# Lecture notes from the course Statistical Mechanics (PHY 29) taught by Leonard Susskind Spring 2009

Bjørn Remseth
rmz@rmz.no

October 30, 2012

# Contents

# Introduction

These are my notes for the course in Statistical Mechanics (Stanford U.) taught by Leonard Susskind.

I usually watched the videos while typing notes in LaTeX. I have experimented with various note-taking techniques including free text, mindmaps and handwritten notes, but I've ended up using LaTeX, since it's not too hard, it gives great readability for the math that inevitably pops up in the things I like to take notes about, and it's easy to include various types of graphics. Also, it fits nicely into the rest of the set of tools I use to follow these lectures: More often than not I'm on a train during may daily commute. My handwriting is bad on any given day, but when combined with a bumpy train it's totally unreadable, even by me. However, having one window with Emacs, another with LaTeX, and a screengrabber program nearby, it is easy to get in "flow" and stay there while producing notes that are possible to read. It's nice :-) The graphics in this document is exclusively screenshots copied directly out of the videos, and to a large extent, but not completely, the text is based on Susskind's narrative. I haven't been very creative, that wasn't my purpose. I did take more screenshots than are actually available in this text. Some of them are indicated in figures stating that a screenshot is missing. I may or may not get back to putting these missing screenshots back in, but for now the are just not there. Deal with it .-)

This document will every now and then be made available on http://dl.dropbox.com/u/187726/statistical-mechanics-notes.pdf. The source code can be cloned on git on https://github.com/la3lma/statistical-mechanics.

A word of warning: These are just my notes. They should't be interpreted as anything else. I take notes as an aid for myself. When I take notes I find myself spending more time with the subject at hand, and that alone lets me remember it better. I can also refer to the notes, and since I've written them myself, I usually find what I'm looking for ;). I state this clearly since the use of LaTeX will give some typographical cues that may lead the unwary reader to believe that this is a textbook or something more ambitious. It's not. This is a learning tool for me. If anyone else reads this and find it useful, that's nice. I'm happy, for you, but I didn't have that, or you in mind when writing this. That said, if you have any suggestions to make the text or presentation better, please let me know. My email address is la3lma@gmail.com.

# 1 Getting started with thermodynamics

## 1.1 Dynamic systems

The first lecture starts with a longish quote, so I'll quote it in full:

> *Statistical mechanics is often thought of as the theory of how atoms comine to form gases liquids solids and even plasmas and black body radiation. But it is both much more and less than thhat. Statistical mechanics is a useful tools in many areas of science where a large number of variables has to be dealt with using statistical methods.*

Here he breaks from reading the quote to interject: "My son who studies neural networks uses, in fact about six months ago he called me up and said "pop, did you ever hear about this thing called the partition function and I'm just learning about it for using it in Neural Networks."

> *I have no doubt that some of the financial wizards of* AIG *and* Lehman brothers *used it. Saying that statistical mechanics is the theory of gases is rather like saying that calculus is the theory of planetary orbits.*

What it really is is a mathematical structure with application. Putting it in a nutshell, one can perhaps say that statistical mechanics is just probability theory. Now Susskind has never understood the difference between statistics and probabilities. It is probabilities under certain specific circumstanses.

It is however a bit tricky to say actually how statistical mechanics really connects to reality.

**Rmz:** *This is perhaps just a way of saying that statistical mechanics is a purely empirical branch of physics :-)*

Let's start with coinflipping (fair coin, equal probabilities of heads and tails that sums to one). A convincing argument that leads to the fairness, is that coins are fairly symmetrical (apart from some small details).

There is a notion of a-priori probability, in this case because it is a symmetry.

Let's take another example, take a dice. There a are six sides, that are named after colors (r, y, b, g, o, p). The six sides of the die are symmetric. There is a symmetry operation of turning the die 90 degrees about any axis. If you believe that the die is symmetric enough, then you are forced to believe that the probability of getting any one of the color is 1/6. However in most situations there aren't such symmetries.

When there aren't such symmetries, is there starting point where you can start thinking about such probabilities? The answer is "not obviously", but let's look at an example. Let's consider a die where we replace the purple color by red.

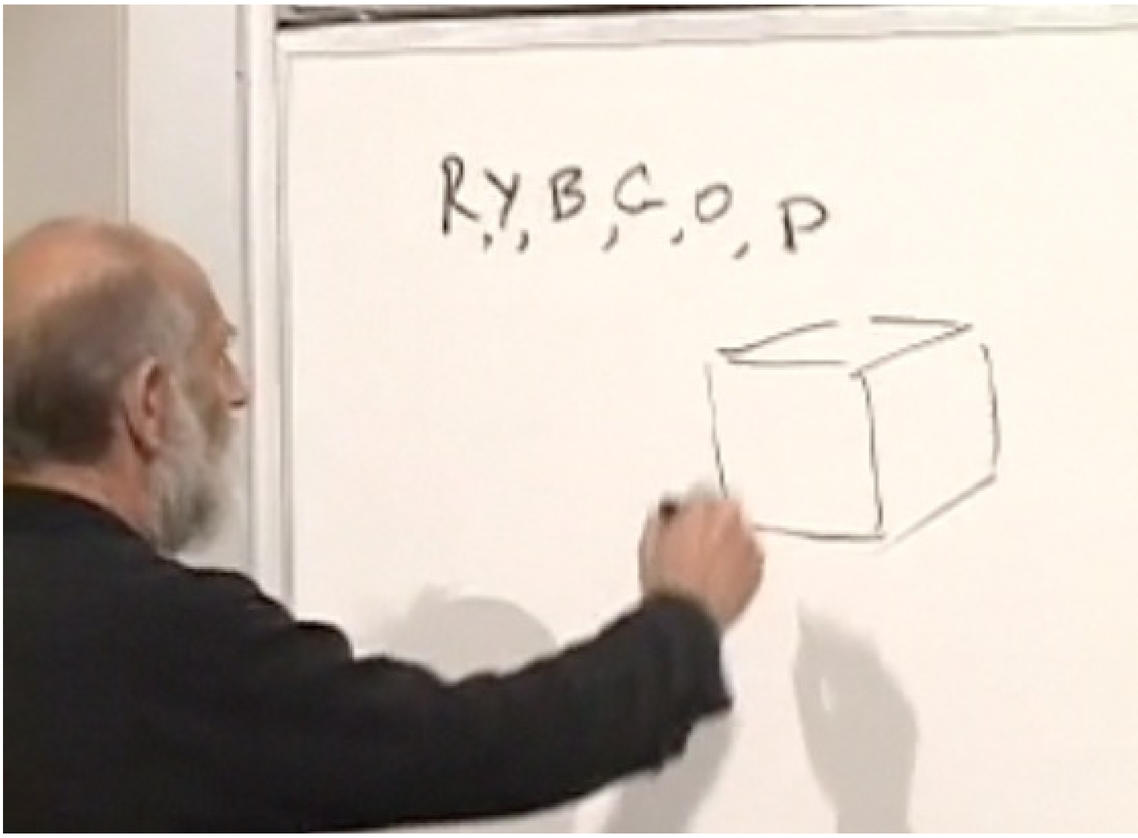


**Fig. 1.1** Purple side of die replaced by red color

It now have five colors. You have no reason to believe there is no symmetry. So what is the probability of flipping a red, so if you didn't know better you would perhaps say that the probability of flipping red was 1/5 but in fact it is (of course) 2/6 (= 1/3). The reason is that the real symmetry of the system acts on the six faces not on the five colors. But it is very easy to assume that a die doesn't have any symmetry at all, that is weighted and off balance (unfair die). And then, where would you get your a priori probabilities from. Well, one way would be to flip it a zillion times and then count how many of the different colors you got, but that's not we're gonna do. So somewhere else we have to get the idea of a prior probabilities. We might go back astep and say “*if the die is not a fair die, then the probabilities may depend on all kinds of things, it may depend on details such as the way the hand flips it, the air currents in the room, the way the surface it may or may not bounce off that are extraneous to the system itself and of course it may be depend on the environment and not the system itself*’. So let’s introduce one element: Let’s think of the die as a dynamical system that changes with time according to some law of motion. If you know what the system is at one instance in time, you will know what it is at the next instance in time. Now, with the motion of particles the motion is smooth and you can divide it into infinitesimal amounts of time. For a die it is (perhaps) a bit difference. Assume that the die performs one operation per *elementary* time interval, and at each time the die rearranges itself only depending on what is shown at the top of the die. You can then represent the motion of a die as a rule:

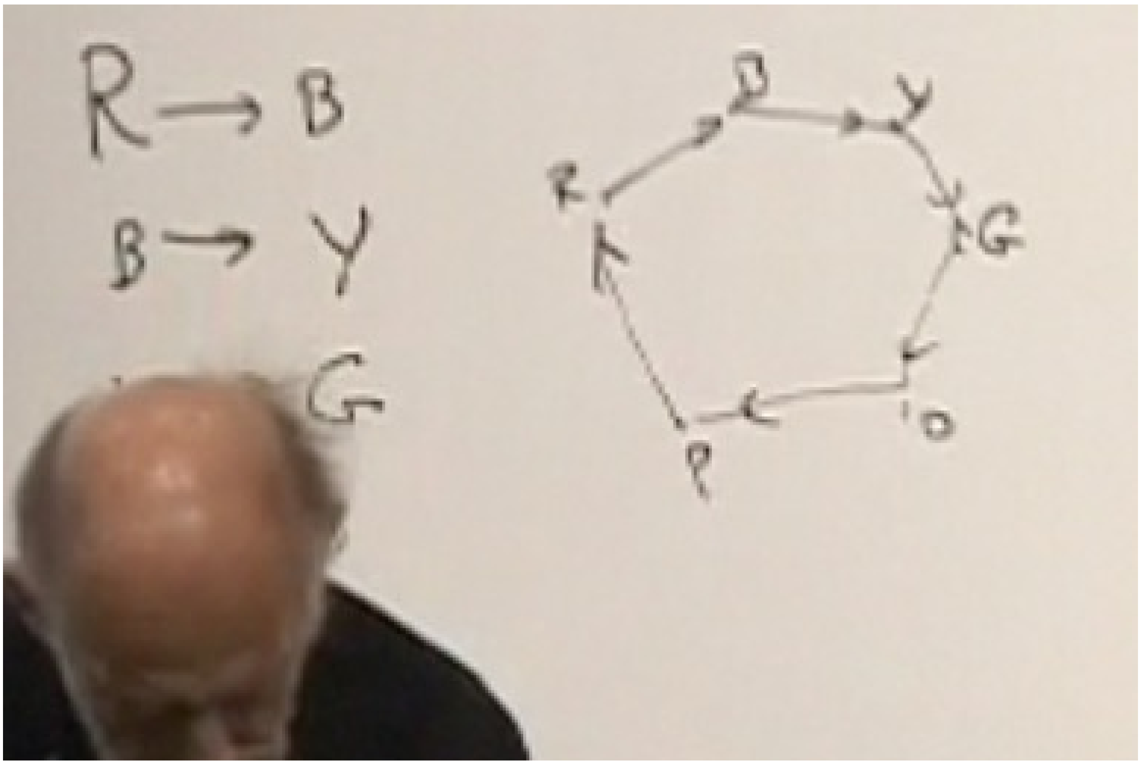


**Fig. 1.2** Purple side of die replaced by red color.

$$
\begin{aligned}
R &\to B\\
B &\to Y\\
Y &\to G\\
O &\to P\\
P &\to R
\end{aligned}
$$

That would be a complete dynamical theory for colors of the dice. Now assuming that the state succession is very fast, then it is very clear that there will be an equal probabilty of any one of the six colors, since they spend the same amount of time being in the same state.

