PROFESSOR: We'll begin this time by looking at some probability distribution that you should be familiar with from this perspective, starting with a Gaussian distribution for one variable.

We're focused on a variable that takes real values in the interval minus infinity to infinity and the Gaussian has the form exponential that is centered around some value, let's call it lambda, and has fluctuations around this value parameterized by sigma. And the integral of this p over the interval should be normalized to unity, giving you this hopefully very family of form.

Now, if you want to characterize the characteristic function, all we need to do is to Fourier transform this. So I have the integral dx e the minus ikx. So this-- let's remind you alternatively was the expectation value of e to the minus ikx. minus x minus lambda squared over 2 sigma squared, which is the probably distribution.

And you should know what the answer to that is, but I will remind. You can change variables to x minus lambda [INAUDIBLE] y. So from here we will get the factor of 2 to the minus ik lambda. You have then the integral over y minus y squared over 2 sigma squared.

And then what we need to do is to complete this square over here. And you can do that, essentially, by adding and subtracting a minus k squared sigma squared over 2. So that if I change variable to y plus ik sigma squared, let's call that z, then I have outside the integral e to the minus ik lambda minus k squared sigma squared over 2. And the remainder I can write as a full square.

And this is just a normalized Gaussian integral that comes to 1. So as you well know, a Fourier transform of a Gaussian is itself a Gaussian, and that's what we've

established. E to the minus ik lambda minus k squared sigma squared over 2m.

And if I haven't made a mistake when I said k equals to 0, the answer should be 1 because k equals to 0 expectation value of 1 just amounts to normalization.

Now what we said was that a more interesting function is obtained by taking the log of this. So from here we go to the log of [INAUDIBLE] of k. Log of [INAUDIBLE] of k is very simple for the Gaussian.

And what we had said was that by definition this log of the characteristic function generates cumulants through the series minus ik to the power of n over n factorial, the end cumulant.

So looking at this, we can immediately see that the Guassian is characterized by a first cumulant, which is the coefficient of minus ik. It's lambda. It is characterized by a second cumulant, which is the coefficient of minus ik squared.

This, you know, is the variance. And we can explicitly see that the coefficient of minus ik squared over 2 factorial is simply sigma squared. So this is reputation.

But one thing that is interesting is that our series has now terminates, which means that if I were to look at the third cumulant, if I were to look at the fourth cumulant, and so forth, for the Guassian, they're all 0. So the Gaussian is the distribution that is completely characterized just by its first and second cumulants, all the rest being 0.

So, now our last time we developed some kind of a graphical method. We said that I can graphically describe the first cumulant as a bag with one point in it. Second cumulant with something that has two points in it. A third cumulant with three three, fourth cumulant with four and four points and so forth.

This is just rewriting. Now, the interesting thing was that we said that the various moments we could express graphically. So that, for example, the second moment is either this or this, which then graphically is the same thing as lambda squared plus sigma squared because this is indicated by sigma squared.

Now, x cubed you would say is either three things by themselves or put two of them together and then one separate. And this I could do in three different ways. And in general, for a general distribution, I would have had another term, which is a triangle. But the triangle is 0.

So for the Gaussian, this terminates here. I have lambda cubed plus 3 lambda sigma squared. If I want to calculate x to the fourth, maybe the old way of doing it would have been too multiply the Gaussian distribution against x to the fourth and try to do the integration. And you would ultimately be able to do that rearranging things and looking at the various powers of the Gaussian integrated from minus infinity to infinity.

But you can do it graphically. You can say, OK. It's either this or I can have-- well, I cannot put one aside and three together, because that doesn't exist. I could have two together and two not together. And this I can do in six different ways. You can convince yourself of that. What I could do two pairs, which I can do three different ways because I can either do one, two; one, three; one, four and then the other is satisfied.

So this is lambda to the fourth plus 6 lambda squared sigma squared plus 3 sigma to the fourth. And you can keep going and doing different things. OK. Question? Yeah?

AUDIENCE: Is the second-- [INAUDIBLE].

PROFESSOR: There?

AUDIENCE: Because-- so you said that the second cumulative--

PROFESSOR: Oh.

AUDIENCE: --x sqaured. Yes.

PROFESSOR: Yes. So that's the wrong--

AUDIENCE: [INAUDIBLE].

PROFESSOR: The coefficient of k squared is the second cumulant. The additional 2 was a mistake. OK. Anything else? All right.

Let's take a look at couple of other distributions, this time discrete. So the binomial distribution is repeat a binary random variable.

