Advanced Thermodynamics and Molecular Simulations

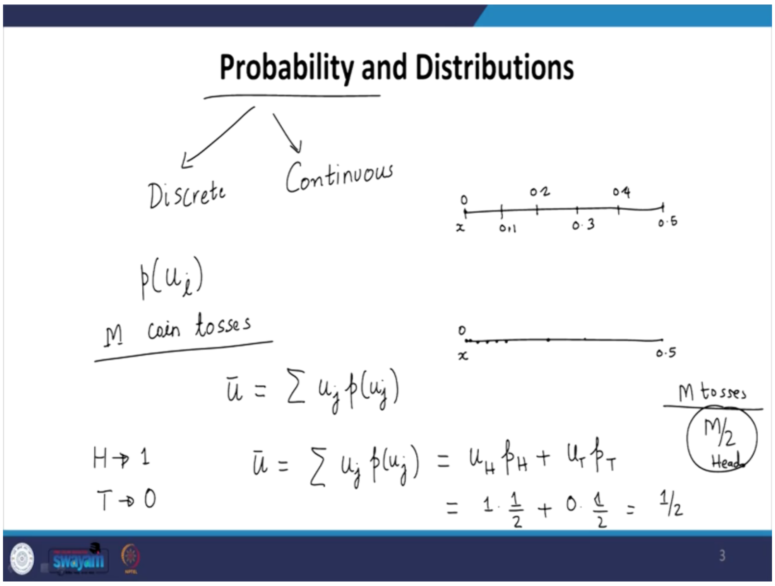
Prof. Prateek Kumar Jha
Department of Chemical Engineering
Indian Institute of Science – Roorkee

Lecture 10
Discrete and Continuous Probabilities; Stirling Approximation

Hello all of you so until so far in this course we have basically, pretty much recapped the thermodynamics you knew the laws of thermodynamics and the idea of thermodynamic functions and try to provide you a molecular description of all of this. So from now on we will get into some more rigorous mathematical derivations that comes in the regime of statistical mechanics and will going over the same concepts but in somewhat more detail and try to establish how can I get actually the thermodynamic variables using that kind of a statistical mechanics framework.

So today I will talk about again the idea of probability and distributions because it will be required in the framework we do from the next lecture and also it is slightly more mathematical in comparison to the intuition based argument we used in the description of entropy earlier. So essentially we already have discussed the idea of probability when we did a coin toss example or the playing card example and so on but let us now just write the same expressions in somewhat more mathematical terms just to set the ground for what we will do in the in the next lecture.

(Refer Slide Time: 01:52)



So essentially when we talk about probability this can be defined for 2 different regimes- one is a discrete system and one is a continuous system, in all the examples that we have done so far it was pretty much discrete we said the number of outcomes can be large but then the outcomes are still finite. When we say it is continuous that means there is no discrete points on which the outcome should be it can be anywhere in the domain.

So let us say for example if I say that my x can be either 0 or 0.1 or 0.2 or 0.3 this is what I mean by a discrete kind of a distribution. So if x can take only those particular values then we say it is a discrete distribution.

On the other hand if x can take any value between the range we are interested in let us say 0 to 0.5 in that case we say it is a continuous distribution it can be any number that we can imagine or anywhere we drop a point it is probable for x to be there. How much probability is there is a different thing altogether but x can be anywhere in the domain. So naturally if the number of points in the discrete distribution increases, we go towards a continuum or continuous distribution and that is what we have seen earlier.

Let us say for example if I am doing 10000 tosses then the fraction of heads can become like 0.0001, 0.000 2 and so on. If I am doing only 2 tosses then it can be only 0.5 and 1. So as the number of points on a discrete distribution increases we start to approach a continuous distribution. So therefore if for example for any thermodynamic system we know that the numbers of molecules are huge.

So let us say for example if I talk about exchange of quanta and things of that sort we need to keep in mind that the number of quanta of energy available is also pretty large. So any energy distribution we are looking at it has too many points for us to start considering the distribution as a continuous distribution and in mathematical terms this gives me one huge advantage when we are doing a discrete distribution, we are basically summing our probabilities when we are doing a continuous distribution we can replace the sums with an integral. So therefore even when what quantum mechanics says that energy has to be quantized and only discrete energy values are possible if we are in the limit where the number of discrete points becomes so large that we can pretty much cover all possible values of energy in that case we can use the framework of continuous distribution and that will really make our math much, much easier. This is the example that we are trying to build on.