If you change some of the colors, each state would still occupy 1/6 of the time, so there are many possible laws. However, there are other laws that don't give an answer altogether, here is an example:

$$
\begin{aligned}
R &\to B\\
B &\to G\\
Y &\to P\\
P &\to O\\
O &\to Y
\end{aligned}
$$

### 1.1.1 Conserved quantities

In every state of the system the system says what the system should do but it has the funny property of having two orbits or cycles. Now, there is no way a priori to konw which cycle you are in. So this is a counterexample that there is equal a priori probability of being in a state. However, the two-cycle example has something called a preserved state. Let's call it Z ("zilch"). It's 1 for the RGB orbit, and and 0 for the YPO orbit. The zilch is preserved, this is a *conservation law.* So in this case, and indeed in any case a conservation law means that the system breaks up into different orbits with each orbit representing some *conserved quantity.*

You now have two possibilities, either you fix the zilch, you then have equal probabilities of the different states with the preserved *quantum number* (zilch), and the other possibility is you only may know statistical information about the states then you use that.

In physics, or at least in thermodynamics, energy is the most important conserved quantity. Momentum isn't so important. The reason is that when thinking about systems contained within containers, and in statistical mechanics that's what we usually do. When a molecule bump into the container it gives a little momentum, but it's not so important. Electric charge can be important. Angular momentum usually don't matter that much.

Now a simple rule could be that you take all the conserved quantities and fix them, then you study the system subject to the constraint that the conserved quantities have certain values. That's the essence of statistical mechanics: Calculating probabilities of things happening subject to constraints that usually take form that some (one or more) conserved quantity is fixed.

Using two dice we can think up an example. We now number the sides from 1 to 6. The rule is now that the dice interact in the sense that when one flip the other flips. They flip in such a way that the sum of the numbers don't change. Each number has a cycle associated with it, so (1,1) must go to (1,1). The 3 cycle flips between (1,2) and (2,1) and so forth. There are a bunch of disconnected orbits. Once you fix the sum, you can throw away all the others and concentrate on the subsystem you're interesting in.

The most important conserved quantity is energy so that is where most of our energy will be conserved. In chemistry there are more, for instance the number of instances of atoms ofthe various elements are conserved. In nuclear physics there are other conserved quantities.

But let us go a step back and look at information. Information plays a very great role. Let's look at a much worse rule of movement than the ones above.

$$\begin{aligned}R &\to R\\ B &\to R\\ B &\to R\\ O &\to R\\ P &\to R\end{aligned}$$

This is a perfectly good deterministic law. There are no conserved quantities here. It's certainly true that over reasonable lengths of time the most likely thing you'll find is red. You'll simply find nothing else after a while. However it is not true that that there is equal a priori probabilities of any colors. This system lacks one property that real systems of movement has that Susskind calls "*conservation of information*", but you could equally well call it the "*conservation of distinctions*". It would mean that distinctions don't merge. Because the paths merge information is lost. This is irreversibility, but it is not thermodynamical irreversibility. The rules of statistical mechanics is fundamentally based on the fact that physics at the deepest level that the laws are consistent with the conservation of information (distinction). This is a strong restriction, without it we'll get nowhere. It is a good physical assumption and it is a consequence of a basic assumption of classical mechanics (*Liouville's theorem*). There is a *quantum mechanical* version (*unitarity*) but we'll not do much using QM in this course.

### 1.1.2 Conditions for applying thermodynamic theory

The classical world is fully deterministic, but it apparently statistical because it is coupled to a much larger system (a heat bath) about which you know very little. You don't know enough about the heatbath to specify the details of it. Things are random not because there is any intrinsic randomness in the laws of physics, but simply because you don't know enough. That's the basic idea of entropy.

This principle of conservation of distinction is so important. It is rarely mentioned because it is so deeply assumed by anyone that does classical physics. Susskind would call it the zeroeth law of thermodynamics, except that that name is used by something else, so perhaps we should call it the -1th law of thermodynamics.

Classical mechanics deal with continous systems with momenta. Each coordinate has associated with a set of momenta. Of course there is deeper idea about the idea of momentum but for the purpose of this class it can be just ordinary momentum.

So what is the state of a system? Well, for a simple die it was just the label color. If it's two dice, it is a pair of colors. For a single point particle, it is a collection of coordinates and the corresponding momenta. The historical symbol for momentum is "p" and for coordinate is "x". For an ordinary particle it would be three coordinates for position and three for momentum, so the phase space for a particle is six dimensional, and the configuration space (the position coordinates) is three dimensional.

What constitutes the state is the pair of the x and the p, this is because if you want to know where a particle is next, you not only need to know where it is, but you have to know where it is moving. So for a single particle the (x,p) is like the color of the surface of the dice, it labels the state of a particle. For a many-particle system there are many x-es and many p-s.

But first, what is a history? What is an orbit? You have some starting point, and then you have laws of motion, Newton's or others, and then you use them to follow the particle in time. Little "ticks" that are of some length. The motion is of course continous but you can divide it up. You will then get the trajectory in the phase space. There are diffent types of trajectories. In some systems you just give a particle a litte push and it goes of into infinity, but for the kinds of systems we are interested in that are contained within finite boxes, there is a rule that the the orbit will usualy wind up coming back or something, perhaps not the same point ;-)

What doesn't happen is that trajectories never, ever merges. This is the analog of saying that distinctions are preserved

- Trajectories never cross.
- What doesn't happen is that trajectories never, ever merges. This is the analog of saying that distinctions are preserved

- A given starting point never splits. It is always determinstic.

**Rmz:** *this should just be two points*

The different trajectories, whatever they do, may be labelled by their energy, and that energy doesn't change. You pick it once and for all and it stays that way forever.

Consider an *harmonic oscillator*. The motion in phase space is just an ellipsis. However, there are many trajectores, and what distinguishes them is the energy of the oscillator and where you started them. They don't cross, but that's a general principle.

**Rmz:** *When physicists don't know what to do, they often look at how the harmonic oscillator will work in whatever conditions are being studied.*

### 1.1.3 Liouville's theorem

Never crossing and never merging is not quite enough for us, because there is a situation that is almost as bad as merging, and that is where the trajectories don't actually merge, but come so closely together that they get asymptotically closer. For practical purposes you would then lose distinction (just wait long enough), but that doesn't happen. Trajectories don't merge in that way either. There is a theorem that states this (*Liouville's theorem*). It says.

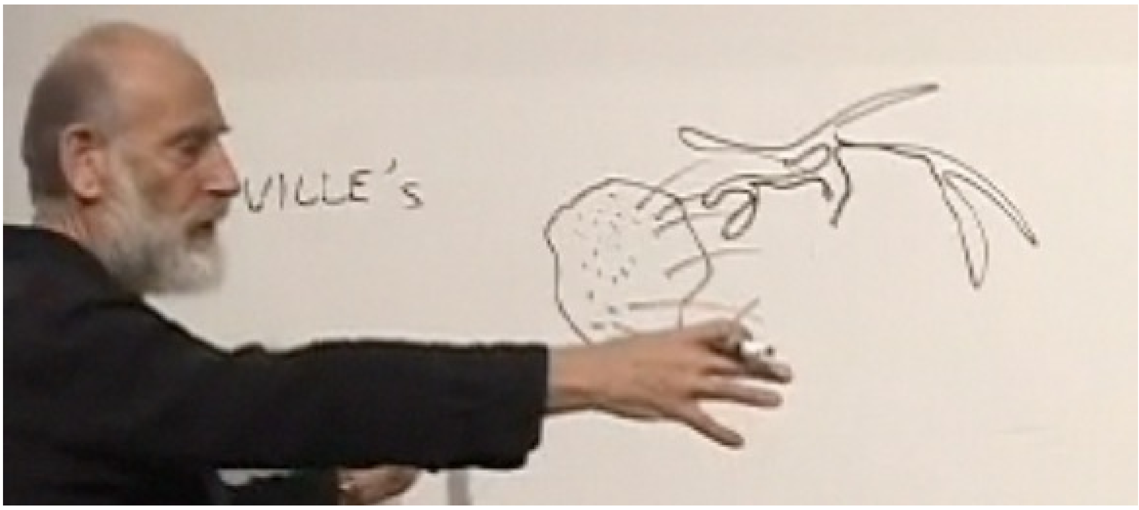


**Fig. 1.3** An illustration of Liouville's theorem.

> If you start with all of the points in a patch of phase space, at time t=0, and follow each one of them for a certain length of time, the volume of the patch of phase space doesn't change. [1]

[1] The unit of momentum is called "*action*", and for a three dimension particle, the action is of dimension $l^3p^3$, so when we're talking about "volumes of phasespace", it is volums over this type of unit.

So if the points contract in one direction then they spread out in another. This is another way to say that distinctions are not lost.

### 1.1.4 Ergodicity

Now an important point: Can it happen that the volume of this space stays conserved, but it branches out in some horrible, fractallated way so that it apparantly fills up the alloted volume in the phasespace. The answer is "yes", and it usually happens. However, it preseves it topology, no merges etc.

Liouville's theorem can be proved with a starting point either in the *Lagrangian* form of classical mechanics, the *Hamiltonian* form or the principle of least action. It all traces back to the *principle of least action*.

If we have a system that is enclosed so that it doesn't escape into infinity, and it shares the property with there are no merges and splittings etc., then if the system moves throught he phase space fast enough (or you wait long enough), then there is equal a priori probability (under the constraint of conserved energy, leading to surfaces in phase-space), are uniform in the phase space. Each volume of the phase space has equal a priori probability.

**Rmz:** *Wow!*

A sidenote about *ergodicity*. *Ergotic* is a bit related to "chaotic", it means that the phase point wanders about thoroughly througout the phasespace and pretty much touches every point int he phase space. In the above, Susskind is assuming that the system is ergodic. When a system is *not* ergodic it means there are extra conserved quantities: When a system is not ergodic that means that the phase space divides up into different pieces which carries different conserved quantities. Then you have to pick a value of the conserved quantity.

If a path in the phasespace doesn't touch every point in the phasespace, it means that there are conserved quantities. For example for the harmonic oscillator the phase space points stay on a given ellipse. The time average for each point on every point on the ellipse will be uniform for the single ellipse subspace.

*Comment from the audience*: Now, you can map the points of the unit circle to a circle with area two, and that does not preserve area but it is one to one. Susskind response that what we're looking at is more than one to one, it is phase-space volume preserving. The basic justification for the principle of equal voume in phase space really is quantum mechanical. Statistical mechanics as it is delt with in this class is fully classical so he's relucant to go to much into it, but if we were to do it we would divide volume into small volumes that are of $\hbar$ length on the sides, defined with the maximum amount of certainty allowable and take it from there. But we wont. However, it is only i quantum mechanics it is true that the number of states in a system is discrete and that you can actually count them and make it more like the "loopy" system described earlier.

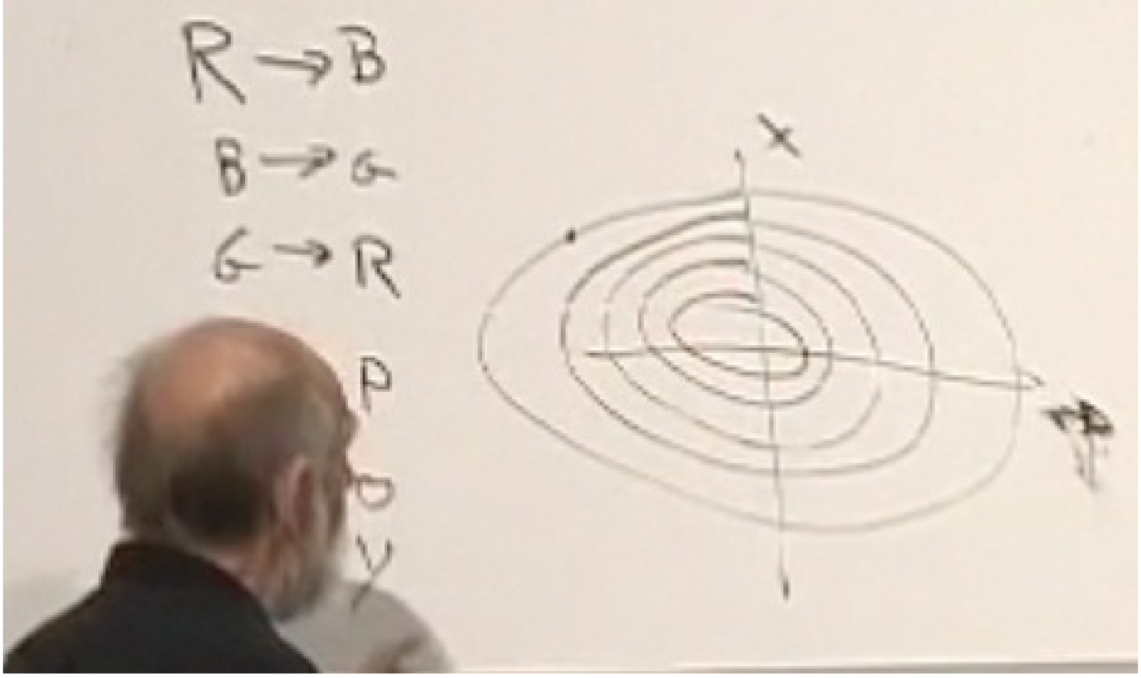


**Fig. 1.4** The phase space of the harmonit oscillator (to the right), orbits separated by energy levels.

## 1.2 Entropy

Now let's get to *entropy.* We've discussed energy. *Energy* is a conserved quantity, we fix it for a system, and within a perticular value of energy the system may hop around the states it is allowed to be in, and you then get the average probabilitiy that the system spend in any given state is equal for all states (average over the statespace).

