

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: We established that, essentially, what we want to do is to describe the properties of a system that is in equilibrium. And a system in equilibrium is characterized by a certain number of parameters. We discussed displacement and forces that are used for mechanical properties. We described how when systems are in thermal equilibrium, the exchange of heat requires that there is temperature that will be the same between them. So that was where the Zeroth Law came and told us that there is another function of state.

Then, we saw that, from the First Law, there was energy, which is another important function of state. And from the Second Law, we arrived at entropy. And then by manipulating these, we generated a whole set of other functions, free energy, enthalpy, Gibbs free energy, the grand potential, and the list goes on.

And when the system is in equilibrium, it has a well-defined values of these quantities. You go from one equilibrium to another equilibrium, and these quantities change. But of course, we saw that the number of degrees of freedom that you have to describe the system is indicated through looking at the changes in energy, which if you were only doing mechanical work, you would write as sum over all possible ways of introducing mechanical work into the system.

Then, we saw that it was actually useful to separate out the chemical work. So we could also write this as a sum of an alpha chemical potential number of particles. But there was also ways of changing the energy of the system through addition of heat. And so ultimately, we saw that if there were n ways of doing chemical and mechanical work, and one way of introducing heat into the system, essentially $n + 1$ variables are sufficient to determine where you are in this phase space.

Once you have $n + 1$ of that list, you can input, in principle, determine others as long as you have not chosen things that are really dependent on each other. So you have to choose independent ones, and we had some discussion of how that comes into play.

So I said that today we will briefly conclude with the last, or the Third Law. This is the statement about trying to calculate the behavior of entropy as a function of temperature. And in principle, you can imagine as a function of some coordinate of your system-- capital X could indicate pressure, volume, anything.

You calculate that at some particular value of temperature, T , the difference in entropy that you would have between two points parametrized by X_1 and X_2 . And in principle, what you need to do is to find some kind of a path for changing parameters from X_1 to X_2 and calculate, in a reversible process, how much heat you have to put into the system.

Let's say at this fixed temperature, T , divide by T . T is not changing along the process from say X_1 to X_2 . And this would be a difference between the entropy that you would have between these two quantities, between these two points. You could, in principle, then repeat this process at some lower temperature and keep going all the way down to 0 temperature.

What Nernst observed was that as he went through this procedure to lower and lower temperatures, this difference-- Let's call it Δs of T going from X_1 to X_2 goes to 0. So it looks like, certainly at this temperature, there is a change in entropy going from one to another.

There's also a change. This change gets smaller and smaller as if, when you get to 0 temperature, the value of your entropy is independent of X . Whatever X you choose, you'll have the same value of entropy. Now, that led to, after a while, to a more ambitious version statement of the Third Law that I will write down, which is that the entropy of all substances at the zero of thermodynamic temperature is the same and can be set to 0. Same universal constant, set to 0.

It's, in principle, through these integration from one point to another point, the only thing that you can calculate is the difference between entropies. And essentially, this suggests that the difference between entropies goes to 0, but let's be more ambitious and say that even if you look at different substances and you go to 0 temperature, all of them have a unique value.

And so there's more evidence for being able to do this for different substances via what is called allotropic state. So for example, some materials can exist potentially in two different crystalline states that are called allotropes, for example, sulfur as a function of temperature. If you lower it's temperature very slowly, it stays in some foreign all the way down to 0 temperature.

So if you change its temperature rapidly, it stays in one form all the way to 0 temperature in crystalline structure that is called monoclinic. If you cool it very, very slowly, there is a temperature around 40 degrees Celsius at which it makes a transition to a different crystal structure. That is rhombohedral. And the thing that I am plotting here, as a function of temperature, is the heat capacity.

And so if you are, let's say, around room temperature, in principle you can say there's two different forms of sulfur. One of them is truly stable, and the other is metastable. That is, in principle, if you rate what sufficiently is of the order of [? centuries ?], you can get the transition from this form to the stable form.

But for our purposes, at room temperature, you would say that at the scale of times that I'm observing things, there are these 2 possible states that are both equilibrium states of the same substance. Now using these two equilibrium states, I can start to test this Nernst theorem generalized to different substances. If you, again, regard these two different things as different substances.

You could say that if I want to calculate the entropy just slightly above the transition, I can come from two paths. I can either come from path number one. Along path number one, I would say that the entropy at this T_c plus is obtained by integrating degree heat capacity, so $\int dT C_x$ of T divided by T .

