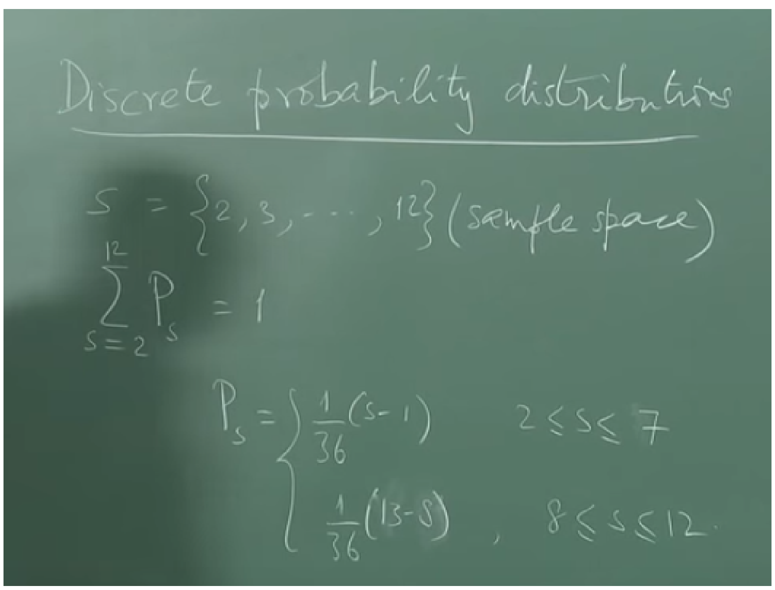
Physical Applications of Stochastic Processes
Prof. V. Balakrishnan
Department of Physics
Indian Institute of Technology-Madras

Lecture - 01
Discrete Probability Distributions (Part 1)

Alright, we will begin this course on stochastic processes by little quick introduction to discrete probability distributions. I will initially talk about probability distributions both for variables which take on a discrete set of values as well as continuous random variables and eventually work our way to random processes of stochastic processes where you have a random variable which varies in time, in continuous time for instance.

So let us start with the discrete probability distributions and I am assuming that you have, I assume that you have some elementary, some familiarity with the elementary notions of probability and statistics, but let us do this very quickly, recapitulate some salient features, the ones that are going to be of interest to us in our future work.

(Refer Slide Time: 00:55)



Let us start with the simplest of these distributions. Let us take a pair of dice and toss these dice and ask what kind of distribution you get for the score from the pair of dice. So each of these we assume these are fair dice and each of them has number on the top face running from 1 to 6 and you have 2 of these dice, you toss it once and you ask what is the probability of receiving various scores and so on.

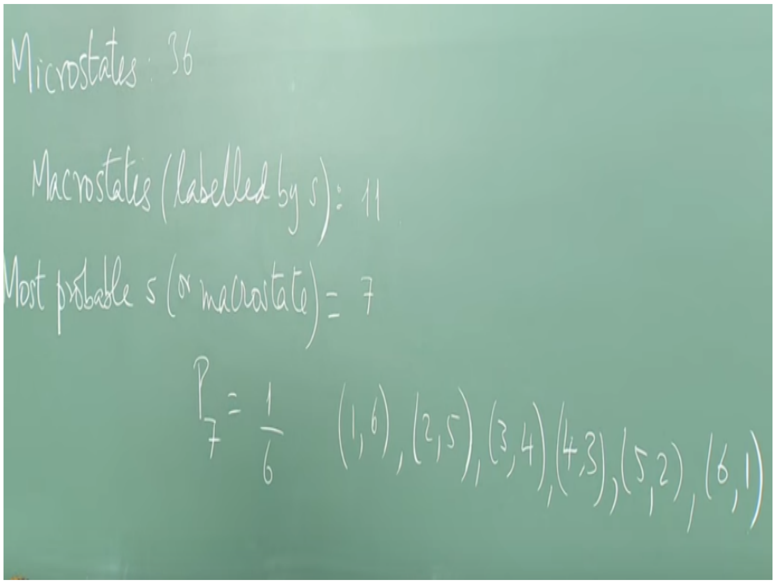
So let us call this random variable the score s, which is the s of the numbers that you see on this pair of dice, each of them can have values, each dice can give you a value from 1 to 6 with equal probability okay. Then what is the sample space of this sum s, the score s. Clearly, the least value you get is 2 and the largest value you get is 12.

So this is the set of integers 2, 3, up to 12, that is the sample space and then you can ask what is the probability of getting a particular score and this probability if I denote it by P sub s, it is clear that it will sum from s = 2 to 12, this must be equal to 1. That is the conservation of total probability okay. So each P sub s is a number running a positive number running from starting from between 0 and 1 and s runs from 2 to 12 here.

Now what is this equal to? By the way what is the, this is a very simple experiment. What is the most probable score to get? It is 7, the most probable score is 7 and what is the probability of getting a 7? Well, this is figured out by saying what is the probability of getting any score at all to start with? What is the probability that the first die gives you say 3 and the second die gives you 4, what is the probability of it?

It is one sixth for the first one, one sixth for the second one, so it is 1 over 36 right? How many such possibilities, distinct possibilities are there. Here is the die on the left, here is the die on the right and how many distinct possibilities are there? This can take 6 possibilities and the other one can take 6 possibilities. So there are 36 possibilities in all right. Every time you toss this die, there are 36 possible sequences that you can get assuming that this die and that die are distinguishable completely.

(Refer Slide Time: 04:18)



So a set of microstates, I am going to use the language of physics here, of statistical physics. There are 36 microstates in all, 6 for each of these and they are independent objects. So we really have 36 microstates. That is the detail configuration of the system as a whole. What is happening to die 1, what is happening to die 2; completely specified 36 such microstates and the probability of any one of them if they are all equally probable because the die is completely, the dice are fair dice, is 1 over 36 okay.

How many macrostates are there? By macrostate, I do not ask the detailed information about the system, but I only ask for what is the score, what is the total score. So s is going to label the macrostates, labeled by s. How many macrostates are there? There are 11 macrostates here right. Most probable, what is the most probable score? 7 is the most probable score and why is that happening. The most number of accessible microstates. Each microstate is equally probable.

This sort of reminds you of the microcanonical ensemble in statistical mechanics where you have a whole set of microstates, accessible microstates, and you postulate that they are equally probable. Then the most probable macrostate is going to be the one that gets contributions from the largest number of microstates okay and this is what is the score. By the way what is P 7 equal to? It is one sixth because the number of, what are the contributing microstates to this, well certainly (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1).

There are 6 possible microstates which contribute to this. So 6 times 1 over 36 gives you one sixth here. What is the formula for P sub 7, what does it look like? Well as you can see that if s takes on its extreme value 2, there is only 1 contributing microstate, 1 and 1 and similarly for 12, it is 6 and 6 and in between it increases and decreases once again once the peak is reached okay.

So it is a very typical behaviour this guy and this is 1 over 36 into s - 1 as long as s is less than equal to 7, it increases and what is it once you exceed the number 7, what is going to happen? It is going to be a downward thing. So it is going to be something minus s, it is 1 over 36, 13 - s. you just have to figure out that at 7 it is going to be one sixth. This is s not 5 okay. So we have a complete distribution of this whole thing.

We can compute all sorts of things, the variants, the mean, the higher moments of this distribution etc., etc. okay? So this is an utterly trivial example, completely trivial example, very elementary example. The only assumption is that all the microstates are equally probable, every one of them is equally probable and the uniform distribution in the microstates still leads to a non-uniform distribution. This is not a constant.

It goes up and comes down in the macro variable here, macroscopic variable here. This is at the root of what is happening around us in statistical mechanics when you discover that system sit in states the most probable macrostate okay even though the microstates are equally accessible or equally probable, the system will sit in the most probable macrostate and there is such a steady equilibrium state in many cases simply because the number of contributing microstates is the largest, exponentially so in many cases as we will see in a simple coin tossing example here.

Now that we have this let us go to a slightly less trivial example. Oh just one question, an incidental question. We assumed these 2 dice were distinguishable objects, were distinguishable objects. Now I would like to constantly come back to the physical application the physics of it and make a make connection with what you know from statistical mechanics. So suppose these 2 dice were made so perfectly that they are indistinguishable from each other.

