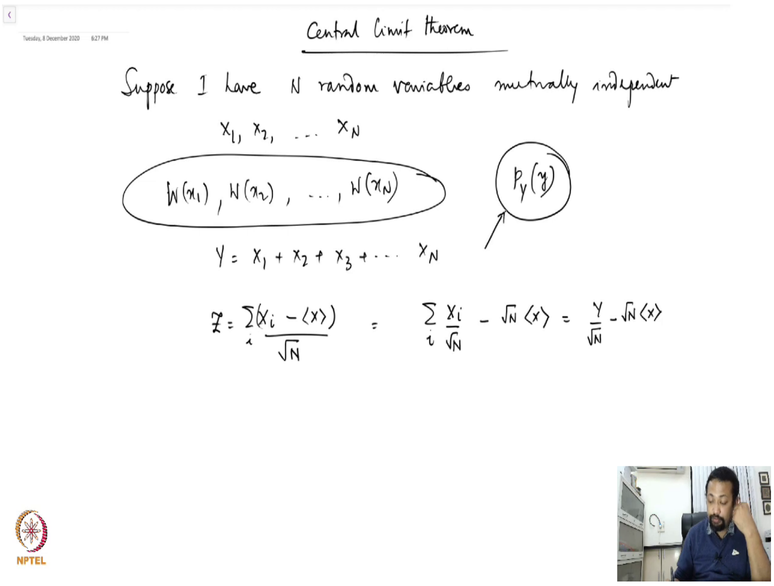
Statistical Mechanics

Prof. Dipanjan Chakraborty
Department of Physical Sciences
Indian Institute of Science Education and Research, Mohali

Lecture – 19
Central Limit Theorem and Statistical Entropy

So, now that we have discussed most of the relevant materials on probability theory that we require in our course.

(Refer Slide Time: 00:23)



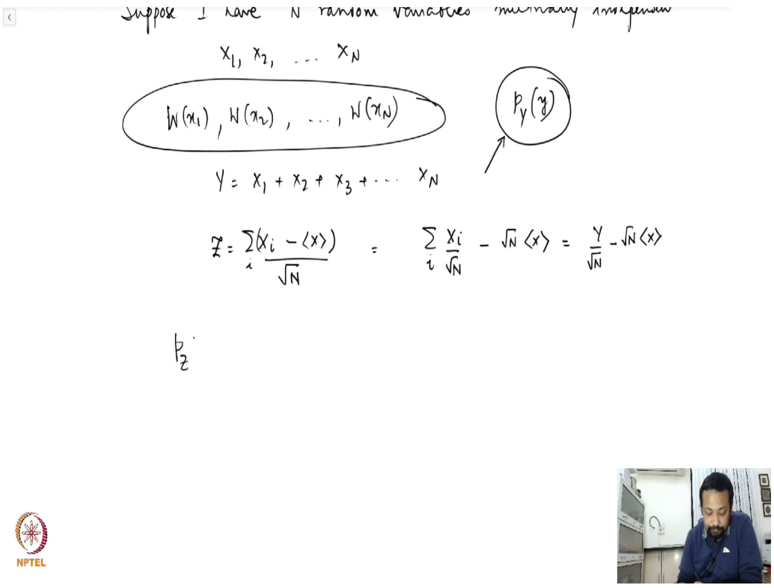
We want to discuss something which is very very significant; which is the Central Limit Theorem. As we shall see later this has a quite application or is 1 of the pillars of statistical mechanics. Now, essentially suppose I have N random variables which are mutually independent. They are mutually independent and let us denote them by X 1, X 2, X N.

The associated probability densities are; W x 1, W x 2, W x N. So, the form of W is same for everybody same. Essentially they are drawn from the same distribution, but they are mutually independent right. So, then let us define a variable which is X 1 plus X 2 plus X 3 all the way up to X N. So, given this probability density is W; I want to know what is the probability density p Y of y right.

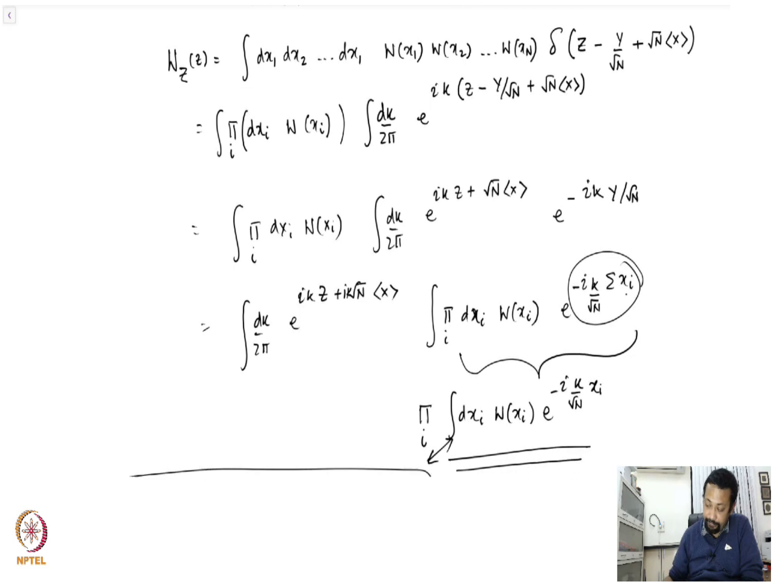
Now, for that let us construct this variable which is sum over i minus average of X divided by square root N. And if you just expand this now this becomes X i over square root N minus square root N times average X which is nothing, but Y over square root N minus square root N times average X.

So, we want we will go over to this quantity we will evaluate this quantity, but slightly in an indirect way. And it will you will see the significance of this. Right now what I want to say is that I have never given you the form of W. And we will see later whether we need it or not and that is kind of an essence in this whole story.