Now *entropy.* You might have thought that the next topic in thermodynamics would be *temperature* at you might think temperature is more intuitive than entropy, but it's really not. You have of course a body sense of what temperature is, but you don't have a sense of what hot and cold is. However, it is *energy* and *entropy* that are the more basic concepts so let's talk about that, but before we do that let's talk about probability distributions.

It has a horizontal axis for the states, and a vertical axis for the probabilities. It may be discrete or continous. The only two requirements for something to be a probability distribution is that it is positive everywhere and that it sums (or integrates) to one. For integrals we talk about probability densities.

In the discrete case, if there is some function of i (the system state), some

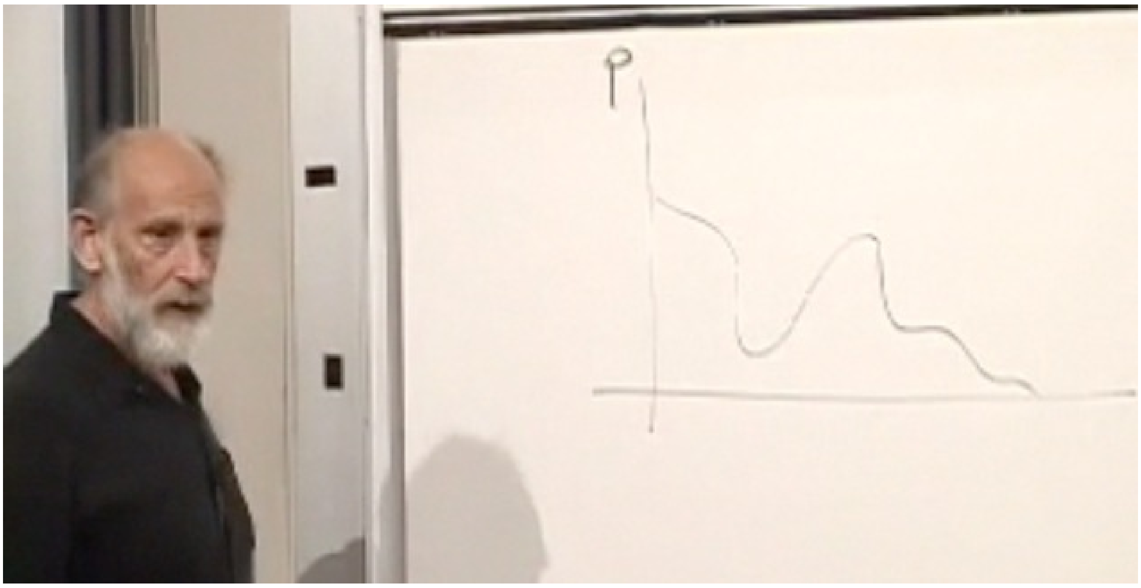


**Fig. 1.5** Just this probability distribution from somewhere

quantity $F$ that depend on the system. What is the average value of $F$ (we'll use the standard fysicist notation for averages and put a bracket around it: $\langle F \rangle$ ). The definition of the average:

$$\langle F \rangle = \sum_i F(i)P(i)$$

Just sum the values, and weight them by their probabilities.

Now keep in mind that the space of states can be multidimensional, but they are enumerated by the index. Suppose that the only thing we know about the system is that is in one of "$m$" possible states. What is the probability of being anywhere else than in one of the m states is zero. The probability of being in one of the m states is uniformly $1/m$. Btw, saying "nothing but", that is another way of saying "equally probably" that they are in any one of the "$m$". A statement of complete ignorance is equivalent to a statement of equal probabilities.

The "$m$" is a measure of our ignorance. The bigger "$m$", the bigger our ignorance. There are many ways of we can be ignorant: The state may be very small (microscopic) and there may be very many of them. All you know is some restrictive information. "$m$" is not the only statement of ignorence. Any monotonically increasing measure of $m$. The one called "entropy" is the logarithm of $m$, called $S$ (for entroy, Susskind don't know why)

$$\begin{aligned} S &= \quad \log m \\ S &= - \log(1/m) \end{aligned}$$

The basis of the logarithm can differ a bit between branches of science. In Physics it is always “e”, but in information theory it is usually “2”. The conversion factor is just a factor log(2).

Why take the logarithm? because it’s useful :-) Let’s look at an example: Assume that you have N coins each of which can be heads or tails. Supposing you know nothing about the system, each of the $2^N$ configurations are equally probably, so the entropy is $N \cdot log(2)$. Since the entropy is proportional to the number of coins, it makes sense to talk about entropy per coin. By taking the logarithm, we change the description from something really large, to something that is additive with respect to the number of components in the system. You have something that is proportional to the degrees of freedom. Entropy like energy is something that adds up for a system. Entropy is measured in bits.

Supposing we know everything about the system, in this case exactly which configuration we’re in. That would mean that ne of the states has probabiity 1 and all the other has zero. That would give an entropy of $zero = \log(1)$. With this definition of entropy, entropy isn’t just defined by the system, but your state of knowledge about it. That is a bit rediculous, because we will treat is as a quantity of the system (we’ll come to why), but really it’s a function of the probability distribution.

Now let’s find to some other definition. Consider a probability that has some width, close to zero (but not zero) and close to one (but not 1) at various places. Can we generalize the the number of states under the probability distribution (the logarithm of it). The definition is simply the average of the $-\langle log(P(x)) \rangle$ (negative since $P < 1$ so $\log(P) < 0$. How do we calculate that? We just use the formula for averages.

$$\begin{aligned} \langle F(i) \rangle &= \quad \sum F(i)P(i) \\ S &= - \sum P(i)log(P(i)) \end{aligned}$$

This is the final definition of entropy. The contributions from the probable regions are more important than the others.

*Bolzmann’s constant* traces back to the definition of temperature. Energy and entropy together determine temperature. In the early days physicists, steam engineers etc, were interested in temperature they really didn’t know what temperature was. They knew how to measure it (with thermometers). They didn’t even have the idea of an absolute zero temperature. The temperature of a gas is a measure of the energy of the individual molecules of the gas (more or less). The conversion factor between entropy and temperature wasn’t known. Bolzmann (and Maxwell)

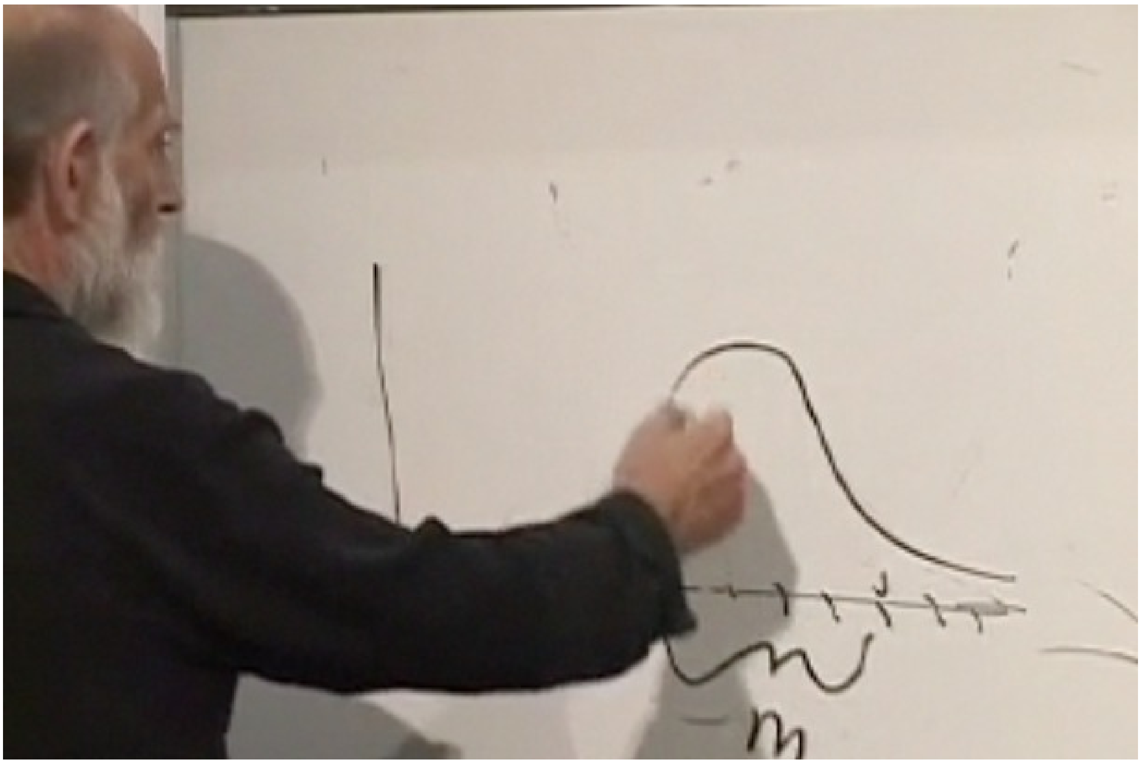


**Fig. 1.6** Entropy based on a general probability distribution

realized for an ideal gas energy is a measurement of the energy of the molecules. The conversion factor was unknown basically because they didn't know how many molecules there were in a volume of gas. Bolzmann's constant was unknown. Today we know, so Bolzmann's constant is known. We can, if we like, work with units of temperature which is really just units of energy. We can measure temperature in Joules. That's just a historical fact. *Bolzmann* did not know the value of his own constant. Bolzmann was so depressed by the fact that he didn't know his own constant that he committed suicide, and the next year Susskind believe *Einstein* figured out what Bolzmann's constant was from the brownian motion (Susskind then says he don't know why Bolzmann committed suicide, but believes it wasn' because he didn't know his own constant). *Newton* btw. didn't know his own constant either. It took many years for Newton's constant to be measured. *Planck* knew his constant ;) One of the hard constants to measure was the electric charge. It was easy to measure the ration of charge to mass, then it took some time before *Milliken* figured out how to measure the charge separate from the mass. Whoever figured out that electrons had charge didn't know the value of the charge (it may have been *Benjamin Franklin*, Susskind thinks).