Incidentally in this problem here what is the probability of getting 3, a score of 3? It is 1 over 18 because you could have done it with 1, 2 or 2, 1; so it is 1 over 18. Now I ask, I make these 2 dice absolutely indistinguishable from each other. What is the probability of getting 3? These 2 die look dice look exactly the same, absolutely alike. Is it still 1 over 18? Would you say it is still 1 over, after all when you toss a pair of dice, they are supposed to be fair dice, they are supposed to be indistinguishable in any case right? So if you take these 2 classical dice and toss them even if they look exactly alike, you can certainly see which one is which and the probability is still 1 over 18 .

Now suppose these 2 dice were quantum particles, identical particles. Both of them were spin one particles, nice identical particles, would the probability still be 1 over 18? If they were indistinguishable in the quantum sense, would it still be 1 over 18? What do you think? They are completely indistinguishable right and then this obey Bose statistics and in the Bose statistics way of counting this is certainly not 1 over 18 because you cannot distinguish a state in which one of them is in level 1 and the other is in level 2 and the situation is reversed.

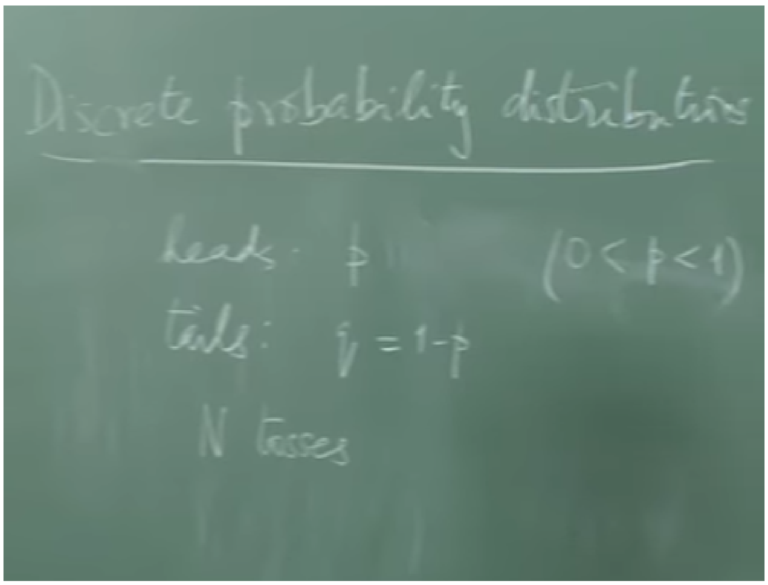
These 2 really form, they are completely part of one state, you cannot distinguish in this fashion. So what is the new feature in quantum physics that has happened. At what stage would you say that these 2 are indistinguishable in the quantum sense and therefore the counting of the number of independent states is different from what it is in classical physics. After all we know from quantum statistics that is true, you do not count in the same way as you do in classical objects right.

Even if these classical objects are completely indistinguishable perfectly, you still cannot put them in what is called an entangled quantum state. On the other hand potentially if the 2 particles are absolutely identical you can put that system in a quantum mechanically entangled state and once you do that once that possibility exists you have to change the rules of counting. It is no longer what it was earlier.

So the short answer to this question of when you use quantum statistics has in part to do with the fact that you can put these indistinguishable objects in what is called an entangled state which you cannot do with classical objects okay and that is the reason why even if the dice are indistinguishable from each other absolutely to whatever degree of perfection you still have exactly classical counting in case of the dice, but in quantum particles you have to change the rules okay, a very profound reason here.

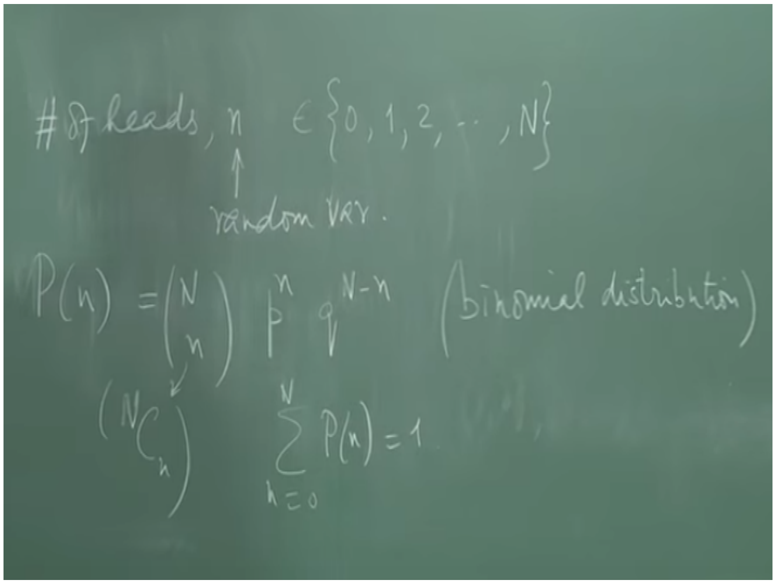
We will come back to this when we talk a little bit about quantum statistics okay. Let us go on now to the first nontrivial distribution which again let us do with the help of an example. Let us take a fair coin and toss this coin and look at the statistics of the coin tosses.

(Refer Slide Time: 13:57)



In fact let us make it a little more general and say the coin is such that the probability of heads is p and tails is q equal 1 - p and p is some number between 0 and 1. I am going to toss this coin 10 times and that is my experiment okay. So I take this coin and toss it N times, lay them all out or take N coins which are identical and toss them and lay them out and ask for the total number of heads okay and that is my random variable, the total number of heads, and I must talk about the probability space, the sample space of this random variable.

(Refer Slide Time: 14:45)



This is my random variable and what is its sample space? Ya it is 0, 1, 2, all the way up to N. There are n + 1 possibilities and they label my macrostate, any given value of little n labels my macrostate. In a given toss, the probability of success if I say heads is success is little p and that of failure is little q which is 1 - p okay. I ask now for the probability that this random variable takes on one value from its sample space.

Now I should be a little careful here. In mathematics you would distinguish between the random variable per se, the symbol for the random variable and the symbol for one of its values from its sample space. I am going to, unless we need it, I'm going to avoid this distinction. I am going to use the same symbol for both. You have to be a little careful here, but in the context it is very clear what I mean.

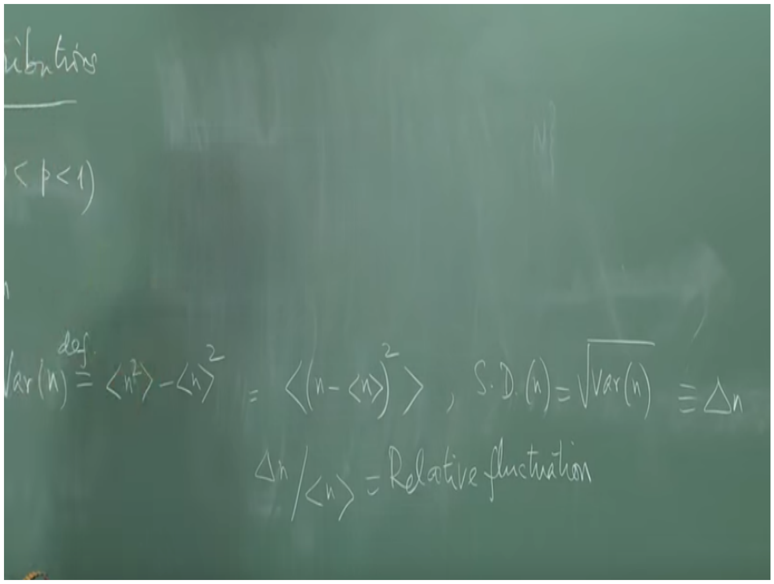
So what is the probability that the random variable takes on one of these values in this here. What is the P of n. These tosses are independent of each other and the coins are all exactly the same okay. They are all identical. So what is the probability of getting n heads? Well, first of all I must get exactly n heads and they are all independent tosses, so this must be proportional to p to the power n and the rest of the tosses must give me tails because I want exactly n heads.