(Refer Slide Time: 02:34)



(Refer Slide Time: 02:39)

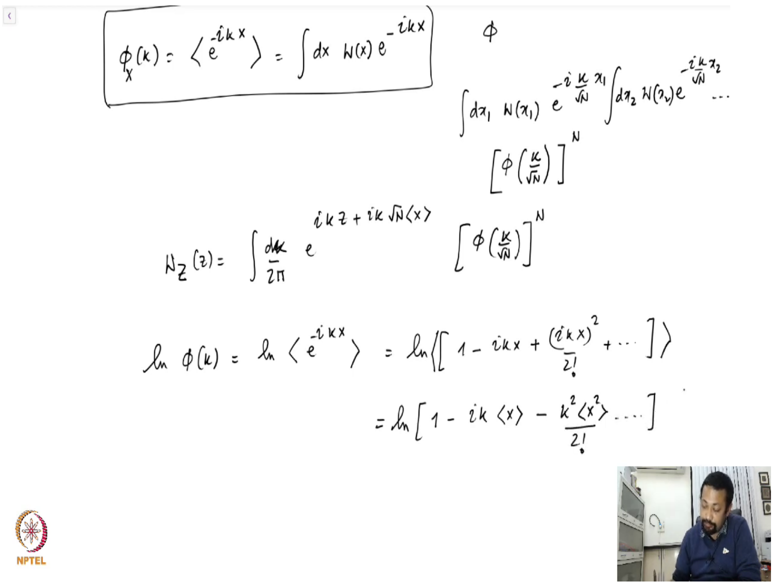


So, let us write down p Z or since I am denoting it with W so W Z of z. Then W Z of z is of course, dx 1, dx 2, dx, W x 1, W x 2, W x N, right. Not only that now I also have to make sure that Z has this particular form which means; Z delta Z minus Y over square root N plus square root N times average X. So, let us compact if our notation a little bit and we write down W x i product right.

And this delta function I am going to replace them in an integral representation e to the power i k Z minus Y over square root N plus square root N times average X right. So, I have product over i d X i W of X i and if I look at this over here square root N times average X and I have minus ik Y over square root N.

Let us change the integrals now, the order of the integrals and we will write down this as dk over 2 pi ikZ plus square root N times average X. Product over i; d of x i, W of x i, e to the power. If you now replace Y by sum over x i and like this. Then if you inspect this more carefully then you see that this. Since this is in the exponential this also becomes a product and therefore, I can write down this part as dx i W x i e to the power minus ik square root N x i.

(Refer Slide Time: 04:45)



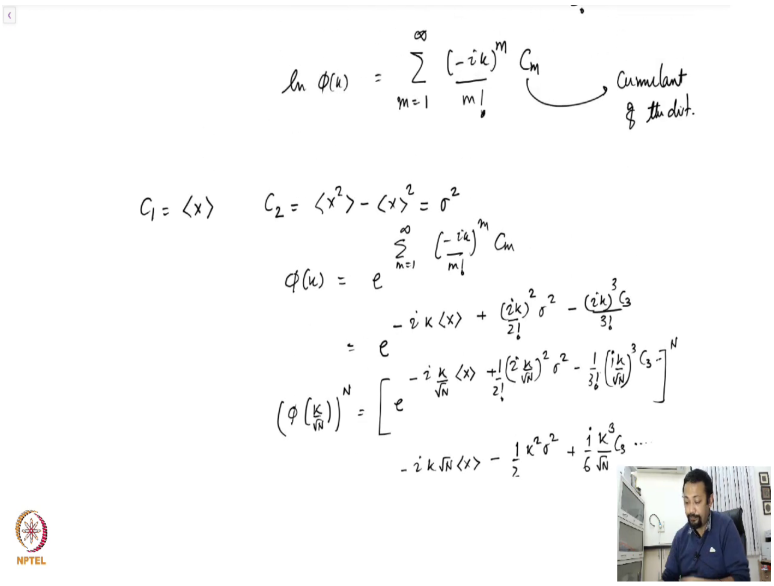
If you recall this; then this is essentially the definition of your characteristic function. We had defined it with the plus 1, but you can as well define it with the minus 1. It all depends on how you define your delta function; I can define with the delta function also with minus ik. So, now therefore phi k the other way down let us say. So, phi of X k is average of e to the power ikX which is integration dx W x e to the power minus ikX right. So, if I do the mapping if you look at it carefully.

Then this integral the product of this integral of W x 1 e to the power i k square root N x 1 dx 2 W x 2 e to the power minus k over square root N x 2 so on and so forth, all the way going up to x n, but the form of W is the same that is one of the advantages that we have in our discussion. So, I can clearly write down this as phi of k over square root N and therefore, this whole integral the product of this integral becomes this raise to the power N.

So, W Z of z is essentially dk over 2 pi e to the power ikZ plus square root N, there is ik missing somewhere, there is an ik here, phi of k square root N raised to the power N right. We will assume now that N is large right.

Before proceeding so, let us just look at ln of phi k then ln of phi k is ln of average e to the power minus ikX right, which is log of 1 minus ikx there is an average outside and then you have plus ikx whole square over 2 factorial so on and so forth average closes and this you can see that this is 1 minus ik average X, i square is minus minus K square X square over 2 factorial and higher order terms right.

(Refer Slide Time: 08:01)



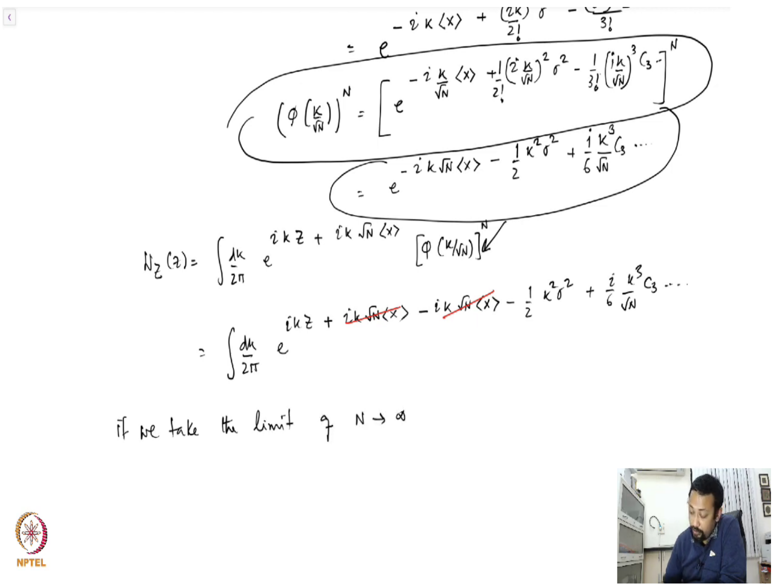
This I can write down as minus ik raised to the power m C m over m factorial, m is equal to 1 to infinity this quantity is called the cumulant of the distribution or the density; so C 1 for example, is average X, C 2 is average X square minus average X whole square which is the variance and you can clearly define the other coefficients as well, other cumulants from this definition right.

So, now, I have this ln of phi k looks like this. So, therefore, if I am careful enough phi k is e to the power sum over m equal to 1 to infinity minus ik raised to the power m over m factorial C m. So, that I can write down this as minus ik, the first one is average X, the second one is plus ik whole square over 2 factorial sigma square minus ik whole cube C 3 over 3 factorial so on and so forth.

