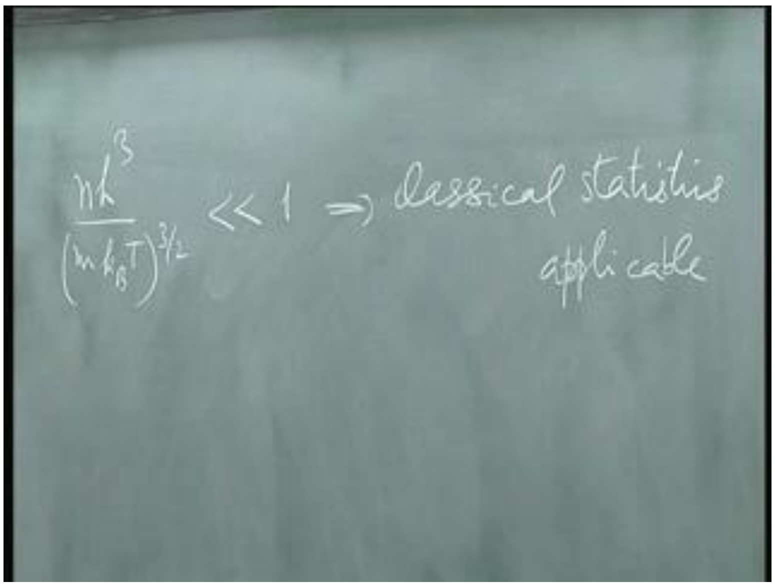
Classical Physics
Prof. V. Balakrishnan
Department of Physics
Indian Institute of Technology, Madras

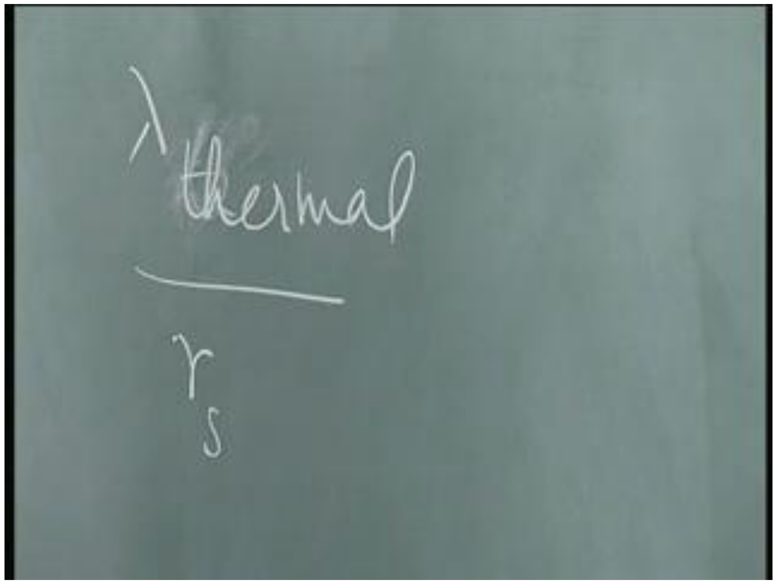
Lecture 26

(Refer Slide time: 01:10)



So, this was the parameter in terms of which we were deciding whether classical statistics should be used in a given situation or quantum statistics and the what we derived was that, if this quantity is less than one implies that classical statistics is applicable and I pointed out that it is indeed much less than 1 for normal gases at room temperature for instance. So, there are two ways in which this could change one of them is, if you have very high temperatures then you are in the classical regime, if you are at very low temperatures then this could go the other way the inequality could be violated and then of course you have to go to quantum physics or if the density is very high, if the number density is very high which is what happens in astrophysical context the number density becomes very high even though the temperature is high, it is not high enough and then you have what is called the very degenerate quantum system or quantum gas and you have to use you different techniques, you have we will do this in the next course little bit. This is a dimensionless number because ultimately it arose from the following.

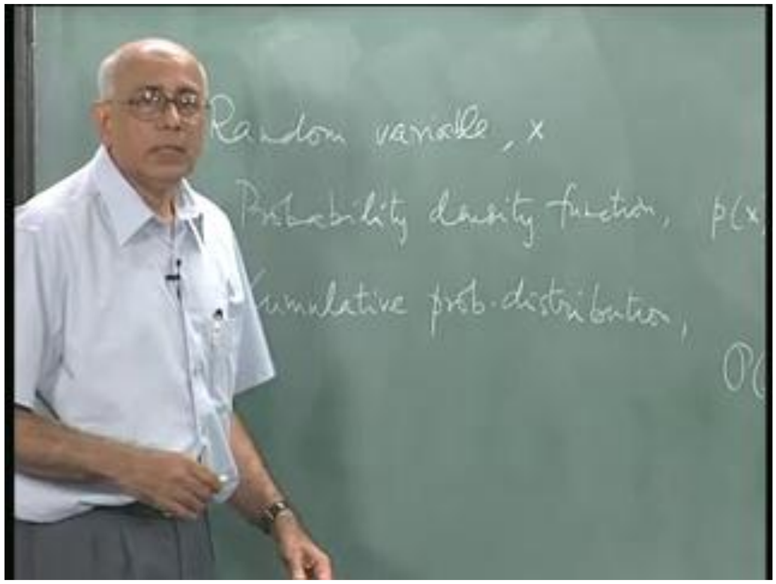
(Refer Slide time: 02:22)

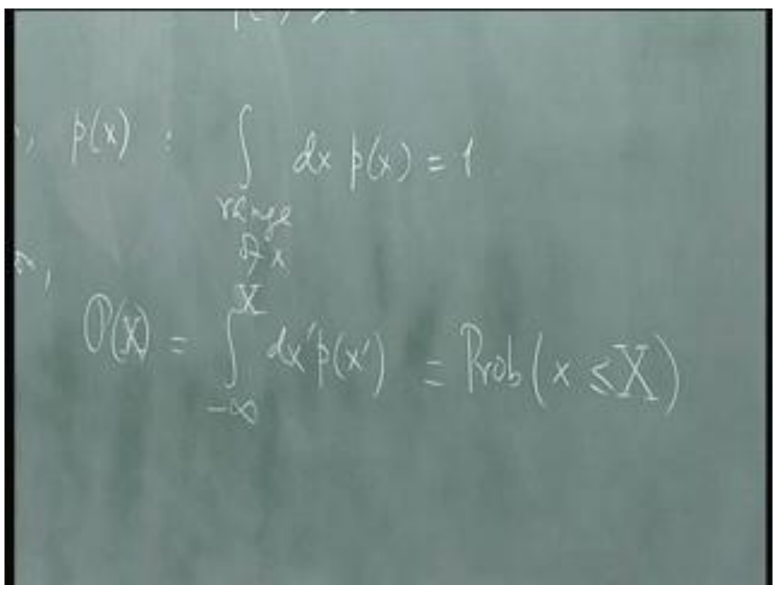


It arose from the de broglie wave length or thermal lambda thermal divided by the mean separation and I just said, if lambda thermal the fuzziness in the position of each particle is much smaller than the inter particle separation, you could get away with classical statistics. On the other hand, if it is very spread out then the quantum effects become important and then the effect of exchange the indistinguishability of particles becomes a plays a predominant role in and the statistics changes. So, that is the idea, a very useful parameter to tell when you should have classical statistics and when you should have quantum statistics for these systems for gases and things fluids and so on. There are other systems like spin systems which we look at in the next course where the system is intrinsically quantum mechanical and then there could be temperature regimes where it becomes effectively classical for various purposes we will look at that too.

I promised earlier that, I would spend a little bit of time on probability distributions and maybe this is a good time to do this, because if we postponed it what I would like to do after this is to go on to real gas like the vanderwaals gas, and then talk a little bit about the Weiss molecular theory of ferromagnetism and then we move on to the next topic. So, this is what the curriculum says the syllabus says and we will stick to that. So, let me spend a few minutes talking about probability distributions you probably already know a lot of this many aspects of it so, this so let us do this in a kind of informal way.

(Refer Slide time: 04:09)



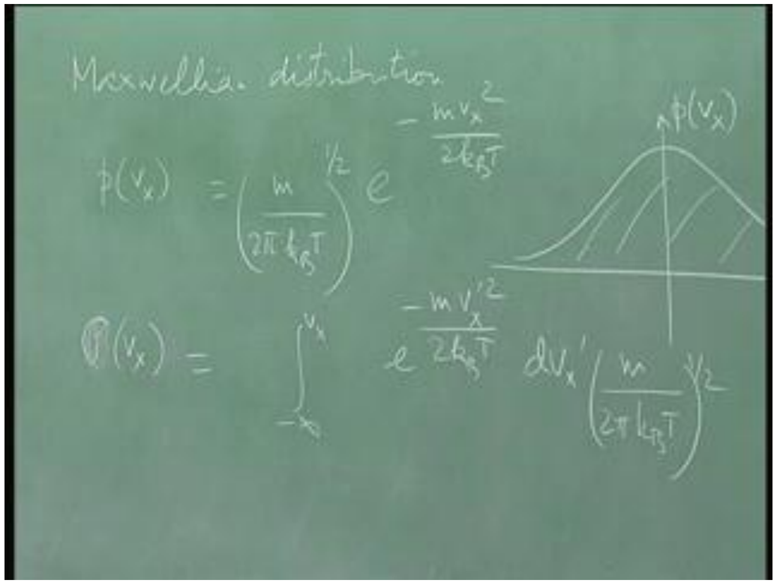


So, if I have a random variable some kind of random variable x could be discrete value in which case all that I am going to write down integrals will be replaced by sums, but if it is continuous which is the more interesting case, in many situations then there exists a probability, we assume that there exists a probability density function p of x, this is an assumption of course all random variables do not have to have well a behaved probability density function, whenever it does we are going to assume some properties on the part of this p of x, the first part of course is p of x is greater than equal to 0 cannot be negative and second thing we are going to assume about it is normalized to unity, in other words this is such that over the range of whatever it be dx p of x equal to 1. So, the total probability is one.

You can always define, also define the cumulative probability distribution, sometimes called the probability distribution it iself, but it is actually the cumulative probability distribution or distribution function we need a name for this symbol. Let me just call it p of x this is an integral up to the value x dx p of x. So let us be careful and put p x prime from minus infinity, assuming that the lower limit of integration is minus infinity or whatever is the lower bound. Now, what is the meaning of this is equal to the probability that your random variable sorry call this a capital X, call this p of capital X, it is the probability that the random variable is less than equal to this given value capital X. So, this it is the area under the curve from minus infinity up to whatever point you stop and that is the total probability that the random variable has some value less than or equal to this prescribed value and p of capital infinity is of course one by normalization physicists most of the time use this talk about this and the cumulative probability distribution is an integral over this density, but it is not as frequently used by in physical applications as the probability density function itself.

I should also caution you that, very often in the physics literature this quantity itself is called the probability density the, this quantity is itself is called the distribution function but it is actually not the distribution, it is the density, but then very loosely one says for instance when you say, Maxwellian distribution of velocities and you write a Gaussian, down that Gaussian is a probability density function it is not the distribution function so that is a typical example.

