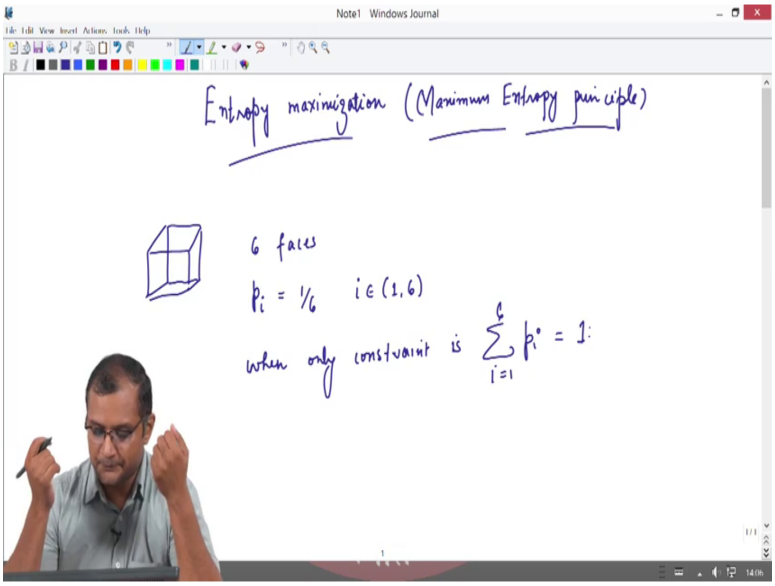
Statistical Mechanics
Prof. Ashwin Joy
Department of Physics
Indian Institute of Technology, Madras

Lecture – 10
Entropy Maximization

Good afternoon students. So, in this lecture, we will continue with the Entropy Maximization. And also a specific case of a dice that is biased the probability distribution comes out to be Maxwell Boltzmann type or it becomes non-uniform ok.

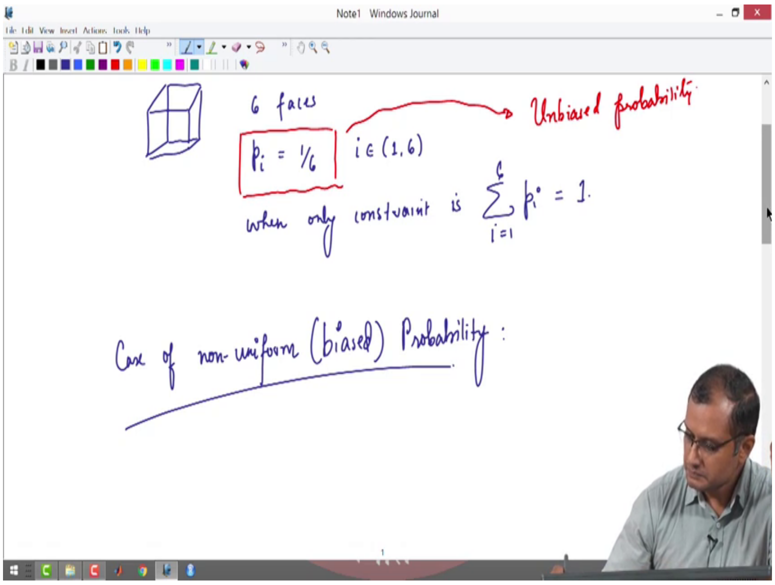
(Refer Slide Time: 00:44)



So, let us continue the discussion on entropy maximization or you may wish to call it as maximum entropy principle ok. Now, we saw earlier that if you take a dice and say that its faces are all unbiased, which means 6 faces. And the probability of each phase, you know if you rule it n times, then it is expected that one-sixth of the n would be the probably; one-sixth of n would be the number of occurrence of any i-th face. So, the probability of any face comes out to be 1 by 6, where i is one of the six outcomes ok.

Now, this was definitely the case, when we discussed that the only constraint as the normalization of the probability.

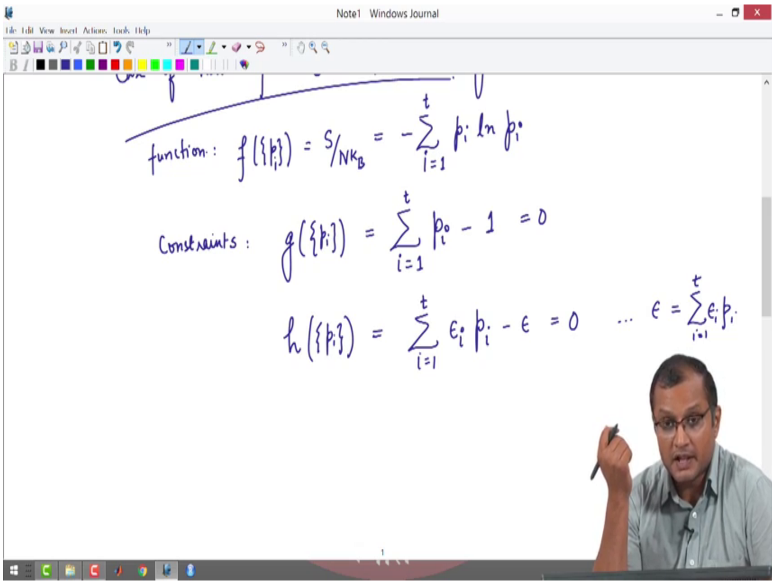
(Refer Slide Time: 02:54)