So rest of it is q to the power N - n and that is the probability of getting any one sequence of letters capital N letters such that little n of them are Hs and the rest of them are tails Ts okay but I do not care about the order in which this these heads are achieved at all. So I must add now all the number of microstates for each of these. I must multiply this by the number of microstates which is just N n. This is this stands for N C n, the binomial the combinatorial coefficient okay.

So this is my probability distribution and for obvious reasons it is called the binomial distribution because this thing here is just the coefficient of little p to the power of n in the binomial expansion of p + q to the power of capital N. The questions of interest are things like what is the average number, what is the average value of little n? What is the variance in this number? What is the scatter? What is the actual, what are the various other properties of this distribution, various factorial moments and so on and so forth, all of which can be deduced from this. It is by the way a trivial matter to verify that summation n = 0 to capital N = 1.

That follows immediately because it is just equal to p + q to the power capital N but p + q is unity. So it is normalized, this probability distribution is normalized. What is the average number? Incidentally, this is called a Bernoulli trial, a trial in which the probability of success is some little p and failure is 1 - p and we are doing n Bernoulli trials okay and the result for this little n this random variable the number of successes is a binomial distribution.

(Refer Slide Time: 19:00)



What we need are quantities like the average value which I will denote with angular brackets. This is the mean value or average value, mean value of n. We are going to look for things like the variance of n. What is the definition of the variance of this variable? How is the variance defined? It is a square the value of the mean square of the square of n minus the square of the mean value square right.

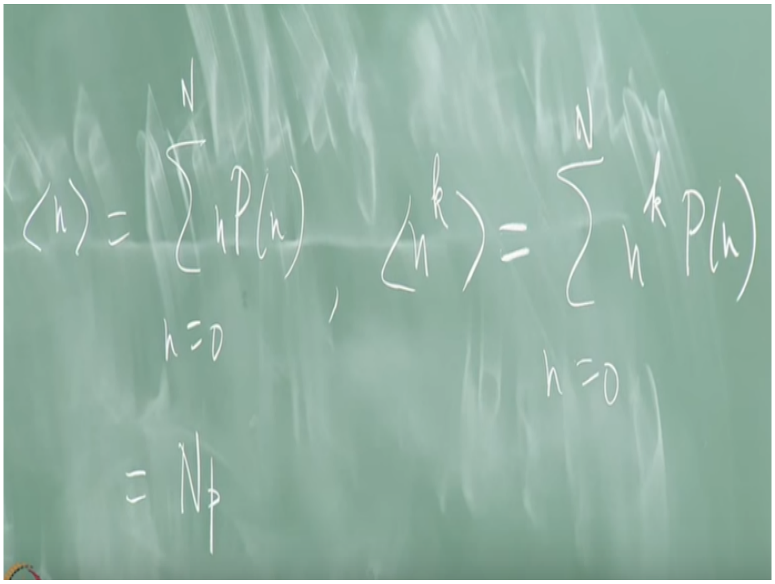
Mean square minus the square of the mean, that is the value, that is the definition of the variance. Can this be negative? Can the variance be negative? No, how do we know this? Is there a way of rewriting this variance so that we know for sure that it is, cannot be negative. Ya, so we can also rewrite this as equal to you start with n minus the mean value so the deviation from the mean is this quantity here.

Square it so it becomes positive and then take its average and it is a trivial matter to verify that this is equal to that okay. So can the variance be negative? Can it be 0, yes. There is only one condition under which it can be 0, only one possibility and that is if the random variable always has the value equal to its mean which means that it is a sure variable, it is no longer a random variable. So the moment you have a random variable which can take on more than one possible value you have a variance which cannot be 0, it is got to be positive okay.

The standard deviation is defined as the square root, the positive square root of the variance and it is sometimes denoted by delta n. So the standard deviation sort of measures in some specific sense the scatter of this variable about its mean value and then of course you would naturally like to know how big it is relative to the mean so delta n/n is called the relative fluctuation.

It has got other names but I like this name relative fluctuation and incidentally if n has physical dimensions then this quantity this relative fluctuation ensures that it is dimensionless, it is an absolute number. We would like to find out what all these quantities are for this binomial distribution. We would like to find out what these various quantities are. So what is the easiest way to do that? First we got to define them right? We have to define what I mean by the mean etc. in terms of the distribution.

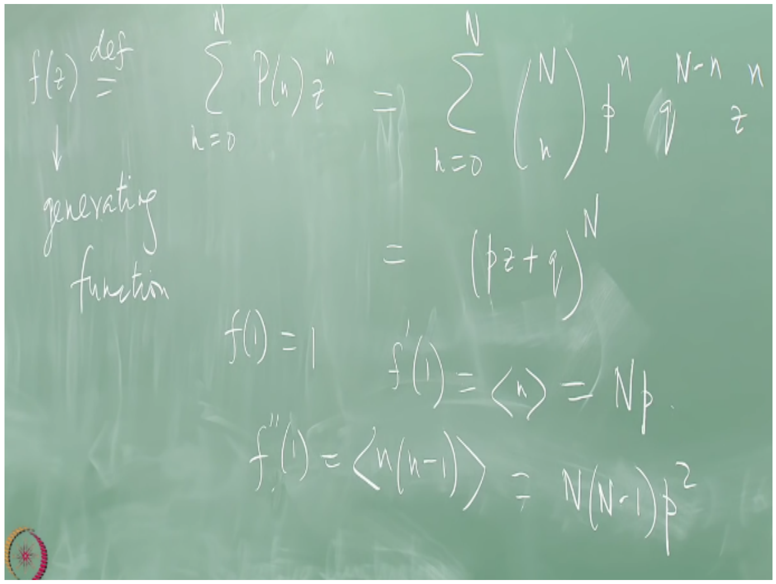
(Refer Slide Time: 22:42)



So n = n P of n because it is a normalized probability distribution so these numbers are all less than 1, positive numbers and this is the weighted s and n square of course n to any power k equal to summation n to the k over the sample space. What is the mean value in the binomial distribution. Ya capital N times p.

So it is not hard to show that this for the binomial distribution thus becomes equal to N p. Now the quickest way to do this is to use what is called a generating function that the moment you have a probability distribution whether it is continuous variable or discrete variable does not matter, we define a suitable generating function something which will generate all the moments for you.

(Refer Slide Time: 23:57)



So let us define an f of z definition and this is the generating function. This thing here equal to a summation over the sample space. P of n z to the power n. So I define a variable z, a continuous variable which could be complex. In general it will be complex, it does not matter because then it has got nice analytical properties, the function of z, multiply P of n by z to the power n and then you have an f of z.

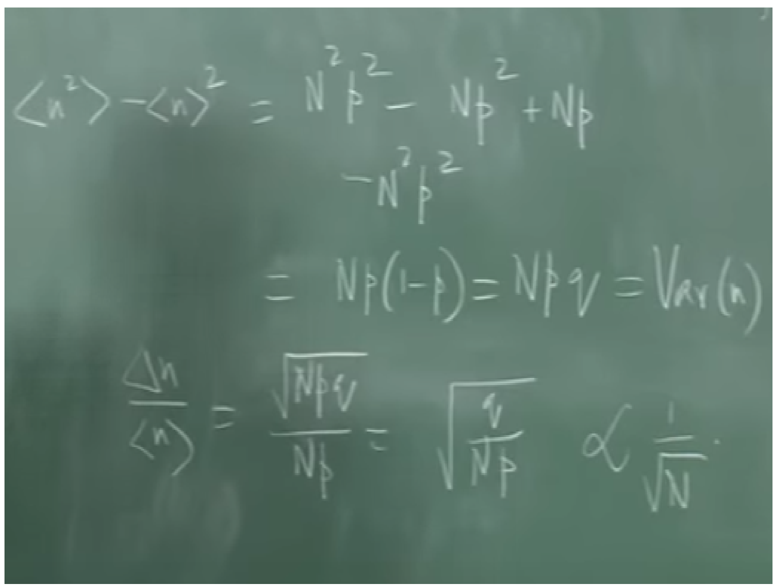
So you can identify P of n as the coefficient of z to the power of n in this function f of z, in the expansion of f of z in a power series in z and this is utterly trivial for the problem we have at hand because this is equal to summation n equal 0 to N; P of n is of course N n p to the power n, q to the power N - n z to the power n and what is this equal to? Ya it is a binomial expansion. So this is trivial. It is p z + q to the power N okay. What should f of 1 be?