So, just to go back to our earlier example of the coin toss what we can say is that we can have certain number of outcomes and the probability of certain outcome is the number of ways in which that outcome is possible divided by total possible number of ways. Now if I for example ask you what is the most probable number of outcomes most probable number of heads when I am doing 5 tosses then I am not talking about 1 particular experiment I am talking about a series of experiments being conducted and for each experiment you can have a head or a tail. So if I go back to our example of M coin tosses and number of heads that we get in those number of tosses it can be 1, 2 until M that is the all possibility that we have discussed.
So now if I ask you what is the most probable number of heads then in that particular case what we are talking about is something that is called expectation in the language of probability and the way an expectation is defined is something like this-

$$\bar{u} = \sum u_j p(u_j)$$

the average value or the expected value is sum over all the possible outcomes with the probability of that particular outcome.

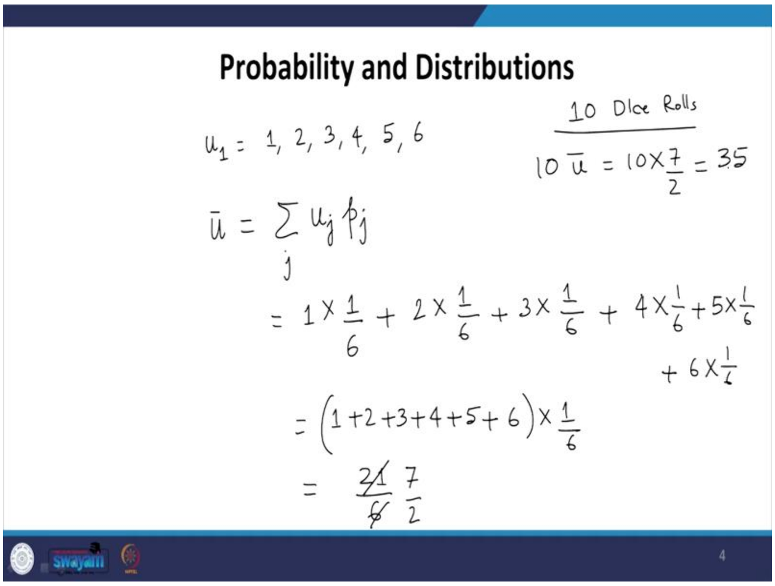
So let us say for example I call I define head as something like 1 and tail as 0 and now I count the number of heads in the in the in the problem. So in that case what I can say is the expected value of the number of head is equal to the sum over the 2 outcomes which are basically probability of having a head multiplied with the head value plus probability of having a tail multiplied with the tail value and the tail value we have assigned as 0 head value we have assigned as 1. So we have-

$$\bar{u} = u_H p_H + u_T p_T = 1.\frac{1}{2} + 0.\frac{1}{2} = \frac{1}{2}$$

So in every toss the mean number of head that we will get is 1 by 2 and why we get this number because we have said tail as 0. So any non-zero number is because of the heads so now if I am doing for example M tosses then for every toss the expected number or the mean number of head is 1 by 2. So for M tosses we should have M by 2 heads that is simply found by multiplying the expected value for 1 toss with m because in every toss there is this mean 1 by 2 chance of getting a head this is what is called an expectation.

Now the coin toss example was pretty trivial because there were only 2 possible outcomes head and tail. Now I will do a slightly more complicated example to demonstrate that this is like can be much more useful when the number of outcomes are much larger. So let us say for example now we are playing a game of dice.

(Refer Slide Time: 09:56)



So a dice can have numbers 1, 2, 3, 4, 5 and 6 so we can have a 6 sided dice and all these outcomes are probable again for an unbiased dice as 1 by 6. Now if I ask what is the mean number we will get again we will multiply this outcomes with their probabilities. So what we have is-

$$\bar{u} = \sum_j U_j p_j$$

$$\bar{u} = 1\,x\frac{1}{6} + 2\,x\frac{1}{6} + 3\,x\frac{1}{6} + 4\,x\frac{1}{6} + 5\,x\frac{1}{6} + 6\,x\frac{1}{6} = (1+2+3+4+5+6)\,x\frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

which is the expected number that we get.

Now for 1 toss it does not have much meaning because you cannot have the expected number to be if I just factor it as 7 by 2 you will never get a 7 by 2 in any dice that is not a even a possible outcome but if I do many of the dice rolls and if I count the outcome or sum over all the outcomes what I will have is 7 by 2 multiplied with the number of dice rolls we did and then it will have a meaning.