So, when the only requirement is that probabilities are well behaved, then you get probability distribution that is unbiased. So, I will call this is in this result that we obtained in the last class as an unbiased probability distribution ok. So, this will be called as unbiased, all faces have equal probabilities. Now, today will see the important result, which is establish in the context of biased probabilities.

So, this will be the case of dice, where some faces I have a higher probability of occurrence compared to the other. Now, what do I mean by this it means that I have now three conditions ok. So, I am going to consider the case of non-uniform or biased PDS or probability distributions ok.

(Refer Slide Time: 04:17)

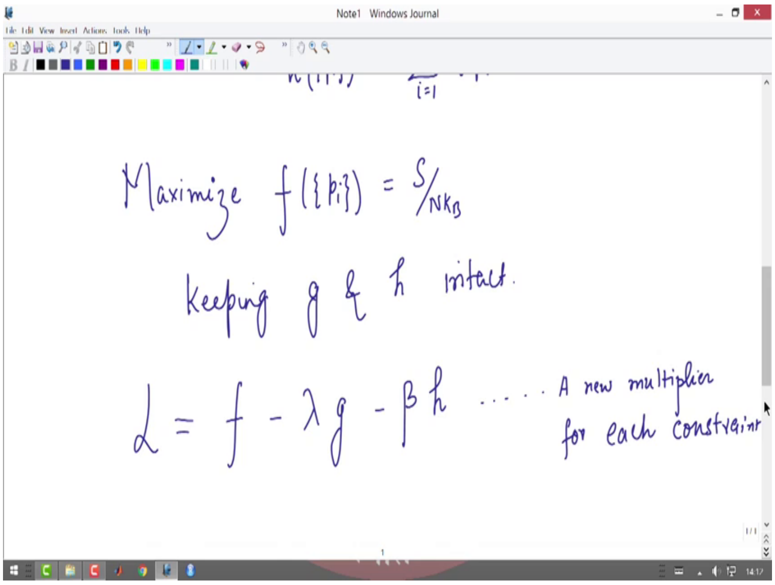


Now, this simply means that I have my function of various probabilities as you know the entropy over N K B, which is given as minus summation i going from 1 to t, I am going to taking a t face die again pi l n pi as before. And my constraint, so I will take constraints now instead of one constraint I will take two constraints. So, the first constraint is my function g again a function of all these probabilities p 1 p 2 all they have to p of t as conservation of the of the norm ok, so sum of all probabilities is 1. So, this is one constrained.

And I have another constrained now I call it as function h it is the function of all these probabilities as the average outcome is fix to some epsilon. So, the average would mean I call the outcome of each faces epsilon i with the weight pi, and i summed over all t possible outcomes. And this is basically nothing but epsilon ok.

So, what this basically means that have taken I have said that the die is biased in the sense that the average outcome you know is epsilon. So, this is like saying the average value of epsilon is fixed which is nothing but I will call it as I can simply write it as ok. So, definitely my second constraint is h, I have two constraints. And I want to maximize my function, which is entropy. So, my function is f which is nothing but entropy per N K B.

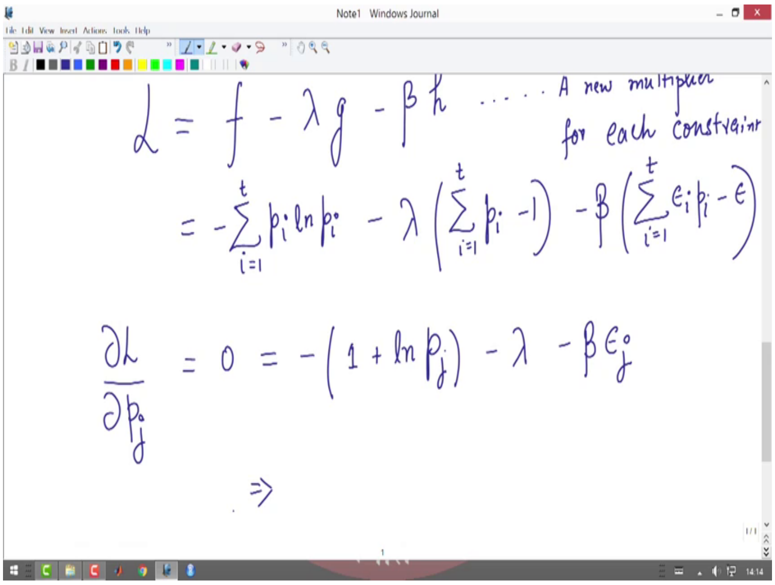
(Refer Slide Time: 07:06)



And I need to maximize my function, which is entropy subject to the constraints ok. Keeping g and h intact. So, under these constraints I want to maximize the entropy, and you shall be surprised to see the outcome. So, our usual prescription as discussed in the last lecture, you can consult the last lecture if where we have derived the where we have discussed the concept of function optimization using Lagrange multipliers.