What should f of 1 be? Should be unity because f of 1 if you said z = 1 is just the s over all the probability so it should be equal to 1. Is it equal to 1 here? Yes indeed because p + q is 1. So this is satisfied, the condition is, normalization condition is satisfied. What is f prime of 1 equal to? Go back to this formula, if I differentiate this with respect to z then I get n P of n z to the power n - 1 and then I said z = 1.

So I get n times P of n which is just the average value right. So it is clear this is equal to n and what is that equal to if I put it in here this closed form, n times p. What happens if I differentiate a second time? I find the second derivative of f with respect to z and then z = 1. Well the first time I am going to pull down an n, the second time I pull down an n - 1 and then I said z = 1, it is the average value of n times (n – 1).

So it is immediately clear that f double prime of 1 = n times (n – 1) and what is that equal to? Well I go back here, this stage. Differentiate once I get N and then a p outside, N p times this fellow to the power N and (n – 1)of x and that goes to 1. I differentiate a second time and then what do I get okay?

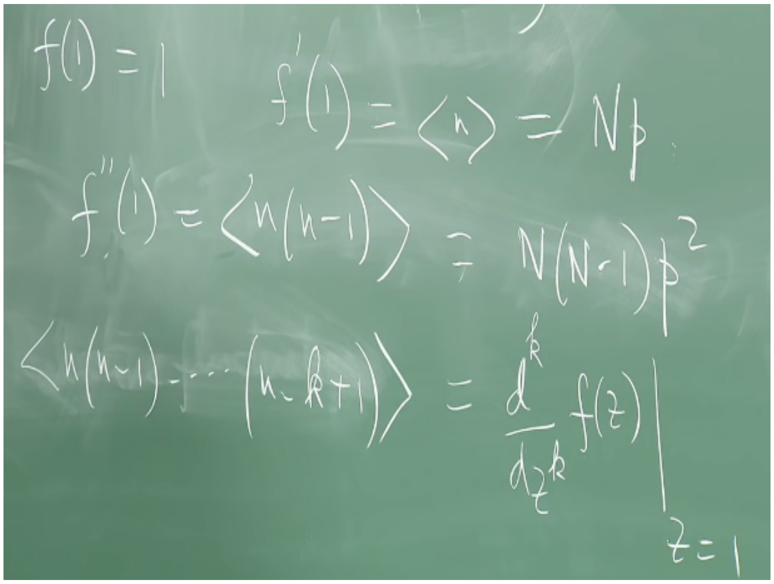
(Refer Slide Time: 28:12)



So we have a first interesting result which says that the average value of n square minus the average value of n that is what this is, that is equal to N square p square - N p square. So the mean square value N square is this plus N p which is the mean value and if I subtract the mean value square it is this minus square p square which is equal to this and that which is N p times 1 - p which is Npq and that is the variance of n.

It is just Npq okay and what is the standard deviation, square root of Npq. So what is the relative fluctuation? Okay. So the important thing is that this is proportional to 1 over the square root of N. So the larger the number the less the scatter about the mean relatively, relative to the mean okay. This is a crucial fact, so very crucial fact. Let us look at a simple physical example just to see where this is going to get us okay. So is this clear the way it comes about.

(Refer Slide Time: 29:48)

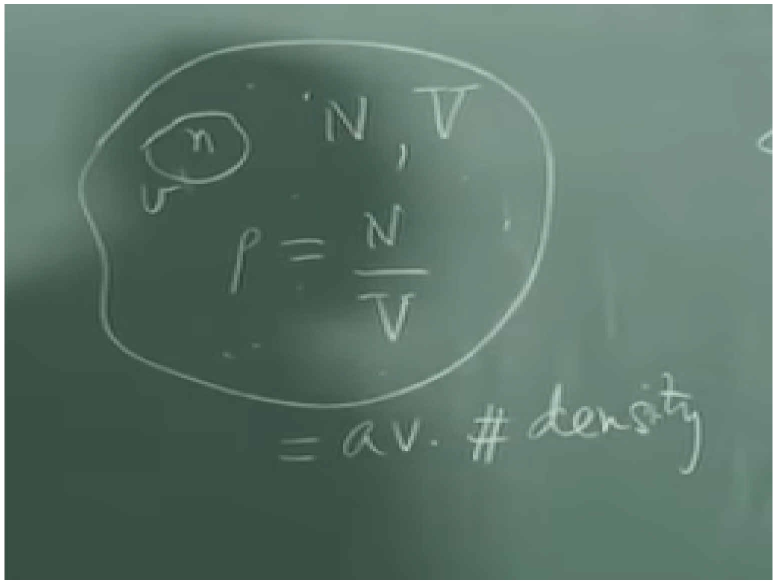


By the way you can generalize this a little bit what would this be n times n - 1, n - k + 1. Ya this would be the derivative. This is equal to d k over d z to the k of f of z at z = 1. It is called the kth factorial moment. Now once you have the kth factorial moment you can with simple algebra write down the kth moment itself okay. The higher moments themselves are not that useful.

Just as in the case of the square the second moment what was much more physically relevant was the square, the average here minus this guy here like the variance. So we need to find similar subtractions for the higher moments and what are those things called when you subtract out all the "irrelevant parts" of the lower moments in the kth moment, what is the higher analog of the variance.

These are called the cumulants, these are called the cumulants of the distribution and we will come to it. It is a very important notion, a very important reason why you have cumulants are more useful than the moments themselves and we will come to that, a very very important significant property of the cumulants which is not shared by the moments. We will see what that is okay. So let us give a physical example of this I think here.

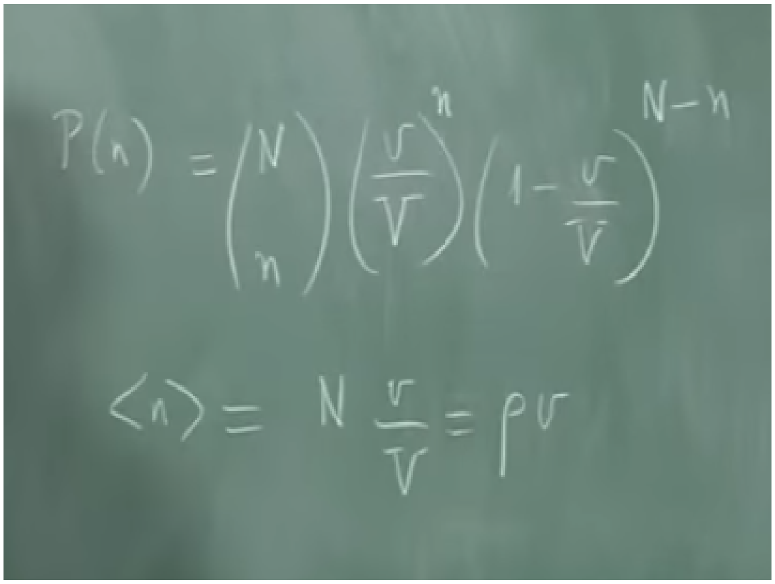
(Refer Slide Time: 31:24)



Let us assume that we have classical ideal gas in a big volume V and there are N particles in a big volume like in this room. Let us assume it is closed and it is in thermal equilibrium okay classical particles. Then there is an average number density rho which is N over V. That is equal to the average number density.

So you have these particles moving about and then there is an average number density in this. Now I ask the following question. Suppose I take a small volume v, smaller than this capital V, little v and focus on just the sub volume. Of course the number in this sub volume is changing. It is changing extremely rapidly. There are rapid fluctuations. But suppose I take an instantaneous snapshot and ask at any given instant of time, what is the probability that there are exactly n particles inside here.

(Refer Slide Time: 32:36)



So I ask for the probability little n P if little n that there are exactly little n particles inside this sub volume little v at any given instant of time. We are going to assume the whole thing is in equilibrium, so the statistical properties do not change with time. What is this going to be and all the particles are assumed to move independent of each other completely. So it is like a whole lot of Bernoulli trials.