Now, recall I have phi k over N, what I have in the expression I have phi k over square root N raised to the power N, which means this is i k over square root N average X plus i k over square root N whole square 1 over 2 factorial sigma square 1 over 3 factorial i k over square root N whole cube C 3 whole raised to the power there is of course, other terms raised to the power N.

So, if I carefully do this becomes minus i k square root N average X minus half k square sigma square this becomes an independent minus one sixth this is i right k cube. So, this becomes i cube square is minus 1. So, this becomes plus and I have N to the power 3 half which I am raising it to the power N becomes square root N C 3 higher order terms.

(Refer Slide Time: 11:19)

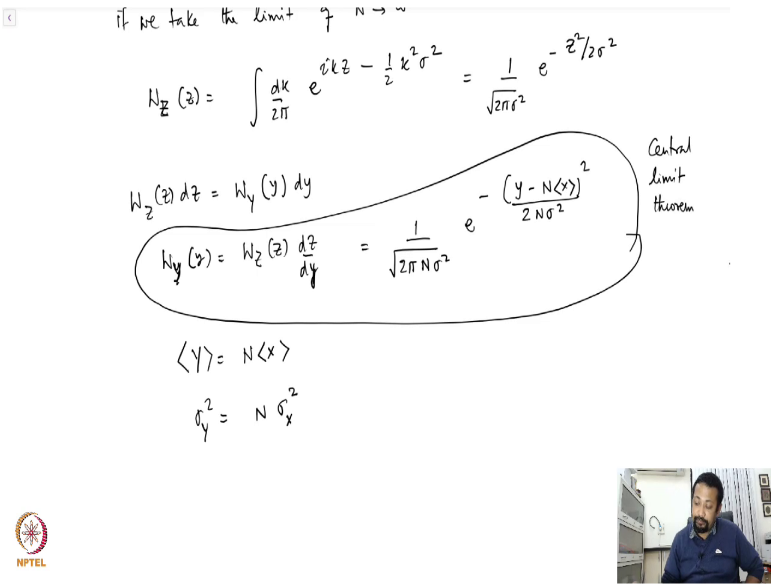


Now, let us come back to this W Z. So, the W Z the density that we are probability density for the variable Z we were calculating had the form dk over 2 pi ikZ plus ik square root N average X and I had phi over k over square root of N raised to the power N. So, this becomes d k over 2 pi ikZ plus ik square root N X.

And then I substitute this expression or rather the 2nd one into this and write it down as square root N times average X minus half k square sigma square plus i by 6 k cube square root N C 3 then higher order terms. Of course, this and this cancel out.

And then now, if we take the limit of N to infinity which means my number of variables the independent random variables that I started off with is really a large set. Then all the other terms vanish because they decay as 1 by square root and then the next one will be N to the power 3 half so on and so forth.

(Refer Slide Time: 12:55)

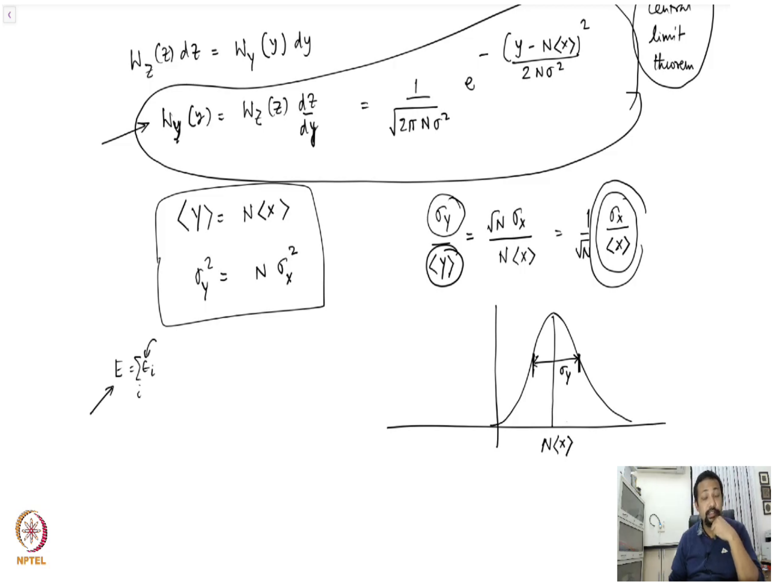


And essentially W Z of z is going to be left out with dk over 2 pi e to the power ikZ minus half k square sigma square all we have to do now is we have to integrate over k and once you do this you will see that this integration is going to give you Z square over 2 sigma square. So, given that I have W Z I want to calculate the p of Y right, but. So, I use that by using the relation because it is a transformation.

So, that W Y capital Y of small y is W Z of z dz dy right. And once you do that you are going to get this as square root 2 pi N sigma square e to the power y minus N average X whole square divided by 2 N sigma square. So, that average of Y is N times average of X and sigma square is N times.

So, sigma square Y is N times sigma square X, but remarkably you see this probability density for Y which is now a sum of random variables comes out to be a Gaussian, with irrespective of the form of W that you choose. So, this is one of the part of the central limit theory. The second part is of course, average of Y is N times average of X, right and sigma square Y is N times sigma square X.

(Refer Slide Time: 15:16)



Therefore, average of Y divided by sigma Y sorry, the other way along it has to be sigma Y divided by average of Y is going to be square root N sigma X divided by N times average of X, which is 1 over square root N sigma X over average of X right.

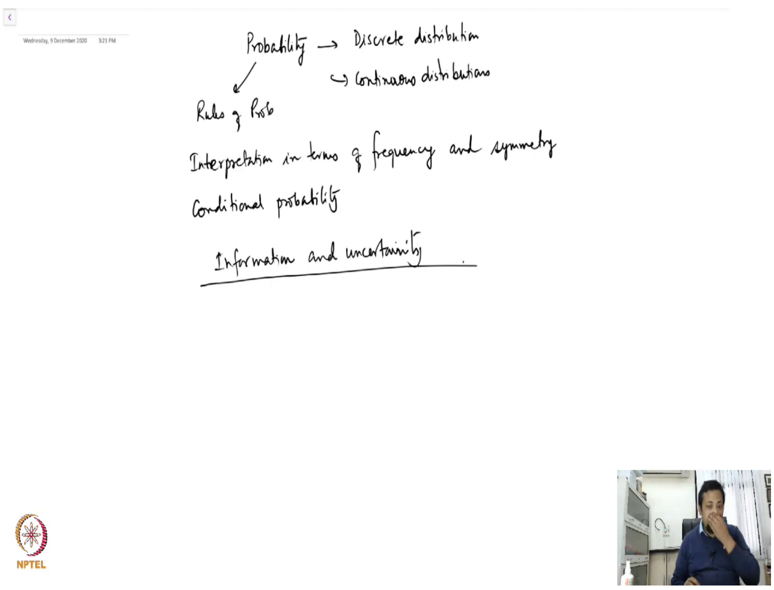
So, this tells you say sigma X and over average of X is a number right, it does not depend on N, but what it essentially tells you that if you keep on adding these numbers N 1, N 2 these random variables X 1, X 2, X 3, X N then the variance of Y to the mean this ratio will decay as 1 by square root N.