So, you can consult that later I am going to just use that concept here. And say that; I am going to defined my Lagrange in L as function f minus lambda times g minus my second Lagrange multiplier beta times h. So, you this way as I have said if you have more constraints for each constraint, you have a new Lagrange multiplier in the problem ok. So, a new multiplier Lagrange multiplier for each constraint each physical constraint is a standard technique of function maximization ok.

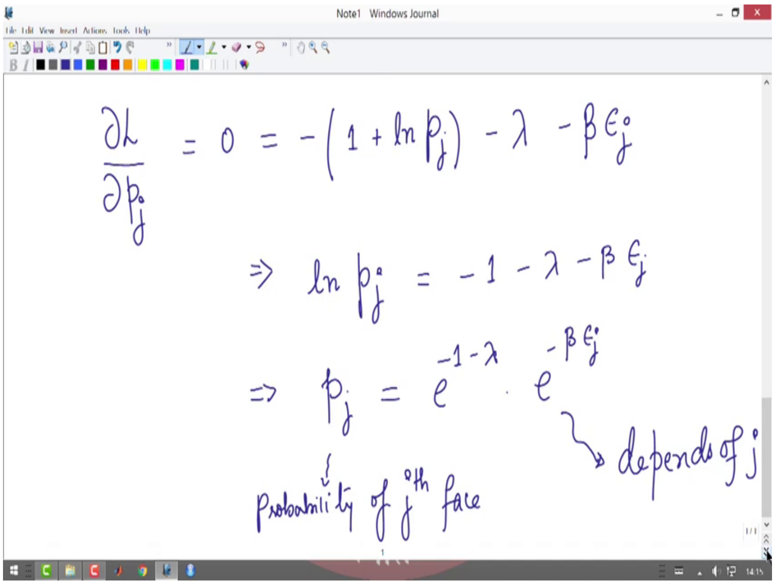
(Refer Slide Time: 08:54)



So, let us put the values of our functions f as you know is nothing but minus summation over all i's going from 1 to t pi l n p i minus lambda times g, which is nothing but summation overall pi minus 1, which is nothing but d conservation of norm. And 2nd Lagrange multiplier multiplied to the fixed value of the expectation of the outcome. So that is nothing but, summation i in from 1 to t epsilon i and the outcome of each face times probability of each face minus the average value itself ok. So, these are the two constraints.

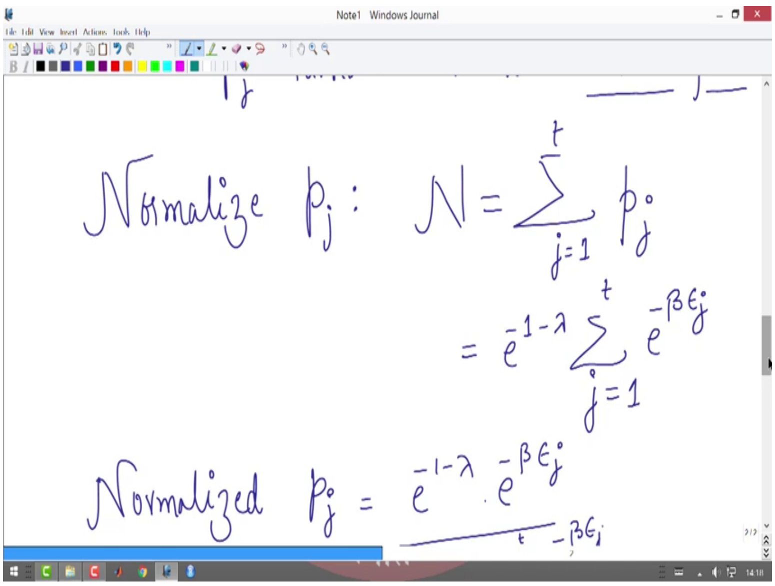
And I know that the solution of optimization is a probability p of j and p j is the solution that is p j optimizes the entropy or the function f, then del L by del p j, it should be 0 should be 0. So, let us compute del L by del p j, so you can see the first term of the Lagrangian only the term p j, I mean I equals to j will contribute, and that is simply gives minus of 1 plus lon p j. And the 2nd term would give minus of minus lambda in to 1, which is just minus lambda. And final term would give minus beta into epsilon j, because only i equal to j term will contribute there, and epsilon is any way constant, the last term is 0. The derivative of last term with respect to p j is 0.