The probability that any given molecule is in this sub volume is just the ratio of this volume to the total volume, completely unbiased, uniform distribution in space. So it is equal to little v over capital V and you want n of them. So that is the probability that n of them are in there right multiplied by the rest of them have to be outside which is 1 - v over capital V to the power N - n out here and you do not care which of the molecules is inside and which of them is outside.

So N and little n runs from 0 to capital N exactly as in the binomial distribution okay. This is therefore a binomial distribution, parameterized by capital N and of course by this ratio. So it is a Bernoulli trial, set of Bernoulli trials in which the probability of success is this and the probability of failure is this out here okay and what is this going to be, average value?

It is n times p which is N times little v over capital v but that is precisely rho v and that is exactly what you would expect because if this is the average number density then the average number of particles in this little volume v is just that number density multiplied by the volume, it is precisely this. So now you can ask what is the probability of fluctuation of a given magnitude from the mean and we already saw that the standard deviation to the mean, the relative fluctuation is proportional to 1 over the square root of capital N and if you have 10 to the 24 particles in this room then of course it is 1 part in 10 to the 12 which is exceedingly small.

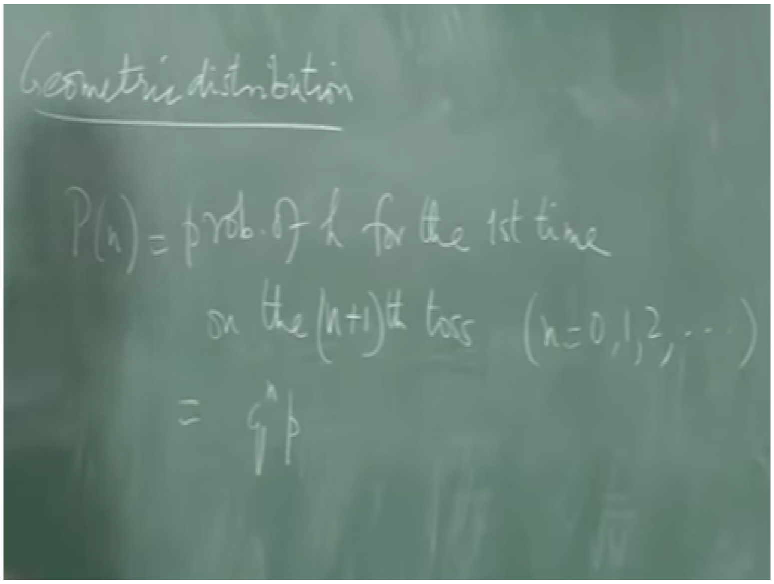
So this is the reason why we do not all suffocate because there is a finite probability that all the molecules will spontaneously congregate in one small sub volume at some instant of time causing us to suffocate but that is extremely unlikely. The probability is vanishingly small. The probability of a scatter about the mean significant scatter become smaller and smaller as the number N increases, capital N increases.

When it goes to Avogadro's number, it is just out of this world completely and that is why thermodynamics works because thermodynamics is a science of averages and we are interested very often in fluctuations about these average and generally in a thermodynamically stable state the fluctuation about the mean the relative fluctuation is of the order of 1 over the square root of the number of degrees of freedom which is already of the order of Avogadro's number okay and that is the reason why it works.

Why you can get away with just the averages without looking at fluctuations except when the system goes near phase transitions or when critical phenomena start playing a role then the fluctuations get bigger and bigger till they become as big as the system itself in some sense. Then the variance, the relative fluctuation, the standard deviation of physical quantities can become as big as the mean itself and then you are in trouble okay.

Then you have a failure of this very robust 1 over square root of n formula and then of course you have to look at it afresh. So that was our simplest example of distribution which is binomial distribution. Now let us look at another distribution which is related to this. We are going to look at number fluctuations and other kinds of gasses also. This is just one kind of classical gas. We will look at it in other gases as well. But before that let us look at another problem which is a related distribution and this is the geometric distribution.

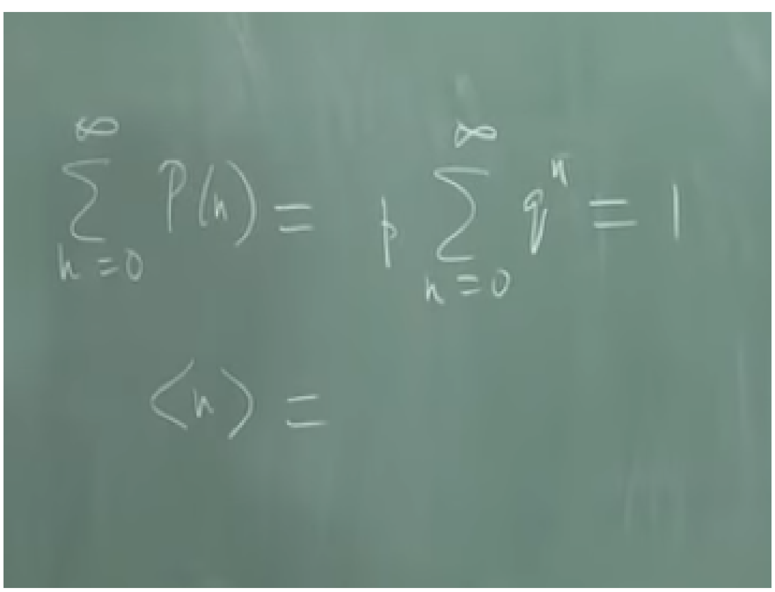
(Refer Slide Time: 37:08)



I go back to my coin toss. I take a single coin, the probability of a head is little p and tail is 1 minus little p and I keep tossing it and I ask what is the probability that I get a head for the first time on the n + 1th toss okay. So probability, I made it n + 1 so that this n could start running from 0, 1, 2, 3, etc. I have a first toss, second toss and so on. The first toss corresponds to n equal to 0 etc.

So what would this be? Well clearly for n of these tosses I must have failure and the probability of that is q each time. So this is equal to q to the power n and the last one must be a success and is just that okay. Is this probability normalized? We need to check that.

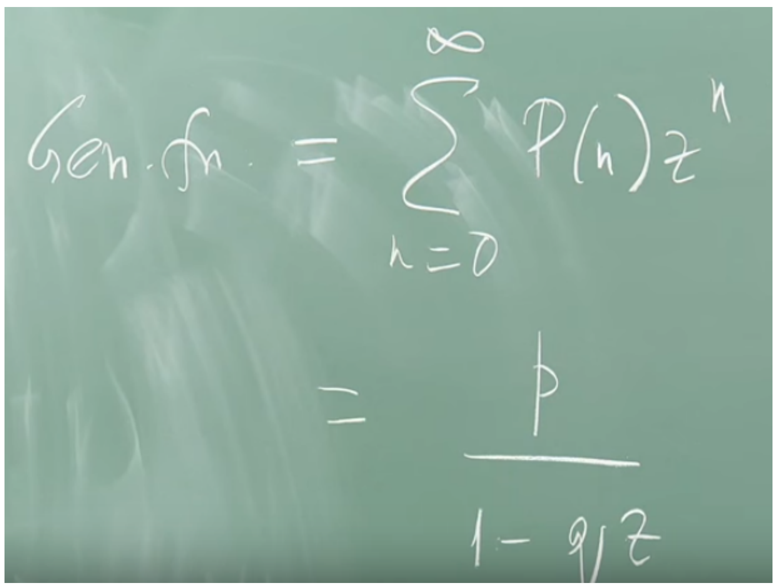
(Refer Slide Time: 38:52)



What is the sample space of n, 0 to infinity because you could have a real bad luck if it keep getting going there so you must have summation n = 0 to infinity P of n equal to summation p comes out n = 0 to infinity q to the power n which is of course 1. That is a geometric series, 1 over 1 - q which is p and the p cancels and you get 1 okay.