So, clearly I know that first of all if this probability density is going to be a Gaussian. So, let us look at it. So, it has an average which is N times average of X and it has a certain width, which is sigma Y. So, if you keep if N is large enough if N is large enough then not only does this average, this ratio decays as 1 by square root N. So, the average also increases, but the sharpness of this which is essentially the spread of this curve also decreases.

So, this curve becomes more and more sharper and sharper right. So, this is the consequence of the central limit theorem. Now, where do we use it in statistical mechanics? Well as we shall see later on see all thermodynamic quantities that we have calculated for a hydrostatic system for example, the energy right extensive quantities the energy I can always write down as energy of the constituent particles right.

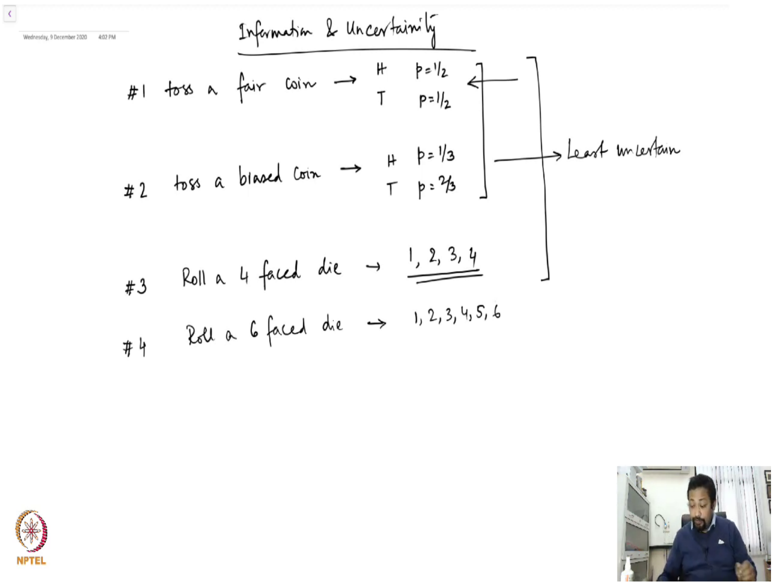
Now, because of that inherence molecule, inherent molecular theorems in such a system these quantities are essentially fluctuating in time and therefore, they are random variables. Consequently it tells you that irrespective of the distribution of the individual energies of the particles. The macroscopic energy e will have a Gaussian distribution. And that is a consequence of your central limit theorem. So, the central limit theorem is extremely powerful and we shall exploit this its power later on in a statistical mechanics.

(Refer Slide Time: 18:13)



So, we learned about probability the parts which are necessary for us, including discrete distributions and continuous distributions. The rules of probability interpretation in terms of frequency and symmetry. The conditional probability and how one goes from, how one defines a continuous random variable that therefore, continuous probability density we also looked at the central limit theorem, but right now what we are going to focus on is information and uncertainty.

(Refer Slide Time: 19:37)



Suppose I give you. So, let us just write information and uncertainty. Now, suppose I give you an experiment for example, the first experiment I do the first experiment is I toss a fair coin, then I know the outcome of this is head and tail with equal probability of half and half.

The second experiment that I do is toss a biased coin. Now, once I toss a biased coin I will still get the same outcomes head and tail, but the probabilities will be slightly different in the sense, let us say p is one third here and p is two third here, probability is two third here.

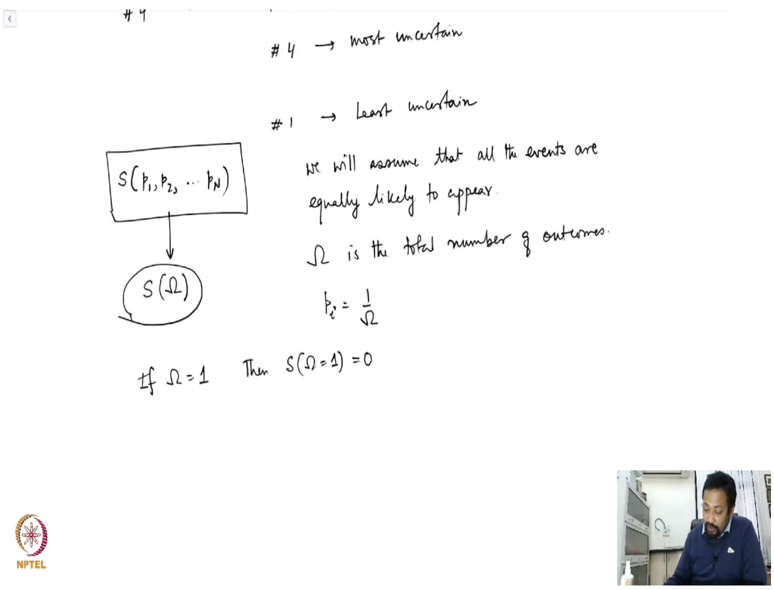
So, clearly of these two if you want to ask, that which in which of this experiment the outcome is more uncertain then you will say experiment number 1, because both of these events are equally likely to happen. In the 2nd experiment of course, this is not true because you know that the tail is most likely to appear more right.

Now, let us do a 3rd experiment where you roll a 4 faced die and the outcomes you know are 1, 2, 3 and 4, if you are now asked to compare between these 3 then you will say experiment number 3 is.

So, now, when you are asked to rate these 3 experiments based on the uncertainty on the outcomes, then you will say look for experiment number 3 the number of outcomes are more, therefore I do not know it is more uncertain for me to predict which outcome is going to come.

So, of these 3 experiments number 2 is the least uncertain. If I now want to do a 4th experiment where I roll a 6 faced die, then you are immediately going to realize that the outcomes are not it does not have 4 outcomes, but there are 5, 6 outcomes. And therefore, this seems to be the most uncertain experiment that we are doing, because we cannot predict very it is very surely which outcome is going to come.