(Refer Slide Time: 11:17)



So, you can re arrange everything and obtain the value of p j now. So, there is a 0 here, so what will do is you can write down a logarithm of p of j as minus 1 minus lambd a minus beta times e of j. And, this will simply give you p of j as e raise to minus 1 minus lambda into e raise to minus beta of j, and look at the right hand side. This time the probability of the j-th face is depended on j, ok. Now you can see it depends on j, because is a term e j, which is in some sense the value of the face j.

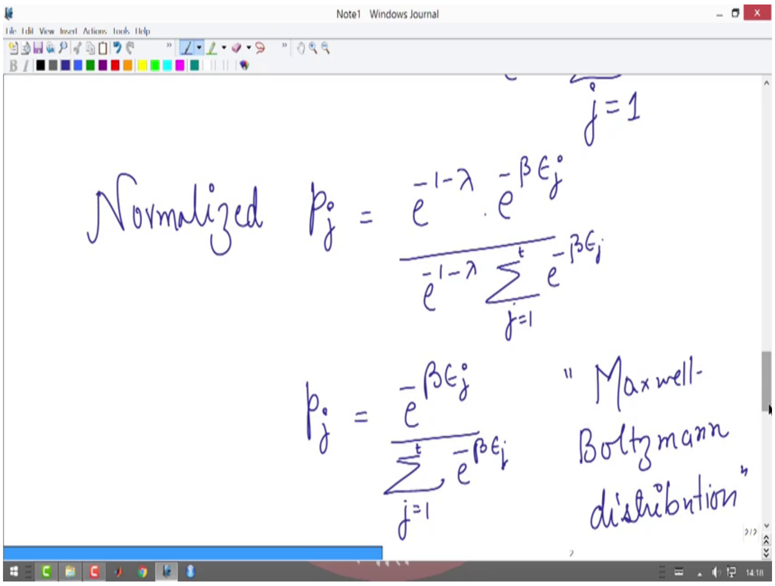
(Refer Slide Time: 12:20)



So, this is called as non-uniform pdf. So, hence we will say that p j turns out to be non-uniform. So, with the extra constraint that we are taken our PDF has now become non-uniform. And this is actually a Boltzmann PDF Maxwell Boltzmann pdf, we shall see so our P j is not yet normalize as can be seen, so will simply normalize p j.

So, normalize p of j and to normalize this p of j, I need to find a normalization constant, which is nothing but summation over all j p of j. So, this will go from 1 to n 1 to t, and because the value of p of j is given here, it simply becomes if I submit t times. This will simply the e raise to minus 1 minus lambda into summation of summation on j p of j.

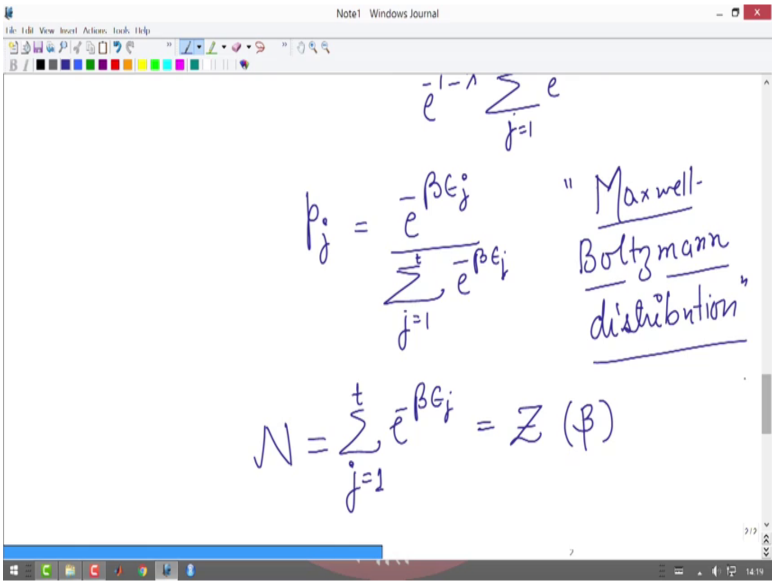
(Refer Slide Time: 13:47)



So, with this normalization constant our final normalized PDF becomes they are normalized pdf, which is simply e raise to minus 1 minus lambda into e raise to minus beta times epsilon j divided by the norm that we have just obtained, so we plug it here. And it is simply becomes right, so I have just (Refer Time: 14:45) small type of here thank you. So, this becomes e to the power minus beta e j of course.

And then you can write down the final answer as your normalized pdf, which is a to the power minus beta e j over summation j going from 1 to t e raise to minus beta e j. And this is nothing but if you say that beta is 1 over K T, and this where e j is nothing but the energy levels of j-th you know state. Then this is nothing but Maxwell Boltzmann distribution, where our normalization constant is also known as partition function.