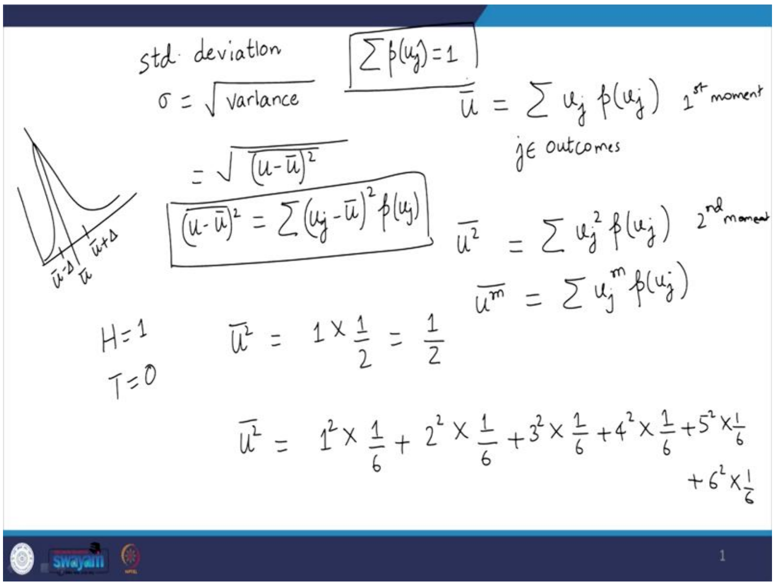
So let us say for example if I have done 10 dice rolls now my number will be-

$$10\,\bar{u} = 10\ x\frac{7}{2} = 35$$

So 35 is the most probable sum that we will get if I am doing 10 dice rolls. So clearly there is not much meaning if I am looking at 1 dice at all but it starts to have a meaning once we are looking at sum over these expectations over a large number of rolls. It tells me where the distribution will peak what is the most probable distribution for this particular problem.

Now as we go from a discrete to a continuous distribution as we have noticed earlier in the coin toss example as well the distributions start to first of all look neater and second become narrower as the number of tosses increase or as the number of dice rolls increase or as a number of any stochastic experiments is increasing the distribution starts to become narrow and this expected value tells me the mean of that.

(Refer Slide Time: 13:50)



Now the second question is let us say if I find the mean then what characterizes the breadth of the distribution and the answer lies in the theory of probability as the following. So just like we had found the mean value as sum over all the possible outcomes multiplied with their probability we can also find what is known as the moments of this distribution-

$$\bar{u} = \sum u_j p(u_j)$$

J=number of outcomes. For example we can define a second moment as-

$$\overline{u^2} = \sum u_j^2 p(u_j)$$

and the second moment characterizes the spread of the distribution first moment characterizes the mean of the distribution.

Now what is that equal to for the dice roll example. So you have so now I am summing over the square of the numbers. So it is-

$$\overline{u^2} = 1^2 x \frac{1}{6} + 2^2 x \frac{1}{6} + 3^2 x \frac{1}{6} + 4^2 x \frac{1}{6} + 5^2 x \frac{1}{6} + 6^2 x \frac{1}{6}$$

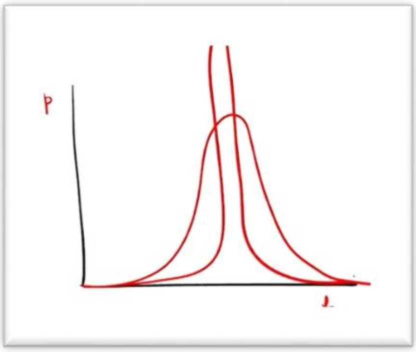
keep in mind that probability remains 1 by 6 for the dice for example only the outcome I am squaring and adding this is what is called the second moment.

Similarly we can define the third and higher moments in general we can define as-

$$\overline{u^m} = \sum u_j^m \, p(u_j)$$

So this second moment is actually related to what you know in statistics as the idea of a standard deviation. So the standard deviation is essentially defined as the square root of what is known as the variance and the variance is defined as not the second moment but the moment of u -.$\bar{u}$ So if I compute the moment of u - $\overline{u^2}$ what it tells me. It tells me about the deviation from the mean value in every possible outcome and if I add all the deviations then what should happen it should go to 0 because both a positive deviation and negative derivation would be equally likely.

So let us say for example I am doing the dice roll example so we have got a mean value we can get a higher value as the same probability as a lower value provided that it is the same departure from the mean and this comes from the symmetry in the distribution. Why is that because if you remember our probability distribution looked something like this-



So let us say if I am here if I find the probability of $\bar{u}$+ some Δ and $\bar{u}$– Δ both are going to be the same if the distribution is symmetric and most distributions we care about are symmetric.