And what does this mean? It means two outcomes that's binary, let's call them A and B. And if I have a coin that's head or tails, it's binary. Two possibilities. And I can assign probabilities to the two outcomes to PA and PB, which has to be 1 minus PA.

And the question is if you repeat this binary random variables n times, what is the probability of NA outcomes of A?

And I forget to say something important, so I write it in red. This should be independent. That is the outcome of a coin toss at, say, the fifth time should not influence the sixth time and future times.

OK. So this is easy. The probability to have NA occurrences of A in N trials. So it has to be index yn. Is that within the n times that I tossed, A came up NA times. So it has to be proportional to the probability of A independently multiplied by itself NA times.

But if I have exactly NA occurrences of A, all the other times I had B occurring. So I have the probability of B for the remainder, which is N minus NA.

Now this is the probability for a specific occurrence, like the first NA times that I threw the coin I will get A. The remaining times I would get B. But the order is not important and the number of ways that I can shuffle the order and have a total of NA out of N times is the binomial factor.

Fine. Again, well known. Let's look at its characteristic function. So p tilde, which is now a function of k, is expectation value of e to the minus ik NA, which means that I have to weigh e minus ik NA against the probability of occurrences of NA times, which is this binomial factor PA to the power of NA, PB to the power of N minus NA.

And of course, I have to sum over all possible values of NA that go all the way from 0 to N.

So what we have here is something which is this combination, PA e to the minus ik raised to the power of NA. PB raised to the complement N minus NA multiplied by the binomial factor summed over all possible values. So this is just the definition of the binomial expansion of PA e to the minus ik plus PB raised to the power of N.

And again, let's check. If I set k equals to 0, I have PA plus PB, which is 1 raised to the power of N. So things are OK. So this is the characteristic function.

At this stage, the only thing that I will note about this is that if I look at the characteristic function I will get N times. So this is-- actually, let's make sure that we maintain the index N. So this is the characteristic function appropriate to N trials. And what I get is that up to factor of N, I will get the characteristic function that would be appropriate to one trial.

So what that means if I were to look at powers of k, the expectation value of some cumulant, if I go to repeat things N times-- so this carries an index N. It is going to be simply N times what I would have had in a single trial.

So for a single trial, you really have two outcomes-- 0 or 1 occurrences of this object. So for a binary variable, you can really easily compute these quantities and then you can calculate the corresponding ones for N trials simply multiplying by N. And we will see that this is characteristic, essentially, of anything that is repeated N times, not just the binomial.

So this form that you have N independent objects, you would get N times what you would have for one object is generally valid and actually something that we will build a lot of statistical mechanics on because we are interested in the [INAUDIBLE]. So we will see that shortly.

But rather than following this, let's look at a third distribution that is closely related, which is the Poisson. And the question that we are asking is-- we have an interval. And the question is, what is the probability of m events in an interval from 0 to T?

And I kind of expressed it this way because prototypical Poisson distribution is, let's say, the radioactivity. And you can be waiting for some time interval from 0 to 1 minute and asking within that time interval, what's the probability that you will see m radioactive decay events?

So what is the probability if two things happen? One is that the probability of 1 and only 1 event in interval dt is alpha dt as dt goes to 0.

OK? So basically if you look at this over 1 minute, the chances are that you will see so many events, so many radioactivities. If you shorten the interval, the chances that you would see events would become less and less.

If you make your event infinitesimal, most of the time nothing would happen with very small probability that vanishes. As the size of the interval goes to 0, you will see 1 event. So this is one condition.

And the second condition is events in different intervals are independent. And since I wrote independent in red up there, let me write it in red here because it sort of harks back to the same condition.

And so this is the question. What is this probability? And the to get the answer, what we do is to subdivide our big interval into N, which is big T divided by the small dt subintervals.

So basically, originally let's say on the time axis, we were covering a distance that went from 0 to big T and we were asking what happens here. So what we are doing now is we are sort of dividing this interval to lots of subintervals, the size of each one of them being dt. And therefore, the total number is big T over dt. And ultimately, clearly, I want to sit dt going to 0 so that this condition is satisfied.

So also because of the second condition, each one of these will independently tell me whether or not I have an event. And so if I want to count the total number of events, I have to add things that are occurring in different intervals. And we can see that this problem now became identical to that problem because each one of these

intervals has two possible outcomes-- nothing happens with probability 1 minus alpha dt, something happens with probability alpha dt.

So no event is probability 1 minus alpha dt. One event means probability alpha dt. So this is a binomial process.