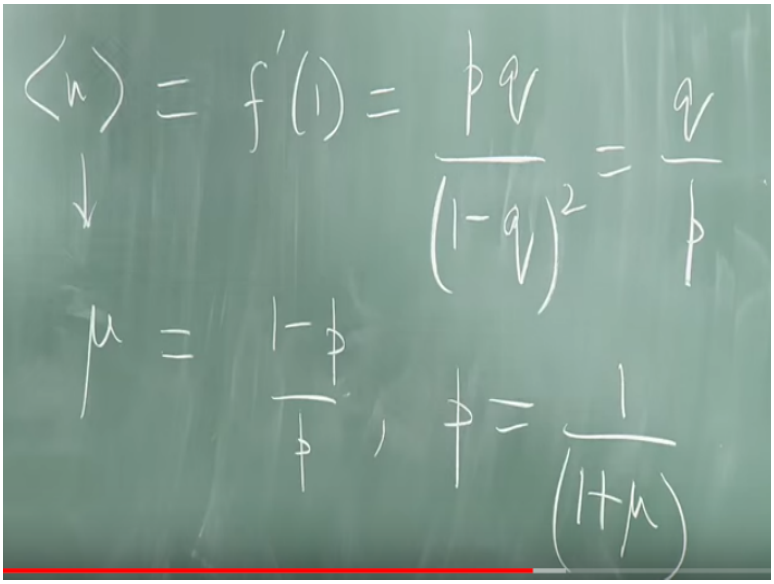
What is the mean value? What is the mean value of n? Oh by the way, you know why this is called a geometric distribution, this is a geometric sequence, just geometric sequence for various ends. Now what is the mean value? What does this give us? Well as before, the obvious thing to do is to find a generating function, then it becomes completely simple right? So let us do that.

(Refer Slide Time: 39:48)



The generating function summation n equal to 0 to infinity, P of n, z to the power n and what is this equal to in this problem it is simplicity itself. Just multiply this by z to n and s over n and then it is p over 1 - q z. It is just a binomial expansion. So it is p over 1 - q z. First thing you do is to check if f of 1 is 1. Is f of 1 1, z equal to 1 and you get q over 1 - q, p over 1 - q which is p.

(Refer Slide Time: 40:47)



So therefore it is normalized, no problem and what is the derivative at the origin? All you got to do is to take the derivative at z equal to 1 but you get, differentiate with respect z which is 1 - qz whole square, z z equal to 1and then. So it is q over p, the average number in this case let us be careful. That is not a great thing . Was my derivation right? This got normalized. So this is okay. If I differentiate I get this 1 - qz whole square and then a pq up there; 1 - q whole square and it is q over p, yes okay.

Now what happens where you can that it is a reasonable answer because if the probability of a head become smaller and smaller, little p become smaller and smaller, the average number when you wait for is going to get is going to increase till p goes to 0 to go to infinity okay and in the converse limit when p is equal to 1 and q is 0 then the first toss because n equal to 0 corresponds to first toss.

You are going to get a head straightaway. So that is perfectly alright as it stands and what about the variance and so on and so forth. You can write this down explicitly right. So one way to do this, to parameterize it slightly differently instead of this n let me call it mu the mean that is the usual symbol for the mean of random variable, let us call it mu and that is equal to 1 - p over p or p is equal to 1 over 1 + mu.

So another way of writing this geometric distribution is to write it as 1 over 1 + mu and then q, q is 1 - p, so it is mu over 1 + mu. So we can write this as mu over 1 + mu to the power n. So that is another way of parameterizing the geometric distribution. So once you identify this form you can read off what mu is, what the average value is. It is a trivial exercise by differentiating a second time and so on, finding out what the mean is, what the variance is and so on and so forth. I leave that to you to do that okay.

Now again let us ask for a physical problem where you have a geometric distribution coming out. The binomial distribution we saw a very simple example of a gas and a classical ideal gas; number fluctuations in a classical ideal gas and thermal equilibrium, typically of a binomial distribution okay. Is there a similar kind of example which you can think of for geometric distribution. You have all seen this but in a slightly different language.

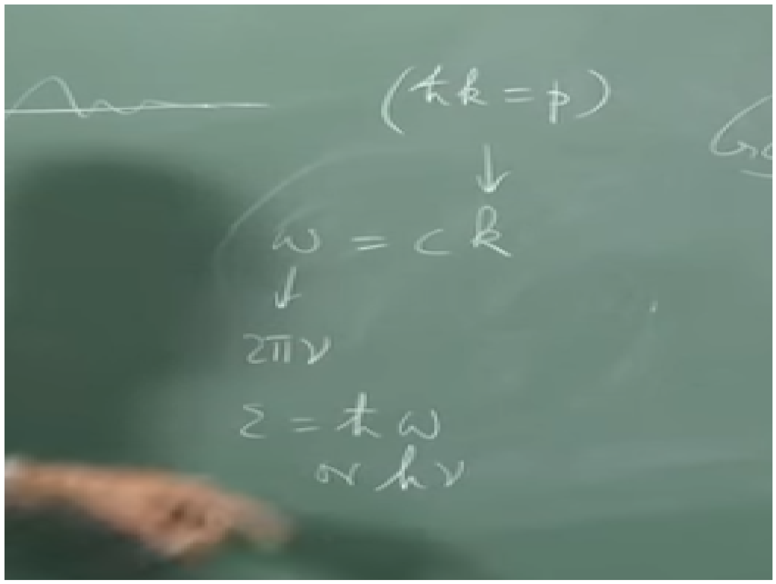
So you would not recognize it as such. Instead of a gas of particles think of a gas of photons inside a cavity. It is a black body cavity. So think of black body radiation. What happens there? It is a collection of photons sitting inside a cavity whose walls are held at a fixed temperature. This gas equally breaks unlike a massive gas of massive particles where systems will be in thermal equilibrium because there are all these elastic collisions going on so if any one particle has too much energy the other fellow is knocked down soon to come near the average and so on.

In the case of a photon gas there is no direct interaction between photons. Direct photon photon interaction is extremely small. There is a small correction due to quantum electrodynamics but the cross-section is negligible for these purposes okay. So here is a gas of non-interacting particles, they do not interact with each other. They do not even undergo elastic collisions with each other.

They are photons, massless particles but they equilibrate because the photons are absorbed by the walls, the atoms in the walls of the cavity and reemitted. So this process of interaction with the walls of the cavity keeps the system in thermal equilibrium and one of the properties of a photon is that it has 0 rest mass so in principle, in principle you can create and destroy any number of photons.

You can absorb and emit any number of photons. So the number of photons is not conserved. There is no energy, there is no number conservation for a gas of photons. In fact the number is fluctuating quite widely and we would like to know what is the probability that at any given instant of time in a black body cavity with a given temperature you have a certain number of photons okay. It is a fluctuating random variable.

(Refer Slide Time: 46:23)



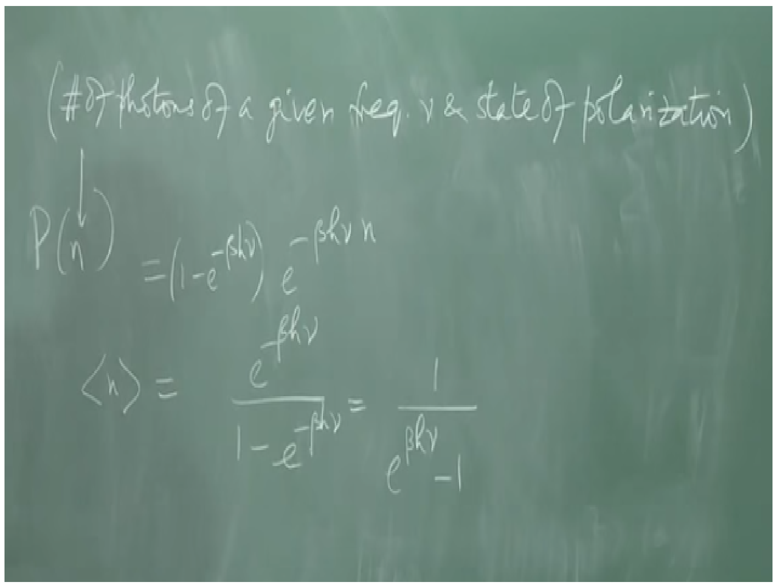
Now of course a photon is also described by specifying its direction of its moment, its wave number which is equivalent also to specifying its frequency because you know that the photon frequency and wavelength or wavenumber would satisfy a relation like omega = c times k where this fellow here is h cross k = p the moment of the photon and this is 2 pi nu the frequency here.