This combination is none other than dQ . Basically, the combination of heat capacity dT is the amount of heat that you have to put the substance to change its temperature. And you do this all the way from 0 to T_c plus. Let's say we go along this path that corresponds to this monoclinic way.

And I'm using this C_m that corresponds to this as opposed to 0 that corresponds to this. Another thing that I can do-- and I made a mistake because what I really need to do is to, in principle, add to this some entropy that I would assign to this green state at 0 because this is the difference.

So this is the entropy that I would assign to the monoclinic state at T close to 0. Going along the orange path, I would say that S evaluated at T_c plus is obtained by integrating from 0. Let's say to T_c minus dT , the heat capacity of this rhombic phase.

But when I get to just below the transition, I want to go to just above the transition. I have to actually be put in certain amount of latent heat. So here I have to add latent heat L , always at the temperatures T_c , to gradually make the substance transition from one to the other. So I have to add here L of T_c .

This would be the integration of dQ , but then I would have to add the entropy that I would assign to the orange state at 0 temperature. So this is something that you can do experimentally. You can evaluate at these integrals, and what you'll find is that these two things are the same. So this is yet another justification of this entropy being independent of where you start at 0 temperature. Again at this point, if you like, you can by [INAUDIBLE] state that this is 0 for everything will start with 0.

So this is a supposed new law of thermodynamics. Is it useful? What can we deduce from that? So let's look at the consequences. First thing is so what I have established is that the limit as T goes to 0 of S , irrespective of whatever set of parameters I have-- so I pick T as one of my n plus one coordinates, and I put some other bunch of coordinates here. I take the limit of this going to 0. This becomes 0.

So that means, almost by construction, that if I take the derivative of S with respect

to any of these coordinates-- if I take then the limit as T goes to 0, this would be fixed T . This is 0. Fine. So basically, this is another way of stating that entropy differences go through 0.

But it does have a consequence because one thing that you will frequently measure are quantities, such as extensivity. What do I mean by that? Let's pick a displacement. Could be the length of a wire. Could be the volume of a gas. And we can ask if I were to change temperature, how does that quantity change?

So these are quantities typically called α . Actually, usually you would also divide by x to make them intensive because otherwise x being extensive, the whole quantity would have been extensive. Let's say we do this at fixed corresponding displacement. So something that is very relevant is you take the volume of gas who changes temperature at fixed pressure, and the volume of the gas will shrink or expand according to this extensive.

Now, this can be related to this through Maxwell relationship. So let's see what I have to do. I have that dE is something like Jdx plus, according to what I have over there, TdS . I want to be able to write a Maxwell relation that relates a derivative of x . So I want to make x into a first derivative. So I look at E minus Jx . And this Jdx becomes minus xdJ .

But I want to take a derivative of x with respect not s , but with respect to T . So I'll do that. This becomes a minus SdT . So now, I immediately see that I will have a Maxwell relation that says dx by dT at constant J is the same thing as dS by dJ at constant T . So this is the same thing by the Maxwell relation as dS by dJ at constant T . All right?

This is one of these quantities, therefore, as T goes 0, this goes to 0. And therefore, the expansivity should go to 0. So any quantity that measures expansion, contraction, or some other change as a function of temperature, according to this law, as you go through 0 temperature, should go to 0.

There's one other quantity that also goes to 0, and that's the heat capacity. So if I

want to calculate the difference between entropy at some temperature T and some temperature at 0 along some particular path corresponding to some constant x for example, you would say that what I need to do is to integrate from 0 to T the heat that I have to put into the system at constant x . And so if I do that slowly enough, this heat I can write as $C_x dT$. C_x , potentially, is a function of T .

Actually, since I'm indicating T as the other point of integration, let me call the variable of integration T' . So I take a path in which I change temperature. I calculate the heat capacity at constant x . Integrate it. Multiply by dT to convert it to T , and get the result. So all of these results that they have been formulating suggest that the result that you would get as a function of T , for entropy, is something that as T goes to 0 , approaches 0 . So it should be a perfectly nice, well-defined value at any finite temperature.