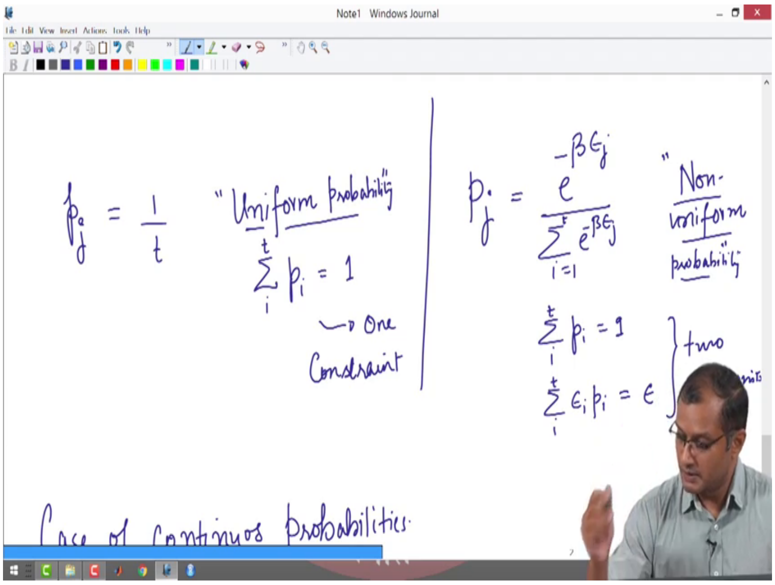
(Refer Slide Time: 15:43)



So, in some sense this is like a partition function for your problem, which depends on beta ok. It does not depend on energy levels per say because, this summed overall the levels that is gone, but it is a parameter here is beta.

So, this is the reason why we see Maxwell Boltzmann distribution, everywhere around as if your system has constraints more than just the conservation of probabilities. So, if you set the average value of energy to some constant, then the distribution will come out to be Maxwell Boltzmann. And the consequence this is a consequence of the entropy maximization that you see Maxwell Boltzmann distribution profound in nature. So, this is one example that we have discussed.

(Refer Slide Time: 17:18)

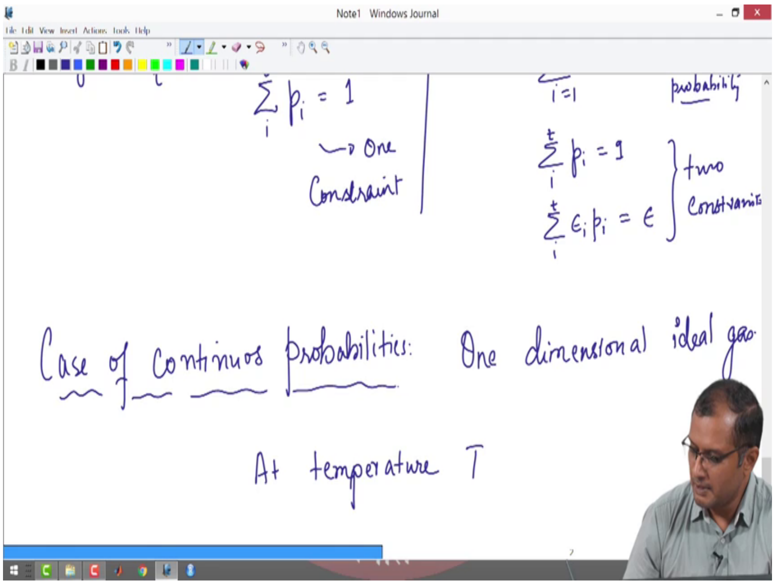


And I am going to give you one more example worked example, where the degrees of freedom are not discrete they are continuous ok. So, what we saw I am just going to summaries it. So, what we saw so far is basically two different types of PDF for the district case, one type of PDF is simply p of j as 1 upon t, I call it as uniform PDF probability density functions. And this comes if there are the only constraint is summation of probabilities, which are conserved or the probabilities are all the add up to one.

And the other case that we have seen basically p of j, which becomes non-uniform. So, I have e raise to minus beta j over some over, so I get non-uniform pdf, if I have an extra constraint. So, above the well behavedness of the probability, I have two constraints here. One that the probability is well behaved the other is the probability the average value of the outcomes is fixed ok.

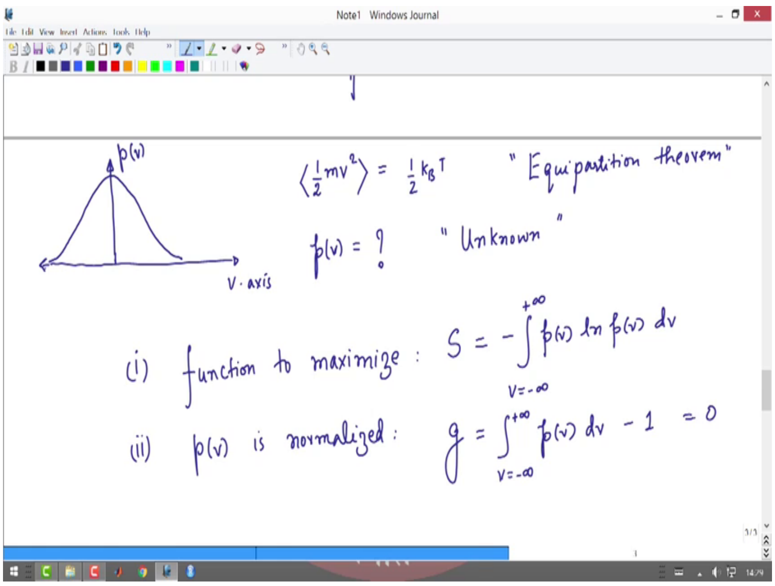
So, I have two constraints here. And I have only one constraint here ok. And these constraints basically manifest in the type of PDF that we obtain ok, so that is the end of discrete probabilities. We will now move on to case of continues probability distributions ok.