### 1.2.1 Thermal equilibrium

Let us now postulate the existence of something called "*thermal equilibrium*" of a system. Let's state a necessary, but not necessarily sufficient condition: Thermal equilibrium is not a property of an isolated system. If you have a truely isolated system, it is *not* in thermal equilibrium. It has an energy, and it's fixed. Thermal equilibrium is a property of a system $A$ in contact with a much bigger system $B$, called the "*heat bath*". It is a very big system with many more degrees of freedom. The combined system $(A + B)$ has a given total energy (that's an assumption). For our purposes the bigger system can be thought of as a closed and isolated system. It is either contained within insulating walls that don't allow any heat energy in or out, or something similar. $A + B$ can be an isolated system but $A$ is not. One more thing. $A + B$ is isolated, but neither A nor B is isolated. They weakly interact. Thies meahns that the energy of the interaction are very small compared to the energy of either $A$ and $B$, but the interactions, whatever they are, allow energy to go back and forth between $A$ and $B$. The whole system has a definite energy, so whole system has a definite value of "zilch", ("zilch" being energy in this case). The whole system will move around its phasespace on a surface of constant energy, but it will move around and hop from pint to point. Among other things, the points will give different ways of partitioning the given amount of energy into the energy of the heatbath of the energy of $A$. Neither of these energies will have a definite predictable value. It will fluctuate, it will have a *probability distribution.*

### 1.2.2 The relation between energy, temperature and entropy

If you wait long enough there will be a probability of the energy of $A$. The various configurations of $A$ may be discrete or continous, but over time the energy will have a time evolution. The probability of an energy of $A$ is a function of both the energy of $A$ and the average energy of $A$.

$$P(i, E_{\text{total}})$$

There are two things you can calculate if the probability distribution. You can calculate the average energy of $A$, and you can calculate the entropy of $A$. $(S_A = \langle \log P_A \rangle)$.

$$T = \frac{\partial E}{\partial S} \log 2$$

In general entropy increases with energy, which leads to the question: "How much change of energy is necessary to change the entropy by an amount of one bit"? This change of energy per unit of entropy, is called "*temperature*". To state it colorfully: *Temperature is the energy needed to hide one more bit of information.*

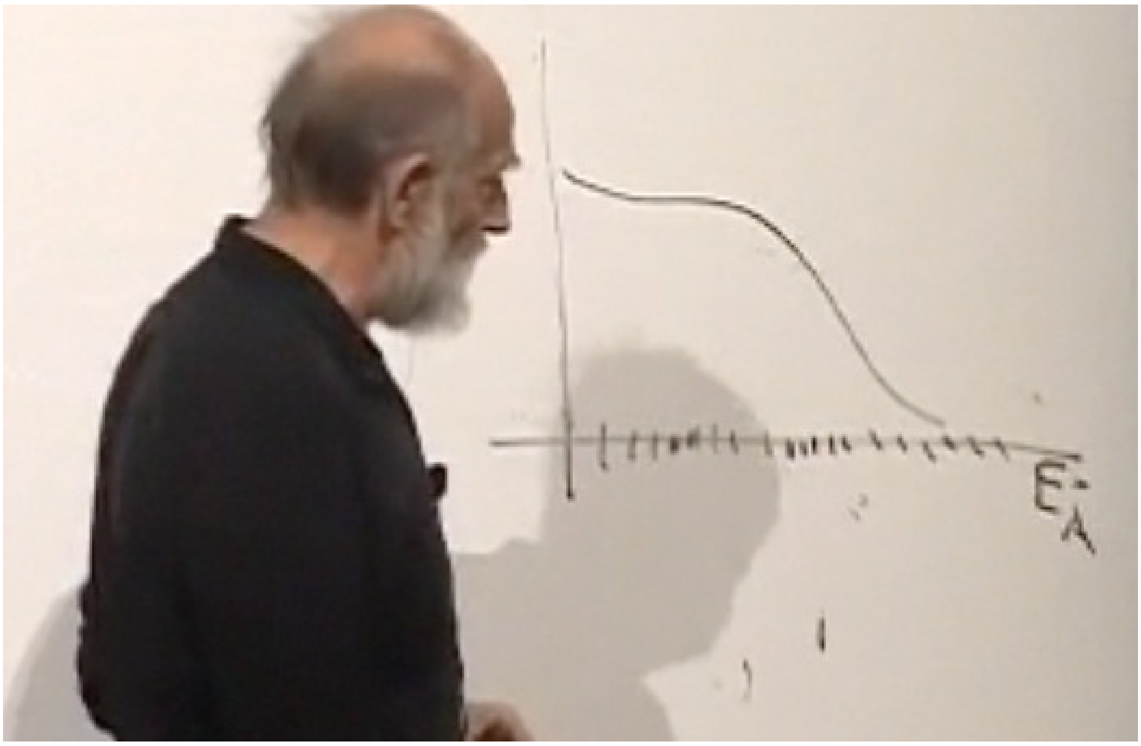


**Fig. 1.7** Time evolution of the energy of A

An example of this is *Landauers principle* in computing: If you want to erase a bit of information in a computer, remembering that we can't destroy information (principle of conservation of distinction), you wind up putting at least that bit into the heatbath surrounding the computer. How much energy do you put out of the and into the heatbath to hide a bit, well it's the temperature times the logarithm of two. That is the energy you put into the environment when you erase a bit.

*Bolzmann's constant* incidentally is basicically the inverse of *Avogadros number* (or something very close to it), this means that when you change a bit of information at a temperature $T$ (which is the temperature of the surroundings).

The standard thermodynamic defintion of entropy also differs by the standard definition of entropy by a factor of $b_k$ (upstairs or downstairs, Susskind can't remember :-).

Einstein realized that the quantities of thermodynamics fluctuated. This would have consequences on small impurities of the system and knock them around. What he did was essentially to calculate the fluctuations using statistical mechanics to various quantities like pressure and so on and demonstrated how that would knock small impurities about in the system.

Next time we will work out and derive the Bolzmann distiribution. The Bolzmann distribution is the probability distribrution for the states in the system $A$.

# 2 Finding Bolzmann's constant

We'll start this presentation by reviewing some math we are going to need. It's nothing fancy. Just stuff many of us has had in college or high school, but we'll need it so we review it and then we'll use it.

## 2.1 Lagrange multipliers

We're constantly minimizing or maximizing things subject to constraints. The thing we're most likely to maximize or minimize is entropy, but we'll get to that later. *Thermal equilibrium* is a state of *maximum entropy* so that means that we want to calculate the maximum as a variable of the functions that define it. Then there are *constraints*. We don't just want to maximize a function, we want to do so subject to some constraints. Some plausible constraints are: That we know the total energy of the system, or the average energy of the system. For example we might have a container full of a lot of molecules, we'll study one small piece of it. If we're told what the energy of a part is. We know what the average energy is, and we want to *maximize the entropy subject to the constraint that total energy is fixed.* That is the kind of problem that occurs over and over. A related problem could be that we know the total amount of electric charge in a region, and we may want to maximize the total amount of entropy given a total amount of electric charge, or the total amount of angular momentum or whatever.

We constantly face problems of maximization and minimization of several variables subject to constraints on those variables. What does that mean? Let's take an example: We want to minimize $x^2 + y^2$. Now anyone can minimize that, it's of course $x = y = 0$). But suppose we add a constraint and say that we want the minimum, give that $x + 2y = 1$. To draw a picture we can draw the contours of $x^2 + y^2$. Right at the center it is minimized, and as we move away the value gets larger. If we then draw the $x + 2y = 1$ (figure ??). By inspection it's not so hard to see that the minimum must be where the straight line is closest to the origin. How do you find that point? One of the ways to do that would be to solve the linear equation with respect to one of the variables, plug that into the quadratic formula and solve. Doing that you get $x = 1 - 2y$ which when substituted into the quadratic formula gives $(1 - 4y + 4y^2) + y^2 = 1 - 4y + 5y^2$. Taking the derivative of this gives $10y - 4$ which is equal to zero at $y = 2/5$, by substituting into the straight line we

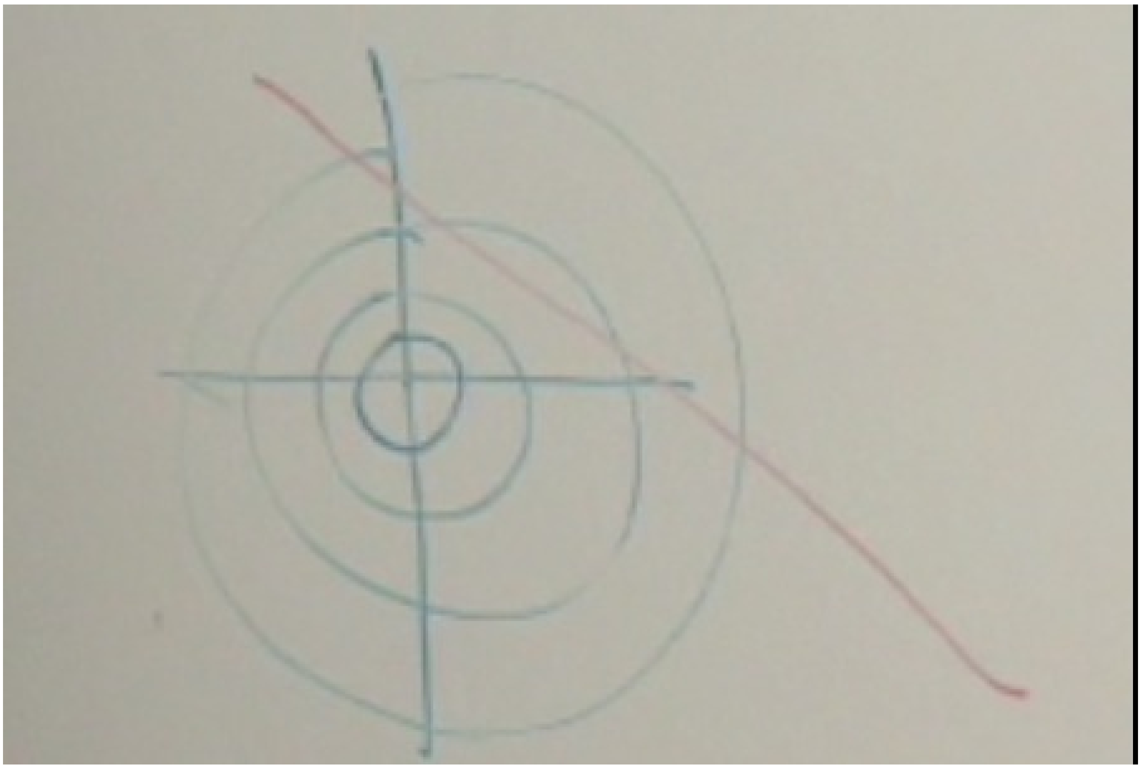


**Fig. 2.1** Contour plot of $x^2 + y^2$ and $x + 2y = 1$

find to be $x + 2 \cdot 2/5 = 1$ which makes $x = 1/5$, so the minimum is at the point $(1/5, 2/5)$. That's one way of doing it, but in many cases the constraint is just to complicated to solve. It might be a very complicated function. However, there is another way of doing it, and that is called the method of *Lagrange multipliers.*

The rule for Lagrange multipliers is: Given that you have some function in some variables, e.g. $F(x, y)$ and we want to minimize it subject to the constraint that some other function $G(x, y) = 0$. The trick you take is to multiply the constraint by a new variable often called $\lambda$ (lambda), called an *lagrange multiplier.* You then add that multiplier timest the constrant to the function you want to minimize, and get a new function:

$$F(x, y) + \lambda G(x, y) = 0$$

What we then do is to minimize the new function, ignoring the constraint. You minimize by differentiating with respect to the variables and you set the result to zero, so we get a set of equations:

$$\begin{array}{rcrcl} \frac{\partial F}{\partial x} & + & \lambda\frac{\partial G}{\partial x} & = & 0 \\ \frac{\partial F}{\partial y} & + & \lambda\frac{\partial G}{\partial y} & = & 0 \end{array}$$

So we have two equations and two unknowns so we can solve it for $x$ and $y$, but what about the $\lambda$? The answer will depend on $\lambda$ . What we do with the $\lambda$ is to adjust it so that the original constraint $G(x, y) = 0$ is really satisfied.

We'll see *how* it works in the example we solved above, then we'll see *why* it works.