(Refer Slide time: 07:33)

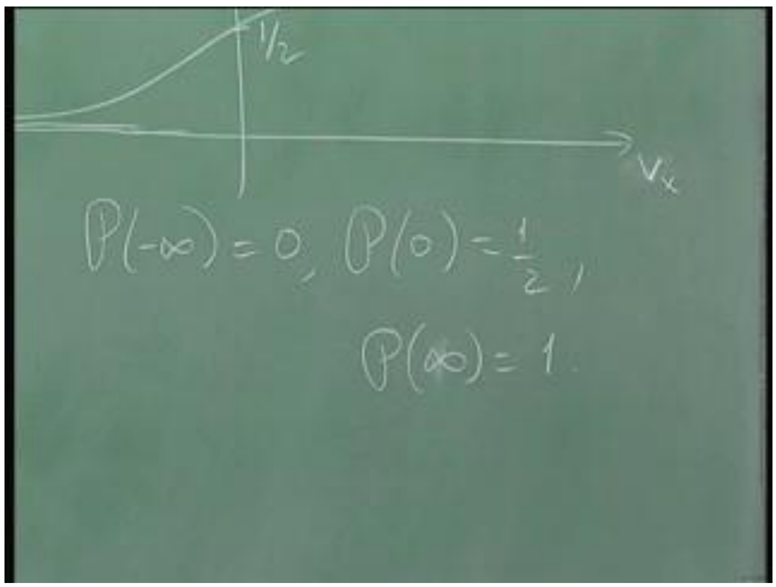


The Maxwellian distribution, it says that the probability density that any Cartesian component say v sub x of the velocity of a particle in thermal equilibrium in a gas in thermal equilibrium a classical gas this quantity is given by e to the minus m v x squared over 2 k Boltzmann T at absolute temperature T and there is a normalization factor which is m over two pi k Boltzmann T to the power half. So, p of v x d v x is a probability that the velocity x component has a value between some v x and v x plus d v x.

And if you plot it this is the famous Gaussian which has a picture like that and this is p sub x this is p v x and similarly for the y and z components ok, this is not the distribution of the speed which is modulus of the velocity vector this is just the x component or the y or the z components.

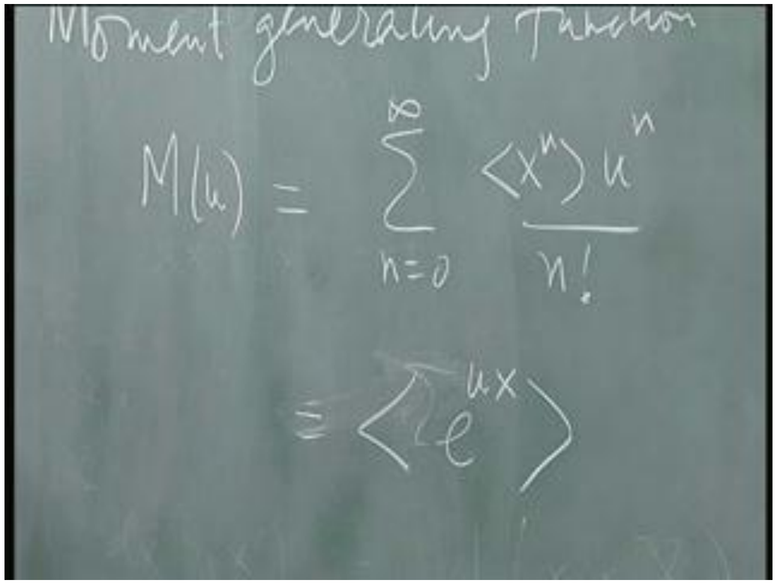
Now, what is the cumulative distribution function p, curly p of v x, this quantity would be an integral over this guy minus infinity up to v x e to the minus m v x prime squared over two k Boltzmann T d v x prime multiplied by the normalization and what does is the graph of this function look like it starts at 0 at minus infinity and goes to one at plus infinity. So, if you plot it is the area under the curve up to whatever points. So, from minus infinity up to this point the area under total area under the this curve is unity. So, if I plot this function here is v v x and here is p of v x it starts at 0 and goes to unity at plus infinity and it crosses the axis at half because half the area is the area to the left of this axis is exactly a half. So, this is what the integral density function looks like that is the cumulative distribution function, there is a name given for this function the integral of a Gaussian up to some point x it is related to what is called the error function this is a special function. We are not going further into this, so there are tables of these error functions. So, it is clear that the this guy is normalized in such a way that p of minus infinity is 0 p of 0 equals to half and p of infinity equal to one.

(Refer Slide time: 09:48)



It is a non negative function and it is an increasing function, it is a non decreasing function certainly as you increase v x it monotonically climbs up towards one that is not as useful the error function is not as useful as the density itself which is a Gaussian and with which you can do things you can do a lot of things with Gaussian. So, this is a very simple physical example of where a Gaussian appears and it appears all over the place and we will see why in a while let me go back and define a few quantities with regard to probability distribution density functions themselves.

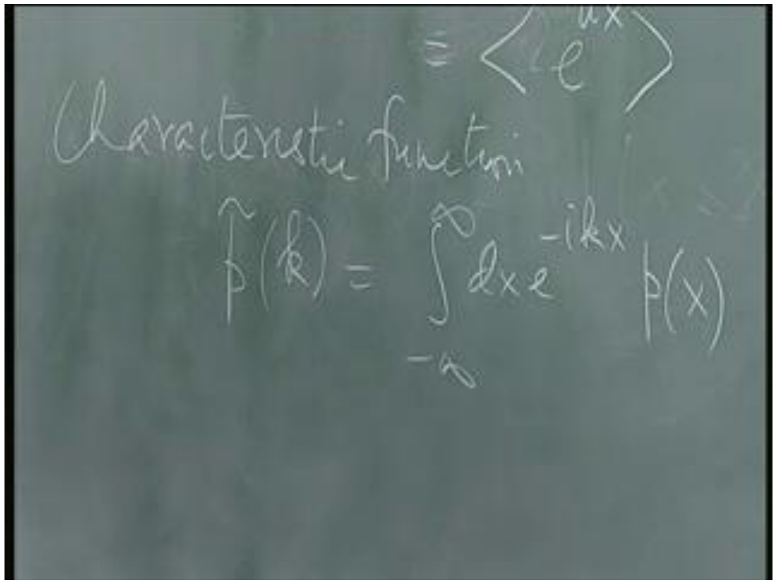
(Refer Slide time: 11:31)



Well, I will denote the mean value of this variable by x ,the mean value by x squared, the mean squared value by value x squared and the n eth moment by x n etcetera. These are the moments of the distribution moments of this random variable x and what I need is a formula to write it and of course, I have a formula because if i x to the power n as you know is integral over the range of x d x p of x, x to the power n assuming this is a normalized density. So, I do not divide by the integral of that, now it is useful to consider not individual moments, but what is called a moment generating function. So, let us define a moment generating function and let me call it M of u there are lots of notations used this is equal to a power series n equal to 0 to infinity whose coefficients are just these quantities by definition this is the generating function it is a power series in u, in this auxiliary variable u and the coefficients of the coefficients of u to the n is the n eth moment of this random variable over n factorial.

It is also clear that you could, well you put this u inside here because u is not a random variable but just an auxiliary variable and then this is an exponential series. So, it is clear that this could also be written as equal to the average value of e to the power u x, because if I expand this term by term the n eth term is just u to the power n average x to the power n over n factorial, it looks like a sort of Fourier transform kind of thing except you got a u here.

(Refer Slide time: 13:42)

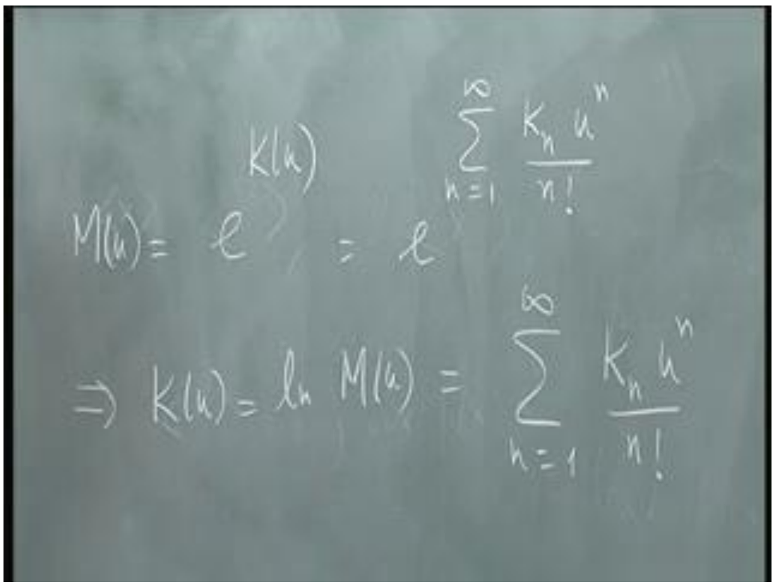


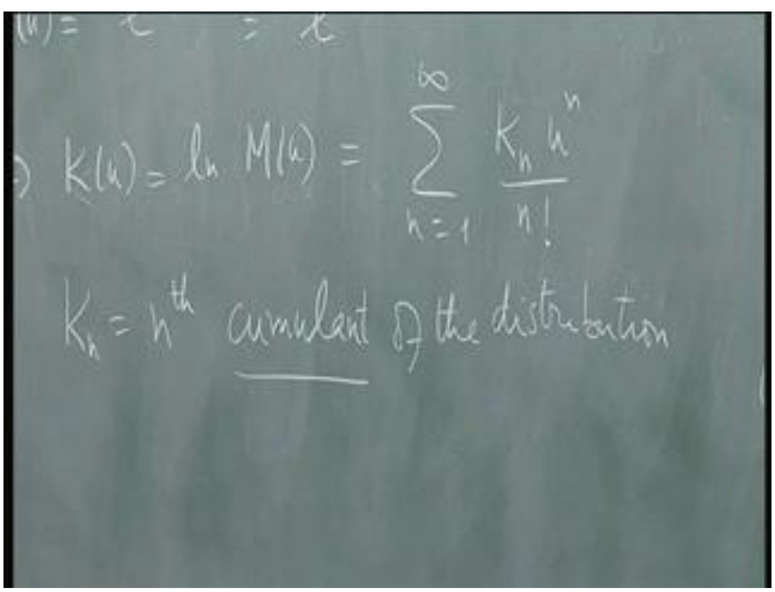
Well you could also define the following you could also define what is called the characteristic function corresponding to the density p of x and let us call it p tilde of k this is an integral minus infinity to infinity d x e to the minus i k x p of x. This is the Fourier transform of the probability density function, but you can also relate it to this because notice that it is also equal to the average value of e to the minus i k x, x is a random variable, k is an auxiliary variable the Fourier transform variable. So, the Fourier transform of incidentally, this I have used a particular convention for the Fourier transform you know that you can put a one over two pi here or a one over root two pi, because this could be a plus sign etcetera.

So, I have defined it such that to go from x to k I have no factor here and a plus and a minus here, so if I want to write p of x it is equal integral d k e to the plus i k x p tilde of k divided by two pi. So, that is my convention and the thing to notice is that the fact the Fourier transform of this distribution could also be regarded as the average value of e to the minus i k x over all realizations of the random variable x, but it is also obvious that it is also immediately obvious that this equal to m of minus i k, so instead of u, I put minus i k i get exactly the same thing.

So, the characteristic function is nothing but the moment generating function evaluated at a pure imaginary value of the argument, now what is it for a Gaussian well we have not come to Gaussian yet, I will do so very shortly very often it turns out that it is useful, since this is the expectation value of an exponential this is a very reminiscent of what we did when we defined the free energy there you had a whole lot of exponentials the Boltzmann factors and I said you could write it as a single exponential with some effective variable there energy there the same idea is actually the idea is borrowed from statistics.