Remember that the energy of a photon epsilon is h cross omega or h nu depends on what you like to use. Physicists like to use the angular frequency as the frequency itself and h cross is h over 2 pi, h is Planck's constant right. So for a photon you have to specify what its wavenumber is actually or its wavelength or frequency; they are all equivalent in this language. In addition you have to say what its state of polarization is.

Now every free photon has 2 degrees, possible degrees of 2 kinds of polarization. It could be either left circularly polarized or right circularly polarized. Correspondingly its spin quantum number in the direction of its motion is either + 1 or - 1. These are the 2 helicity states of the photon or the 2 states of polarization of photons.

Now let us simplify matters and say we are not worried about any of these things. We are simply asking for a given frequency nu and a given state of polarization, what is the number of probability that you have n photons in this cavity okay. How do we go about it? Well this system is in thermal equilibrium with a cavity.

(Refer Slide Time: 48:00)



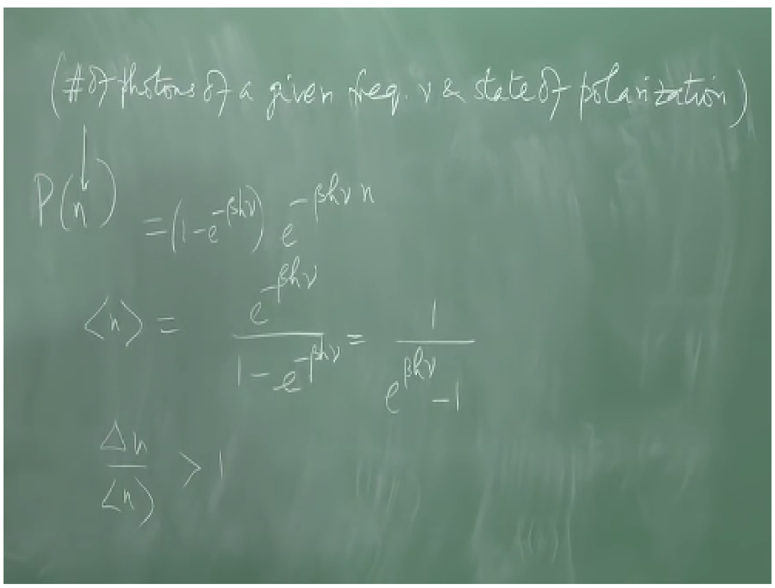
So I know that the probability that you have n photons number of photons of a given frequency nu and state of polarization. That is what this n stands for and I would like to know and n has no limit on it. It can go from 0 to infinity because as I said the number of photons is not conserved okay. What is this proportional to? What is this going to be here?

It is clearly going to be proportional to if it is going to be in equilibrium with the heat bath at temperature t, some temperature t then this thing here has to be proportional to, proportional to e to the power minus the energy of these n photons which is n h nu/Boltzmann's constant multiplied by the absolute temperature. That is what equilibrium statistical mechanics prescribes for us in the canonical ensemble that it has got to be proportional to e to the power the energy of that state/kT okay.

Now there is a standard symbol for this 1 over kT which is e to the minus beta h nu times n; beta stands for 1 over kT, k Boltzmann times t; k Boltzmann is Boltzmann's constant. It is value in standard international units is 1.38 10 to the minus 23 whatever it is Joules per degree or something like that. So this guy here is just a conversion factor to convert from temperature to energy and you have got this formula here.

But already you see this is a geometric distribution because e to the minus beta h nu is a number less than 1 and that is raised to the power n and is exactly like the q to the power n that we had. So this is basically a geometric distribution. All we need to do is to multiply it by the p to normalize it.

(Refer Slide Time: 50:19)



So if I want to write an equality I must multiply it by the p which is 1 - q. So 1 - e to the beta h nu and that is it. That is a geometric distribution okay. What is the average number? We have a formula and that was q over p right. So this was e to the minus beta h nu over p which is 1 - e to the minus beta h nu which is also equal to 1 over e to the beta h nu - 1. Do you recognize this factor?

It is the Bose factor from Bose statistics, it is there, it is right there I mean you know these are very simple argument. I push certain things under the rug but this is the Bose factor, exactly the Bose factor. You begin to see where it comes from now. Just from the canonical ensemble normalized and that is it. Of course there were assumptions made here which are correct in this case. The photons do not interact with each other.

It is an ideal gas and there was one more assumption made which is that it does not cost any energy to add one more photon to a statistical collection so the chemical potential here is 0. Otherwise you must have beta times energy minus the chemical potential out here. But the chemical potential of a photon gas is 0 because the photon has 0 rest mass okay. For bosons the chemical potential cannot be positive, it has got to be negative or 0 and in the extreme case of photons it is 0. So we have the Bose factor here.

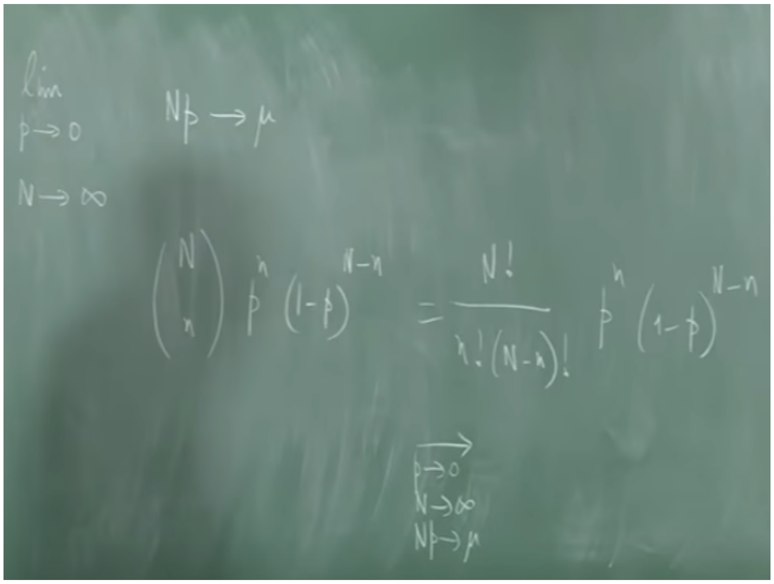
This is in fact the average number for a given frequency nu for a certain state of polarization. Of course if you want to find the actual total photon number of all frequencies you must integrate this with respect to frequency and then you must multiply by 2 for the 2 polarization states okay but we already have the Planck formula in some sense and we have now an expression for what the mean number is, what the variance is, etc., etc.

What is the variance going to be? Well all you have to do is to take the generating function, differentiate it, plug it in etc., etc. and find out what the variance is and you discover for this probability I leave it as a simple exercise for you to show that delta n over n this guy is greater than 1 in this problem. So it is a huge scatter whereas normally we saw in the classical ideal gas that there was a scatter the relative fluctuation was not that large.

It was going like 1 over the square root of n. Here the guy is actually larger than unity. So this fellow is bigger than that. Shows that this is tremendously fluctuating system in which the photon number is really fluctuating enormously. The relative fluctuation is huge compared to the classical ideal gas. So the effect of statistics are shown up here in this very indirect fashion.

There are other examples of the geometric distribution that we will come across as we go along but I thought that it is a good idea to have one physical problem here, black body radiation here. We are going to look at radiation in other states like inside a laser cavity and then the statistics is going to change completely as we will see.