Now, if you integrate a constant divided by T , divided by dT , then essentially the constant would give you a logarithm. And the logarithm would blow up as we go to 0 temperature. So the only way that this integral does not blow up on you-- so this is finite only if the limit as T goes to 0 of the heat capacities should also go to 0 . So any heat capacity should also essentially vanish as you go to lower and lower temperature. This is something that you will see many, many times when you look at different heat capacities in the rest of the course.

There is one other aspect of this that I will not really explain, but you can go and look at the notes or elsewhere, which is that another consequence is unattainability of T equals to 0 by any finite set of operations. Essentially, if you want to get to 0 temperature, you'll have to do something that cools you step by step. And the steps become smaller and smaller, and you have to repeat that many times. But that is another consequence. We'll leave that for the time being.

I would like to, however, end by discussing some distinctions that are between these different laws. So if you think about whatever could be the microscopic origin, after all, I have emphasized that thermodynamics is a set of rules that you look at substances as black boxes and you try to deduce a certain number of things based

on observations, such as what Nernst did over here.

But you say, these black boxes, I know what is inside them in principle. It's composed of atoms, molecules, light, quark, whatever the microscope theory is that you want to assign to the components of that box. And I know the dynamics that governs these microscopic degrees of freedom. I should be able to get the laws of thermodynamics starting from the microscopic laws.

Eventually, we will do that, and as we do that, we will find the origin of these different laws. Now, you won't be surprised that the First Law is intimately connected to the fact that any microscopic set of rules that you write down embodies the conservation of energy. And all you have to make sure is to understand precisely what heat is as a form of energy. And then if we regard heat as another form of energy, another component, it's really the conservation law that we have.

Then, you have the Zeroth Law and the Second Law. The Zeroth Law and Second Law have to do with equilibrium and being able to go in some particular direction. And that always runs a fall of the microscopic laws of motion that are typically things that are time reversible whereas the Zeroth Law and Second Law are not. And what we will see later on, through statistical mechanics, is that the origin of these laws is that we are dealing with large numbers of degrees of freedom.

And once we adapt the proper perspective to looking at properties of large numbers of degrees of freedom, which will be a start to do the elements of that [? prescription ?] today, the Zeroth Law and Second Law emerge.

Now the Third Law, you all know that once we go through this process, eventually for example, we get things for the description of entropy, which is related to some number of states that the system has indicated by g . And if you then want to have S going through 0, you would require that g goes to something that is order of 1-- of 1 if you like-- as T goes to 0.

And typically, you would say that systems adopt their ground state, lowest energy

state, at 0 temperature. And so this is somewhat a statement about the uniqueness of the state of all possible systems at low temperature. Now, if you think about the gas in this room, and let's imagine that the particles of this gas either don't interact, which is maybe a little bit unrealistic, but maybe repel each other.

So let's say you have a bunch of particles that just repel each other. Then, there is really no reason why, as I go to lower and lower temperatures, the number of configurations of the molecules should decrease. All configurations that I draw that they don't overlap have roughly the same energy.

And indeed, if I look at say any one of these properties, like the expansivity of a gas at constant pressure which is given in fact with a minus sign. dV by dT at constant pressure would be the analog of one of these extensivities. If I use the Ideal Gas Law-- So for ideal gas, we've seen that PV is proportional to let's say some temperature.

Then, dV by dT at constant pressure is none other than V over T . So this would give me 1 over V , V over T . Probably don't need it on this. This is going to give me 1 over T . So not only doesn't it go to 0 at 0 temperature, if the Ideal Gas Law was satisfied, the extensivity would actually diverge at 0 temperature as different as you want. So clearly the Ideal Gas Law, if it was applicable all the way down to 0 temperature, would violate the Third Law of thermodynamics. Again, not surprising given that I have told you that a gas of classical particles with repulsion has many states.

Now, we will see later on in the course that once we include quantum mechanics, then as you go to 0 temperature, these particles will have a unique state. If they are bosons, they will be together in one wave function. If they are fermions, they will arrange themselves appropriately so that, because of quantum mechanics, all of these laws would certainly breakdown at T equals to 0.

You will get 0 entropy, and you would get consistency with all of these things. So somehow, the nature of the Third Law is different from the other laws because its validity rests on being able to be living in a world where quantum mechanics

applies. So in principle, you could have imagined some other universe where \hbar equals to 0, and then the Third Law of thermodynamics would not hold there whereas the Zeroth Law and Second Law would hold. Yes?