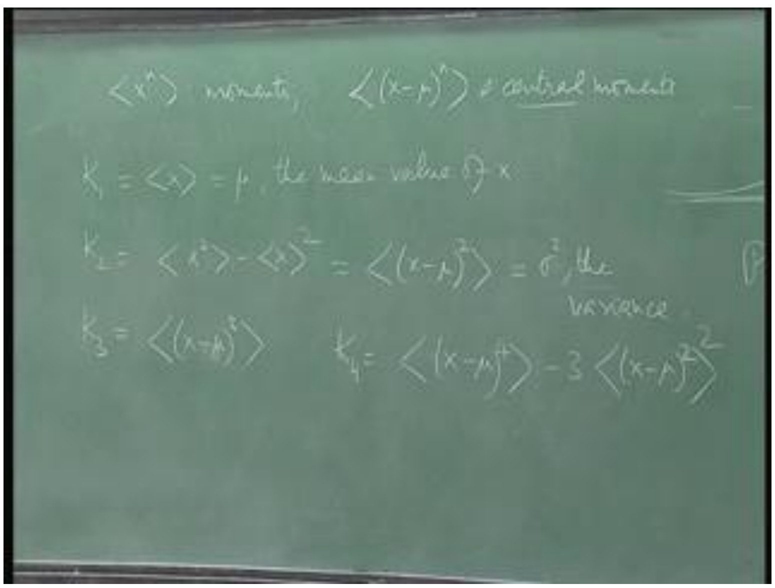
(Refer Slide time: 16:03)





Because you could also write m of u, could also be written as e to the power some other function k of u which is e to the power summation n equal one to infinity some k power sub n u to the n over n factorial, because it is clear that this thing here is going to have all powers of u appearing in it which you could also write in this form in this representation then it would immediately imply that k of u equal to log m of u and this is equal to n equal to one to infinity k power sub n due u to the n over n factorial. Formally, so it is again a power series and the question is what are these coefficients k power sub n and this number k power n equal to n eth cumulant of the distribution in the older statistics literature these things used to be called semi invariants, I will explain the meaning the reason for introducing this quantity, but it is called the cumulant of the distribution and they are related to the moments through this through this relationship, in fact you can write the cumulants in terms of the moments of the distributions.

(Refer Slide time: 17:43)



So, let us see what that relationship is just a little bit of playing around, with this I will tell what these quantities are k 1 the first cumulant turns out to be the average value it iself the first moment it iself. So, let me use the a symbol for this equal to mu the mean value of x the first cumulant is just the first moment it iself and that is very trivial to see from here the second moment k two, the second cumulant turns out to be x squared minus x the whole squared which is the variance. So, this is also equal to the expectation value of x minus mu the whole squared it is the variance. So, it is the mean value of the square of the deviation from the mean also called the variance.

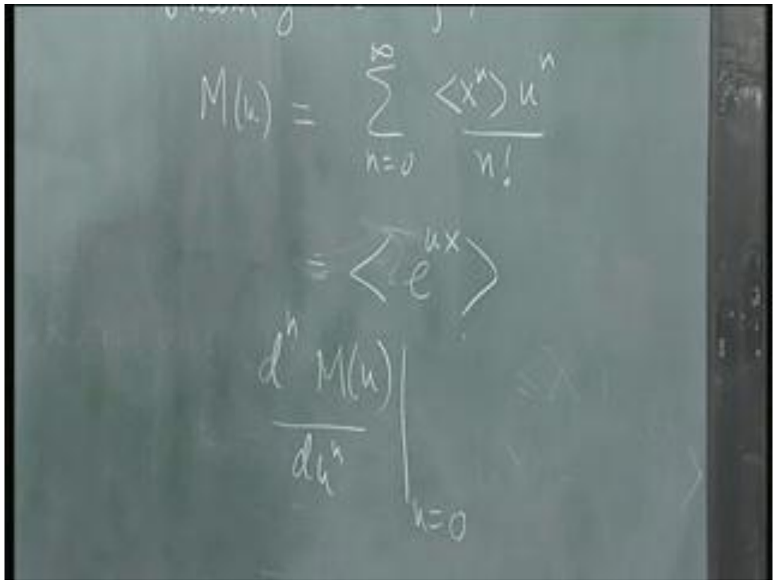
So, let us use a symbol for it so this is beginning to tell us what these cumulants are doing and I will explain what the significance of these quantities is the third moment, k three the third cumulant k power three is equal to x minus mu the whole cubed it is the mean value of the cube of the deviation from the mean incidentally these quantities x minus mu, so these quantities x to the power n are the moments these quantities x minus mu to the power n are the so called central moments they are the moments about the mean.

So, it is very often convenient to shift to the mean if the mean is systematically changing with time for instance it is convenient to remove that secular variation and then look at only the fluctuations. So, this is what the central moments are and the third cumulant is the third central moment also the fourth is not k power four equal to x minus mu to the power four minus three times x minus mu the whole squared the whole thing squared this is the variance and it is squared once gain so that the dimensions come out right and you have to subtract three times that from the fourth central moment and that is equal to the fourth cumulant and so on.

So, there are standard formulas which will tell you what k power n is in terms of the n eth moment the n minus one eth moment and so on, all the way down which you get simply by comparing on both sides so write M of u in that form takes it log and then compare it power term by term in the power series of obviously this guy has some radius of convergence in u in the u plane, so everything is valid inside the radius of convergence when you have a power series you must treat the function defined by a power series as function of a complex variable and it is absolutely convergent inside some circle of convergence and then you can integrate differentiate term by term etcetera. So, inside the circle of convergence a power series converges absolutely and you can do all sorts of manipulations with it and you can compare coefficients etcetera.

So, I have assumed that there is a finite radius of convergence of this quantity. So, within that radius of convergence this is perfectly all right and I have kept u real I mean, I do not have to as you can see I can go to pure imaginary values and so on the point is this a function of some this is an analytical function in the neighborhood of u equal to 0 it is guaranteed to be an analytic function with some radius of convergence.

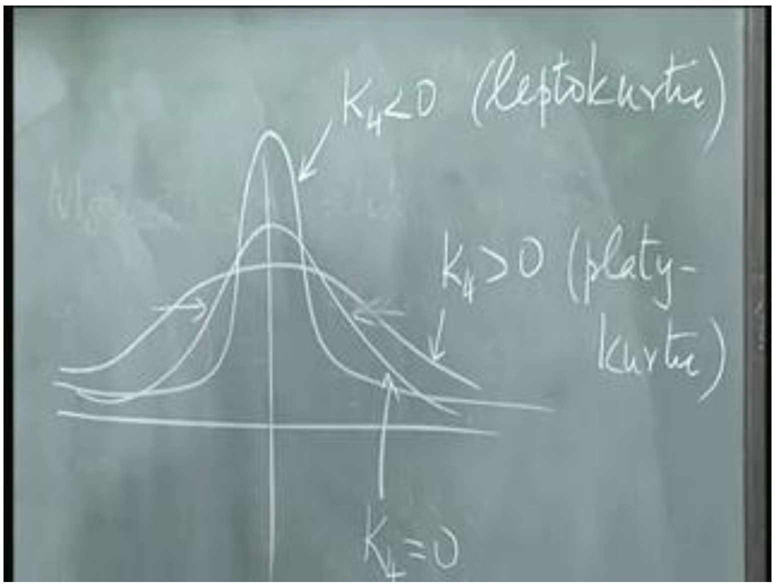
(Refer Slide time: 22:12)



Now what is the formula, how to invert it well, that is not very hard what is the advantage of finding this, but it is obvious that if I differentiate n times M of u over d u to the power n, if I differentiate this guy all the terms less than power n will disappear, if I put u equal to 0. So, if I differentiate this and put u equal to 0, I differentiate the first time with respect to u, I get n next time n minus one and that keeps canceling this. So, it is clear that this quantity is in fact equal to x to the power n, so that is the inversion formula you could also write it as a contorential round the origin in the u plane in the usual way of inverting a Taylor's series this is like a power series representation of this analytic function, similarly here too k power n could also be written as equal to d n over d u to the power n k of u the cumulant generating function evaluated at u equal to 0, so formally you can always invert the power series in this form.

Now what is the physical meaning of these quantities here the moment of course the first moment tells you where the mean value is the second moment, the second the variance tells you the spread about this central value first measure of a spread the third gives you the skewness, because if this is symmetric if this distribution is symmetric about it is mean value then this cube and the average value vanishes, but if it is not then this tells you the skewness, how much it is to one side or the other a distribution suppose the mean is 0, for example a distribution which looks like this is a skewed distribution. So, if this is the mean value if this is the mean value there could be more of it on this side it should not be so fat it could look like this and maybe this side tapers of more slowly or something like that these are skewed distributions and the third moment this quantity, here is a measure of the skewness or asymmetry of the distribution, if of course it is a symmetric distribution about it is mean then this is 0 and all odd central moments should vanish, if it is a symmetric distribution about it is mean.

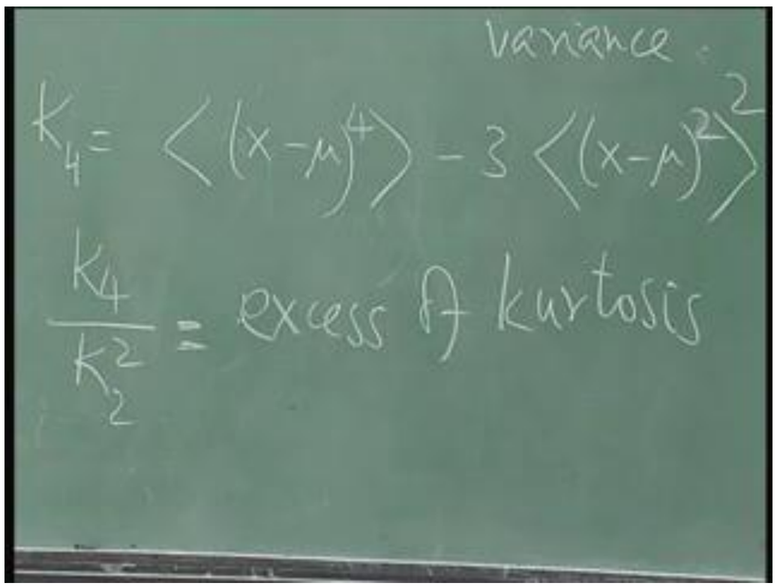
(Refer Slide time: 24:49)



The physical meaning of the fourth is not so obvious, what it does is the following and with reference to a symmetric distribution it is got the following meaning. So, suppose you have a symmetric distribution this is the mean value say and it is about the mean, so now no skewness here then this width here is proportional to the square root of the variance the standard deviation measures how fat it is, what this guy does is to measure the departure from a Gaussian distribution this quantity as we will see is exactly 0 for a Gaussian distribution the fourth cumulant is exactly 0 for a Gaussian distribution. So, if that number k power four is positive what it implies is that as you can see this is the fourth power sitting here so large values of x are going to dominate in the fourth power.

So, this means that your distribution is such that the dominance of large values of x is greater than that of smaller values of x which means that it is in some sense fatter than a Gaussian more importance is being given due to larger values this is if a let me schematically say this is, if k power four equal to 0 this is, if k power is greater than four 0 positive but it could also be negative there is nothing that says that this fourth cumulant has to positive or negative it could be negative if it is negative it could mean means that this distribution is leaner than a Gaussian more importance is given is being given to smaller values here it is leaner and that the reason why the quantity here becomes negative. So, this is leaner than a Gaussian and this has a name it is called leptokurtic and this guy has a name it is called platy-kurtic distribution.