(Refer Slide Time: 54:15)



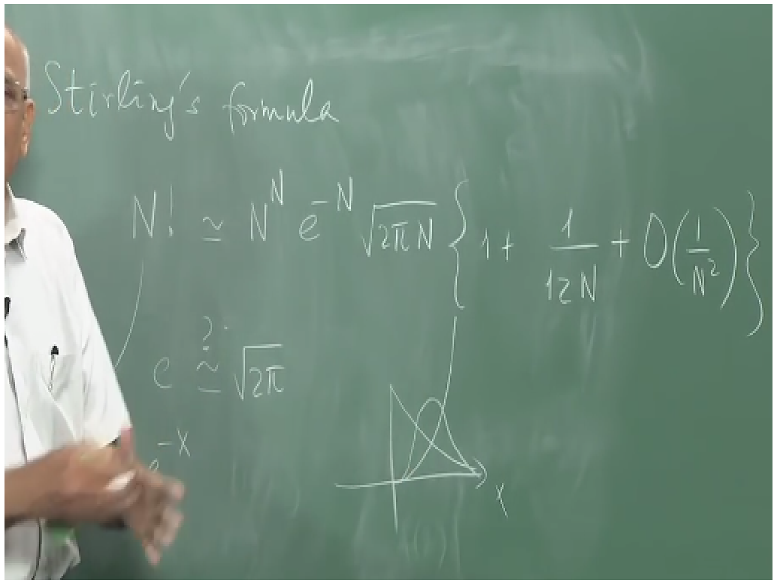
The next question is to ask what happens in a Bernoulli trial, what happens in the binomial distribution if the probability p goes to 0 and the number of trials goes to infinity in such a way that the product of these goes to some finite number in the limit such that N p goes to some number mu, some finite number mu. So the mean does not change but the probability of success in every given trial is vanishingly small.

The number of trials is increased so that the mean remains the same as before. What happens to the binomial distribution? So if you start with N n p to the power n, 1 - p to the power N - n. What does this do in this limit here and this is not a very hard question to answer. What we need to recognize is that now the sample space of this little n will go all the way to infinity because earlier it went from 0 to capital N for the binomial distribution.

But capital N is going to go to infinity so little n will now run all the way from 0 to infinity all the non-negative integers and what would you do well what you would use is a very interesting formula because this fellow here is equal to N!/n! (N – n)! p to the power n (1 – p) to the N - n and this will tend in this limit, it tends. What should I do when you have the factorial, very large number what do you do?

Use Stirling's so called approximation. I would like to call it Stirling's formula because it is a very good formula. It is not it is not much of an approximation, it is an extremely good approximation, extremely good approximation already because what does it say.

(Refer Slide Time: 56:39)



It says Stirling, it says that if you have a number N, N! then N! is of the order of N to power N, in the leading approximation when you replace all the factors here by N then that is of course an over estimation, so you got to kill that over estimation and what you get is a correction, it goes like e to the minus N and then if you do it a little more carefully and normalize it you get 2 pi N okay.

That is the leading term in here and what you actually have is 1 plus corrections which go down as N increases which die to 0 and the first correction is of order 1 over 12 N plus order 1 over N square. So you get a whole infinite series, an asymptotic series, but the first term already is extremely good alright. As N becomes bigger and bigger it gets better and better because you can see that the first correction goes like 1 over 10 times N already.

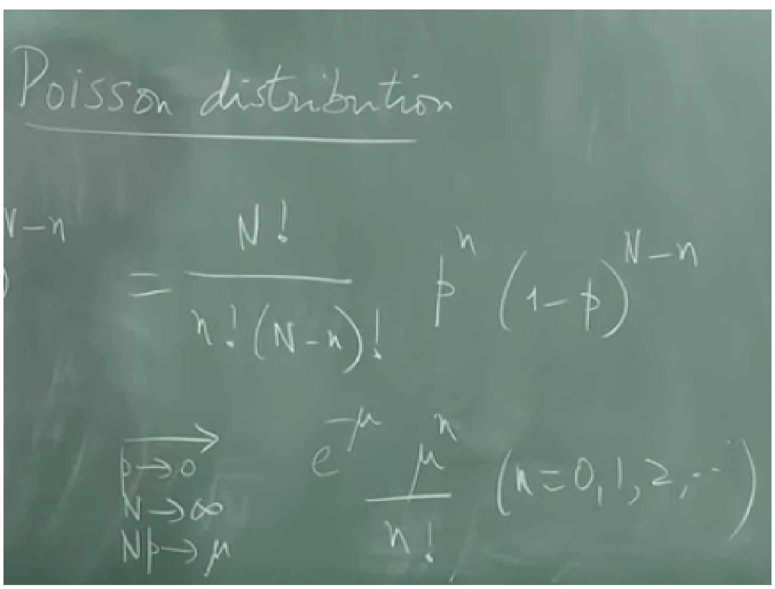
So already when N is 10 that is a 1% correction, one twelfth. When N is 1 what happens? It is a very bad thing to do. When N is 1 what happens? Well if this formula is true it says 1 is approximately equal to e to the power - 1 square root of 2 pi. So the question you are asking is, is this so? Is square root of 2 pi approximately equal to e and so on but we know exactly how good it is going to be. It is going to be correct to one part in 12, 92% accuracy already.

So the error is just 8% or something like that. When N is 10, it is already of the order 1%, when N is 1000 it gets much better. When N is Avogadro's number forget about it, so almost exact. So it is an extremely good formula. The reason of course is you write this as an integral, you write this as integral 0 to infinity dx x to the power N e to the minus x write it in this form and then use this, this guy is an increasing function of x as N becomes larger and larger increases more and more steeply that fellow comes down more and more steeply exponentially so.

So you have an integral which has a function of x, this one factor starts like this and goes up and the other fellow starts at 1 and comes down very rapidly. The result is the product of the 2 will be 0 almost everywhere except in a small region here and now do an Gaussian approximation like this is an Gaussian integral and you get Stirling's formula the leading term including the square root of 2 pi N okay.

So I leave that as an exercise to you and the corrections are going to be as good as this Gaussian integral is. So they are sort of extremely small. That is why you have these terms here okay. So this is a sort of Gaussian approximation to an integral and it is very powerful.

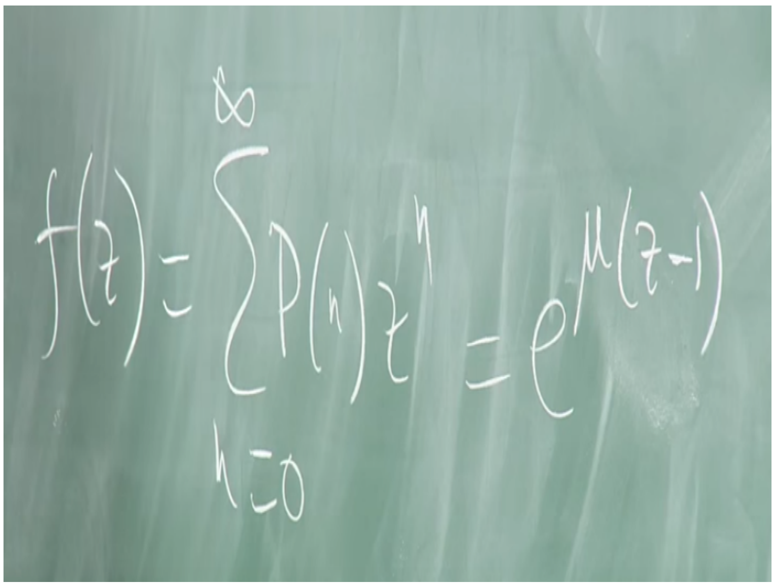
(Refer Slide Time: 1:00:12)



So use Stirling's formula here in this and a little bit of algebra and the result that emerges is that this becomes e to the - mu. You can see where the exponential is going to come from. This little p can be written as mu/N. So you have (1 – mu)/N to the power n where N goes to infinity that e to the - mu. So that is where this comes, this formula and then the remaining portions are mu to the n over n!

What you would call this distribution? It is a Poisson distribution. So this is the Poisson distribution. It is a one parameter distribution specified by the value of mu and mu is the mean value. That is where it started and that is where it remains. It is the mean value. What is the generating function of the Poisson distribution?

(Refer Slide Time: 1:01:16)



So once you have that f of z equal to summation P of n z to the power n, n is 0 to infinity and that is equal to e to the - mu + mu z. We will write this as e to the mu times (z – 1). I have a z to the n so it is mu z to the power n over n! summed is e to the mu z multiplied by this and that is it exponential okay. This exponential form of the generating function for the Poisson distribution has all sorts of miraculous properties as we will see.

The product of 2 exponentials is again an exponential the s that is going to play a big role okay. Is f(1) = 1? Yes, indeed. What is f prime of 1? That is the mean value okay. So we will stop here and I am going to take it up from this point next time.