### 2.1.1 How it works

We assume that $F(x, y) = x^2 + y^2$ and we want to minimize that subject to the constraint that $G(x, y) = x + 2y - 1 = 0$. We follow the procedure above and get that we concoct the function:

$$Z(x, y) = F(x, y) + \lambda G(x, y)$$

Then we find the partial derivatives and set them to zero:

$$\begin{array}{rcrcl} \frac{\partial F}{\partial x} & + & \lambda\frac{\partial G}{\partial x} & = & 0 \\ 2x & + & \lambda & = & 0 \\ & & x & = & -\lambda/2 \end{array}$$

and

$$\begin{array}{rcrcl} \frac{\partial F}{\partial y} & + & \lambda\frac{\partial G}{\partial y} & = & 0 \\ 2y & + & \lambda 2 & = & 0 \\ & & y & = & -\lambda \end{array}$$

We now choose $\lambda$ so that the constraint $x + 2y - 1 = 0)$ is satisfied. We do this by substituting the solutions for $x)$ and $y$ in terms of $\lambda$ and get $-\lambda/2 - 2\lambda - 1 = - - \lambda 5/2 - 1 = 0$ which implies that $\lambda = -2/5$. We can then use this value of $\lambda$, substitute back, and get that $x = 1/5$ and $y = 2/5$, which is exactly the same result as we got above, as it should be :-)

### 2.1.2 Why it works

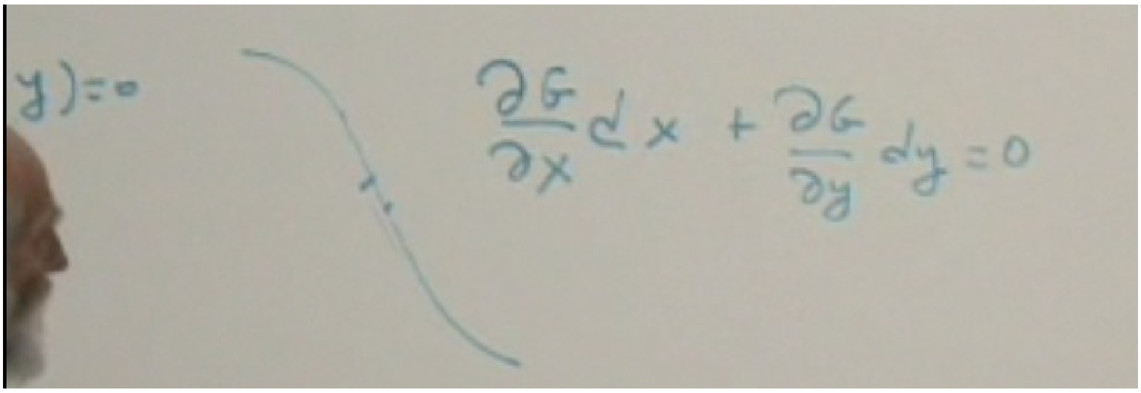
It only takes a little bit of algebra to show why this works. This is in fact a very general function, the constraint didn't have to be linear, and the function to optimize on certainly didn't have to be quadratic.

We'll proceed by solving a general problem both with substitutions and the Lagrange multiplier method.

### Using substitutions

We start with a function $F(x, y)$ subject to some constraint $G(x, y) = 0$, and we want to minimize $F$. We can go about this by solving $G$ with respect to one of the variables, so we get some curve e.g. $y(x)$. We then plug that into $F$, so that we get $F(x, y(x)$. This is now a function of a single variable $x$ since $y$ is now a definite function of $x$. The next is to differentiate with respect to $x$ to find the minimum with respect to $x$:

$$\begin{array}{lcl} F(x,y) & = & F(x,y(x)) \\ D(F(x,y)) = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} & & \end{array}$$



y)=0
$\frac{\partial G}{\partial x} dx + \frac{\partial G}{\partial y} dy = 0$

**Fig. 2.2** A small differential on G does not change G

We then set $D(F(x, y)) = 0$. But what about $\frac{dy}{dx}$, let's see what we can figure out about it. The curve $G(x, y))$ is a curve of "constant $G$". Let's assume we're moving along it and make small change along it, by definition this will not change $G$. So, we can also write:

$$\frac{\partial G}{\partial x}dx + \frac{\partial G}{\partial y}dy = 0$$

this says "G doesn't change when we make a small differential displacement". The reason we did this that it allows us to solve for the $\frac{dy}{dx}$ in terms of things involving the constraint $G$:

$$\frac{dy}{dx} = \frac{-\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial y}}$$

Now we plug that in into the equation above and get:

$$\begin{array}{rcrcl} \frac{\partial F}{\partial x} & + & \frac{\partial F}{\partial y}\frac{dx}{dy} & = & 0 \\ \frac{\partial F}{\partial x} & - & \frac{\partial F}{\partial y}\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial y}} & = & 0 \\ \frac{\partial G}{\partial y}\frac{\partial F}{\partial x} & - & \frac{\partial F}{\partial y}\frac{\partial G}{\partial x} & = & 0 \end{array}$$

That's not a bad looking equation. After eliminating $y$, this is the equation that solves for the minimum of $F$ subject to $G$. We're not gonna use this, but we will see that we get exactly the same result when using Lagrange multipliers.

### Using Lagrange multipliers

Recall the two equations that followed from applying the method of Lagrange multipliers:

**Rmz:** *The signs in the calculations below are a bit bogus, needs to be checked again*

$$\begin{array}{rcrcl} \frac{\partial F(x,y)}{\partial x} & + & \lambda\frac{\partial G(x,y)}{\partial x} & = & 0 \\ \frac{\partial F}{\partial y} & - & \lambda\frac{\partial G}{\partial y} & = & 0 \end{array}$$

Now we're going to eliminate $\lambda$ from this equation, we do that by solving for $\lambda$ in one equation and plug it into the other:

$$\lambda = -\frac{\partial F}{\partial x}/\frac{\partial G}{\partial x}$$

We then plug it in:

$$\begin{array}{rcrcl} \frac{\partial G}{\partial x}\frac{\partial F}{\partial y} & - & \frac{\partial F}{\partial x}\frac{\partial G}{\partial y} & = & 0 \end{array}$$

By comparing, we see that it's the same equation (modulo a factor of -1 :) as we got with the substitution method.

### Summing up what we've got so far

There is an intuitive geometric picture of what's happening, but it takes longer to draw that than to do the arithmetic. The method of Lagrange multipliers is completely equivalent to using substitutions, and is usually easier. The rules are:

**Rmz:** *Perhaps find some artwork online and paste it in here?*

- Add the function to be optimized to the constraint times a factor $\lambda$.
- Then differentiate the whole thing (the sum) as if you are trying to minimize the sum.
- You then get two (or more equations), solve them for the variables.
- Then you chose the $\lambda$ so that the constraint $G(x, y) = 0$ is satisfied.

**Rmz:** *Doing a few of these examples is good for you, do them :-)*

### 2.1.3 Multiple variables and multiple constraints

One of the nice things about the method of Lagrange multipliers, is that it doesn't force you to single out a single variable for special treatment, it treats the variables symmetrically. Suppose we had a function of several variables and a collection of constraints.

$$F(_1x, \ldots, X_n)G_1 = 0, \ldots G_m = 0$$

It's a precondition that you have more variables than constraints. In the general case, there is no solution if you have more constraints than variables.

You then introduce a Lagrange multiplier for each of the constraints:

$$Z = F + \lambda_1 G_1 + \ldots + \lambda_1 G_m$$

Then you make a new set of equations:

$$\frac{\partial Z}{\partial x_1} = 0, \ldots, \frac{\partial Z}{\partial x_n} = 0$$

Solve that system, and then choose all of the $\lambda$s you have so that all the constraints are satisfied.

That is the method of Lagrange multipliers, and it is central to statistical mechanics.

The typical thing to minimize is the entropy. Some typical constraints are total energy, total whatever (charge, spin, ...)

## 2.2 Stirling's approximation to the factorial function

$N$ factorial $N!$ occurs over and over in all sorts of probability contexts. Every time you calculate permutations in order to calculate probabilities, $N$ factorial occurs all over the place. Generally the $N$s that occurs are very large numbers, so it is important to have a reliable approximation which becomes better and better as $N$ becomes large.

$$N! = 1 \cdot 2 \cdot \ldots \cdot N$$

We will start by estimating $\log(N!)$, and then we'll exponentiate that to get an estimate of $N!$.

So what we're doing is to add up the area under the logarithm curve in figure **??**. When $N$ becomes large, it's a good approximation of the sum to replace it by the integral, so let's consider then:

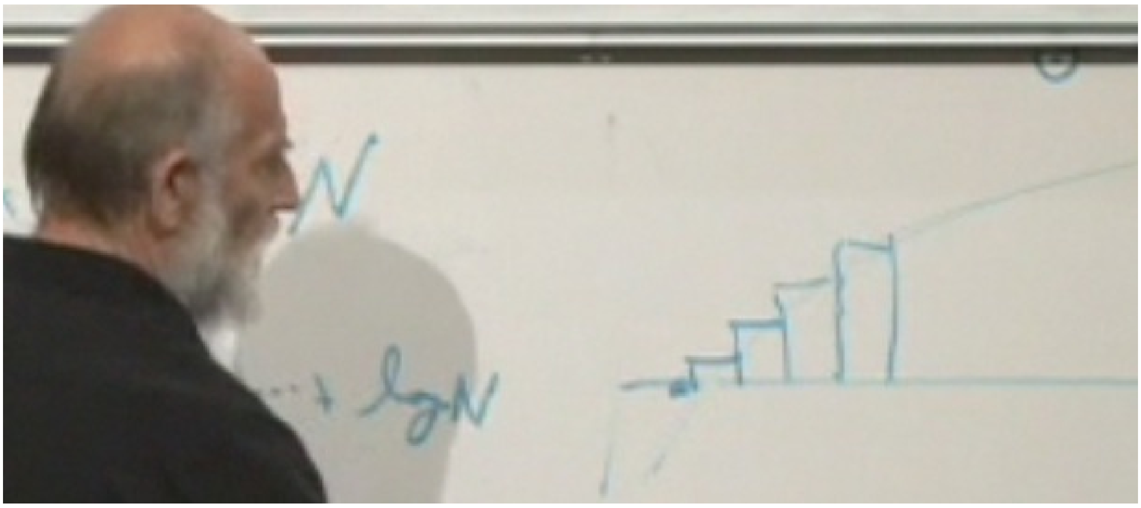


**Fig. 2.3** A rendition of the logarithm function

$$\begin{aligned}\log(N!) &= \log(1)+\log(2)+\log(3)+\ldots+\log(N)\\ &= \textstyle\sum_{i=1}^{N}\log(i)\\ &\approx \textstyle\int_{i=1}^{N}\log(i)dx\end{aligned}$$