(Refer Slide Time: 22:19)



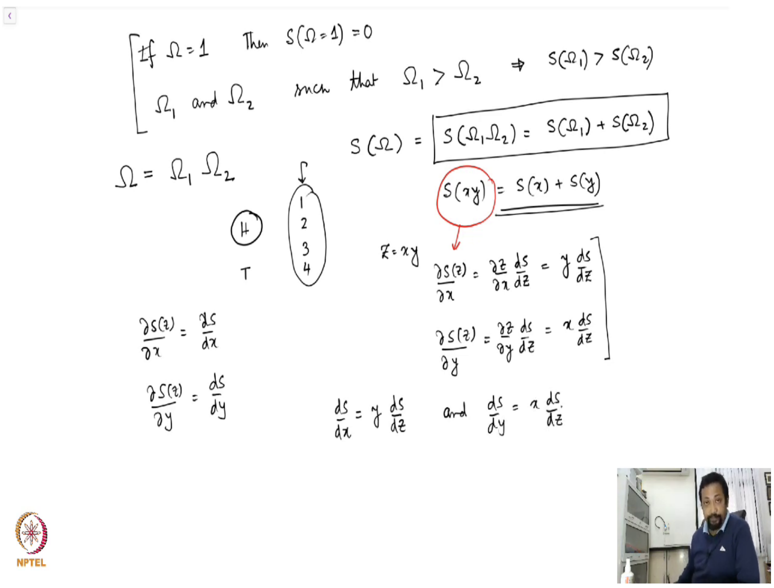
So, if you were to rank these experiments then experiment number 4 would be most uncertain. And experiment number 1 would be least uncertain. I do not know how to rank 3, I do not know how to rank experiment number 3 and experiment number 1. So, the idea is now therefore, to develop a measure for this uncertainty and let us call this measured as S and let us denote the probabilities p 1, p 2, p N if there are N outcomes of the event.

The idea now is, in the very first case that we discuss or we try to figure out an expression for S, we will assume that all the events are equally likely to appear. If this is the case then if omega is the total number of outcomes, then the probability since there the probabilities of these outcomes are equally likely therefore, p i is 1 by omega the associative probability for each of this outcomes is just 1 by omega.

For example for a fair coin we saw that the probability for head was half tail was half, for a fair die we wrote four phase time we know that the probabilities are 1 by N. So, that is the same thing that we are applying over here. We first assume that all these events are likely to appear and therefore, all the outcomes are equal a probability and that probability is given by 1 by omega. So, this expression therefore now reduces to S of omega correct.

So, now it is clear let us what characteristics of that S will have, if omega is equal to 1 then S of omega is equal to 1 is 0, what does this mean omega is equal to 1 means. Essentially there is only 1 outcome and for that event you know that, that is the only outcome that can appear. So, there is absolute certainty there is no uncertainty associated with that right. And as we illustrated over here when we looked at experiment number 3 and experiment number 4.

(Refer Slide Time: 25:12)



If there are 2 events whose outcomes are omega 1 and omega 2, such that omega 1 is greater than omega 2 this would imply that S of omega 1 is greater than S of omega 2. So, the more the number of outcomes you have more uncertainty you have in that experiment right. This is just intuitive knowledge of the phenomena right.

Now, let us now say that with these tools with this information that I have in my hand. Let us say I have 2 events with outcomes omega 1 and omega 2, and I do them simultaneously. So, the total number of outcomes are omega 1 times omega 2 right. For example, I can toss a coin and roll a die right.

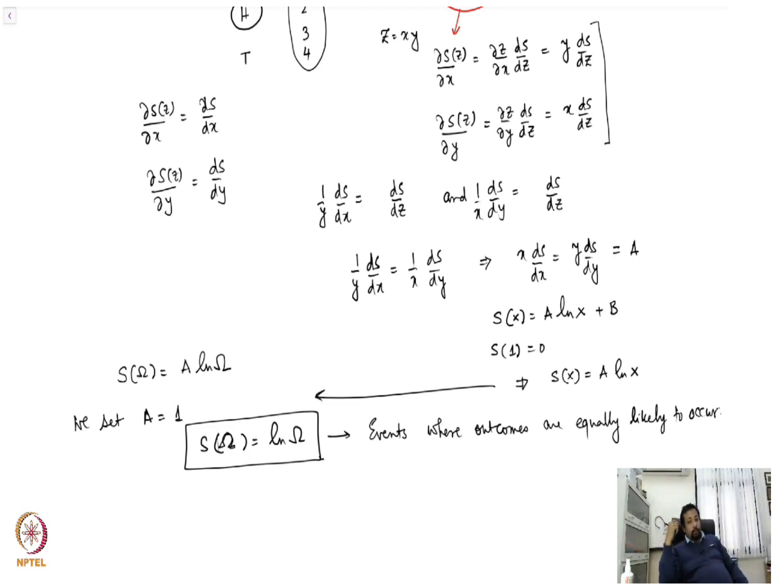
So, now I want to know what is S omega. Clearly if I know the outcome of one of the event, then the uncertainty in the other is not completely gone. For example, if I roll, if I toss a coin then I have head and tail and if I roll a die, I have four phase die. Let us say I have these 4 outcomes right ok. So, even if I know the outcome of the tossing, event tossing of the coin I still do not know the outcome of the rolling of the die there is uncertainty which is contained over here.

So, it follows that I have must be equal to S of omega 1 plus S of omega 2 right. Now, if I just look at this functional expression that I have written down over here based again on intuitive analysis, then it follows that there is a very definite functional form for S which will obey this. So, if I have x and y is equal to S of x plus S of y, then there is a very definite functional form of S. So, I want to know what that is that is not very difficult to do.

First of all we write z is equal to x y, right then del S z del x, which essentially means that I am just looking at the left hand side and trying to calculate these quantities is del z del x dS dz right. And del z del x is y dS dz. If I do the same thing and write down del x del S of del y, just looking at the left hand side then I know that this is del z del y dS dz which is equal to x dS dz right good. So, we will keep this in our hand, but now if you are operating del del x on S of z is like operating del del x on the right hand side also.

So, del of S z del of x by this expression that we have written down over here is dS d x and del of S z del of y is dS dy right. So, it follows that dS dx is equal to y dS dz and dS dy is equal to x dS dz. So, which means I can just write down dS dz.