So we can calculate, for example, the characteristic function. And I will indicate that we are looking at some interval of size T and parameterized by this straight alpha, we'll see that only the product will occur. So this is this before Fourier variable.

We said that it's a binary, so it is one of the probabilities plus e to the minus ik plus the other probability raised to the power of N. Now we just substitute the probabilities that we have over here.

So the probability of not having an event is 1 minus alpha dt. The probability of having an event is alpha dt. So alpha dt is going to appear here as e to the minus ik minus 1. So alpha st, e to the minus ik, is from here. From PB I will get 1 minus alpha dt.

And I bunched together the two terms that are proportional to alpha dt. And then I have to raise to the power of N, which is T divided by dt. And this whole prescription is valid in the limit where dt is going to 0.

So what you have is 1 plus an infinitesimal raised to a huge power. And this limiting procedure is equivalent to taking the exponential. So basically this is the same thing as exponential of what is here multiplied by what is here. The dt's cancel each other out and the answer is alpha T e to the minus ik minus 1.

So the characteristic function for this process that we described is simply given by this form. You say, wait. I didn't ask for the characteristic function. I wanted the probability.

Well, I say, OK. Characteristic function is simply the Fourier transform. So let me Fourier transform back, and I would say that the probability along not the Fourier axis but the actual axis is open by the inverse Fourier process.

So I have to do an integral dk over 2 pi e to the ikx times the characteristic function. And the characteristic function is e to minus-- what was it? e to the alpha T. E to the minus ik k minus 1.

Well, there is an equal to minus alpha T that I can simply take outside the integration. I have the integration over k e to that ikx. And then what I will do is I have this factor of e to the something in the-- e to the alpha T to the minus ik. I will use the expansion of the exponential.

So the expansion of the exponential is a sum over m running from 0 to infinity. The exponent raised to the m-th power. So I have alpha T raised to the m-th power, e to the minus ik raised to the m-th power divided m factor here.

So now I reorder the sum and the integration. The sum is over m, the integration is over k. I can reorder them. So on the things that go outside I have a sum over m running from 0 to infinity e to the minus alpha T, alpha T to the power of m divided by m factorial. Then I have the integral over k over 2 pi e to the ik.

Well, I had the x here and I have e to the minus ikm here. So I have x minus m.

And then I say, OK, this is an integral that I recognize. The integral of e to the ik times something is simply a delta function. So this whole thing is a delta function that's says, oh, x has to be an integer. Because I kind of did something that maybe, in retrospect, you would have said why are you doing this. Because along how many times things have occurred, they have either occurred 0 times, 1 times, 2 decays, 3 decays. I don't have 2.5 decays.

So I treated x as a continuous variable, but the mathematics was really clever enough to say that, no. The only places that you can have are really integer values. And the probability that you have some particular value integer m is simply what we have over here, e to the minus alpha T alpha T to the power of m divided by m factorial, which is the Poisson distribution. OK.

But fine. So this is the Poisson distribution, but really we go through the root of the

characteristic function in order to use this machinery that we developed earlier for cumulants et cetera.

So let's look at the cumulant generating function. So I have to take the log of the function that I had calculated there. It is nicely in the exponential, so I get alpha T e to the minus ik minus 1.

So now I can make an expansion of this in powers of k so I can expand the exponential. The first term vanishes because this starts with one. So really I have alpha T sum running from 1 to infinity or N running from 1 to infinity minus ik to the power of n over n factorial.

So my task for identifying the cumulants is to look at the expansion of this log and read off powers of minus ikn to the power of n factorial. So what do we see? We see that the first cumulant of the Poisson is alpha T, but all the coefficients are the same thing.

The expectation value-- sorry. The second cumulant is alpha T. The third cumulant, the fourth cumulant, all the other cumulants are also alpha T.

So the average number of decays that you see in the interval is simply alpha T. But there are fluctuations, and if somebody should, for example, ask you what's the average number cubed of events, you would say, OK. I'm going to use the relationship between moments and cumulants. I can either have three first objects or I can put one of them separate in three different factions.

But this is a case where the triangle is allowed, so diagrammatically all three are possible. And so the answer for the first term is alpha T cubed. For the second term, it is a factor of 3. Both this variance and the mean give me a factor of alpha T, so I will get alpha T squared. And the third term, which is the third cumulant, is also alpha T. So the answer is simply of this form.

Again, m is an integer. Alpha T is dimensionless. So there is no dimension problem by it having different powers. OK. Any questions?

All right. So that's what I wanted to say about one variable. Now let's go and look at corresponding definitions when you have multiple variables.