We then have the problem of integrating $\log(x)$. It's easy to do, you just look it up in a table of integrals and find. However, for very large values, $\log(x)$ is almost constant, so an approximation of the approximation might be $x\log(x)$. Let's see what the derivative of $x\log(x)$ is just to see how far off we are:

$$(x\log(x))' = \log(x)+1$$

That's not entirely right, but it's easily fixable; we just add a term $-x$ to the presumed derivative and we're where we want to be (the integral of $\log(x)$:

$$(x\log(x)-x)' = \log(x)$$

to continue the train of thought started above, we get:

$$\log(N!) \approx N\log(N)-N$$

and that's Sterling's approximation for $\log(N!)$, and it's in this form we'll usually use it. However, just to complete the exercise we exponetialize this value to find the approximation of $N!$:

$$N! \approx e^{N\log(N)-N} = N^N e^{-N}$$

### 2.2.1 Number of ways a partitioning into states can be achieved

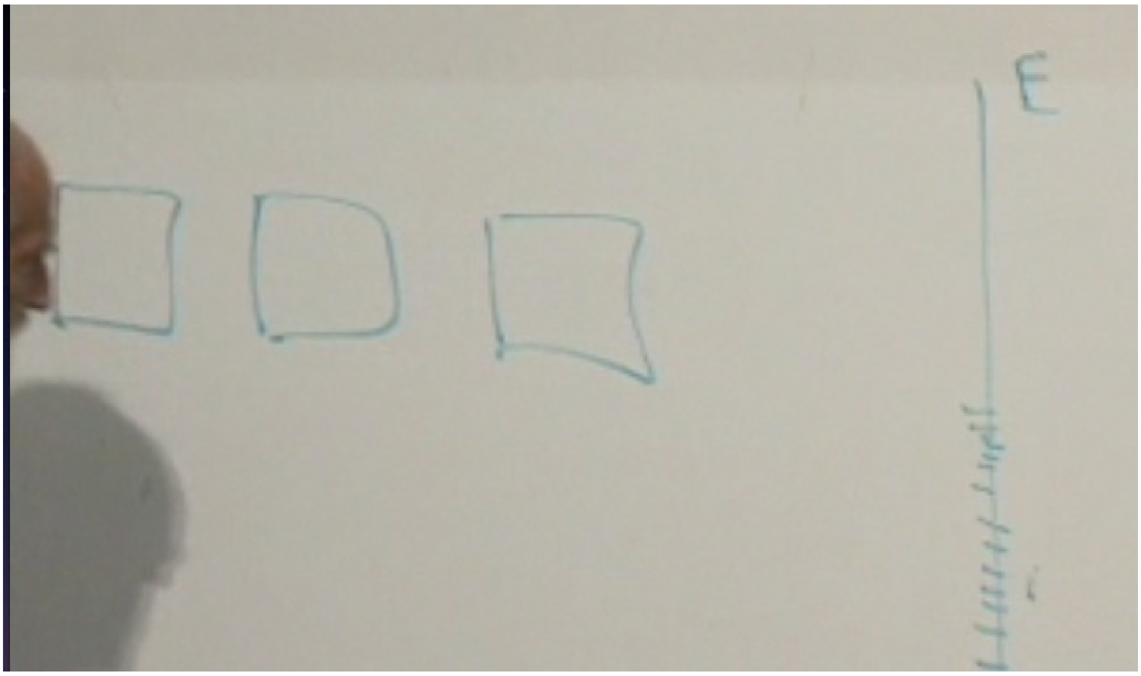


**Fig. 2.4** Boxes in energy states

If we have en copies of the same system that can be in a state (corresponding to the energy state of the system). Each of these boxes can be identified with a state on the $E$ scale, (a particular $i$). The boxes can be assigned to any any of the $i$s, and any $i$ may be mapped to by many boxes (or none). There are $N$ systems (boxes) each with its own state.

We can now ask the question: How many boxes are there in state $i$? In general the answer will be an integer $n_i$ for all possible values of $i$. There will be a constraint that the sum of all the $n_i$s add up to number of boxes:

$$\sum_i n_i = N$$

A question we can now ask is: *How many ways are there to coosing a state for each box such that there are $n_1$ boxes in state 1, $n_2$ boxes in state 2 and so on forth? How many ways are there to get a given set of ns?*

The answer

$$\frac{N!}{n_1!n_2!\ldots}$$

This only works because $0! = 1$ by definition, so all the unoccupied states gives 1.

### Why?

Why? Well, I'm guessing that

$$\left(\begin{array}{c} N \\ n_1 \end{array}\right) \cdot \left(\begin{array}{c} N - n_1 \\ n_2 \end{array}\right) \ldots \left(\begin{array}{c} N - \sum_k^{i-1} n_k \\ n_i \end{array}\right) \ldots = \frac{N!}{n_1!n_2!\ldots}$$

And given that I believe that, I really should prove it, so here goes :-).

We know that $\left(\begin{array}{c} n \\ k \end{array}\right) = \frac{n!}{k!(n-k)!}$ so we can try substituting into the first pair of "k of N"s and we boy are we lucky, because:

$$\begin{array}{rcl} \left(\begin{array}{c} N \\ n_1 \end{array}\right) \cdot \left(\begin{array}{c} N - n_1 \\ n_2 \end{array}\right) & = & \frac{N!}{n_1!(N-n_1)!} \cdot \frac{(N-n_1)!}{n_2!(N-n_1-n_2)!} \\ & = & \frac{N!}{n_1 n_2!(N-n_1-n_2)}!) \end{array}$$

which is really useful as a basis for an induction proof, since the result above can immediately generalized (through simple variable substitution) to

$$\left(\begin{array}{c} N \\ n_i \end{array}\right) \cdot \left(\begin{array}{c} N - n_i \\ n_{i+1} \end{array}\right) = \frac{N!}{n_i n_{i+1}!(N - n_i - n_{i+1})!}$$

Now we already know that $\sum_k n_k = N$ (for all possible $k$s), so that if $k$ gets large enough, we can deduce that $(N - \sum_k n_k)! = 0! = 1$ which will serve as the termination of our induction.

tbd.

## 2.3 The Bolzmann distribution

In an ensamble of particles there will be an energy distribution, and that distribution is, given that the particles are in a thermal equilibrium *heat bath*, is given by the *Bolzmann distribution* (sometimes called the *Maxwell, Bolzmann distribution*, but it was really Bolzmann who got it right.

It's simplest for a system $A$to let the heatbath. The system has a system of states. We have a bunch of discrete states, there is a lowest energy level but no highest level, an infinite number of states.

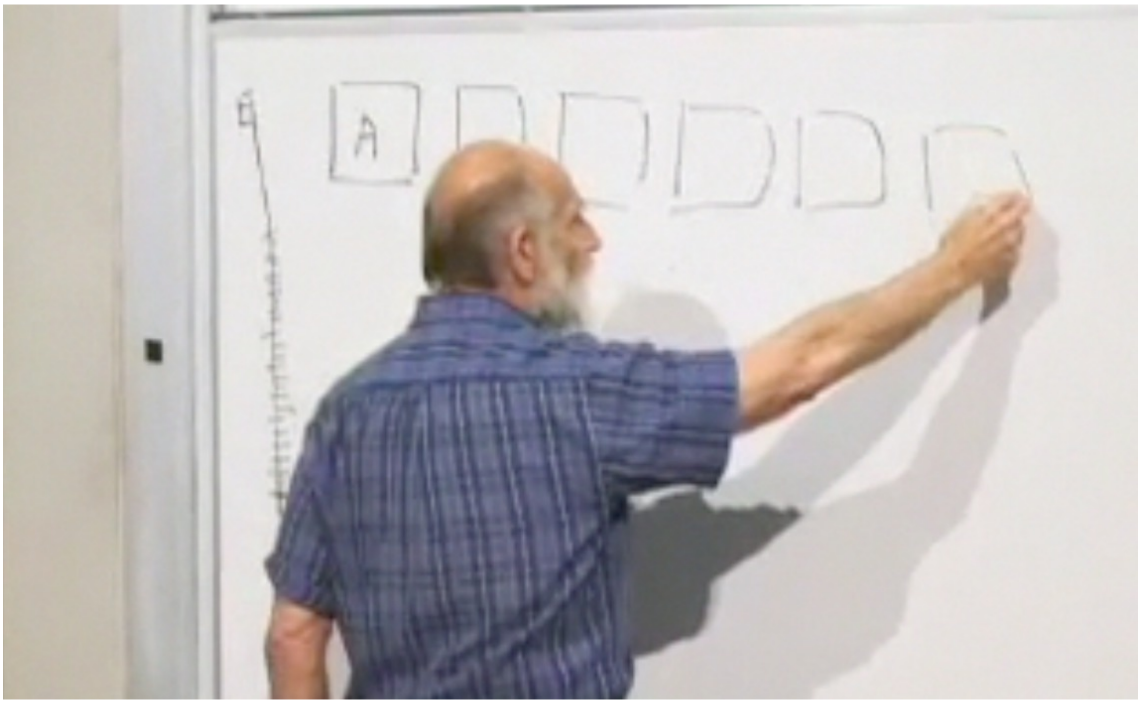


**Fig. 2.5** A bunch of copies of a system A

In a special case of heat bath, we assume a bunch of copies of the same system over and over again. When $A$comes to thermal equilibrium in a heatbath, it is essential that heat can be exchanged between them. That can happen in a variety of ways: Molecules moving about, radiation going back and forth etc. We can think of it as pipes connecting all the copies. There are $N$ copies, and $N$ is very large.

We will now talk about something called *occupation numbers*. These are not the occupation numbers from quantum field theory, is something else. The question we are answering is "How many boxes occupy a particular energy lever $i$". There can in principle be any number of boxes in any individual states, so we introduce a notation. The number of boxes, drawn from big $N$ that occupies state $i$ is denoted $n_i$. Now, most of these states will have zero boxes assigned to them, since there is an infinite number of states and only a finite number or boxes. We will also assume that the total energy of the collection is fixed, let us call it $E_{\text{total}}$. We won't even define it, since there is an *average energy* that each box has, and they are all identical to each other. When they come to thermal equilibrium they will all have the same energy just by symmetry. So since all the boxes has the same energy, the total energy will simply be $N$ times the average energy for each box:

$$E_{\text{total}} = N \cdot E$$

Notationalwize Susskind will sometimes use the “overbar”, sometimes the “angle brackets” to indicate average energy, and sometimes (like above), just ignore any special notation.

So we have the total energy as one fact. Another fact we have is the set of $n_i$s. They are subject to two constraints:

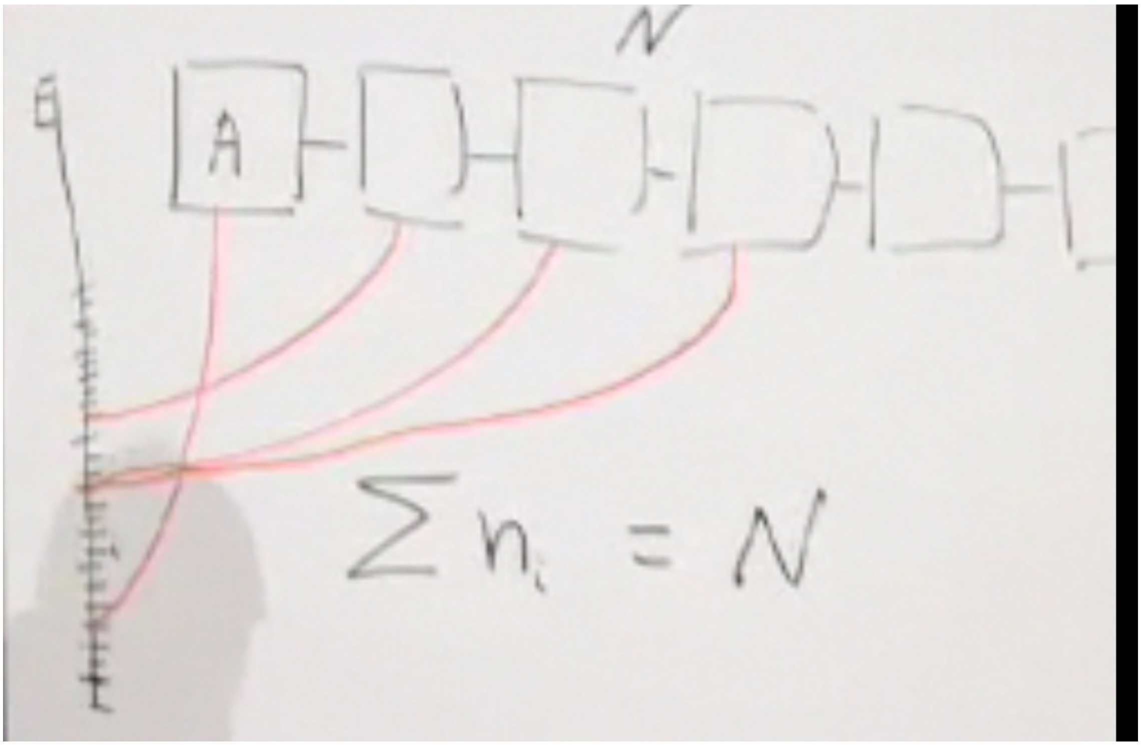


**Fig. 2.6** The number of boxes doesn't change

1. The number of boxes doesn't change (by assumption).
2. $\sum n_i = N$

Since $N$ is bounded, most of the $n_i$ states must be unoccupied.