For a dice roll the deviation to the positive side is same as deviation to the negative side when we talk in terms of probability. So for example to the coin toss example of course for 10 tosses we should have 5 heads as the most probable number but having 4 heads should we have same probability as having 6 heads because 6 heads correspond to 4 tails. Since heads and tails are identical the probability of having 4 heads should be equal to probability of having 4 tails. So therefore the distribution is symmetrical so whenever we have a symmetrical distribution we have equal probability of having a positive deviation and negative deviation so if I simply add up all the deviations what we will get is 0 but on the other hand if I sum over the squares then it will not go to 0 because positive and negative both will square to a positive number.

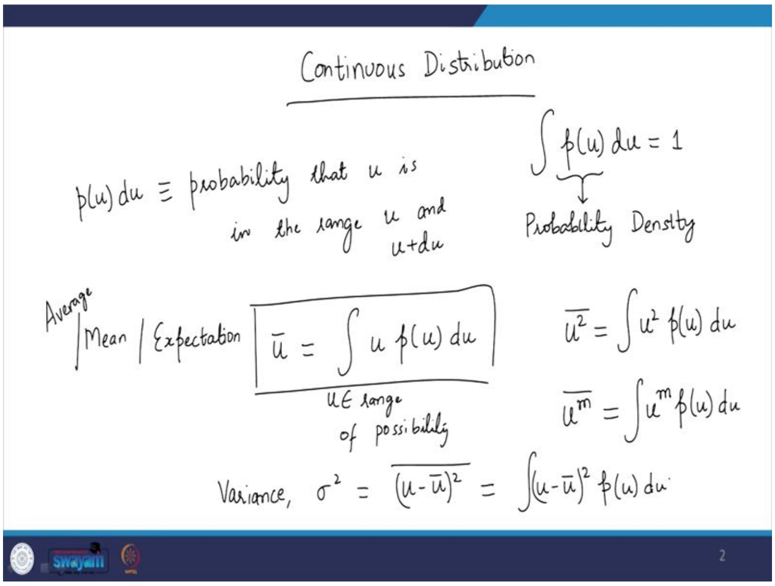
So in this case u - $\overline{u^2}$ would be defined as-

$$\overline{u-\overline{u^2}} = \sum (u_j - \bar{u})^2 p(u_j)$$

so whenever u_j is higher than $\bar{u}$ you will have a positive number whenever u_j is less than $\bar{u}$ you will have a positive number. It does not matter which side you are in the squares already always add up but the absolute value can be positive or negative. So this is what defines the idea of variance.

So now if I have a continuous distribution I will simply replace all these sums with the integrals and what also changes is how we are normalizing the distribution. So in this case for a discrete distribution you may recall that the probability of all the outcome should sum to 1 and this is going to change when we are doing the continuous distribution.

(Refer Slide Time: 20:15)



So for a continuous distribution the condition of normalization is that the probability integrated over all possible outcomes is equal to 1 and now we do not have probabilities anymore what we have is what we refer as the probability density that is-

$$\int p(u)\,du = 1$$

The probability that u is in the range of u and u + du.

So provided that the probability is normalized we can again define the mean and other movements of distribution. For example the mean or the expectation or you may call it an average this is given as now instead of summation we have integration-

$$\bar{u} = \int u\,p(u)du$$

Keep in mind that we had u_j earlier because we can enumerate the outcomes now we have a continuous variable u that goes over the range of possibilities.

Similarly I can define the other movements here let us say if I am interested in the second moment this I can define as-