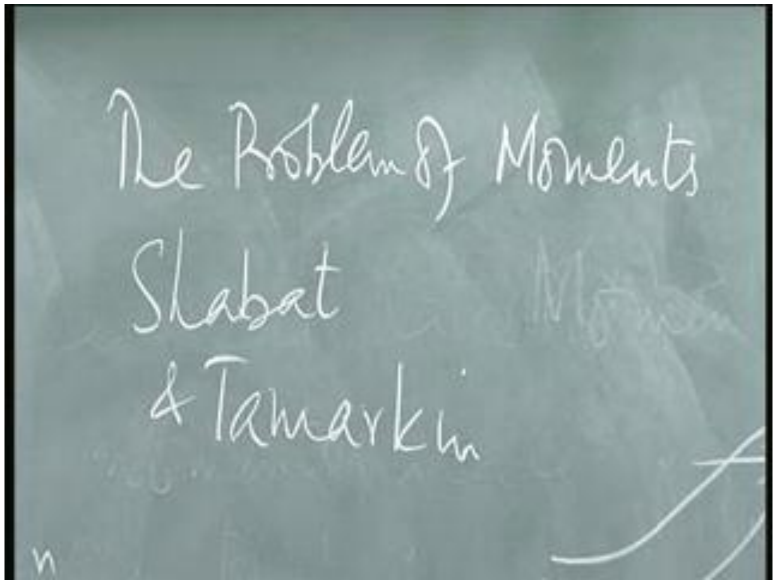
(Refer Slide Time: 26:53)



And this quantity k power four divided by k power two whole squared this is the square of the variance on this side it normalizes by dividing by this square this quantity, here is called the excess of kurtosis for a probability distribution.

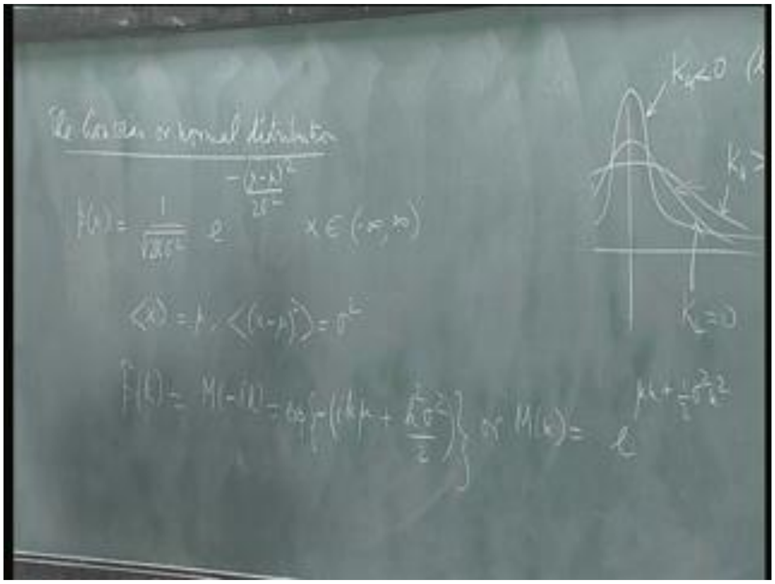
So, in a sense in empirical usage the first four moments can tell you more or less all that you need to know at a first approximation at least about the distribution of course in principle you need to know the entire set of moments you need to know an infinite number of piece of information to be able to reconstruct the distribution and this is a central problem of statistics given the moments can you reconstruct the distribution or not and there are conditions necessary and sufficient conditions when you can do this it is a little bit like asking, if I give you all the Fourier coefficients of a periodic function can you reconstruct the function I give you discrete set of piece of information can you reconstruct a continuous function and the answer is in the circumstance is yes and it is the same sort of thing this is called the problem of moments.

(Refer Slide Time: 28:19)



There is a very excellent text have forgotten the name, now it is called the problem of moments and it was probably written in the nineteen twenties or nineteen thirties or something and it is by Shabat and Tamarkin and it is called Problem of Moment. It is worth repays reading because it is a very classic book and in statistics many such very good books and this is a particularly good one soagain to repeat myself, the first four moments or the first four cumulants essentially tell you more or less what the distribution is like in practice, now the question is where does the Gaussian come in what does that look like and that is an interesting thing, we already already mentioned that Gaussian has the fourth cumulant equal to 0 and k three this is about the centre of course it is distributed symmetrically. So it is also 0 so only k power one and k power two are relevant so let us see what the Gaussian is.

(Refer Slide time: 29:16)

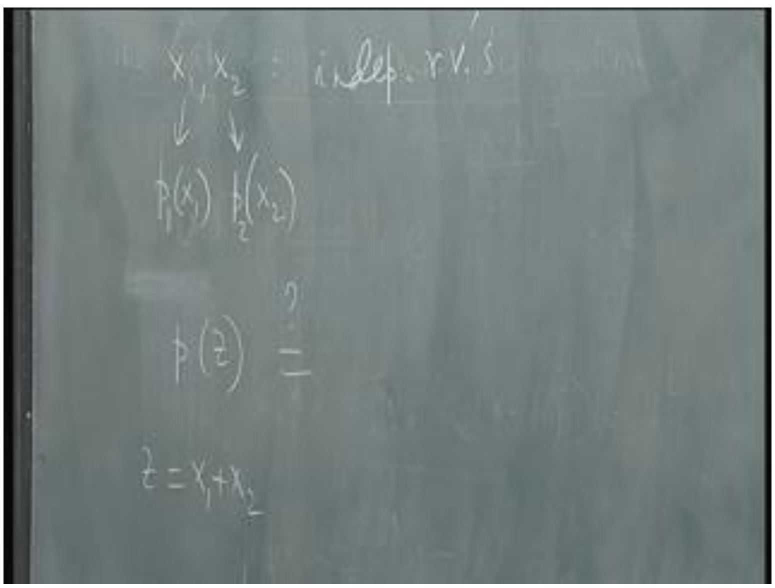


The Gaussian or the normal distribution, it is a two parameter distribution the normalized probability distribution for a random variable running from minus infinity to infinity is the following it is one over square root of two pi sigma squared e to the power minus x squared over two sigma squared and x is on the real axis this is normalized to one, but I have put the mo mean equal to 0. So, for a general mean it is actually this x minus mu whole squared over two sigma squared and x equal to mu x minus mu whole squared equal to sigma squared the variance is sigma squared, one can take it is moment generating function or it is compute what the Fourier transform of this quantity is and it turns out that p tilde of k equal to m of minus i k, this case is not hard to do you need a Fourier transform of a Gaussian and many of you may be aware that the Fourier transform of Gaussian is also Gaussian in it is turn and this is equal to minus i k mu minus k squared sigma squared over two or if you like m of u equal to e to the power exponent solve that this thing M of u is e to the power mu plus half sigma squared x squared.

So, the Gaussian is a particularly simple moment generating function, it is a exponential of a polynomial in u which stops at the quadratic level. So, this immediately tells you this implies that K of u equal to log M of u and that is equal to mu u plus half sigma squared mu squared. So, for a Gaussian and only for a Gaussian the first moment is mu the first cumulant is mu the second cumulant is a variance sigma squared and all the higher cumulants vanish identically all of them not just the third or fourth but everything vanishes identically and that is a great help that is a very great help. Now, why do I need the cumulants, why not work with the moments it itself the reason has to do with the fact if you are looking at a single variable single random variable then whether you use moments or cumulants is not very not very significant does not matter and the cumulants are helpful because you give some interpretation to them, as I did a few minutes a ago, but you see most of the time you are concerned with sums of random variables linear combinations of random variables, so probably even more complicated functions of random variables then look at what happens.

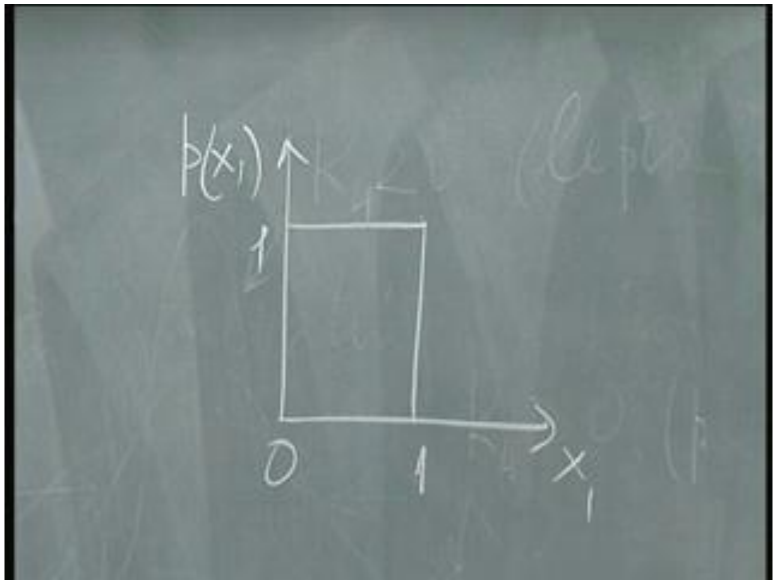
Suppose, you have two random variables and you want to ask and they are independent of each of other and you want to ask what is the probability distribution of the sum of these random variables then what would you do well.

(Refer Slide time: 33:05)



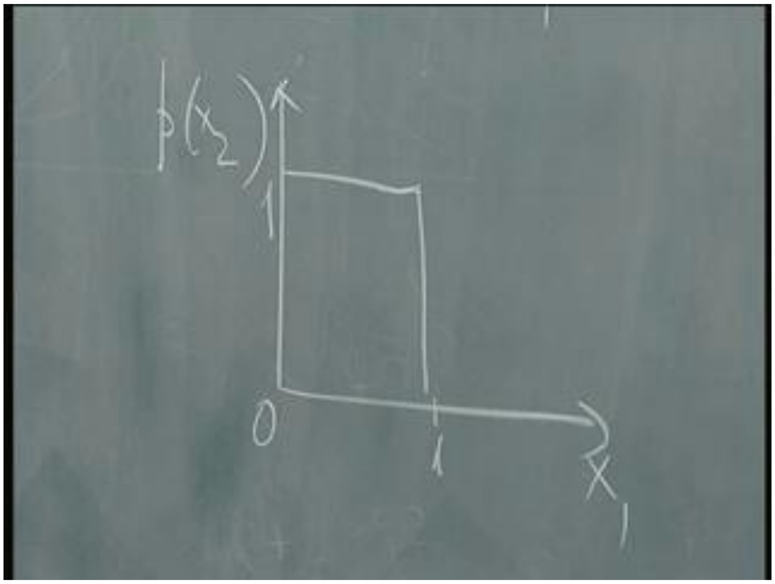
So, let us suppose you have two variable**s** x 1 and x 2 and they are independent random variables independent random variables, so they are very very dependent of on each other and I would to know what is the probability distribution of the sum of these two random variables, some linear liner combination of the two random variables, so suppose this has a probability distribution or density p one of x 1 and this is p two of x two, then I ask what is the probability density p of z, z equal to x 1 plus x 2 let us look at the sum to start with what should I do remember these are independent of each other completely x 1 and x 2 vary independently if we look at several examples let me ask the following question.