(Refer Slide Time: 29:25)



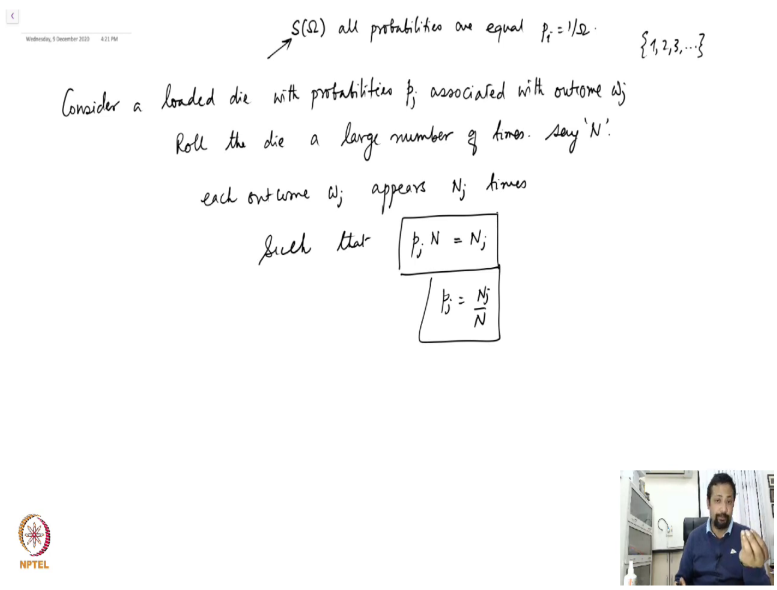
And put 1 over y over here and 1 over x over here. So, that 1 by y dS dx is equal to 1 over x dS dy, which would imply x dS dx is equal to y dS dy. Now, the left hand side and the right hand side the left hand side is a function of X the right hand side is a function of X and therefore, this can only be a constant which means that S of X can be A ln X right plus B.

Since S of 1 is equal to 0, the one that we written down over here that is omega is equal to 1 the uncertainty associated with it becomes vanishes, it follows this would imply that S of X is

A ln X, hence our measure for this entropy becomes A ln omega. And since A is arbitrary we set A to we set A equal to 1.

So, that our expression of this uncertainty becomes ln omega this is the expression for the uncertainty for events where outcomes are equally likely to occur. So that means, all the outcomes have equal A priority probability of 1 by omega right.

(Refer Slide Time: 31:25)



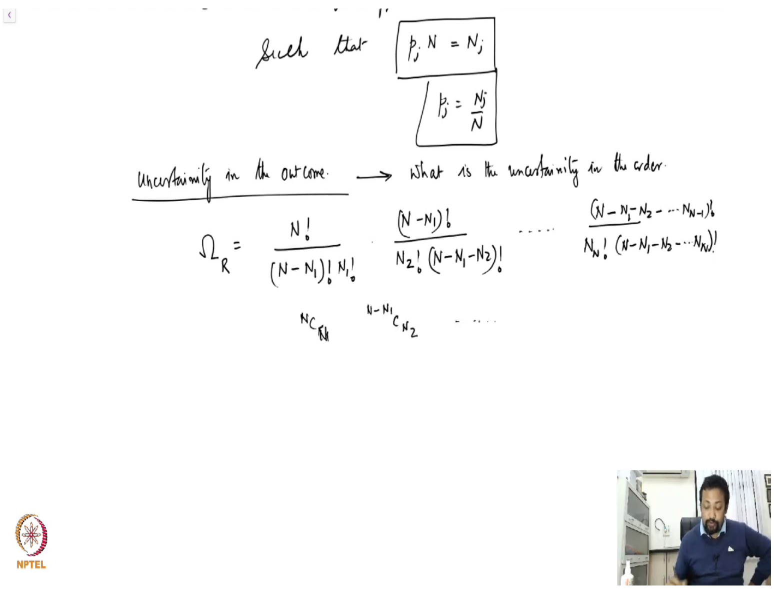
So, now that we have developed an expression for the uncertainty in the outcomes of an experiment, where all probabilities where all probabilities are equal are equal. And p i is 1 over omega right, but we want to now focus on the cases where my probabilities of the outcomes are different what happens to this S, how do I write down an S.

For example consider a loaded die essentially; that means, that a die which is not which is not fair, but a biased die, where probabilities with probabilities p j associated with outcome omega j right, which the set of omega j S are the numbers 1, 2, 3.

So, if there are N it is an investor you know that the there are N outcomes. Now, what we do is we roll this die a large number of times. So, roll the die a large number of times, say N right. So, what is going to happen? Then each outcome is going to appear, you have an outcome omega j appears, N j times such that you can p j times N is N j, this is what follows from the frequency interpretation of the probability or alternatively you can write down p j as N j over N we shall use this later right.

Now, the outcomes can occur in different order right. But, so, when you roll the die N number of times you can expect that the outcomes that come will come in different order. Therefore, the original uncertainty about rolling and a die for one single event when you are rolling the die only once has now, been transformed into the uncertainty of the order. However, all the orders are equally likely to appear.

(Refer Slide Time: 34:18)



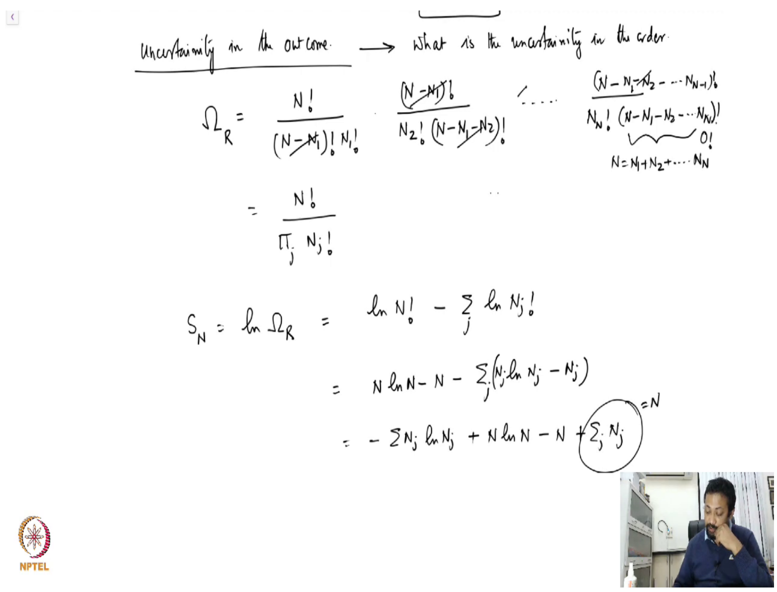
So, all the orders; so what you had started off is essentially asking the question that what is the uncertainty in the outcome right and that now, I have transformed into a different question that, what if I roll it N number of time what is the uncertainty in the order, because my outcomes can appear in different order and therefore, you expect that if you roll it say 10 times.