$$\overline{u^2} = \int u^2 p(u)du$$

the m^{th} moment can be defined as-

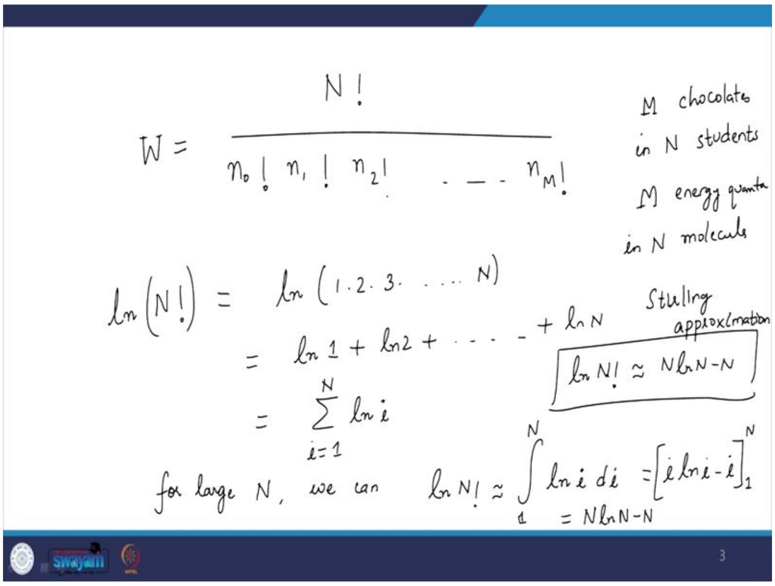
$$\overline{u^m} = \int u^m p(u)du$$

and my definition of variance which is the square of the standard deviation is going to be-

$$variance, \sigma^2 = \overline{(u-\bar{u})^2} = \int (u-\bar{u})^2\, p(u) du$$

and will use this concept over and over but keep in mind that we first have to ask ourselves whether it is a discrete distribution or a continuous distribution. Even if it is a discrete distribution are the numbers of outcomes very large or are the numbers of experiments very large in that case can we make an approximation of a continuous distribution and that will greatly simplify our analysis.

(Refer Slide Time: 23:52)



So, just to give an example of how this can be helpful, we are talking about the number of ways of distributing some N things into different number of bins when I was talking about the chocolates this is distributing N chocolates among certain students. If we are talking about distribution of quanta of energy we said that we are distributing so many quanta in so many people and the formula for that was the number of ways was given as-

$$W = \frac{N!}{n_0!\, n_1!\, n_2! \ldots \ldots \ldots \ldots .. n_M!}$$

where you are either doing m chocolates in an students or m energy quanta in N molecules and N_0 is the number of students having 0 chocolates or number of molecules having 0 quanta N_1 is the number of students having 1 chocolate or number of molecules having 1 quanta and so on.

So now the ultimate objective is we want to find the most probable distribution or probably the spread of that distribution and we are talking about a very large number of N so the objective is to maximize this particular ways of distribution because when we maximize it we will automatically get the most probable distribution.

Now the problem is the way we do the maximization at least that is what comes naturally to our mind is by taking derivatives. Now derivatives itself are defined for continuous functions not for discrete functions, when we have discrete functions then derivatives itself are not defined.

So one way to do that that is only true where the N is very large is to assume as if this is a continuous distribution and if I assume that then we can use the machinery of calculus we can apply derivatives and find the most probable distribution in that particular way not only this it comes at a big advantage also in the evaluation of those factorials because if you try to plug in N very large in a calculator and compute its factorial you will start to see that the calculator blows up in no time after certain time there are so many integer, so, many so many numbers coming in the coming in the final expression that the calculator cannot handle it and therefore we need to find not only a way to find or estimate the factorial but also to maximize this W and the way to do that is to assume first that N is very large and then we can apply the following approximation.

What we can say is instead of the N factorial we can look at the ln value or the log value of the N factorial and what is this equal to it is equal to-

$$ln(N!) = ln(1,2,3 \dots \dots \dots \dots \dots . . N)$$
$$= ln1 + ln2 + ln3 \dots \dots \dots \dots . . \ln N$$
$$= \sum_{i=1}^{N} \ln i$$

So provided that N is very large we can replace this summation with an integration. So for large N we can approximate-

$$\ln N! \approx \int \ln i \, di$$

where i is no longer a discrete variable that takes a value from 1 to N but I is now a continuous variable that can take any value in the range 1 to N that point is very important and we knowwhat the integration of ln of i is that is-

$$\ln N! \approx \int_{i}^{N} \ln i \, di = [i \ln i - i]_1^N$$

and the lower limit will be 1 ln 1 that is equal to 0 then we put 1 more 1 there but the point is that N is so large that this that 1 is not really important.