AUDIENCE: Are there any known exceptions to the Third Law? Are we going to [? account for them? ?]

PROFESSOR: For equilibrium-- So this is actually an interesting question. What do I know about-- classically, I can certainly come up with lots of examples that violate. So your question then amounts if I say that quantum mechanics is necessary, do I know that the ground state of a quantum mechanical system is unique.

And I don't know of a proof of that for interacting system. I don't know of a case that's violated, but as far as I know, there is no proof that I give you an interacting Hamiltonian for a quantum system, and there's a unique ground state. And I should say that there'd be no-- and I'm sure you know of cases where the ground state is not unique like a ferromagnet.

But the point is not that g should be exactly one, but that the limit of $\log g$ divided by the number of degrees of freedom that you have should go to 0 as n goes to infinity. So something like a ferromagnet may have many ground states, but the number of ground states is not proportional to the number of sites, the number of spins, and this entity will go to 0.

So all the cases that we know, the ground state is either unique or is order of one. But I don't know a theorem that says that should be the case.

So this is the last thing that I wanted to say about thermodynamics. Are there any questions in general? So I laid out the necessity of having some kind of a description of microscopic degrees of freedom that ultimately will allow us to prove the laws of thermodynamics. And that will come through statistical mechanics, which as the name implies, has to have certain amount of statistic characters to it.

What does that mean? It means that you have to abandon a description of motion that is fully deterministic for one that is based on probability. Now, I could have told

you first the degrees of freedom and what is the description that we need for them to be probabilistic, but I find it more useful to first lay out what the language of probability is that we will be using and then bring in the description of the microscopic degrees of freedom within this language.

So if we go first with definitions-- and you could, for example, go to the branch of mathematics that deals with probability, and you will encounter something like this that what probability describes is a random variable. Let's call it X , which has a number of possible outcomes, which we put together into a set of outcomes, S .

And this set can be discrete as would be the case if you were tossing a coin, and the outcomes would be either a head or a tail, or we were throwing a dice, and the outcomes would be the faces 1 through 6. And we will encounter mostly actually cases where S is continuous.

Like for example, if I want to describe the velocity of a gas particle in this room, I need to specify the three components of velocity that can be anywhere, let's say, in the range of real numbers. And again, mathematicians would say that to each event, which is a subset of possible outcomes, is assigned a value which we must satisfy the following properties.

First thing is the probability of anything is a positive number. And so this is positivity. The second thing is additivity. That is the probability of two events, A or B , is the sum total of the probabilities if A and B are disjoint or distinct. And finally, there's a normalization. That if you're event is that something should happen the entire set, the probability that you assign to that is 1.

So these are formal statements. And if you are a mathematician, you start from there, and you prove theorems. But from our perspective, the first question to ask is how to determine this quantity probability that something should happen. If it is useful and I want to do something real world about it, I should be able to measure it or assign values to it.

And very roughly again, in theory, we can assign probabilities two different ways.

One way is called objective. And from the perspective of us as physicists corresponds to what would be an experimental procedure. And if it is assigning p of e as the frequency of outcomes in large number of trials, i.e. you would say that the probability that event A is obtained is the number of times you would get outcome A divided by the total number of trials as n goes to infinity.

So for example, if you want to assign a probability that when you throw a dice that face 1 comes up, what you could do is you could make a table of the number of times 1 shows up divided by the number of times you throw the dice. Maybe you throw it 100 times, and you get 15. You throw it 200 times, and you get-- that is probably too much. Let's say 15-- you get 35. And you do it 300 times, and you get something close to 48. The ratio of these things, as the number gets larger and larger, hopefully will converge to something that you would call the probability.

Now, it turns out that in statistical physics, we will assign things through a totally different procedure which is subjective. If you like, it's more theoretical, which is based on uncertainty among all outcomes. Because if I were to subjectively assign to throwing the dice and coming up with value of 1, I would say, well, there's six possible faces for the dice. I don't know anything about this dice being loaded, so I will say they are all equally alike.

Now, that may or may not be a correct assumption. You could test it. You could maybe throw it many times. You will find that either the dice is loaded or not loaded and this is correct or not. But you begin by making this assumption. And this is actually, we will see later on, exactly the type of assumption that you would be making in statistical physics.

Any question about these definitions? So let's again proceed slowly to get some definitions established by looking at one random variable. So this is the next section on one random variable. And I will assume that I'll look at the case of the continuous random variable. So x can be any real number minus infinity to infinity.