So for many random variables, the set of possible outcomes, let's say, has variables x1, x2. Let's be precise. Let's end it a xn.

And if these are distributed, each one of them continuously over the interval, to each point we can characterize some kind of a probability density. So this entity is called the joint probability density function.

And its definition would be to look at probability of outcome in some interval that is between, say, x1, x1 plus dx1 in one, x2, x2 plus dx2 in the second variable. xn plus dxn in the last variable. So you sort of look at the particular point in this multi-dimensional space that you are interested.

You build a little cube around it. You ask, what's the probability to being that cube? And then you divide by the volume of that cube.

So dx1, dx2, dxn, which is the same thing that you would be doing in constructing any density and by ultimately taking the limit that all of the x's go to 0. All right. So this is the joint probability distribution.

You can construct, now, the joint characteristic function. Now how do you do that? Well, again, just like you would do for your transform with multiple variables. So you would go for each variable to a conjugate variable.

So x1 would go to k1. x2 would go to k2. xn would go to kn. And this would mathematically amount to calculating the expectation value of e to the minus i k1x1 x1, k2x2 and so forth, which you would obtain by integrating over all of these variables, e to the minus ik alpha x alpha, against the probability of x1 through xn.

Question?

AUDIENCE: It's a bit hard to read. It's getting really small.

PROFESSOR: OK.

[LAUGHTER]

PROFESSOR: But it's just multi-dimensional integral. OK? All right.

So this is, as the case of one, I think the problem is not the size but the angle I see. I can't do much for that. You have to move to the center. OK.

So what we can look at now is joint moment. So you can-- when we had one variable, we could look at something like the expectation value of x to the m. That would be the m-th moment.

But if you have two variable, we can raise x1 to some other, x2 to another power, and actually xn to another power. So this is a joint moment.

Now the thing is, that the same way that moments for one variable could be generated by expanding the characteristic function, if I were to expand this function in powers of k, you can see that in meeting the expectation value, I will get various powers of x1 to some power, x2 to some power, et cetera. So by appropriate expansion of that function, I can generate all of-- read off all of these moments.

Now, a more common way of generating the Taylor series expansion is through derivatives. So what I can do is I can take a derivative with respect to, say, ik1. If I take a derivative with respect to ik1 here, what happens is I will bring down a factor of minus x alpha. So actually let me put the minus so it becomes a factor of x alpha.

And if I integrate x alpha against this, I will be generating the expectation value of x alpha provided that ultimately I set all of the k's to 0. So I will calculate the derivative of this function with respect to all of these arguments. At the end of the day, I will set k equals to 0. That will give me the expectation value of x1.

But I don't want x1, I want x1 raised to the power of m1. So I do this. Each time I take a derivative with respect to minus ik, I will bring down the factor of the corresponding x. And I can do this with multiple different things. So d by the ik2 raised to the power of m2 minus d by the ikn, the whole thing raised to the power of

mn.

So I can either take this function of multiple variables-- k1 through kn-- and expand it and read off the appropriate powers of k1k2. Or I can say that the terms in this expansion are generated through taking appropriate derivative. Yes?

AUDIENCE: Is there any reason why you're choosing to take a derivative with respect to ikj instead of simply putting the i in the numerator? Or are there-- are there things that I'm not--

PROFESSOR: No. No. There is no reason. So you're saying why didn't I write this as this like this? i divided?

AUDIENCE: Yeah.

PROFESSOR: I think I just visually saw that it was kind of more that way. But it's exactly the same thing. Yes. OK. All right.

Now the interesting object, of course, to us is more the joint cumulants. So how do we generate joint cumulants? Well previously, essentially we had a bunch of objects for one variable that was some moment. And in order to make them cumulants, we just put a sub C here. So we do that and we are done.

But what did operationally happen was that we did the expansion rather than for the characteristic function for the log of the characteristic function. So all I need to do is to do precisely this set of derivatives applied rather than to the joint characteristic function to the log of the joint characteristic function. And at the end, set all of the case to 0. OK?

So by looking at these two definitions and the expansion of the log, for example, you can calculate various things. Like, for example, x1x2 with a C is the expectation value of x1x2. This joint moment minus x1x2, just as you would have thought, would be the appropriate generalization of the variance. And this is the covariance. And you can construct appropriate extensions. OK.

Now we made a lot of use of the relationship between moments and cumulants. We

just-- so the idea, really, was that the essence of a probability distribution is characterized in the cumulants. Moments kind of depend on how you look at things.