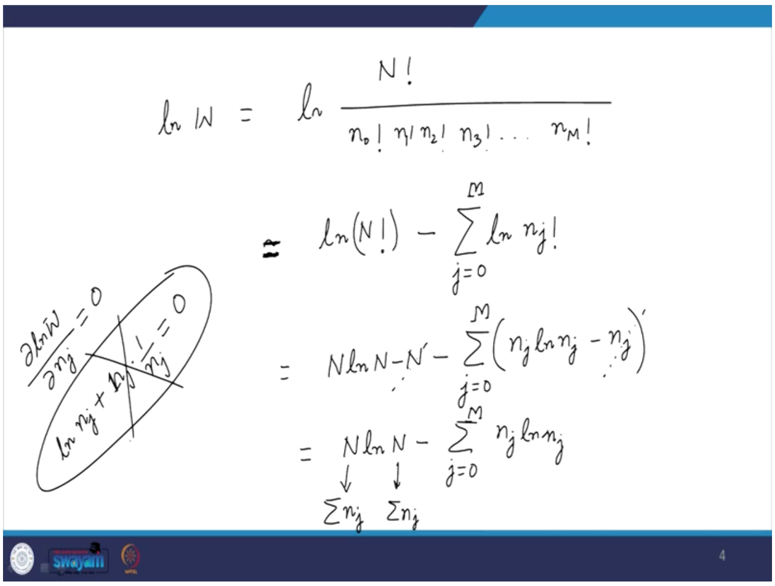
So essentially we can approximate-

$$\ln N! = N \ln N - N$$

which we have been able to do if I am doing a continuous distribution provided that N is very large.

So this particular idea is referred as the Stirling Approximation that is extremely important in thermodynamics and why is that because as I was telling you for an isolated system what exactly is the problem we are trying to solve we want to maximize the entropy and entropy essentially is given by this, even when we are doing a closed system ultimately we are maximizing some energy wherein entropy is already present. So this kind of an expression will always find place in whatever function we are trying to minimize and to do the minimization of that we can resort to the idea of calculus I just described and instead of maximizing w we can maximize ln of W because ln of W increases as w increases ln of w decreases as W decreases so the maxima of W is the maxima of ln W as well. So instead of like solving for W we can solve for ln W and when we are doing that we can use the Stirling approximation. So let us see how it works.

(Refer Slide Time: 31:22)



So now-

$$\ln W = \ln \frac{N!}{n_0!\, n_1!\, n_2!\, n_3! \ldots \ldots n_M!}$$

$$\approx \ln N! - \sum_{j=0}^{M} \ln n_j!$$

this is not really an approximation. The approximation will come now. So now we can use the Stirling approximation and this will be-

$$= N \ln N - N - \sum_{j=0}^{M} (n_j \ln n_j - n_j$$

Now you can look at the second term on the inside the summation. The summation of the number of molecules having so much quanta of energy if I sum over all the possibilities that should be equal to the number of molecules N. So what essentially we have is-

$$= N \ln N - \sum_{j=0}^{M} n_j \ln n_j$$

and now N is anyway constant in this particular problem if I am doing a isolated system. What is changing however is N_j for different distribution. So we can then use the derivative of ln W with respect to N_j and compute its extremum by setting derivative equal to 0 and this can be done for all possible value of N_j and this will give me in this case something like this-

$$\ln n_j + n_j . \frac{1}{n_j} = 0$$

and this will give me a value of N_j.

Now here is something that we have done wrong while doing that so although we have taken N as constant we have to be careful when we are taking a partial derivative with respect to N_j because when I take a partial derivative with respect to N_j what I mean to say is that for every other value of N_k where k is not equal to j. Let us say j is equal to 2 so I am looking at partial derivative with respect to N_2 then I am assuming that N_0, N_1, N_3 and N_4 and everything else are constant. So in there, there is a small detail that comes into picture and that is we cannot really consider N as really constant because I have assumed all the variables except N_2 as constant. So if that is true then N_2 will also become constant if capital N is constant so I cannot take the derivative. So the shortcut of doing that is that I will replace N with sum over N_j in the first expression and that will give me a more meaningful answer that is what we will do later.

Just to repeat this particular point whenever we take this partial derivative we cannot take the derivative keeping capital N is equal to constant because when we take the partial derivative with respect to N_j to be 0 what I mean is that all the values of N_k where k is not equal to j are constant that is N_0, N_1, N_2 everything except N_j are constant. As soon as we assume that if we also assume capital N to be constant then the N_j also become constant. So our problem of taking a derivative with respect to N_j does not make any sense. So therefore for the purpose of the partial derivative we need to replace capital N with summation over and j and that will become a very important point in the arguments that we will follow.

So basically to conclude the discussion of today I have discussed the idea of discrete and continuous distribution. I have said that for large number of outcomes for large number of experiments we can always assume a continuous distribution. We can define the moments of distribution and we did it for both discrete and continuous case and finally we discussed the idea of Stirling Approximation that I claimed will be very handy in our treatment of thermodynamics.

So with that I conclude here, thank you.