(Refer Slide time: 34:04)



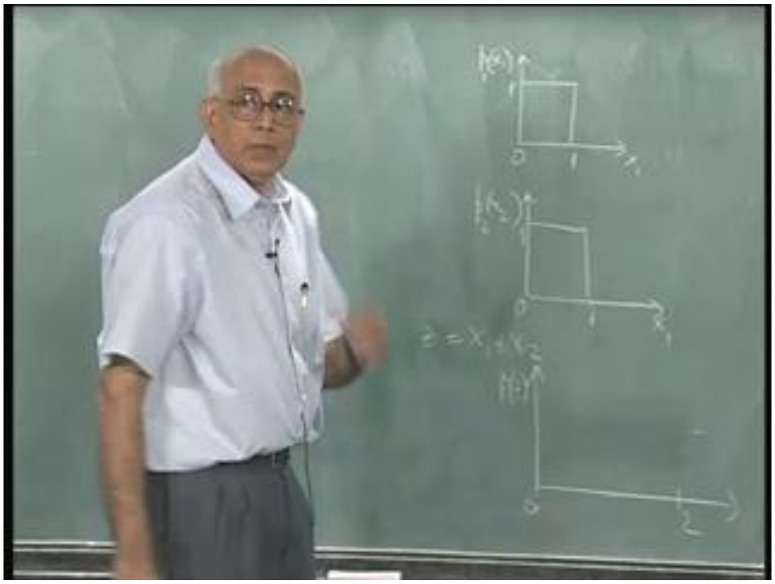
Suppose, I have a variable x, x 1 p of x 1 let us suppose this variable runs only between 0 and one and it runs uniformly between 0 and 1, so here is a random variable which can take on uniformly any value between 0 and 1 or be in any small interval d x 0 1 with equal probability, so uniform distribution normalized to unity, so that is the probability density function it is 0 outside here 0 outside there and it is one in between.

(Refer Slide time: 34:36)



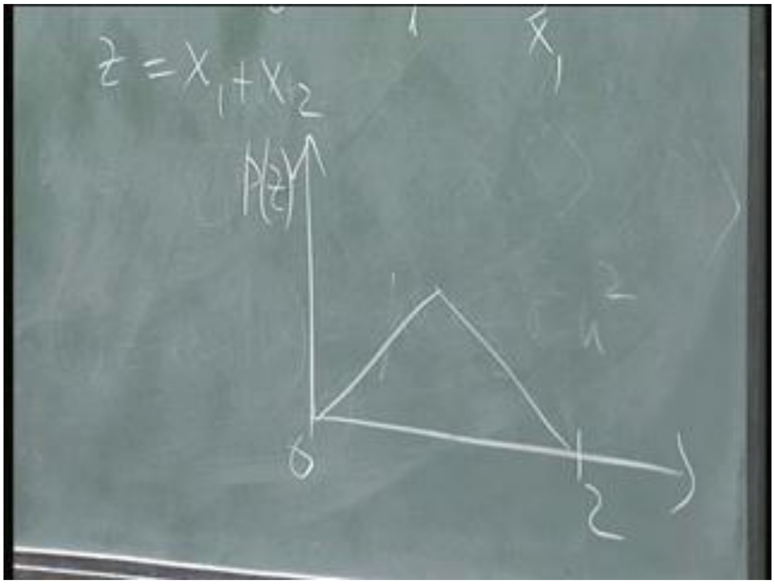
And let us suppose, x 2 is another such variable it is also identically distributed so p of x 2 is also small 0 one x 1 it stops at one, what is the distribution of the variable x 1 plus x 2 would it be uniform no, it is not uniform well Gaussian is not going to happen now because it is just two such variables right so would it be uniform, first of all this random variable z x 1 plus x 2 it runs from 0 to 2 now, because you have a sum and it is maximum value is actually two but you could take x 1 plus x 2 divided by two and then of course that brings it back to 0 to 1 what would it look like think of it this way, I have two coins I toss one the probability of head or tail is half if toss the other one the probability of head or tail is half, but now if I toss both, I have four possibilities, I have head head, tail tail, head tail, tail head and what does the distribution look like it looks like a quarter, quarter half, you certainly have a greater probability that you have a head and a tail and a tail and a head does not because you do not care which one is the head and which one is the tail so at once you have more accessible microstates in between than at the ends.

(Refer Slide time: 36:10)



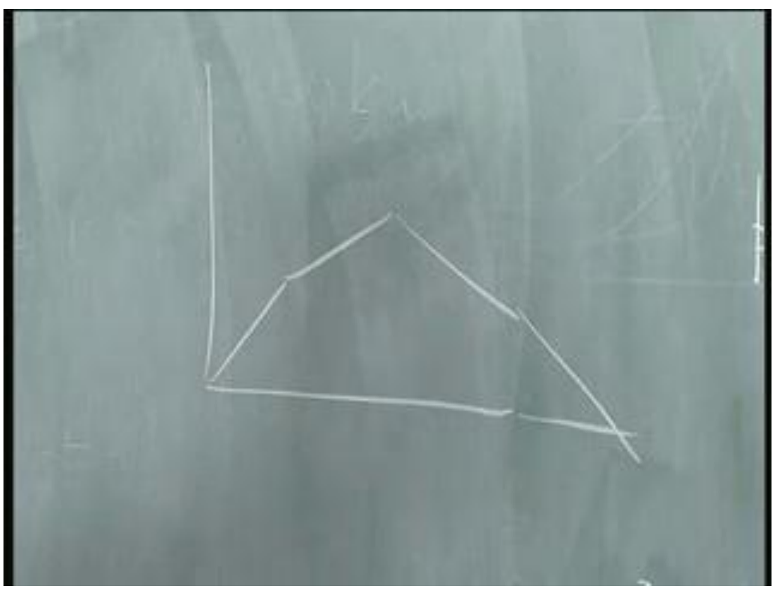
So here, if you say that the sum of the two z equal to x 1 plus x 2 and you ask what is the distribution of it look like now p of z, I am sorry, I am using the m symbol p here I should use another function it is a different function here, so we can let us call this p one this is p two and then this p of z this will go up to two and will start at 0 and the area under the curve must be equal to one. It is clear immediately that the region near 0 is not going to be weighted as much as the region near one, because for it to be near 0 both x 1 and x 2 have to be near 0 and for it to be near two both x 1 and x 2 have to be at the extreme at one, so it is a lot more likely that when this fellow is around here the other fellow is somewhere here and so on.

(Refer Slide time: 37:03)



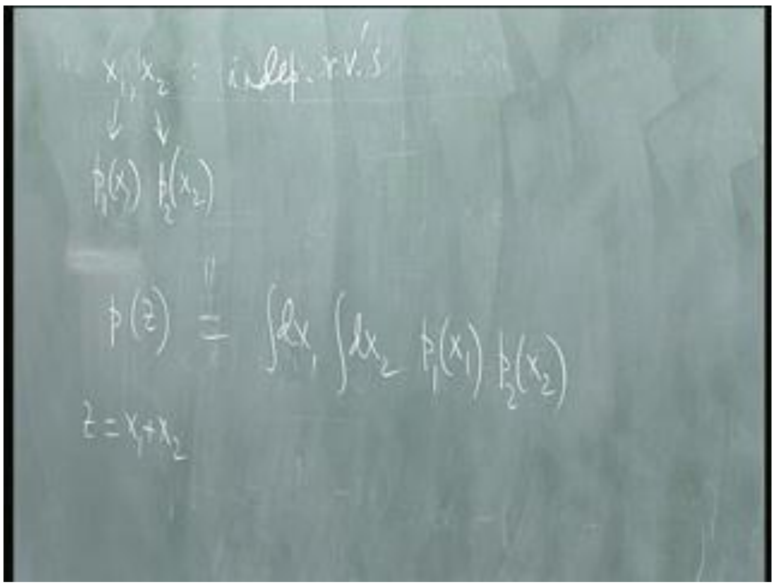
So, it is clear that it is distribution is going to go look like this you add one more variable then the range is from 0 to 3 and what would then it would start looking like would be you had three sets such variables you would do this up to 3.

(Refer Slide time: 37:19)



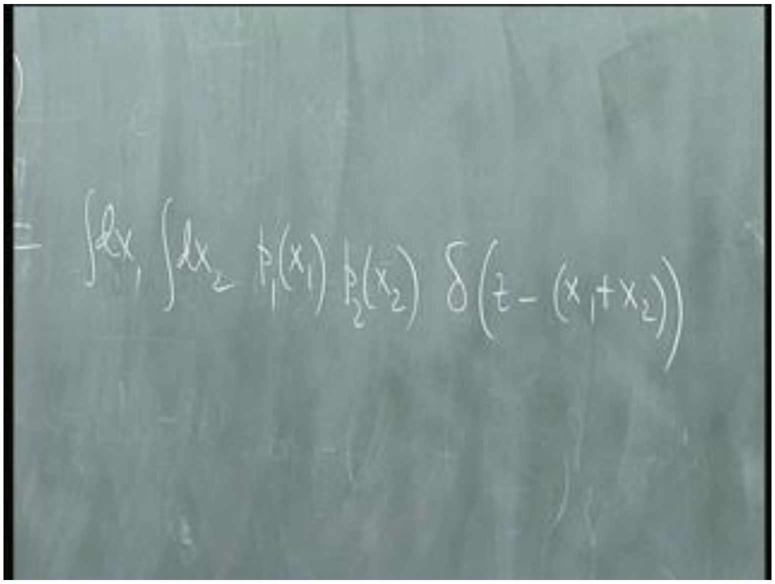
You add one more it is going start broadening out and at the same time it is going to get more and more smooth as you go along and after a while you cannot tell the difference between this and a smooth curve and you have a very large number of them then the whole average also shifts towards n or n over two, but if you subtract that n over two then this thing starts looking like a Gaussian provided you scale the variable properly and that is what the central limit theorem says, I will explicitly state this here so you can already see that the building up from here the fact that when you add a whole lot of random variables the probability that all of them are at one extreme or at the other extreme is small, but then lots and lots of accessible micro states in between and therefore the density sort of peaks up this is the gist of all these statistical theorems.

(Refer Slide time: 38:15)



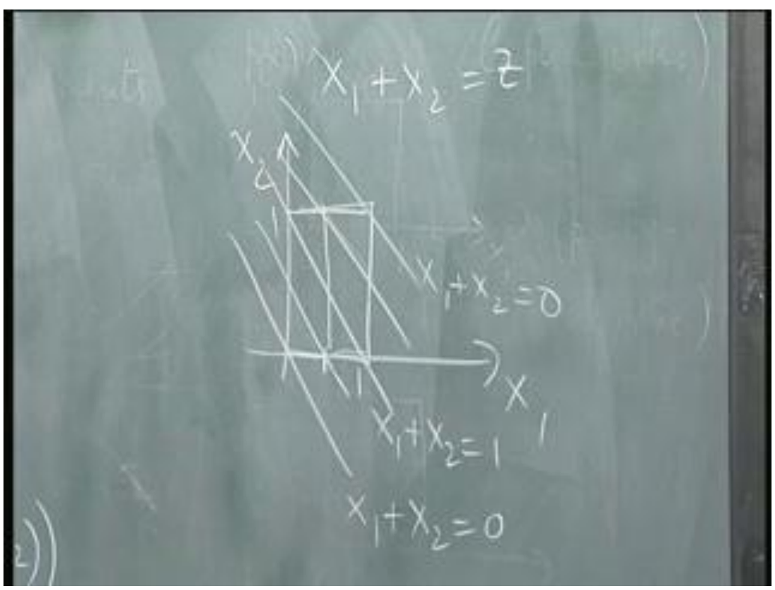
So, let us write this expression down we will see the usefulness of cumulants here this quantity is equal to a sum over all possible x ones a sum over all possible x 2 times the probability, that it is x 1 between x 1 and d x 1 that is the value of the probability density at some point p 2 of x 2, but then there is a constraint which says that x 1 plus x 2 must be equal to your given number z. So, what should I put, I should integrate over x 1 and x 2 subject to the statement that the sum of x 1 plus x 2 is some given number z and then of course as z changes it becomes a function. So, how do I impose this constraint, how will I impose a constraint that x 1 plus x 2 must be equal to z?