The essence is in the cumulants, but sometimes the moments are more usefully computed, and there was a relationship between moments and cumulants, we can generalize that graphical relation to the case joint moments and joint cumulants. So graphical relation applies as long as points are labeled by appropriate or by corresponding variable.

So suppose I wanted to calculate some kind of a moment that is x1 squared. Let's say x2, x3. This may generate for me many diagrams, so let's stop from here.

So what I can do is I can have points that I label 1, 1, and 2. And have them separate from each other. Or I can start pairing them together. So one possibility is that I put the 1's together and the 2 starts separately.

Another possibility is that I can group the 1 and the 2 together. And then the other 1 starts separately. But I had a choice of two ways to do this, so this comes-- this diagram with an overall factor of 2.

And then there's the possibility to put all of them in the same bag. And so mathematically, that means that the third-- this particular joint moment is obtained by taking average of x1 squared x2 average, which is the first term. The second term is the variance of x1. And then multiplied by x2.

The third term is twice the covariance of x1 and x2 times the mean of x1. And the final term is just the third cumulant. So again, you would need to compute these, presumably, from the law of the characteristic function and then you would be done.

Couple of other definitions. One of them is an unconditional probability. So very soon we will be talking about, say, probabilities appropriate to the gas in this room. And the particles in the gas in this room will be characterized where they are, some position vector q, and how fast they are moving, some momentum vector p. And there would be some kind of a probability density associated with finding a particle

with some momentum at some location in space.

But sometimes I say, well, I really don't care about where the particles are, I just want to know how fast they are moving. So what I really care is the probability that I have a particle moving with some momentum p, irrespective of where it is.

Then all I need to do is to integrate over the position the joint probability distribution. And the check that this is correct is that if I first do not integrate this over p, this would be integrated over the entire space and the joint probabilities appropriately normalized so that the joint integration will give me one. So this is a correct normalized probability.

And more generally, if I'm interested in, say, a bunch of coordinates x1 through xs, out of a larger list of coordinates that spans x1 through xs all the way to something else, all I need to do to get unconditional probability is to integrate over the variables that I'm not interested. Again, check is that it's a good properly normalized.

Now, this is to be contrasted with the conditional probability. The conditional probability, let's say we would be interested in calculating the pressure that is exerted on the board. The pressure is exerted by the particles that impinge on the board and then go away, so I'm interested in the momentum of particles right at the board, not anywhere else in space.

So if I'm interested in the momentum of particles at the particular location, which could in principle depend on location-- so now q is a parameter p is the variable, but the probability distribution could depend on q.

How do we obtain this? This, again. is going to be proportional to the probability that I will find a particle both at this location with momentum p. So I need to have that. But it's not exactly that there's a normalization involved. And the way to get normalization is to note that if I integrate this probability over its variable p but not over the parameter q, the answer should be 1.

So this is going to be, if I apply it to the right-hand side, the integral over p of p of p and q, which we recognize as an example of an unconditional probability to find

something at position 1. So the normalization is going to be this so that the ratio is 1.

So most generally, we find that the probability to have some subset of variables, given that the location of the other variables in the list are somewhat fixed, is given by the joint probability of all of the variables x1 through xn divided by the unconditional probability that that applies to the parameters of our fixed. And this is called Bayes' theorem.

By the way, if variables are independent, which actually does apply to the case of the particles in this room as far as their momentum and position is concerned, then the joint probability is going to be the product of one that is appropriate to the position and one that is appropriate to the momentum.

And if you have this independence, then what you'll find is that there is no difference between conditional and unconditional probabilities. And when you go through this procedure, you will find that all the joint cumulants-- but not the joint moments, naturally-- all the joint cumulants will be 0.

OK. Any questions? Yes?

AUDIENCE: Could you explain how the condition of p--

PROFESSOR: How this was obtained? Or the one above?

AUDIENCE: Yeah. The condition you applied that the integral is 1.

PROFESSOR: OK. So first of all, what I want to look at is the probability that is appropriate to one random variable at the fixed value of all the other random variables. Like you say, in general I should specify the probability as a function of momentum and position throughout space. But I'm really interested only at this point. I don't really care about other points.

However, the answer may depend whether I'm looking at here or I'm looking at here. So the answer for the probability of momentum is parametrized by q. On the

other hand, I say that I know the probability over the entire space to be a disposition with the momentum p as given by this joint probability.

But if I just set that equal to this, the answer is not correct because the way that this quantity is normalized is if I first integrate over all possible values of its variable, p. The answer should be 1, irrespective of what q is.