(Refer Slide Time: 19:15)



So, I am going to talk about the case of continuous probabilities. So, I am going to call this as not as pdf, but actually just probabilities here, because there discrete probabilities. And similarly here to sort of the more accurate, I will call it as probability ok. So, let us look at the case of one dimensional gas ideal gas ok. So, let say the ideal gas is some temperature T. Now, I know by equipartition theorem that per each a degree of freedom the energy contribution to an ideal gases half k of T half k T ok.

(Refer Slide Time: 21:13)



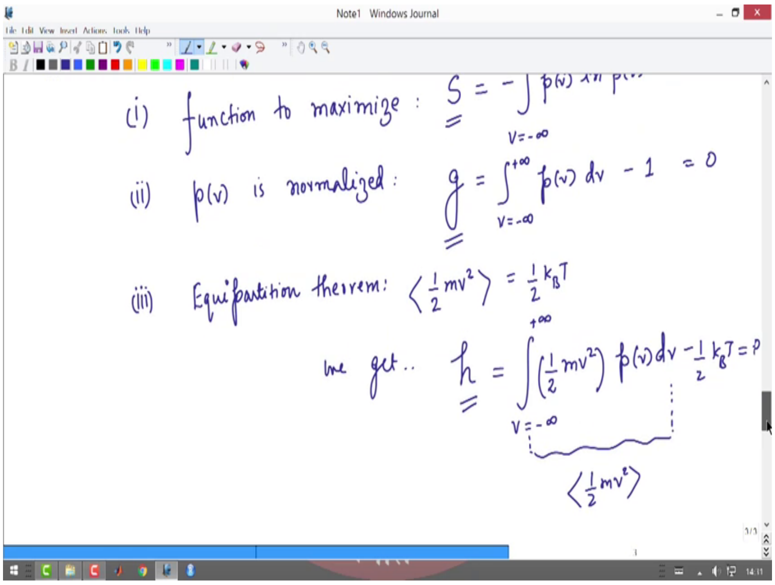
So, I can take the velocities for the case an ideal gas to be just along the line x axis. So, this is the v axis, v is the velocity ok. So, I have some distribution of v and I know that the distribution is Gaussian or Maxwell Boltzmann like. And I know the answer to be so will derived this distribution. And so what we really know here in these cases that my kinetic energy average kinetic energy per degree of freedom is basically half k T by equipartition theorem ok. And I want to know what is my PDF, let us say is unknown although we know from the standard literature that the PDF is supposed to be Maxwell Boltzmann like, but there is no formal group presented to you so far.

So, the task of this section is to show that the only PDF that is satisfied by the particles here is Maxwell Boltzmann PDF or Gaussian PDF ok. So, we will prove that. Now, what are the ingredients needed for this proof? Well, first is definitely I need to know the function of that function that I am going to optimize or maximize. And that function is basically my entropy which is basically my entropy. And just as you know in the case of a discrete case, I will take my entropy is to be instead of taking summation I will take it as a negative integral of p of p ln p of p dv ok. And the limits here are going from minus infinity to plus infinity ok.

Now, if you are bothered by the fact that this p of p is dimensional and I have taken logarithm of it, where it is not dimensional, because I have divided by p upscale of PDF which takes care of non-dimensionalization ok. So, this PDF are all non-dimensional and consequently this velocity scale is also non-dimensional. So, the all this non-dimensionalization leads to a constant of entropy hear that have taken to be conveniently 0, because at the end of the day I do not require exact values of entropy, I required a shift or the change in entropy. So, when you take for any practical purpose delta S, you can simply change this is fact simply go to 0, because it is a constant.

So, I am not writing it, but I am aware that this p of p non-dimensional ok. So now, this is my entropy or the function that I am going to maximize. The second ingredient that I must observed or I must you know take into consideration is the fact that my PDF that I am changing are normalized ok. So, p of v is normalized, very important ingredient. So which means, my constrain g is basically integral v going from minus infinity to plus infinity p of v dv minus 1 which is equal to 0 my PDF is normalized that is my constraint.