Now, a number of definitions. I will use the term Cumulative-- make sure I'll use the - Cumulative Probability Function, CPF, that for this one random variable, I will

indicate by capital P of x. And the meaning of this is that capital P of x is the probability of outcome less than x.

So generically, we say that x can take all values along the real line. And there is this function that I want to plot that I will call big P of x. Now big P of x is a probability, therefore, it has to be positive according to the first item that we have over there. And it will be less than 1 because the net probability for everything toward here is equal to 1.

So asymptotically, where I go all the way to infinity, the probability that I will get some number along the line-- I have to get something, so it should automatically go to 1 here. And every element of probability is positive, so it's a function that should gradually go down. And presumably, it will behave something like this generically.

Once we have the Cumulative Probability Function, we can immediately construct the Probability Density Function, PDF, which is the derivative of the above. P of x is the derivative of big P of x with respect to the x. And so if I just take here the curve that I have above and take its derivative, the derivative will look something like this. Essentially, clearly by the definition of the derivative, this quantity is therefore ability of outcome in the interval x to x plus dx divided by the size of the interval dx.

couple of things to remind you of, one of them is that the Cumulative Probability is a probability. It's a dimensionless number between 0 and 1. Probability Density is obtained by taking a derivative, so it has dimensions that are inverse of whatever this x is. So if I change my variable from meters to centimeters, let's say, the value of this function would change by a factor of 100. And secondly, while the Probability Density is positive, its value is not bounded. It can be anywhere that you like.

One other, again, minor definition is expectation value. So I can pick some function of x. This could be x itself. It could be x squared. It could be sine x, x cubed minus x squared. The expectation value of this is defined by integrating the Probability Density against the value of the function.

So essentially, what that says is that if I pick some function of x-- function can be

positive, negative, et cetera. So maybe I have a function such as this-- then the value of x is random. If x is in this interval, this would be the corresponding contribution to f of x . And I have to look at all possible values of x .

Question? Now, very associated to this is a change of variables. You would say that if x is random, then f of x is random. So if I ask you what is the value of x squared, and for one random variable, I get this. The value of x squared would be this. If I get this, the value of x squared would be something else.

So if x is random, f of x is itself a random variable. So f of x is a random variable, and you can ask what is the probability, let's say, the Probability Density Function that I would associate with the value of this. Let's say what's the probability that I will find it in the interval between small f and small f plus df . This will be Pf of f .

You would say that the probability that I would find the value of the function that is in this interval corresponds to finding a value of x that is in this interval. So what I can do, the probability that I find the value of f in this interval, according to what I have here, is the Probability Density multiplied by df . Is there a question? No. So the probability that I'm in this interval translates to the probability that I'm in this interval. So that's probability p of x dx .

But that's boring. I want to look at the situation maybe where the function is something like this. Then, you say that f is in this interval provided that x is either here or here or here. So what I really need to do is to solve f of x equals to f for x . And maybe there will be solutions that will be x_1, x_2, x_3 , et cetera.

And what I need to do is to sum over the contributions of all of those solutions. So here, it's three solutions. Then, you would say the Probability Density is the sum over i P of x_i , the x_i by df , which is really the slopes. The slopes translate the size of this interval to the size of that interval.

You can see that here, the slope is very sharp. The size of this interval is small. It could be wider accordingly, so I need to multiply by dx_i by df . So I have to multiply by dx by df evaluated at x_i . That's essentially the value of the derivative of f .

Now, sometimes, it is easy to forget these things that I write over here. And you would say, well obviously, the probability of something that is positive. But without being careful, it is easy to violate such basic condition. And I violated it here.

Anybody see where I violated it. Yeah, the slope here is positive. The slope here is positive. The slope here is negative.

So I am subtracting a probability here. So what I really should do-- it really doesn't matter whether the slope is this way or that way. I will pick up the same interval, so make sure you don't forget the absolute values that go accordingly. So this is the standard way that you would make change of variables. Yes?

AUDIENCE: Sorry. In the center of that board, on the second line, it says Pf. Is that an x or a times?

PROFESSOR: In the center of this board? This one?

AUDIENCE: Yeah.