Then you are going to 100 times a, 1000 times you are going to get different order orders in which this is the outcomes are going to appear, but all therefore, your original as uncertainty about the outcome has now, transformed into the uncertainty in the order. But all the orders are equally likely to appear they appear with the same probability.

So, all you have to do is you have to figure out the total number of possibilities that you can have right and we will call this as omega R clearly N factorial N minus N 1 factorial, N 1 factorial times N 1 sorry this has to be N minus N 1 factorial, N 2 factorial N minus N 1 minus N 2 factorial all this way it is going to go.

So finally, you are going to have if there are N such events and such outcomes you are going to have, N minus N 1 minus N 2. And here you are going to have N 1 minus N 2 minus N minus factorial right. This is like N C N 1 and once you have chosen N C N 1 number of balls you are left out with N minus N 1 of them and then you choose N minus N 1 C N 2, so on and so forth right.

(Refer Slide Time: 36:41)

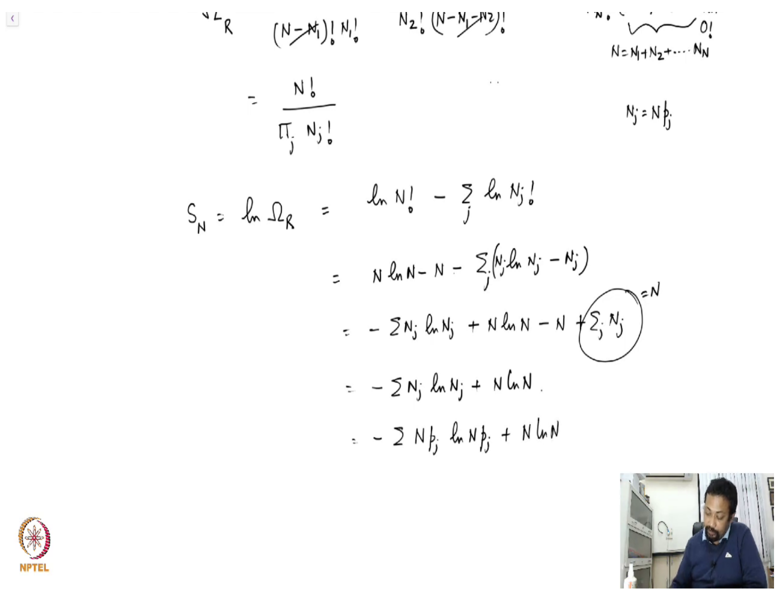


So, it follows that this is going to cancel with this, this is going to cancel with the next term and this is going to cancel with the preceding term and N is N 1 plus N 2 and N. So, that this is essentially a 0 factorial. So, you are going to be left out with product over N product over j N j factorial right, these are the total possible orders that you can have right.

Think about this is just a combinatorics problem it is not very difficult to understand, but all of these orders are equally likely to appear right. So, therefore the associated uncertainty we just calculated is omega R we will just put a subscript N here to say that this is not actually for a single event, but for a N rolls of the die correct.

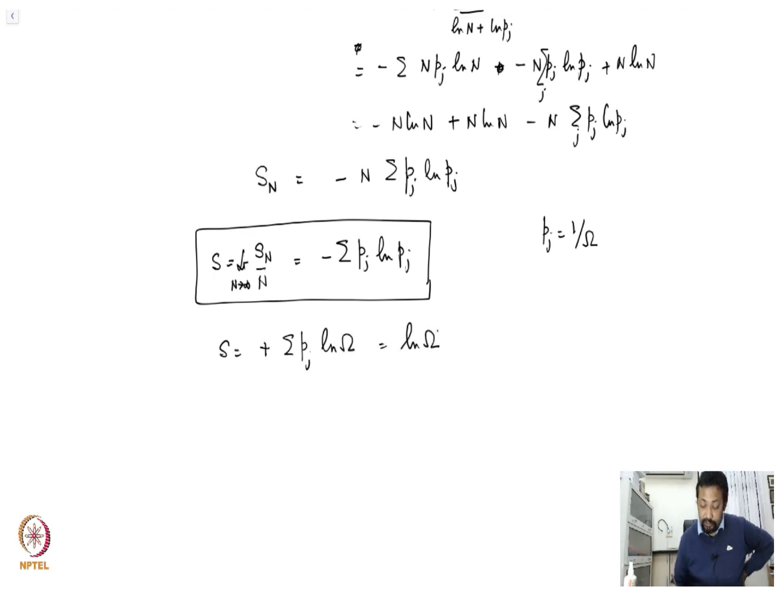
So, therefore this becomes ln N factorial minus sum over j ln N j factorial right. And if I use sterlings approximation then this is N ln N minus N minus sum over j N j ln N j minus N j right. So, I am going to have minus N j ln N j. I am going to have plus N ln N minus N minus sum over j N j and this is equal to. So, this is going to be plus and this is going to be N.

(Refer Slide Time: 38:23)



So, that you are going to be left out with minus sum over N j ln N j plus N ln N right N j is N times p j. So, I can just. So, let us N times p j ln N times p j plus N ln N. Now, you can immediately see where this expression is going to go.

(Refer Slide Time: 39:02)

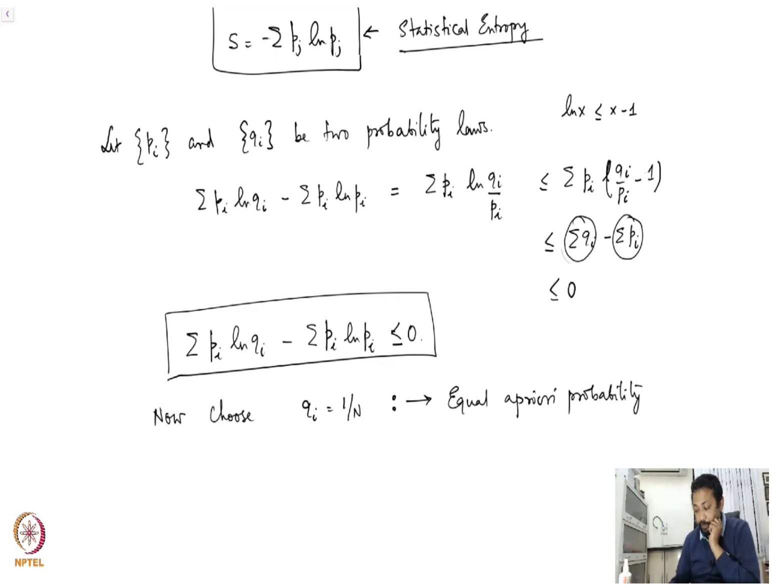


I can open up the log here to write down ln plus ln p j. So, that the first term is going to be minus N p j ln N plus sorry this is minus N p j ln p j with the sum of j plus N ln N, this becomes minus N ln N sum over j p j is 1, the normalization condition and you have a plus N ln N and you left out with minus j this last term.