Furthermore, if we look at the total energy, that doesn't change either:

$$\sum n_i E_= E_{\text{total}} = N\,[E]$$

We must remember that these sums are not over the boxes, they are sums over the energy levels. Furthermore the number boxes in a state multiplied with the energy of a specific energy level, for all energy levels, must add up to the total energy. The total energy never changes.

These are constraints that apply whatever we do.

Now, the next question to ask is:

> How many ways are there, given a set of occupation numbers, that those occupation numbers can be realized?

If we have a number of states and two boxes, where $n_1 = 2$ and $n_2 = 0$, how many ways can this add up? Well, just one. Both boxes has to be in state 1. In general there may be many ways to ensure that a given set of occupation numbers is fulfilled. If for example wa say $n_1 = 1$ and $n_2 = 1$. In general there will be many ways, and that number of ways to arrange the boxes is in general, the number of way to partition the energy into different sets of boxes (or something ;) is:

$$\sharp = \frac{N!}{n_1! \cdot n_2! \cdot \ldots} = \frac{N!}{\prod n_i!}$$

That's the answer to that combinatorical problem.

In practical applications these things should be thought of as functions of the $n_i$s. All the "n"s in the problem is going to be large. The capital $N$ is a huge huge number. So huge that the occupation numbers will be really big. Technically they will be proportional to $N$ itself.

We can think if the object $\sharp$ as a function of the little $n$s. It is also ha highly peaked function. When the $n$s get very large it will have a sharp peak, and it doesn't fluctuate very much. That means that when the number of boxes get large, you can be surer and surer what the little $n$s are. On the face of it this seems a little rediculous, but it simply means that the number of $n$ is narrow with respect to $N$ itself, the number $n_i/N$ don't fluctuate very much. This suggest that we should give them a name, so we'll call it $P(i) = n_i/N$. Why? because it makes a lot of sense to think of it as a probability of any given box to be in the state $i$. That is the basic tool. Let us now rewrite the two constraints we had.

1. First we get $\sum P(i) = 1$, not a big surprise, the sum of the probabilities adds up to one.
2. The other constraint: $\sum n_i E_i = N\bar{E}$. Now, if we divide this by capital $N$, we get:

$$\bar{E} = \frac{\sum n_i E_i}{N} = \sum P(i) E_i$$

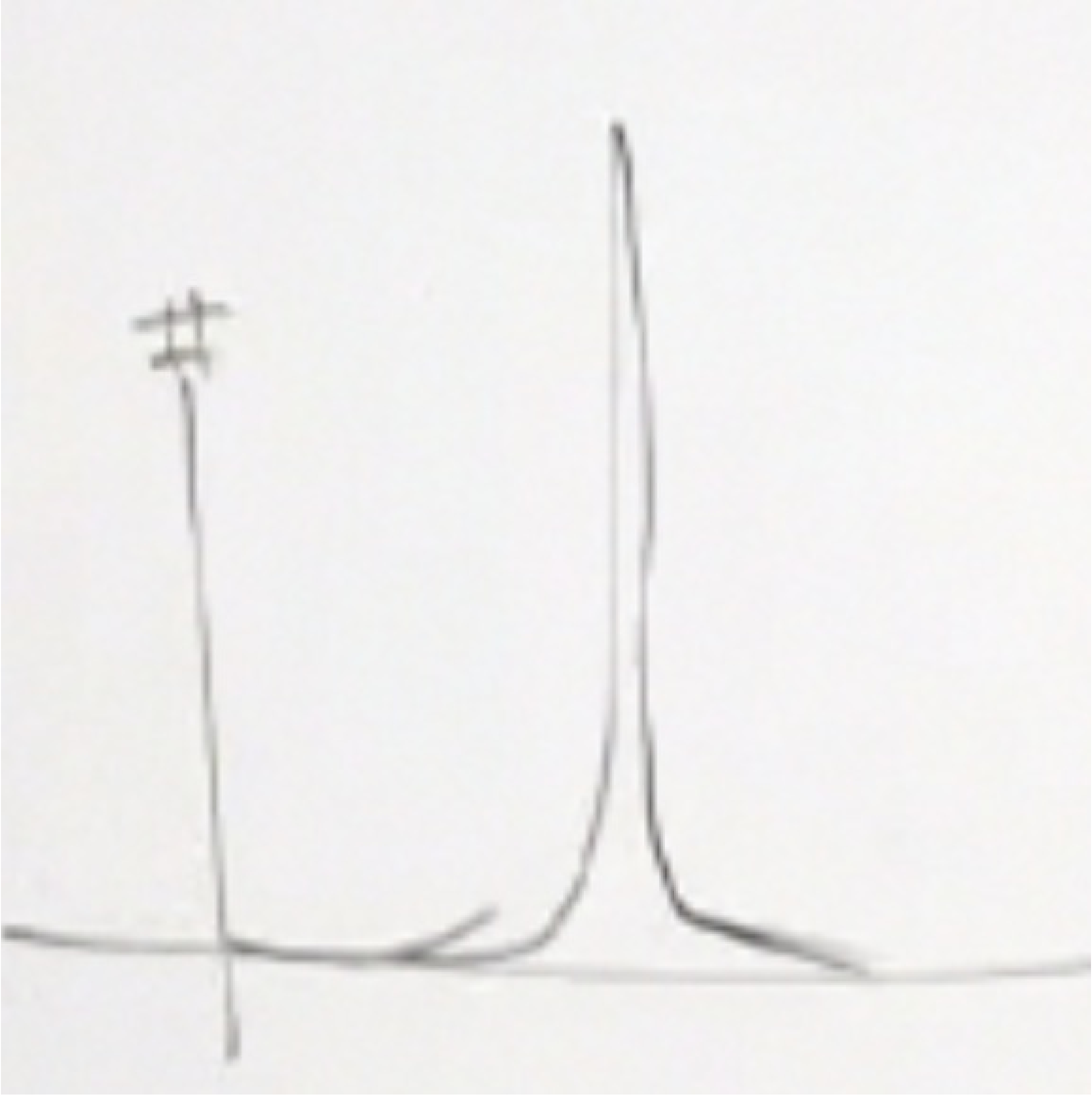


**Fig. 2.7** The number ♯ as a function of the number of $n$s

> This is the very familiar statement that for some quantity (in this case $E$, the energy) for a system that is made up of states, the quantity of the system is the sum for all the components in shstem of the product of the quantity for a state, times the probability of the system being in that state.

nn Now, what are the $P(i)$ values? What we have to do is to find the maximum value of the ♯ function, as a function of the $n$s, subject to the two constraints listed above.

This may look as a bad problem to solve, but fortuntatly it is surprisingly easy to solve. There are many way to do it, but the most efficient procedure is to say that you are maximizing $\sharp$, you can instead say that you should maximize its logarithm. This is interesting since it is simpler to work with sums than with products. So we need

$$\log(\sharp) = \log\left(\frac{N!}{n_1! \cdot n_2! \cdot \ldots}\right) = \log\left(\frac{N!}{\prod n_i!}\right) = \log(N!) - \sum \log n_i!$$

But this simplifies when using the *Stiring approximation*, which can be written in two ways, either $N! \approx N^N e^{-N}$ or you can take logarithms on both sides and get $\log N! \approx N \log N - N$.

So using the latter we get, by substitution:

$$\log(\sharp) = N \log N - N - \sum (n_i \log n_i - n_i)$$

We can work a bit on this. We can use the assumption that for each $i$ we have that $n_i = NP(i)$. We'll plot it in and grind it through, and then we get:

$$\log(\sharp) = NlogN - N - N \sum (P(i) \log (P(i)N)) - \sum n_i$$

Now, $\sum n_i = N$, so we can use that and get:

$$\log(\sharp) = NlogN - N - N \sum (P(i) \log (P(i)N)) + N$$

That cancels out the leading capital N.

$$\log(\sharp) = \log(\sharp) = N \log N - N \sum (P(i) \log (P(i)N))$$

and we can simplify even more:

$$\begin{array}{rcl} \log(\sharp) & = & N \log N - N \sum P(i) \log(P(i)) + N \left(\sum P(i)\right) \log N \\ & = & -N \sum P(i) \log(P(i)) \\ & = & -\sum_i NP(i) \log(P(i)) \end{array}$$

(using the fact that $\sum P(i) = 1$ ). We have seen this before. It is just $N$ times the entropy. Why $N$ times? Well, there are $N$ boxes.

Incidentally, we can get rid of the $N$ since we're trying to maximizing something here, and maximizing anything times $N$ is essentially the same problem as maximizing the same thing multiplied by 1 instead of $N$, and since the latter is simpler we'll do that.

Our mathematical problem now is to figure out the number of way to satisfy a given occupation number, is simply $\log(\sharp) = -N \sum_i P(i) \log P(i)$. Now what we want to do is to maximize this thing, including the minus sign. We want to find the most likely distribution of occupation numbers ($n_i$s), subject to the two constraints listed above ($\sum P(i) = 1$, and $\sum P(i)E(i) = \bar{E}$. How do we do this? Well, we use the method of *lagrange multipliers*. We gather the lagrange multipliers from each constraint, put them all together and get a set of equations that can be solved $\log(\sharp)$ can be viewed either as a function of occupation numbers or probabilities, and in this case we're interested in the probabilities. So the formula we want to optimize is this:

$$-\sum_i P(i)\log P(i) - \alpha \sum P(i) - \beta \sum E_i P(i)$$

Since everything here is summing over the same $i$, we can move the sum to the outside:

$$-\sum_i \left(P(i)\log P(i) - \alpha P(i) - \beta E_i P(i)\right)$$

To ensure that the energies don't get too large, we introduce constraint on the maximum number of state[1]. What is always true is that you can always make the boxes of larger than the number of states, as long as the number of states is bounded. In practice the number of states is always bounded.

The first sum is a sum of individual tuple of parameters. You maximize it by maximizing it for each $i$ separately, you then differentiate with respect to $P(j)$, the expression above be becomes (the "=0" comes from the recipie for Lagrange multipliers):

$$-\log P(j) - 1 - \alpha - \beta E_j = 0$$

We can remove the minus signs, and

$$\log P(j) + 1 + \alpha + \beta E_j = 0$$

and what do you know? We now have a probability distribution

$$logP(j) = -(1+\alpha) - \beta E_j$$

To get $P$ itself we exponentiate:

$$P_j = e^{-(1+\alpha)} e^{-\beta E_j}$$

The first term is ju st a number, it depends on a lagrange multiplier, but it doesn't depend on $j$. The dependence on $j$ is that the probability that we're in the $j$th state is proportional to $e^{-\beta E_j}$. The probability for different energy levels

[1] I'm glossing over a couple of minutes of handwaving here ;-)

falls off exponentially with different energy levels. That's the basic property of the *Bolzmann distribution*: When maximizing the entropy, subject to the onstraint of a constant total energy tells us uniquely that the probability is proportional to the negative exponential of some constant times the energy of the state in question.

Now, what's $\beta$? Well, it's just:

$$\beta = \frac{1}{k_B T}$$

where $T$ is temperature and $K_B$ is *Bolzmann's constant.* Susskind prefers to work in units where $K_B = 1$, and temperature has units of energy, in which case then $\beta = 1/T$, but *why* that is so we have yet to establish.

We have an independent definition of temperature from a previous lecture, but first we'll introduce some new notation. HIstorically the term $e^{-(1+\alpha)}$ has been denoted $1/Z$. Why? It's a historical notation, so the Bolzmann distribution then takes the canonical (and very famous) form of:

$$\frac{1}{Z} e^{-\beta E_j}$$

Now what is $Z$? Well, it's just a Lagrange multiplier, so now that we have solved the problem, we must go back and fix the multipliers. To do that we have go back to the constraints and say what they satisfy the constraints. Let's look at the constraint $\sum P(i) = 1$. When using the formula above, it translates to:

$$\frac{1}{Z} \sum e^{-\beta E_i} = 1$$

This tells us what $Z$ as a formula of $\beta$

$$\sum e^{-\beta E_i} = Z(\beta)$$

Now $Z$ is called the *partition function.* This doesn't seem like very much, it's just a normalization factor. However, we'll see that if you know the partition function, you know *e*verything.