(Refer Slide time: 39:10)



Delta. So, I should put delta here delta z minus x 1 to x 2 subject to that constraint, so that is my definition of the probability density and I have to do this integral and of course this depends on what the ranges of these fellows are the delta function may fire may not fire, I have to be careful in doing this, so in the example I gave between 0 and one I urge you to do that.

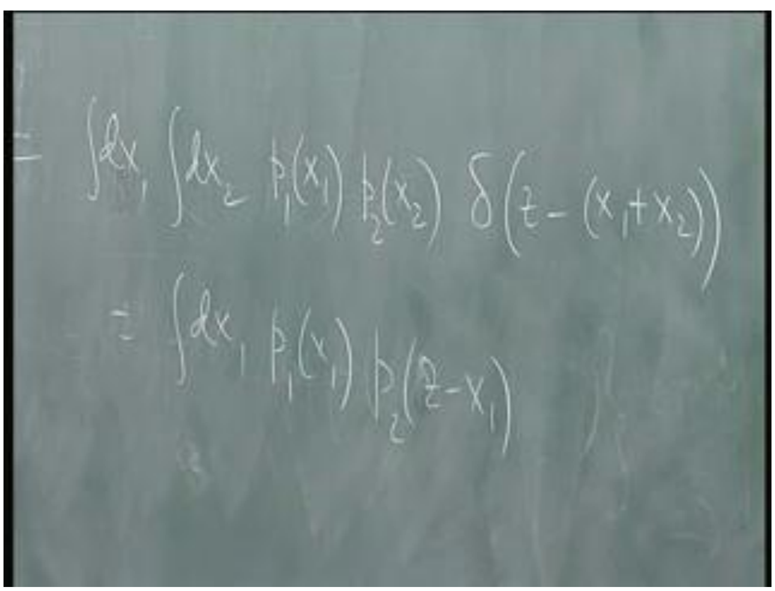
(Refer Slide time: 39:40)



I urge you to do that, because there what happens is if you plotted x 1 here and x 2 here and I plot x 1 plus x 2 equal to z this is a straight line whose slope is minus one, so in fact this is the straight line x 1 plus x 2 equal to 0 this is the straight line x 1 plus x 2 equal to one if this is one and the other guy is two and so on. So, now when you are integrating between from x 1 from 0 to 1 x 2 from 0 to 1 you have a square over which you have to integrate but the constraint is that x 1 plus x 2 must be equal to z. So, if you take sum z between 0 and one this is the constraint then it is clear that when you do the integral over one of the variables and get rid of the delta function the other variable must be in such a range that the delta function fires, so this is the cut off in your x 1 when you are doing the x 2 integration and as this straight line moves up to 2 this is x 1 plus x 2 is 2 you can see the range over this is smaller then it increases and it again become smaller and that is the reason for the triangular shape and after it crosses the diagonal it is a different range.

So, earlier the range in x 1 was this but at this stage the range from x 1 is from here to there, so there is a little jump in the slope and this is the reason why this jump occurred. So, it is clear that you can formally use the delta function but you have got to be very careful to make sure that it really does fire that it stays within a range of integration.

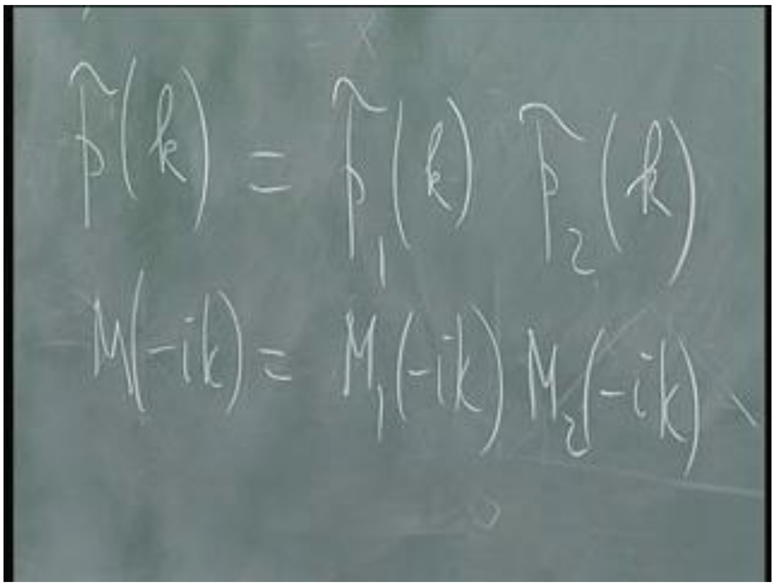
(Refer Slide time: 41:38)



But formally, if these fellows go from minus infinity to infinity for example then the delta function always fires and I could write this as integral d x 1 p 1 of x 1 p 2 of z minus x 1.

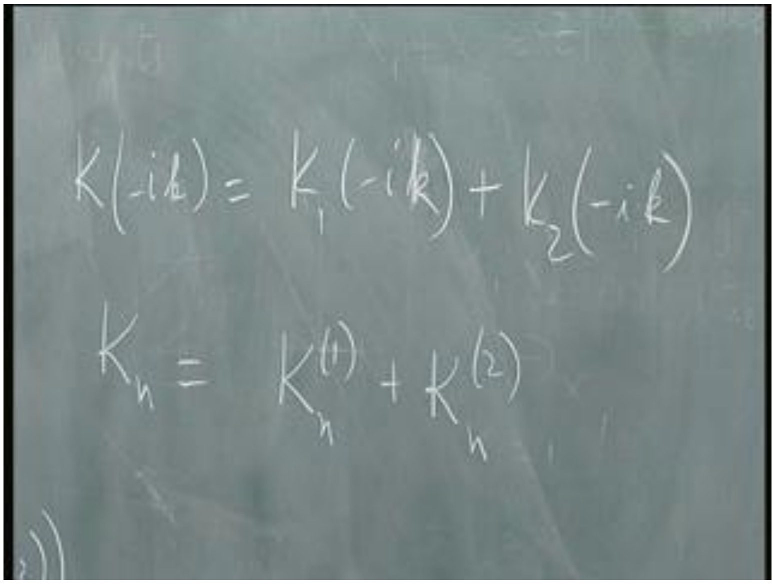
Now, what does that look like what sort of integral is that it is a convolution integral, so it is obvious immediately the corresponding Fourier transforms will multiply.

(Refer Slide time: 42:01)



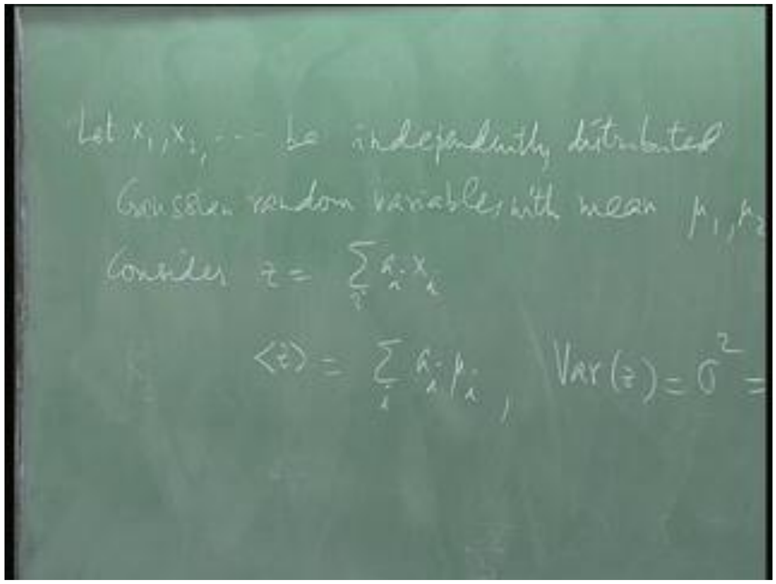
So, it is immediately obvious that if it implies that p tilde of k equal to p 1 tilde of k p 2 tilde of k where these are the characteristic functions of the two distributions or densities of p one and p two or if you like say this as m of minus i k equal to m 1 of minus i k times m 2 minus i k where m 1 and m 2 are the moment generating functions of the variables x 1 and x 2 and m for the sum now that is a product what should you to do get this product simplified take logs then it adds up.

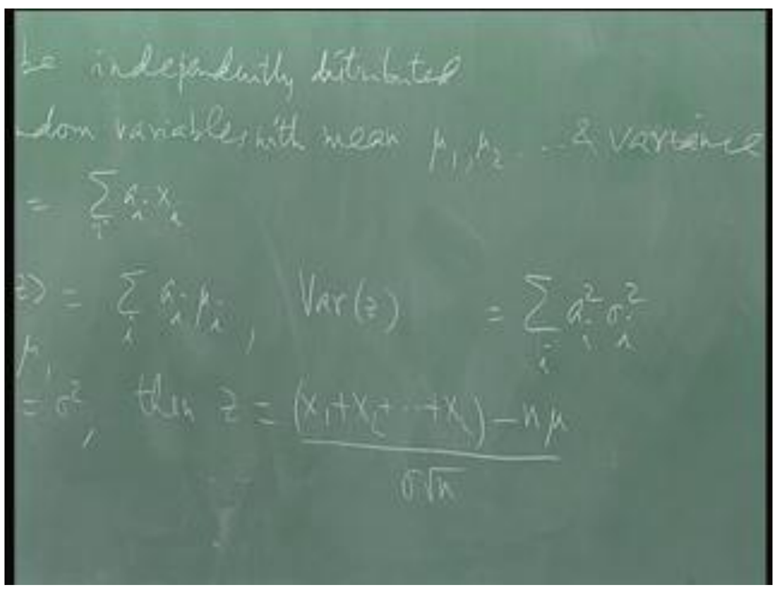
(Refer Slide time: 42:52)

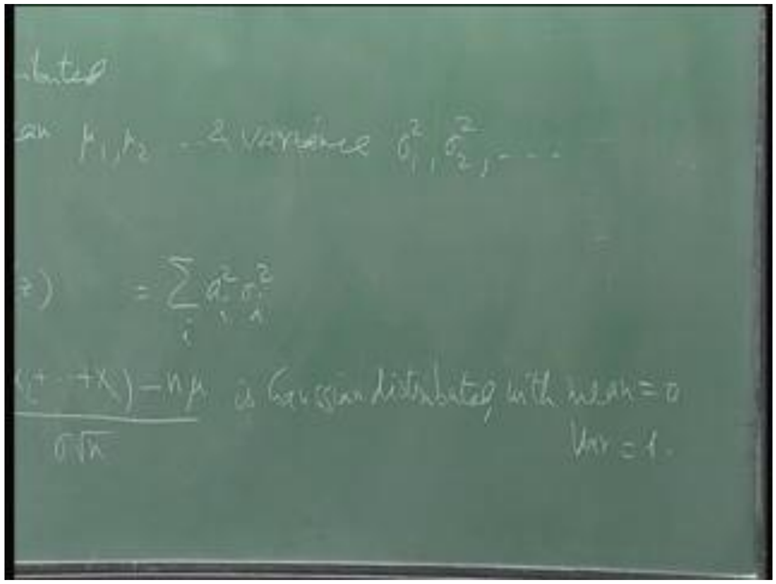


So, it immediately implies this thing here implies that the moment generating function of the cumulants or the cumulant generating functions k of a minus i k equal to k 1 of minus i k plus k 2 minus i k each of these is a power series in k or u therefore you can compare coefficients and it immediately tells you that k power n for the sum is equal to the sum of k power n for the first variable plus k power n for the second variable so the cumulants add up the moments do not you got to take a log first then it does that, so this is a great advantage of cumulants when you have independent random variables and you look at linear combinations of them then the corresponding cumulants would add up that is one of the reasons why you use cumulants at all you can generalize this to higher variables more and more variables of course it would not be in such a simple form, but if each of these variables is itself a sum of random variables and so on then this thing keeps on telescoping and you have a sum of the cumulants finally in the case of the Gaussian the following magic happens.