So I can define a conditional probability for momentum here, a conditional probability for momentum there. In both cases the momentum would be the variable it. And integrating over all possible values of momentum should give me one for a properly normalized probability distribution.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Given that q is something. So q could be some-- now here q can be regarded as some parameter. So the condition is that this integration should give me 1.

I said that on physical grounds, I expect this conditional probability to be the joint probability up to some normalization that I don't know.

OK. So what is that normalization? The whole answer should be 1. What I have to do is an integration over momentum of the joint probability. I have said that an integration over some set of variables of a joint probability will give me the unconditional probability for all the others. So integrating over all momentum of this joint probability will give me the unconditional probability for position.

So the normalization of 1 is the unconditional probability for position divided by n. So n-- this has to be this. And in general, it would have to be this in order to ensure that if I integrate over this first set of variables of the joint probability distribution which would give me the unconditional, cancels the unconditional in the denominator to give me 1.

Other questions? OK.

So I'm going to erase this last board to be underneath that top board in looking at the joint Gaussian distribution. So that was the Gaussian, and we want to look at the

joint Gaussian.

So we want to generalize the formula that we have over there for one variable to multiple variables. So what I have there initially is a factor, which is exponential of minus 1/2, x minus lambda squared. I can write this x minus lambda squared as x minus lambda x minus lambda.

And then put the variance. Let's call it is 1 over sigma rather than a small sigma squared or something like this. Actually, let me just write it as 1 over sigma squared for the time being. And then the normalization was 1 over root 2 pi sigma squared.

But you say, well, I have multiple variables, so maybe this is what I would give for my B variable. And then I would sum over all N, running from 1 to N. So this is essentially the form that I would have for an independent Gaussian variables.

And then I would have to multiply here factors of 2 pi sigma squared, so I would have 2 pi to the N over 2. And I would have product of-- actually, let's write it as 2 pi to the N square root. I would have the product of sigma i squared.

But that's just too limiting a form. The most general form that these quadratic will allow me to have also cross terms where it is not only the diagonal terms x1 and x1 that's are multiplying each other, but x2 and x3, et cetera.

So I would have a sum over both m and n running from 1 to n. And then to coefficient here, rather than just being a number, would be the variables that would be like a matrix. Because for each pair m and n, I would have some number. And I will call them the inverse of some matrix C.

And if you, again, think of the problem as a matrix, if I have the diagonal matrix, then the product of elements along the diagonal is the same thing as the determinant. If I were to rotate the matrix to have off diagonal elements, the determinant will always be there. So this is really the determinant of C that will appear here. Yes?

AUDIENCE: So are you inverting the individual elements of C or are you inverting the matrix C

and taking its elements?

PROFESSOR: Actually a very good point. I really wanted to write it as the inverse of the matrix and then peak the mn [INAUDIBLE]. So we imagine that we have the matrix. And these are the elements of some-- so I could have called this whatever I want.

So I could have called the coefficients of x and n. I have chosen to regard them as the inverse of some other matrix C. And the reason for that becomes shortly clear, because the covariances will be related to the inverse of this matrix. And hence, that's the appropriate way to look at it.

AUDIENCE: Can [INAUDIBLE] what C means up there?

PROFESSOR: OK. So let's forget about this lambdas. So I would have in general for two variables some coefficient for x1 squared, some coefficient for x2 squared, and some coefficient for x1, x2.

So I could call this a11. I could call this a22. I could call this 2a12. Or actually I could, if I wanted, just write it as a12 plus a21 x2x1 or do a1 to an a21 would be the same. So what I could then regard this is as x1 2. The matrix a11, a12, a21, a22, x1, x2. So this is exactly the same as that. All right?

So these objects here are the elements of this matrix C inverse. So I could call this x1, x2 some matrix A x1 x2. That A is 2 by 2 matrix. The name I have given to that 2 by 2 matrix in C inverse. Yes?

AUDIENCE: The matrix is required to be symmetric though, isn't it?

PROFESSOR: The matrix is required to be symmetric for any quadrant form. Yes. So when I wrote it initially, I wrote as 2 a12. And then I said, well, I can also write it this fashion provided the two of them are the same. Yes?

AUDIENCE: How did you know the determinant of C belonged there?

PROFESSOR: Pardon?

AUDIENCE: How did you know that the determinant of C [INAUDIBLE]?

PROFESSOR: OK. How do I know the determinant of C? Let's say I give you this form. And then I don't know what the normalization is.