So, therefore, your S N is minus N sum over p j ln p j right. It follows that right now, that I know that the uncertainty associated with N rolls of the die is this, what is the uncertainty associated with one roll of a die that becomes S N over N? The original answer that we are looking for except one has to take care that N is very large and you finally, come up with the answer p j ln p j.

So, this is the measure of your uncertainty when the probabilities for the outcomes are all different. If I now want to see whether I am writing doing that, I can always take p j as omega. So, that S becomes minus we will keep sum of a p j here ln omega which is ln omega. So, we recovered the earlier result for equal a priori probability that we had derived before.

(Refer Slide Time: 40:48)



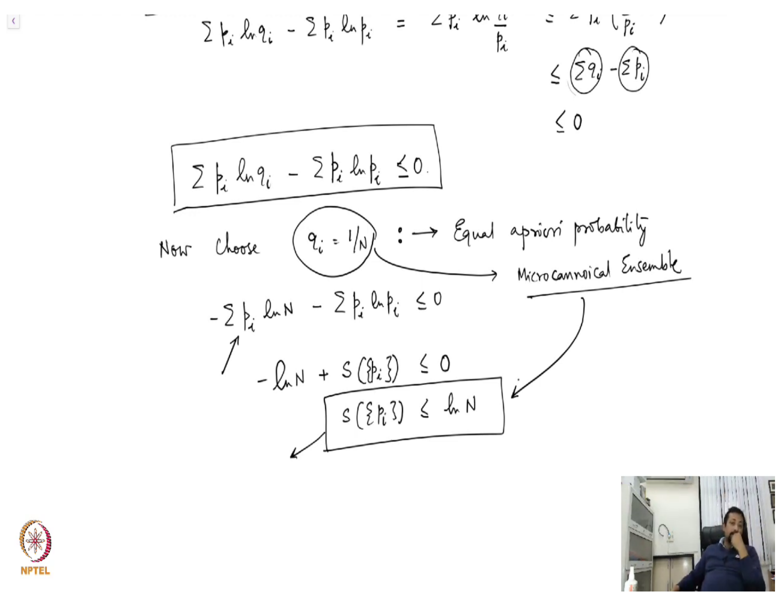
So, the general expression for an uncertainty associated with the outcomes of this is minus p j ln p j right. As we shall see that this one has a very very deep connection to statistical mechanics right. Now, the statistical entropy; so, what we call this is we call this as the statistical entropy and this statistical entropy has certain characteristics right.

So, first let us say let p i and q i be two probability laws right. So, then we consider p i ln q i minus sum of a p i ln p i correct. So, this quantity Now, I can very nicely write down this as sum over ln q i over p i you will see in a moment why I judiciously chose my left hand side it is essentially I am motivated towards a certain thing I want to show. So, p i ln q i.

Now, I know that ln X is less than equal to X minus 1. Therefore, this sum is less than equal to sum of a p i X minus 1 would be q i over b i minus 1 which is sum over q i minus sum over p i, but p i and q i are two probability laws. Therefore, they must satisfy normalization therefore, the both the sums that you see over here this one and this one are 1.

So, that this is less than equal to 0. Therefore, it follows from this little derivation that we did that p i ln q i minus p i ln p i is less than equal to 0 right, for two probability set of laws. Now, choose q i as 1 by N which means that this corresponds to equal apriori probability right.

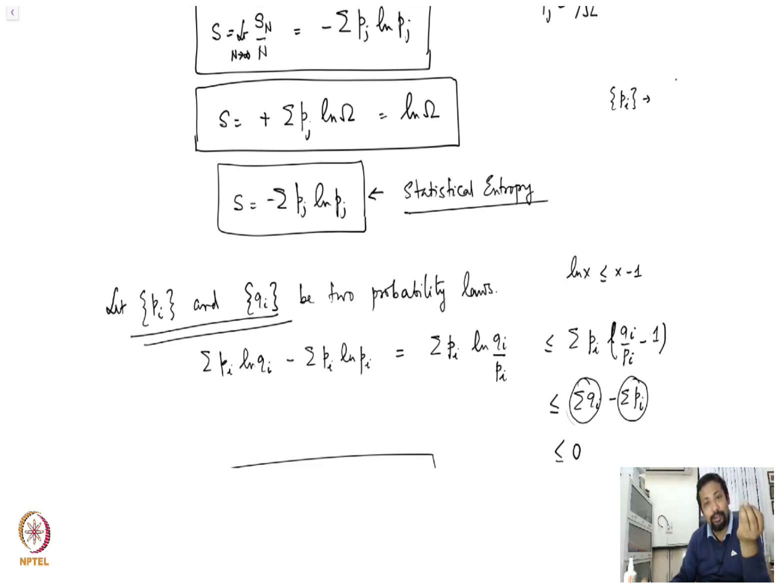
(Refer Slide Time: 43:34)



If that is the case then I have sum over p i minus ln N minus sum over p i ln p i is less than or equal to 0, but sum over p i is 1 in this particular case. Therefore, I have minus ln N plus S of p i the entropy the statistical entropy associated with the probabilities p i must be less than 0, which means S of p i must be less than ln N.

Therefore, there is an upper bound to the statistical entropy that we have calculated and the upper bound corresponds to the case where all probabilities are equal, all the outcomes are equally likely. So, for a fixed, but one has to be careful that the number of events or the outcomes are the same right. So, which means for both these probabilities sets for both these probability sets the number of outcomes are same.

(Refer Slide Time: 44:47)



So, for example, one can think of let p i correspond to the probabilities associated with the rolling of a biased die, it can be enfaced. And q i corresponds to the rolling of a unbiased fair die right. So, in this case q i is 1 by N and that gives you the upper bound of the statistical entropy.

As we learn thermodynamics we will see that this particular choice is what is called a micro canonical and symbol right. So, we and when we do micro canonical and symbol we shall use this expression for the statistical entropy to build on the thermodynamics.