(Refer Slide time: 44:23)







It turns out that if let x 1 x 2 etcetera be independent independently distributed they are independent random variables Gaussian random variables, so suppose they are all Gaussian random variables they are all independently distributed then it turns out we work out through this algebra it will turn out that if you consider z equal to summation a i x i over i, so a i's are some numbers. So, a general linear combination of all these guys things then it turns out that z is also a Gaussian random variable, so a linear combination of any number of independent Gaussian random variables is also Gaussian it is probability distribution is also Gaussian. Now, how is a Gaussian specified, it has that exponential distribution form with two parameters the mean and the variance.

So, if I tell you the mean and the variance of this sum that the job is done finished so suppose it turns out with mean mu 1 mu 2 and variance sigma 1 squared sigma 2 squared suppose means of all these are mu one mu two etcetera. The variances are sigma 1 sigma 2 squared etcetera, then the question is what is this guy here it turns out the average value of z that is obvious is equal to summation i a i mu i and the variance of z equal to sigma squared equal to summation over i a i sigma i squared so that is it the answer is also Gaussian and it is specified by it is mean value and it is variance in particular and this is one statement of the central limit theorem in particular suppose there are all identically distributed variables with the same mean value mu and the same variance sigma squared.

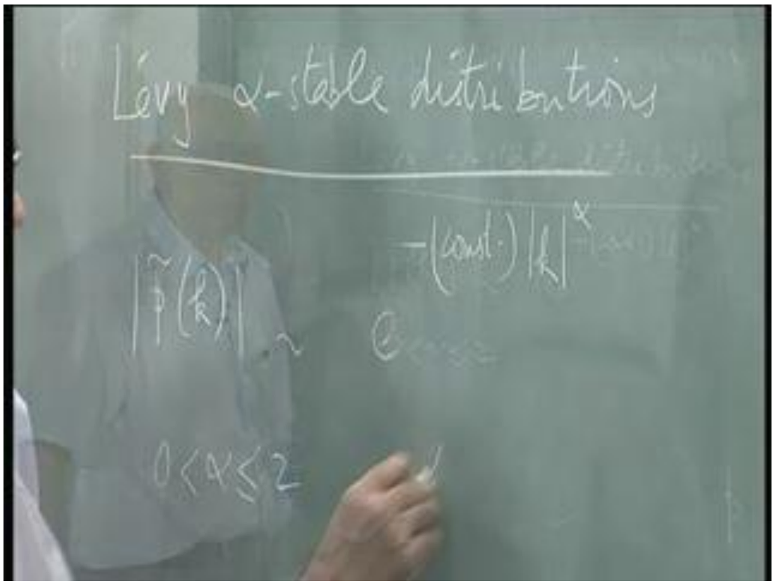
Then if mu one equal to mu 2 dot dot equal to mu and sigma 1 squared equal to sigma 2 squared dot dot dot equal to sigma squared bad notation.

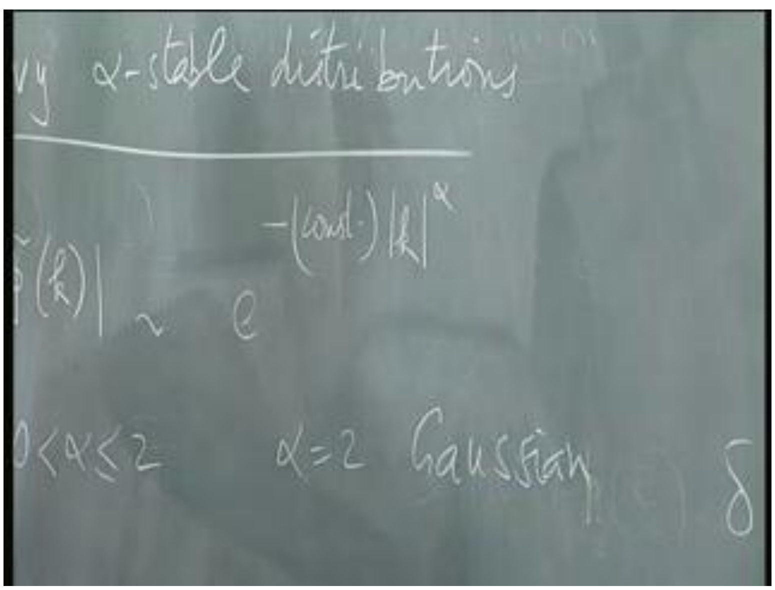
So, I should not do this variance equal to that, let us not call it sigma squared sigma squared is a common variance here then z equal to x 1 plus x 2 plus x n there are n of these fellows mine n mu, so you subtract out the common mean it is clear that this variable would have 0 mean, now because I removed that thing here but you have to scale it by dividing by sigma root n here this quantity here is Gaussian distributed with mean equal to 0 and variance equal to one that is why normalize it by doing this and the remarkable thing is that if the variables x 1 x 2 x 3 are not Gaussian distributed but they have some means and all their variances are finite and their means are finite then a combination of this kind a suitable combination of this kind in the limit in which n goes to infinity becomes a Gaussian even if these guys did not have Gaussian distributions just like I said they had uniform distribution between 0 and 1 but you add up a whole lot of them and the answer will turn out to be Gaussian suitably shifted and that is the content of the central limit theorem.

I will at some stage show you how a random box generates a Gaussian but this is the whole idea here it is a very powerful theorem it is the centre of statistics, if you like sort of crown jewel of statistics is the central limit theorem because it says if you have some effect which is caused by a number of causes they are all independent of each other they are all random working incoherently with each other then under suitable conditions fairly general conditions the effective random variables that you get finally is really going to have a Gaussian distribution and that is the reason why the normal distribution appears so often in physical applications almost everywhere. So, this root of it and there are generalizations of this whole thing here notice that I said you could ask is this the only kind of distribution that is so central to everything namely we add a whole lot of things I get a Gaussian are other distributions where, if I take identically distributed random variables they would also give exactly the same distribution when I sum them such distributions are known there are a family called infinitely divisible distributions such that if you have a whole lot of random variables all identically distributed by some distribution their sum also has the same kind of distribution.

These things for arbitrary values of sums the number of variables you sum over and they are called infinitely divisible and a sub class of infinitely divisible distributions are the so called stable distributions and the Gaussian is one member an extreme member of the family of stable distributions let me explain what a stable distribution is…

(Refer Slide Time: 50:57)

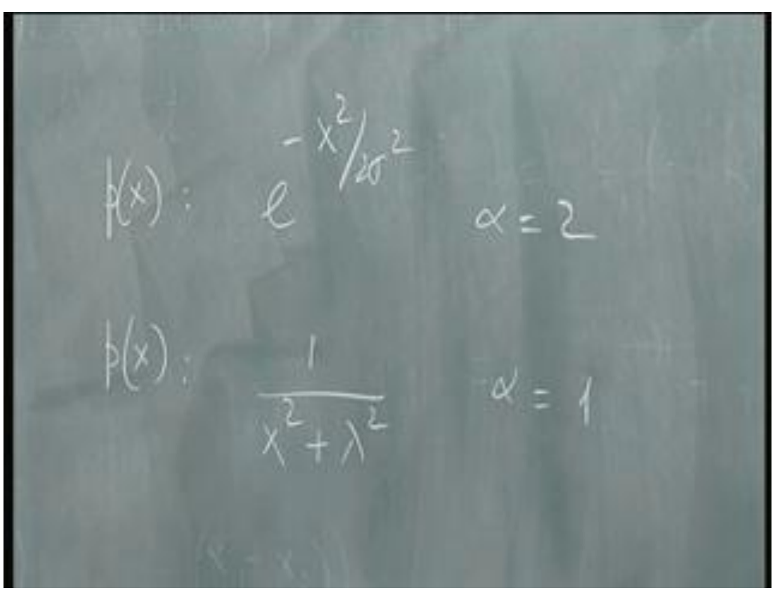




These are called Levy alpha stable distributions, there is a technical definition of these distributions but let me give a simple puristic definition here these are distributions for which the characteristic function of p tilde of k the Fourier transform of the probability density function it is magnitude goes like e to the power minus, in the case of the Gaussian remember the characteristic function was e to the power minus i k whatever it was plus k squared plus i k mu and there was a minus k squared sigma k over two so the magnitude was e to the minus k squared the magnitude here is like some constant modulus k to the power alpha the distribution the characteristic function it itself has a face factor there is an k to the i k whatever it is but it is a complicated expression well known expression, but what we need to know here is that the magnitude goes like e to the minus mod k to the power alpha.

Then this makes sense in the range 0 less than alpha less than equal to two, because you can show that it is only in that range where the inverse Fourier transform of p tilde of k will be non negative and therefore will qualify to be a probability distribution remember we want p of x to be a probability distribution, so that it should be non negative that happens one can show when alpha runs between 0 and two alpha equal to two is the Gaussian that is the most well known distribution here the most familiar and the most famous, one you can also show that it is only when alpha equal to two that the variance is finite the mean is finite the variance is finite as you keep decreasing the value of alpha the moments start diverging and for alpha less than two the variances are all infinite for alpha less than one even the mean value is infinite alpha equal to one also got a name it is called the Cauchy distribution and what it looks like is the following.

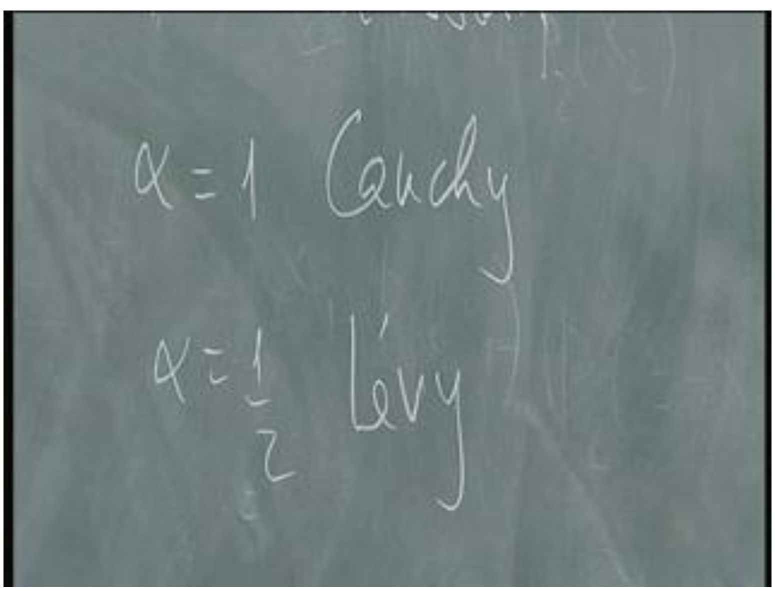
(Refer Slide time: 53:31)



So, if I look at p of x this is p to the minus x squared over two sigma squared this is alpha equal to two p of x for alpha equal to one the Cauchy distribution apart from a normalization factor it looks like sum x squared plus lambda squared it is what in physics you call the Laurentian shape and the second moment of this is infinite it is a symmetric distribution you have integral x squared over lambda squared minus infinity to infinity which diverges. So, the mean value the mean, value the variance of this distribution is infinite it is too broad it is a also symmetric by the way this is a member of a family of non symmetric distribution and the extreme case is this alpha equal to one skew parameter equal to 0 and then you end up with this you could also say the mean itself is infinite because you have to do integral x dx over dx over lambda squared even though you could say a diverges at the ends still an odd function and therefore it vanishes.