What I can do is I can do a change of variables from x1 x2 to something like y1 y2 such that when I look at y1 and y2, the matrix becomes diagonal. So I can rotate the matrix. So any matrix I can imagine that I will find some U such that A U U dagger is this diagonal matrix lambda.

Now under these procedures, one thing that does not change is the determinant. It's always the product of the eigenvalues. The way that I set up the problem, I said that if I hadn't made the problem to have cross terms, I knew the answers to be the product of that eigenvalues.

So if you like, I can start from there and then do a rotation and have the more general form. The answer would stay as the determinant. Yes?

AUDIENCE: The matrix should be positive as well or no?

PROFESSOR: The matrix should be positive definite in order for the probability to be well-defined and exist, yes. OK. So if you like, by stating that this is a probability, I have imposed a number of conditions such as symmetry, as well as positivity. Yes. OK.

But this is just linear algebra. I will assume that you know linear algebra. OK.

So this property normalized Gaussian joint probability. We are interested in the characteristic function. So what we are interested is the joint Gaussian characteristic. And so again we saw the procedure was that I have to do the Fourier transform.

So I have to take this probability that I have over there and do an integration product, say, alpha running from 1 to N dx alpha of e to the minus ik alpha x alpha. This product exists for all values. Then I have to multiply with this probability that I have up there, which would appear here. OK.

Now, again maybe an easy way to imagine is what I was saying to previously. Let's imagine that I have rotated into a basis where everything is diagonal. Then in the rotated basis, all you need to do is to essentially do product of characteristic functions such as what we have over here.

So the corresponding product to this first term would be exponential of minus i sum over N running from 1 to N k alpha lambda alpha. kn lambda n. I guess I'm using n as the variable here. And as long as things would be diagonal, the next ordered term would be a sum over alpha kn squared the corresponding eigenvalue inverted. So remember that in the diagonal form, each one of these sigmas would appear as the diagonal.

If I do my rotation, essentially this term would not be affected. The next term would give me minus 1/2 sum over m and n rather than just having k1 squared k2 squared, et cetera. Just like here, I would have km kn.

What happened previously was that each eigenvalue would get inverted. If you think about rotating a matrix, all of its eigenvalues are inverted, you are really rotating the inverse matrix. So this here would be the inverse of whatever matrix I have here. So this would be C mn. So I did will leave you to do the corresponding linear algebra here, but the answer is correct.

So the answer is that the generator of cumulants for a joint Gaussian distribution has a form which has a bunch of the linear terms-- kn lambda n. And a bunch of second order terms, so we will have minus 1/2 sum over m and n km kn times some coefficient.

And the series terminates here. So for the joint Gaussian, you have first cumulant. So the expectation value of nm cumulant is the same thing as lambda m. You have covariances or second cumulants xm, xn, C is Cmn. And in particular, the diagonal elements would correspond to the variances. And all the higher orders are 0 because there's no further term in the expansion.

So for example, if I were to calculate this thing that I have on the board here for the

case of a Gaussian, for the case of the Gaussian, I would not have this third term. So the answer that I would write down for the case of the third term would be something that didn't have this. And in the way that we have written things, the answer would have been x1 squared x2-- would be just a lambda 1 squared lambda 2 plus sigma 1 squared, or let's call it C11, times lambda 2 plus 2 lambda 1 C12. And that's it.

So there is something that follows from this that it is used a lot in field theory. And it's called Wick's theorem. So that's just a particular case of this, but let's state it anyway.

So for Gaussian distributed variables of 0 mean, following condition applies. I can take the first variable raised to power n1, the second variable to n2, the last variable to some other nN and look at a joint expectation value such as this. And this is 0 if sum over alpha and alpha is odd and is called sum over all pairwise contraction if a sum over alpha and alpha is even.

So actually, I have right here an example of this. If I have a Gaussian variable-- jointly distributed Gaussian variables where the means are all 0-- so if I say that lambda 1 and lambda 2 are 0, then this is an odd power, x1 squared x2. Because of the symmetry it has to be 0, but you explicitly see that every term that I have will be multiplying some power of [INAUDIBLE].

Whereas if for other than this, I was looking at something like x1 squared x2 x3 where the net power is even, then I could sort of imagine putting them into these kinds of diagrams. Or alternatively, I can imagine pairing these things in all possible ways. So one pairing would be this with this, this with this, which would have given me x1 squared C x2 x3 C.

Another pairing would have been x1 with x2. And then, naturally, x1 with x3. So I would have gotten x1 with x2 covariance x1 with extreme x3 covariance. But I could have connected the x1 to x2 or the second x1 to x2.