PROFESSOR: So the value of the function is a random variable, right? It can come up to be here. It can come up to be here. And so there is, as any other one parameter random variable, a Probability Density associated with that. That Probability Density I have called P of f to indicate that it is the variable f that I'm considering as opposed to what I wrote originally that was associated with the value of x.

AUDIENCE: But what you have written on the left-hand side, it looks like your x [? is random. ?]

PROFESSOR: Oh, this was supposed to be a multiplication sign, so sorry.

AUDIENCE: Thank you.

PROFESSOR: Thank you. Yes?

AUDIENCE: CP-- that function, is this [INAUDIBLE]?

PROFESSOR: Yes. So you're asking whether this-- so I constructed something, and my statement is that the integral from minus infinity to infinity df Pf of f better be one which is the

normalization. So if you're asking about this, essentially, you would say the integral $\int dx p(x)$ is the integral $\int dx dP$ by dx , right? That was the definition $p(x)$.

And the integral of the derivative is the value of the function evaluated at its two extremes. And this is one minus 0. So by construction, it is, of course, normalized in this fashion. Is that what you were asking?

AUDIENCE: I was asking about the first possibility of cumulative probability function.

PROFESSOR: So the cumulative probability, its constraint is that the limit as its variable goes to infinity, it should go to 1. That's the normalization. The normalization here is that the probability of the entire set is 1. Cumulative adds the probabilities to be anywhere up to point x . So I have achieved being anywhere on the line by going through this point. But certainly, the integral of $P(x)dx$ is not equal to 1 if that's what you're asking. The integral of small $p(x)$ is 1. Yes?

AUDIENCE: Are we assuming the function is invertible?

PROFESSOR: Well, rigorously speaking, this function is not invertible because for a value of f , there are three possible values of x . So it's not a function, but you can certainly solve for $f(x) = f$ to find particular values.

So again, maybe it is useful to work through one example of this. So let's say that you have a probability that is of the form $e^{-\lambda|x|}$. So as a function of x , the Probability Density falls off exponentially on both sides.

And again, I have to ensure that when I integrate this from 0 to infinity, I will get one. The integral from 0 to infinity is $1/\lambda$, from minus infinity to zero by symmetry is $1/\lambda$. So it's really I have to divide by 2λ -- to $1/(2\lambda)$. Sorry.

Now, suppose I change variables to F , which is x^2 . So I want to know what the probability is for a particular value of x^2 that I will call f . So then what I have to do is to solve this. And this will give me $x = \pm\sqrt{f}$. If I ask for what f of-- for what value of x , $x^2 = f$, then I have these

two solutions.

So according to the formula that I have, I have to, first of all, evaluate this at these two possible routes. In both cases, I will get minus lambda square root of f. Because of the absolute value, both of them will give you the same thing. And then I have to look at this derivative.

So if I look at this, I can see that df by dx equals to $2x$. The locations that I have to evaluate are at plus minus square root of f. So the value of the slope is minus plus to square root of f. And according to that formula, what I have to do is to put the inverse of that. So I have to put for one solution, 1 over 2 square root of f.

For the other one, I have to put 1 over minus 2 square root of f, which would be a disaster if I didn't convert this to an absolute value. And if I did convert that to an absolute value, what I would get is lambda over 2 square root of f e to the minus lambda root f.

It is important to note that this solution will only exist only if f is positive. And there's no solution if f is negative, which means that if I wanted to plot a Probability Density for this function f, which is x squared as a function of f, it will only have values for positive values of x squared.

There's nothing for negative values. For positive values, I have this function that's exponentially decays. Yet at f equals to 0 diverges. One reason I chose that example is to emphasize that these Probability Density functions can even go all the way infinity. The requirement, however, is that you should be able to integrate across the infinity because integrating across the infinity should give you a finite number less than 1.

And so the type of divergence that you could have is limited. 1 over square root of f is fine. $1/f$ is not accepted. Yes?

AUDIENCE:

I have a doubt about [? the postulate. ?] It says that if you raise the value of f slowly, you will eventually get to-- yeah, that point right there. So if the prescription that we have of summing over the different roots, at some point, the roots, they converge.

PROFESSOR: Yes.

AUDIENCE: So at some point, we stop summing over 2 and we start summing over 1. It just seems a little bit strange.

PROFESSOR: Yeah. If you are up here, you have only one term in the sum. If you are down here, you have three terms. And that's really just the property of the curve that I have drawn. And so over here, I have only one root. Over here, I have three roots. And this is not surprising. There are many situations in mathematics or physics where you encounter situations where, as you change some parameters, new solutions, new roots, appear.