How do we choose all $\beta$. The other constraint is that the probabilities of all boxes times the energy in that box has to add up to the total energy:

$$\sum_i P_i E_I = \bar{E}$$

If we assume that we know the average energy. Let's write down the formula:

$$\frac{1}{Z} \sum_i e^{\beta E_i} E_i = \bar{E}$$

This does epress $\beta$ , but e can do a slick thing by this by expressing it in terms of the partition function: The trick is replace $E_i$ with the derivative with respect to $\beta$. This is a standard trick (according to Susskind ;-): If I have $e^{-\beta E}$ and I want to multiply it by $E$, one way of doing it is just to differentiate it with respect to $\beta$, $\frac{d}{d\beta}e^{-\beta E}$ and then we have to change the sign and we get this equality:

$$-\frac{d}{d\beta}e^{-\beta E} = Ee^{-\beta E}$$

We then use this trick to rewrite the expression for average energy:

$$-\frac{1}{Z}\sum_i \frac{\partial}{\partial\beta}e^{\beta E_i} = \bar{E}$$

This can be simplified:

$$-\frac{\partial}{\partial\beta}\frac{1}{Z}\sum_i e^{\beta E_i} = \bar{E}$$

but $\sum_i e^{\beta E_i}$ is jsut the partition function $Z$, so that means that we have a formula:

$$\bar{E} = -\frac{1}{Z(\beta)}\frac{\partial Z(\beta)}{\partial\beta} = -\frac{\partial \log Z}{\partial\beta}$$

This gives a relationship between energy and temperature. $\beta$ is both the lagrange multiplier and it is also the temperature. This formula, together with the formula $Z(\beta) = \sum_i e^{-\beta E_i}$ contains a great deal of information.

Thelogarithm of $Z$ occurs so often that it is given its own name.

$$A = -T\log Z = -\frac{\log Z}{\beta}$$

It is called the *Helmholtz free energy.* It's just a definition.

What is not a definition is the relationship between the Helmholtz free energy, the energy, and the entropy.

$$\begin{array}{rcl} A & = & -\frac{\log Z}{\beta} \\ E & = & -\frac{\partial \log Z}{\partial\beta} \end{array}$$

These are the two important quantities, let's calculate the *entropy* based on these. The entropy is:

$$S = -\sum P_i \log P_i$$

but now we know what $P_i = \frac{1}{Z}e^{-\beta E_i}$ and also that $\log P_i = -\beta E_i - \log Z$ so let's start cranking on this thing and see what we get:

$$S = -\sum \frac{1}{Z}e^{-\beta E_i}\left(-\beta E_i - \log Z\right)$$

The minus sign goes away

$$S = \sum \frac{1}{Z}e^{-\beta E_i}\left(\beta E_i + \log Z\right)$$

We can then express this in terms of the average energy:

$$S = \sum \frac{1}{Z}e^{-\beta E_i}\left(\beta E_i + \log Z\right)$$

So, what's this? If we look at the things in the parenthesis and multi ply them with the exponential, we see that this is just $\beta$ times the average energy $sume^{-\beta E_i} E_i = \beta E$. The second term has no dependency on $i$, so since $Z = \sum e^{-\beta E_i}$, all that we are left with in the second term is $\log Z$. Thus we have:

**Rmz:** *Did I get that right? at 1:08:30 in lecture 3*

$$S = \beta E + \log Z = \beta(E - A)$$

So the difference between the energy and the *Helmholtz free energy* is the entropy. Assuming that $\beta = 1/T$ we can write this as:

$$T \cdot S = E - A$$

We can now express the *entropy* in terms of the *partition function.*

$$S/\beta = \frac{\log Z}{\beta} - \frac{\partial \log Z}{\partial \beta}$$

or

$$S = \log Z - \beta\frac{\partial \log Z}{\partial \beta}$$

There are multiple formulas we can use, but the main upshot is that when we know $Z$ we can calculate the energy, entropy, Helmholtz free energy. There is some power in knowing the partition function. It's always the $\log Z$ that comes into play, the logarithm is more interesting than the $Z$ itself.

Let's take a look at the theory of statistical fluctuations. Let's take a value $X(i)$ that has some value for each state $i$of the system. We'd like to know the uncertainty or fluctuation of $X$. The definition is given by centering (subtracting the average), and then square it and then take the average of the whole thing:

$$\left\langle \left(X(i) - \bar{X}\right)^2 \right\rangle$$

This gives us the width of the distribution, it's called the *root mean square* or the *variance* of $X$.

This can be simplified a bit:

$$\begin{aligned}\left\langle \left(X(i)-\bar{X}\right)^2 \right\rangle &= \langle X^2 - 2X\bar{X} - \bar{X}^2 \rangle \\ &= \langle X^2 \rangle - 2\bar{X}^2\bar{X}^2 \\ &= \langle X^2 \rangle - \bar{X}^2 \qquad = \langle X^2 \rangle - \langle X \rangle^2\end{aligned}$$

(remember not to confuse the *average of the square*, and the *square of the average* in the expression above :-) So the variance is the difference between the average of the square and the square of the average. The smaller it is the narrower the variance.

Let's now see how we can calulate the variance in the energy. If the variance is sharply peaked, it means that the variance of the energy is much smaller than the energy itself.

How far can we do to calculate the variance of the energy $\langle E \rangle = -\frac{\partial \log Z}{\partial \beta}$? Let's see if we can calculate the average of the square of the energy?

Let's first calculate the average of the square of the energy:

$$\sum P(i)E_i^2 = \frac{1}{Z}\sum e^{-\beta E_i}E_i^2$$

To get to the squared of energy, hit the second derivative:

$$\langle E^2 \rangle = \frac{1}{Z}\frac{\partial^2 Z}{\partial \beta^2}$$

The variance is:

$$\frac{\partial^2 Z}{\partial \beta^2} - \left(\frac{\partial \log Z}{\partial \beta}\right)^2$$

This can also be written as

$$\frac{\partial^2 Z}{\partial \beta^2} - \left(\frac{1}{Z}\frac{\partial Z}{\partial \beta}\right)^2$$

this can be simplified further:

$$\frac{\partial^2 \log Z}{\partial \beta} = \frac{\partial}{\partial \beta}\frac{1}{Z}\frac{\partial Z}{\partial \beta}$$

Let's examine this a bit further:

$$\frac{\partial \log Z}{\partial \beta} = -E$$

The partition function is an *generating function* that generates a lot of things, that in turn describe the full thermodynaics of a system. It's also the *Laplace transform* of the density of states, if that is something that helps the intuition. It doesn't have a simple interpretation. $Z$ is dimensionless. Things in exponents are always dimensionless. The partition function is not in itself directly measurable. It's not it's physical interpretation that is its main function, it is the things you generate with it that has physical interpretations that you do know.

**Rmz:** *No :-)*

Anyway, we can now calculate the fluctuation of the energy.

$$(\Delta E)^2 = \frac{\partial^2 \log Z}{\partial \beta^2} = \partial_\beta E$$

Now, $\beta = 1/T$ so this can be related to the derivative of the energy with respect to the temperature, so

$$\partial_\beta = \frac{dE}{dT}\frac{dT}{d\beta} = T^2 \frac{dE}{dT} = T^2 C$$

since $T = 1/\beta$ and $\frac{dT}{d\beta} = -\frac{1}{\beta^2} = -T^2$

the quantity $\frac{dE}{dT}$ has a name, it's the *heat capacity* $C$. The heat capacity grows roughly proportional with the size of the system (the mass, the number of particles, etc.). The energy itself is also proportional to the number of particles. The square of the energy is proportional to the square of the number of particles. We can take the square root on both sides to get:

$$\Delta E = T\sqrt{C}$$

The fluctuations of the system becomes smaller as the system gets larger. So the size of the fluctuations is small in comparison to the size of the system itself.

The standard undergraduate textbook definition of temperature etc., goes something like this:

$$tk_B = T$$

$$s/k_B = S$$

When doing theoretical work, $k_b$ is never used, but it's necessary to get sizes that can be matched to measurement results.

$$\begin{aligned} dE &= TdS \\ &= tds \end{aligned}$$

Now let's look at the formula:

$$(\Delta E)^2 = T^2 \frac{dE}{dT} = T^2 C$$

It does make dimensional sense, but where does $k_B$ go into it? The trick is to substitute $t$ for $T$ and multiply by $k_b$ (call it "*undergraduate physics notation*":

$$(\Delta E)^2 = t^2 \frac{dE}{dt} k_B = t^2 C k_B$$

This is in fact the first place $k_B$ appears in anything ;-) Now $k_B$ is a very small number, in the order of $1.4 \cdot 10^{-23}$, so we would also expect the fluctuations to be very small. So for a mole of gas, the fluctuation of temperature will be very small. The probability distributions becomes very peaked whtn the numbers becomEs large.

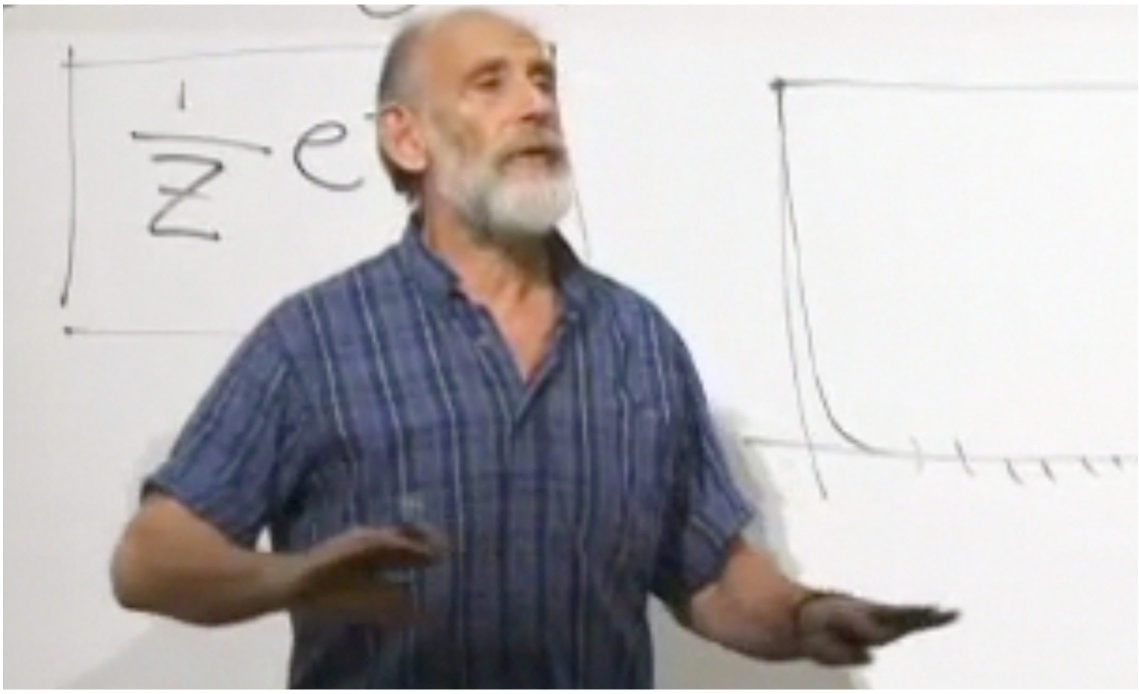


**Fig. 2.8** A flat occupancy distribution at infinite energy

A note from a question: At infinite temperature, all states becomes equally probable, so the distribution curve becomes flat (look at the curves to the left in figure ??. The entropy will be absolutely maximal. If you plot for energies, it will be tilted upwards since there are just a lot more states at high energy than low energies.

Lecture 4 is next ;)

# Index