(Refer Slide Time: 26:24)



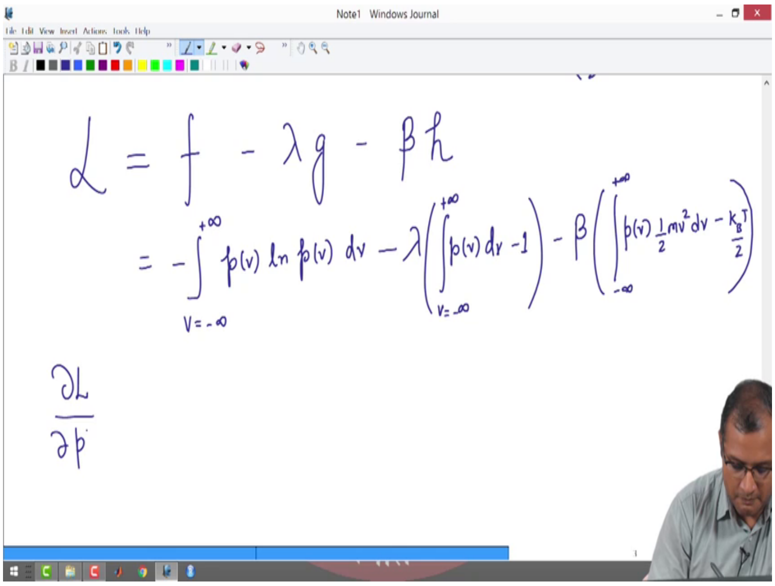
My third ingredient is one more constraint. I know from equipartition just now we discussed I know that half m v square average is just half k T. So, this basically gives me my constraint h I am just going to write down my constraint h as simply integration v minus infinity to plus infinity the average value of half m v square which is nothing but half mv square a measured in the distribution minus half k T is 0 ok. So, average value of any function of v is nothing but the like we said this is nothing but this quantity is integral is nothing but average value of half mv square ok. So that is what I have written here right.

So, now I have the function that I want to maximize. I have my two constraints g and h under which I am going to maximize my function S. So, I will use the concept of Lagrange multipliers and that is straightforward, it is just the same Lagrange multiplier business that we have done in the context of discrete probability. And it is very interesting. At the end of this optimization, you will see and natural emergence of Maxwell Boltzmann distribution.

So, let us do that. And instead of firmly established in our heads why nature why mother nature always prefers Gaussian distribution or Maxwell Boltzmann distribution for velocity for a system which is isolated has been given enough time to evolve and reach thermal equilibrium at temperature T that is only reason for this observation is that a

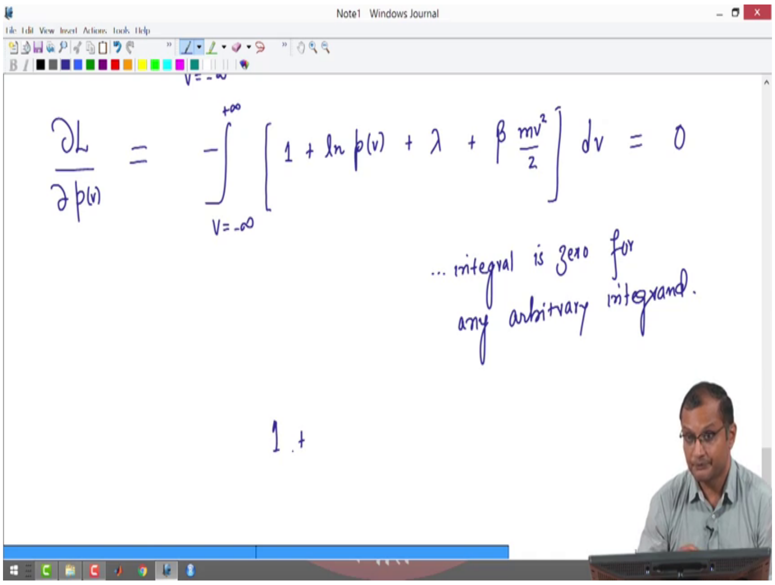
Gaussian distribution is the distribution of maximum entropy. So, let us do that. So, same principle maximization principle leads us to the following equation.

(Refer Slide Time: 29:11)



I will write down the following equation. I will write down our Lagrangian as function f this f is the entropy minus lambda times g, where lambda is the Lagrange multiplied corresponding in the constraint g, another Lagrange multiplier which is beta corresponding to the second constraint h. Let us write down what are these quantities. So, my function f is nothing but the entropy which is minus of integral v going from minus infinity to infinity p of v ln p of v d v minus lambda times g which is nothing but use the space wisely. And now you simply take derivatives and say that this Lagrangian is maximized by some PDF which is p and call it as so this is the p that maximizes it.