And so if this was really some kind of a physical system, you would probably encounter some kind of a singularity of phase transitions at this point. Yes?

AUDIENCE: But how does the equation deal with that when [INAUDIBLE]?

PROFESSOR: Let's see. So if I am approaching that point, what I find is that the f by the x goes to 0. So the x by df has some kind of infinity or singularity, so we have to deal with that. If you want, we can choose a particular form of that function and see what happens. But actually, we have that already over here because the function f that I plotted for you as a function of x has this behavior that, for some range of f , you have two solutions.

So for negative values of f , I have no solution. So this curve, after having rotated, is precisely an example of what is happening here. And you see what the consequence of that is. The consequence of that is that as I approach here and the two solutions merge, I have the singularity that is ultimately manifested in here.

So in principle, yes. When you make these changes of variables and you have functions that have multiple solution behavior like that, you have to worry about this. Let me go down here.

One other definition that, again, you've probably seen, before we go through

something that I hope you haven't seen, moment. A form of this expectation value-- actually, here we did with x squared, but in general, we can calculate the expectation value of x to the m . And sometimes, that is called m th moment is the integral $\int_0^\infty dx x^m p(x)$.

Now, I expect that after this point, you would have seen everything. But next one maybe half of you have seen. And the next item, which we will use a lot, is the characteristic function. So given that I have some probability distribution p of x , I can calculate various expectation values.

I calculate the expectation value of e^{-ikx} . This is, by definition that you have, I have to integrate over the domain of x -- let's say from minus infinity to infinity-- $p(x)$ against e^{-ikx} . And you say, well, what's special about that? I know that to be the Fourier transform of $p(x)$. And it is true.

And you also know how to invert the Fourier transform. That is if you know the characteristic function, which is another name for the Fourier transform of a probability distribution, you would get the $p(x)$ back by the integral over k divided by 2π , the way that I chose the things, into the $e^{ikx} \tilde{p}(k)$. Basically, this is the standard relationship between these objects. So this is just a Fourier transform.

Now, something that appears a lot in statistical calculations and implicit in lots of things that we do in statistical mechanics is a generating function. I can take the characteristic function $\tilde{p}(k)$. It's a function of this Fourier variable, k . And I can do an expansion in that. I can do the expansion inside the expectation value because e^{-ikx} I can write as a sum over n running from 0 to infinity $\frac{(-ikx)^n}{n!} x^n$.

This is the expansion of the exponential. The variable here is x , so I can take everything else outside. And what I see is that if I make an expansion of the characteristic function, the coefficient of k^n up to some trivial factor of $n!$ will give me the n th moment.

That is once you have calculated the Fourier transform, or the characteristic

function, you can expand it. And you can, from out of that expansion, you can extract all the moments essentially. So this expansion generates for you the moments, hence the generating function. You could even do something like this.

You could multiply e to the ikx_0 for some x_0 p tilde of k . And that would be the expectation value of e to the ikx minus x_0 . And you can expand that, and you would generate all of the moments not around the origin, but around the point x_0 .

So simple manipulations of the characteristic function can shift and give you other set of moments around different points. So the Fourier transform, or characteristic function, is the generator of moments. An even more important property is possessed by the cumulant generating function.

So you have the characteristic function, the Fourier transform. You take its log, so another function of k . You start expanding this function in powers of k . Add the coefficients of that, you call cumulants. So I essentially repeated the definition that I had up there.

I took a log, and all I did is I put this subscript c to go from moments to cumulants. And also, I have to start the series from 1 as opposed to 0. And essentially, I can find the relationship between cumulants and moments by writing this as a log of the characteristic function, which is 1 plus some n plus 1 to infinity of $\frac{(-ik)^n}{n!}$, the n th moments.

So inside the log, I have the moments. Outside the log, I have the cumulants. And if I have a log of $1 + \epsilon$, I can use the expansion of this as $\epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3} - \frac{\epsilon^4}{4} + \dots$, et cetera. And this will enable me to then match powers of $-ik$ on the left and powers of $-ik$ on the right.

You can see that the first thing that I will find is that the expectation value of x -- the first power, the first term that I have here is $-ik$ to the mean. Take the log, I will get that. So essentially, what I get is that the first cumulant on the left is the first moment that I will get from the expansion on the right. And this is, of course, called

the mean of the distribution.