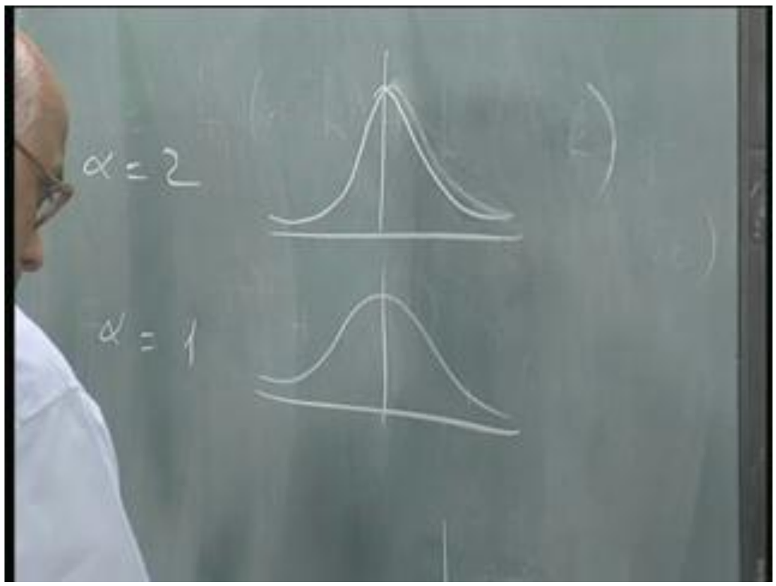
So, if you took the principle value for this integral for the Cauchy distribution you could say it still if alpha is less than one this is not true even the mean is infinite but it is got practical implications and I will mention in a minute in a minute what they are, I should say away that there are only three cases for which you can write the inverse Fourier transform of this distribution this characteristic function in a closed form in a simple closed form in terms of simple functions like this or this and…

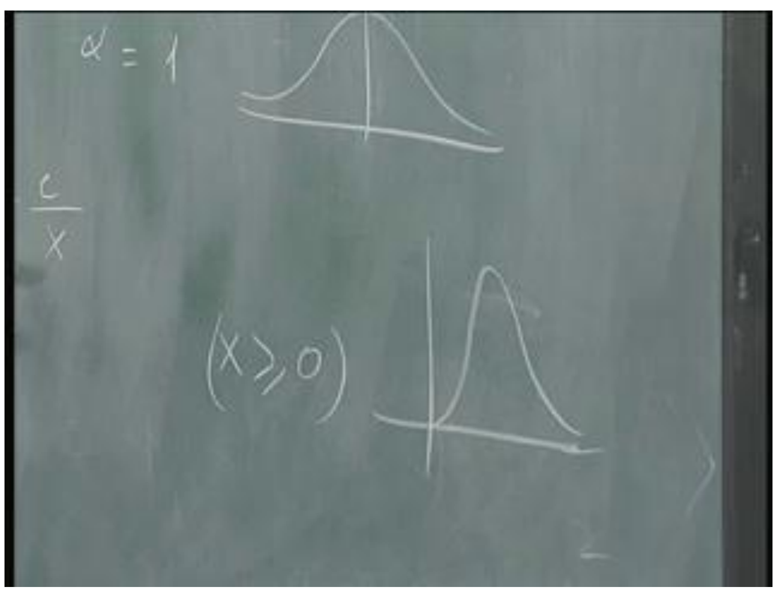
(Refer Slide time: 55:16)



So, on the other one is alpha equal to half and sometimes and very often it is called as Levy distribution itself. Not to be confused with the whole family of this distributions stable distributions and p of x for alpha equal to half looks like one over x to the three halves e to the minus constant over x and this is for x greater than equal to 0 one sided distribution.

(Refer Slide time: 55:43)

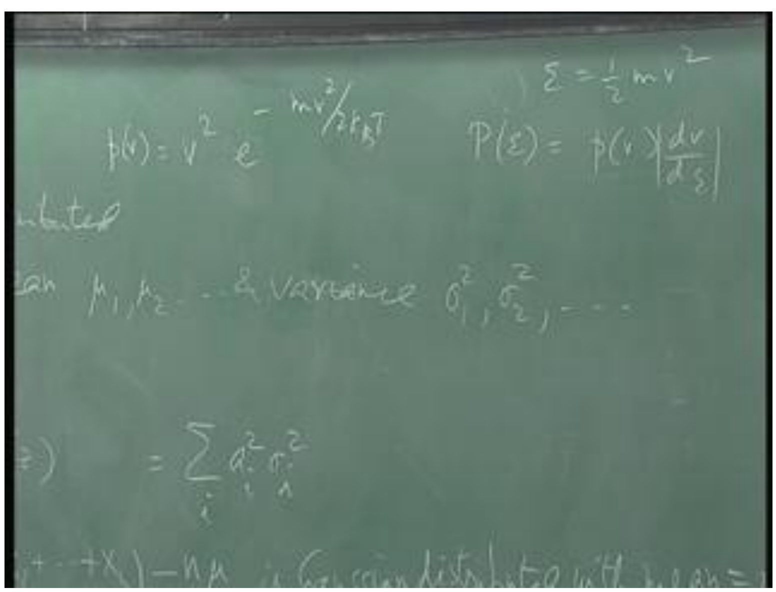


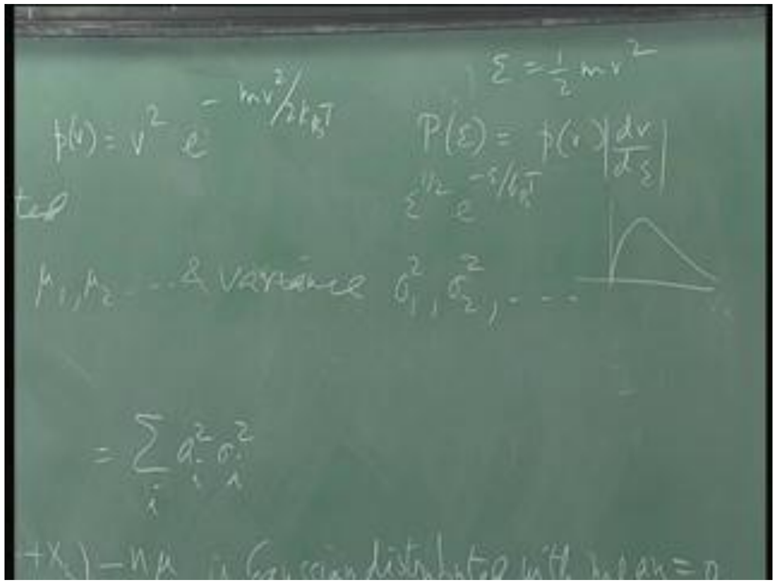


So, while this guy here looks like this the Gaussian this fellow is a Laurentian. So, it is a little more spread out this fellow here 0 diverges it is very flat at the origin, because this thing here goes to 0 very rapidly even though it diverges and at infinity it goes like one over x to the three halves so we can see the mean value of this distribution is infinite because you got to take the x over x and three halves up to infinity in that divergence this also appears in physical applications very often and as I said these are members of whole family and these distributions have many interesting properties among which is the fact that, if you have a set of identically distributed random variables with this with any of Levy's distribution as a common distribution linear combinations have exactly the same functional form as these distributions.

So, the law of distribution does not change and there is a generalized central limit theorem the counter part of that for alpha equal to two for general alpha there is such a central limit theorem, as I said these are also infinitely divisible in the sense that each such random variable could be broken up into other any number of pieces arbitrary number of identically distributed pieces for every end so that is why these are also called infinitely divisible and they have many other properties, now where would this occur well you know already that if I give you a random variable with a certain distribution a function of this random variable does not have to have the same distribution, for example we looked at the Maxwellian distribution there and that was for v x v squared, but if I asked you what is the distribution of the energy of the particle that does not look like this at all.

(Refer Slide time: 57:45)





Then what I have to do is to take the fact that the velocity the speed goes like e to the minus m v squared over two k Boltzmann T and I have to change variables form here. So, this is the distribution of the speed in three dimensions that goes like this, but if I now want to look at what is p of the energy epsilon, epsilon equal to half m v squared then I have changed variables and this will become equal to p of v dv over de and I have to express the right hand side in terms of energy and so and that Jacobian of the transformation also plays a role and what does this look like what does this guy look like, well it will look like epsilon to the power half e to the minus epsilon over k Boltzmann density of states actor multiplied by the Boltzmann factor what would be this guy look like well this distribution will look like this.

So, looks very different and is not a Maxwellian distribution and it is not a Gaussian or anything in the same way, if I give you two random variables each of which if Gaussian and I ask what is the ratio of the distribution of these two random variables that also runs minus infinity to then that answer turns out to be a Cauchy or I give you a random variable which is Gaussian and I ask what is the distribution of one over the square of this random variable. So, x is Gaussian I ask what is the distribution of one over x squared now one over x squared cannot be negative it runs from 0 to infinity and it turns out to be precisely a Levy distribution so there are deep connections between the members of the family of stable distributions there are deep dualities between these distributions etcetera. So, this a subject of considerable development lot of mathematics has been done by the statisticians studying these distributions and it has other implications as well.

Regarding the Gaussian, I should make a mention that it rapidly gets into very interesting things because you could ask what happens if the number of variables is infinite and then you could ask what happens, if the number of variables is continuously infinite and then you have distributions of functional you have Gaussian measures and so on, you have distributions of Gaussian fields as they are called and that leads to various other applications including applications in quantum physics areas etcetera. So, this takes off in a different direction all together, but they all start with this very humble initial Gaussian the fact that sums of Gaussian are all also Gaussian and so on. We have talked only about continuous variables look at discrete variables discrete value random variables then you have a parallel theory to this and then you can define the analogs of stable distributions in certain circumstances on the integers just for example.

For instance, a Poisson distribution is it means a random variable takes on only non negative values now you can easily show that the sum of two independently distributed Poisson random variables is also Poisson with means that add up and so on. So, certain theorems can be expended to these discrete values but you should not assume that is automatically true for arbitrary liner combinations because a difference of two Poisson variables cannot be a Poisson because the difference of two positive integers could be negative, I leave with as an exercise for you to find out what the distribution of two Poisson random variables each with the same mean is what is the distribution of the difference of these two variables this that difference could take on values of minus infinity to infinity over the set of integers. So, your exercise is to find out by doing this kind of a trick the delta function trick except now you should put chronic delta and summations instead of (( )). So, let me stop here and will come back when necessary to do some more distributions.