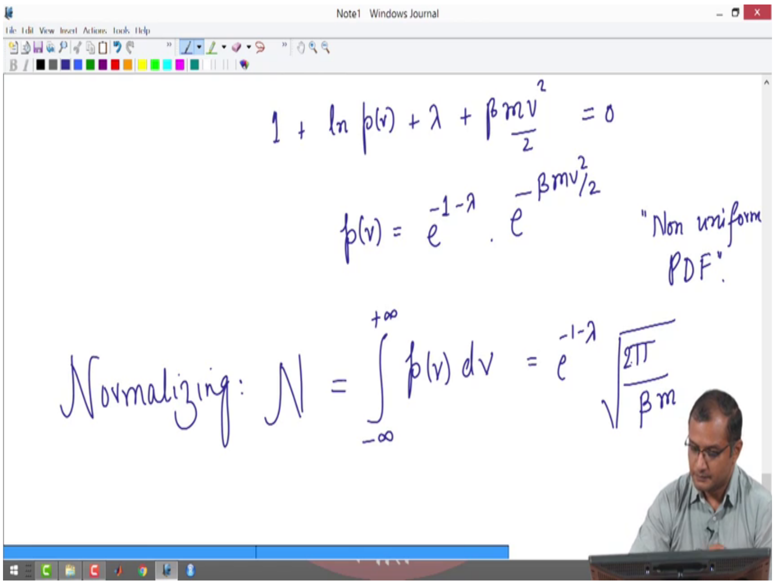
(Refer Slide Time: 31:08)



And so I have to set this to 0, and if you take the derivative this simply becomes I am going to write down the entire thing under the integral. So, from the first term I get 1 plus ln p of v. So, I am going to take derivatives respect to p of v fine. And for the second term I get lambda times 1; and for the third term I will get beta times mv square by 2 that is it this is an integral on v which is 0, I have already written 0. So, it is not required now you can actually write down 0 on the right hand side to sort of and we shall remove a 0 from here ok.

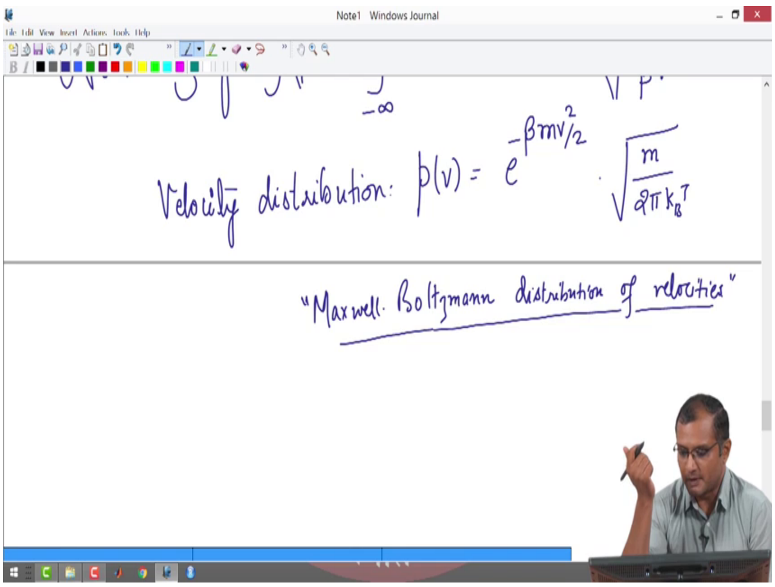
Now, we can see that the integral is 0 for all values of integrand for any arbitrary integrand cannot specify that a p of v has to be a certain type only then this is 0 ok. So, I can say that for this entire integral to be 0, my integrand has to be 0 ok.

(Refer Slide Time: 33:30)



So, which means I will write on my integrand as 1 plus ln p of v plus lambda plus beta mv square by 2 as 0 ok. And so, you can simply right down p of v as e raise to minus 1 minus lambda into minus beta m v square by 2. Now, we can see that this is also a non uniform PDF ok. So, let us normalize it. So, I am going to find the normalization constant ok. This is nothing but the integral minus infinity to plus infinity p of v d v and that will come out to be e raise to minus 1 minus lambda integral of a Gaussian is square root of pi by a pi by a, a itself is beta m the 2 pi fine.

(Refer Slide Time: 35:05)



So, then our normalize PDF I will call it is as a velocity distribution becomes p of v as simply the PDF that we have written divided by the normalization constant it is e raise to minus beta mv square upon 2 into square root of m upon 2 pi k B T. So, this is basically the Gaussian or Maxwell Boltzmann distribution. So, this is the Maxwell Boltzmann distribution of velocities. You can also obtain Maxwell Boltzmann distribution of speeds by simply writing the distribution in terms of molecular speed not velocities.

So, the fact that here is the reason why we get a normal distribution or a Maxwell Boltzmann distribution is the fact that this is the distribution of maximum entropy. So, this basically is a formal proof of why you see certain types of distribution more often than the others, and especially for equilibrium system this is the preferred distribution or in fact this is the only distribution not a preferred distribution.

So, we end the lecture here. And this basically completes the chapter 1 of our syllabus. In the next chapter, we will talk about basic postulates of statistical mechanics. And in this context we will discuss various important ensembles, and discuss examples physical example where the ensembles are relevant, ok. So, we end it here.