The second cumulant, I will have two contributions, one from ϵ , the other from ϵ^2 . And if you go through that, you will get that it is expectation value of x^2 minus the average of x , the mean squared, which is none other than the expectation value of x around the mean squared, which is clearly a positive quantity. And this is the variance.

And you can keep going. The third cumulant is x^3 minus 3 average of x^2 average of x plus 2 average of x itself cubed. It is called the skewness. I don't write the formula for the next one which is called a [? cortosis ?]. And you keep going and so forth.

So it turns out that this hierarchy of cumulants, essentially, is a hierarchy of the most important things that you can know about a random variable. So if I tell you that the outcome of some experiment is some number x , distribute it somehow-- I guess the first thing that you would like to know is whether the typical values that you get are of the order of 1, are of the order of million, whatever.

So somehow, the mean is something that tells you something that is most important is zeroth order thing that you want to know about the variable. But the next thing that you might want to know is, well, what's the spread? How far does this thing go? And then the variance will tell you something about the spread.

So the next thing that you want to do is maybe if given the spread, am I more likely to get things that are on one side or things that are on the other side. So the measure of its asymmetry, right versus left, is provided by the third cumulant, which is the skewness and so forth. So typically, the very first few members of this hierarchy of cumulants tells you the most important information that you need about the probability.

Now, I will mention to you, and I guess we probably will deal with it more next time around, the result that is in some sense the backbone or granddaddy of all graphical expansions that are carrying [INAUDIBLE]. And that's a relationship

between the moments and cumulants that I will express graphically.

So this is graphical representation of moments in terms of cumulants. Essentially, what I'm saying is that you can go through the procedure as I outlined. And if you want to calculate minus μ_k to the fifth power so that you find the description of the fifth cumulant in terms of the moment, you'll have to do a lot of work in expanding the log and powers of this object and making sure that you don't make any mistakes in the coefficient.

There is a way to circumvent that graphically and get the relationship. So how do we do that? You'll represent n th cumulant as let's say a bag of endpoints. So let's say this entity will represent the third cumulant. It's a bag with three points. This-- one, two, three, four, five, six-- will represent the sixth cumulant.

Then, the n th moment is some of all ways of distributing end points amongst bags. So what do I mean? So I want to calculate the first moment x . That would correspond to one point. And really, there's only one diagram I can put the bag around it or not that would correspond to this. And that corresponds to the first cumulant, basically rewriting what I had before.

If I want to look at the second moment, the second moment I need two points. The two points I can either put in the same bag or I can put into two separate bags. And the first one corresponds to calculating the second cumulant. The second term corresponds to two ways in which their first cumulant has appeared, so I have to squared x .

if I want to calculate the third moment, I need three dots. The three dots I can either put in one bag or I can take one of them out and keep two of them in a bag. And here I had the choice of three things that I could've pulled out. Or, I could have all of them in individual bags of their own.

And mathematically, the first term corresponds to $x^3 c$. The third term corresponds to three versions of the variance times the mean. And the last term is just the mean cubed. And you can massage this expression to see that I get the

expression that I have for the skewness. I didn't offhand remember the relationship that I have to write down for the fourth cumulant.

But I can graphically, immediately get the relationship for the fourth moment in terms of the fourth cumulant which is this entity. Or, four ways that I can take one of the back and maintain three in the bag, three ways in which I have two bags of two, six ways in which I can have a bag of two and two things that are individually apart, and one way in which there are four things that are independent of each other.

And this becomes x to the fourth cumulant, the fourth cumulant, 4 times the third cumulant times the mean, 3 times the square of the variance, 6 times the variance multiplied by the mean squared, and the mean raised to the fourth power. And you can keep going.

AUDIENCE: Is the variance not squared in the third term?

PROFESSOR: Did I forget that? Yes, thank you. All right. So the proof of this is really just the two-line algebra exponentiating these expressions that we have over here. But it's much nicer to represent that graphically. And so now you can go between things very easily.

And what I will show next time is how, using this machinery, you can calculate any moment of a Gaussian, for example, in just a matter of seconds as opposed to having to do integrations and things like that. So that's what we will do next time will be to apply this machinery to various probability distribution, such as a Gaussian, that we are likely to encounter again and again.