So this comes in 2 variance. And so the answer here would be C11 C23 plus 2 C1 2

C13. Yes?

AUDIENCE: In your writing of x1 to the n1 [INAUDIBLE]. It should be the cumulant, right? Or is it the moment?

PROFESSOR: This is the moment.

AUDIENCE: OK.

PROFESSOR: The contractions are the covariances.

AUDIENCE: OK.

PROFESSOR: So the point is that the Gaussian distribution is completely characterized in terms of its covariances. Once you know the covariances, essentially you know everything. And in particular, you may be interested in some particular combination of x's. And then you use to express that in terms of all possible pairwise contractions, which are the covariances.

And essentially, in all of field theory, you expand around some kind of a Gaussian background or Gaussian 0 toward the result. And then in your perturbation theory you need various powers of your field or some combination of powers, and you express them through these kinds of relationships.

Any questions? OK. This is fine. Let's get rid of this. OK.

Now there is one result that all of statistical mechanics hangs on. So I expect that as I get old and I get infirm or whatever and my memory vanishes, the last thing that I will remember before I die would be the central limit theorem.

And why is this important is because you end up in statistical physics adding lots of things. So really, the question that you have or you should be asking is thermodynamics is a very precise thing. It says that heat goes from the higher temperature to lower temperature. It doesn't say it does that 50% of the time or 95% of the time. It's a definite statement.

If I am telling you that ultimately I'm going to express everything in terms of probabilities, how does that jive? The reason that it jives is because of this theorem. It's because in order to go from the probabilistic description, you will be dealing with so many different-- so many large number of variables-- that probabilistic statements actually become precise deterministic statements.

And that's captured by this theorem, which says that let's look at the sum of N random variables. And I will indicate the sum by x and my random variables as small x's. And let's say that this is, for the individual set of things that I'm adding up together, some kind of a joint probability distribution out of which I take these random variables.

So each instance of this sum is selected from this joint PDF, so x itself is a random variable because of possible choices of different xi from this probability distribution. So what I'm interested is what is the probability for the sum? So what is the p that determines this sum?

I will go by the root of these characteristic functions. I will say, OK, what's the expectation value of-- well, let's-- what's the Fourier transform of this probability distribution? If we transform, by definition it is the expectation of e to the minus ik this big X, which is the sum over all off the small x's.

Do I have that definition somewhere? I erased it. Basically, what is this? If this k was, in fact, different k's-- if I had a k1 multiplying x1, k2 multiplying x2, that would be the definition of the joint characteristics function for this joint probability distribution.

So what this is is you take the joint characteristic function, which depends on k1 k2, all the way to kn. And you set all of them to be the same. So take the joint characteristic function depends on N Fourier variables. Put all of them the same k and you have that for the sum.

So I can certainly do that by adding a log here. Nothing has changed. I know that the log is the generator of the cumulants. So this is a sum over, let's say, n running

from 1 to infinity minus ik to the power of n over n factorial, the joint cumulant of the sum.

So what is the expansion that I would have for log of the joint characteristic function? Well, typically I would say have at the lowest order k1 times the mean of the first variable, k2 times the mean of the second variable. But all of them are the same. So the first order, I would get minus i the same k sum over n of the first cumulant of the N-th variable.

Typically, this second order term, I would have all kinds of products. I would have k1 k3 k2 k4, as well as k1 squared. But now all of them become the same, and so what I will have is a minus ik squared. But then I have all possible pairings mn of xm xn cumulants.

AUDIENCE: Question.

PROFESSOR: Yes?

AUDIENCE: [INAUDIBLE] expression you probably should use different indices when you're summing over elements of Taylor serious and when you're summing over your [INAUDIBLE] random variables. Just-- it gets confusing when both indexes are n.

PROFESSOR: This here, you want me to right here, say, i?

AUDIENCE: Yeah.

PROFESSOR: OK. And here I can write i and j. So I think there's still a 2 factorial. And then there's higher orders.

Essentially then, matching the coefficients of minus ik from the left minus ik from the right will enable me to calculate relationships between cumulants of the sum and cumulants of the individual variables. This first one of them is not particularly surprising. You would say that the mean of the sum is sum of the means of the individual variables.

The second statement is that the variance of the sum really involves a pair, i and j,

running from 1 to N. So if these variable were independent, you would be just adding the variances. Since they are potentially dependent, you have to also keep track of covariances.

And this kind of summation extends to higher and higher cumulants, essentially including more and more powers of cumulants that you would put on that side.

And what we do with that, I guess we'll start next time around.