}{\partial U}\right)_{\text{Normalization=1}} \tag{59}$$

However, I did not prove this to you. I will leave demonstrating this as a homework problem.