Probability and Statistics

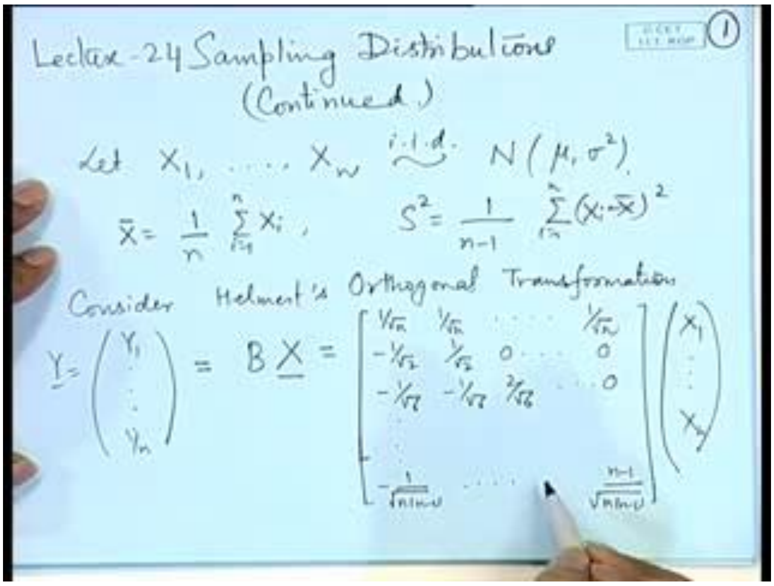
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 47
Chi-Square Distribution (contd.), t – distribution

We have introduced a sampling distribution called Chi-Square Distribution, and then I showed that if we are doing the sampling from a normal distribution then the distribution of the sample variance is a Chi-Square distribution therefore, chi square distribution is a sampling distribution. We used a moment generating function technique to derive the distribution of S square. Firstly, by proving that sample mean and sample variance are independently distributed, when we are sampling from a normal population.

Today I will give a alternative derivation by the method of transformations for the sampling distribution of X bar and S square when we are sampling from a normal population

(Refer Slide Time: 01:06)



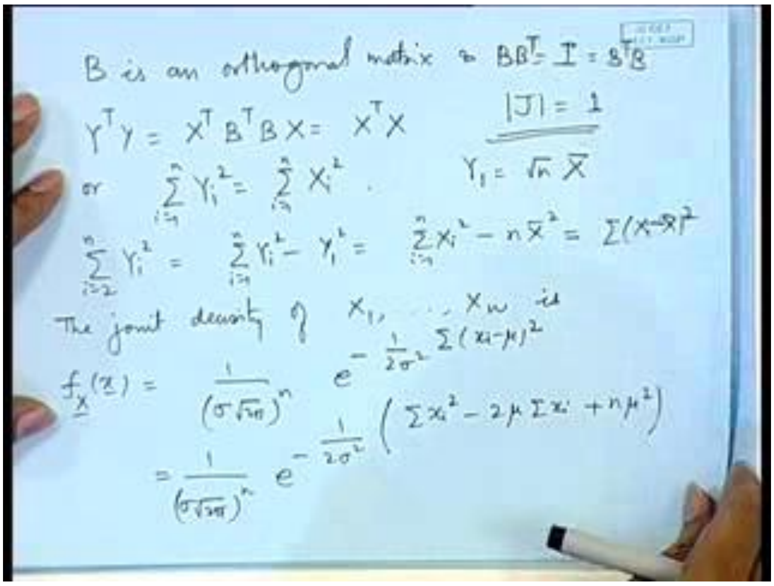
So, let us consider that X 1, X 2, X n is a independently and identically distributed normal mu sigma square random variables. So, we want to derive the distribution of X bar that is sample mean 1 by n sigma X i, i is equal to 1 to n and S square that is 1 by n minus 1 sigma X i minus X bar whole square. So, the method of proof will be through transformation. So, we will consider a transformation of the set X 1, X 2, X n in the following fashion

So, let us consider Helmert's orthogonal transformation; now this is a special transformation given in the following fashion that we define Y is equal to Y 1, Y 2, Y n as B X where B is the matrix of coefficients the first row is 1 by root n, 1 by root n and so on 1 by root n. The second row it is minus 1 by root 2 and 1 by root 2 and remaining terms are 0. The third one is minus 1 by root 6. minus 1 root 6 and 2 by root 6 and the remaining terms are 0; likewise if we continue in the last row we have minus 1 by square root n into 2 n minus 1 and so on finally, the last term is n minus 1 by root n into n minus 1 multiplied by X 1, X 2, X n.

So, first of all let us observe this B matrix this matrix is a special matrix which is called a Helmert's orthogonal matrix, the terms of the first row are same and if you take every next row then it is defined in such a way that if I multiply any 2 rows then the product will be 0, that is the scalar product of any 2 rows is 0 for example, if I take first and second then here it is minus 1 by root 2, and here it is plus 1 by root 2, so if I multiply the sum will give me 0. Suppose I take this with this then again the same thing because these 2 terms are same, so if I multiply here and add then this will become 0.

Similarly, if I talk these and multiply by the first row then minus 1 by root 6 n, minus 1 by root 6 n plus 2 by root 6 n so again the sum is equal to 0. So, this is a special matrix which is constructed for this purpose. Now let us see the effect of this, what we have done is that we have transformed the X 1, X 2, X n variables to new variables called Y 1, Y 2, Y n by means of this.

(Refer Slide Time: 04:32)



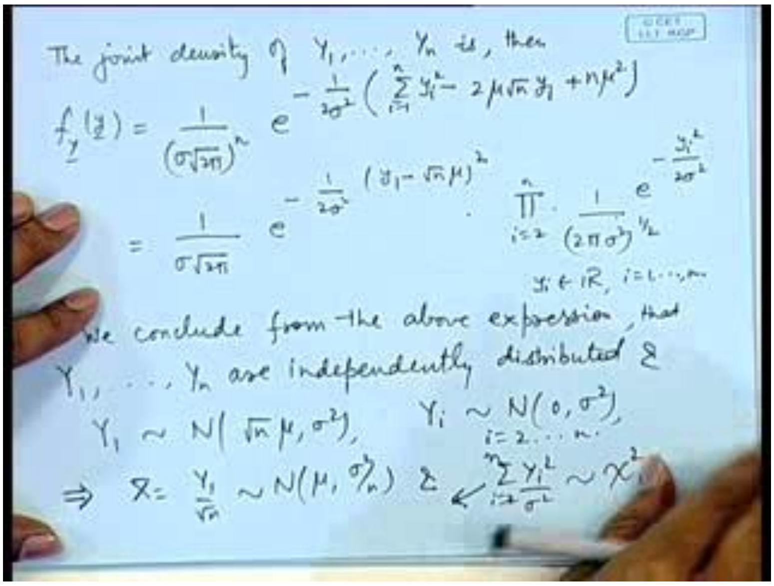
So, we have B is an orthogonal matrix and we have BB transpose is equal to identity matrix. So, if I consider Y transpose Y then that is equal to X transpose B transpose B, X that is equal to X transpose. So, this is also B transpose B that is equal to X transpose X; that means, sigma of Y i square is equal to sigma of X i square i is equal to 1 to n. That means, the original sum of squares is equivalent to the new sum of squares, also we can see here by this transformation that Y 1 is equal to root n X bar; because if I consider the multiplication of the this matrix with this vector and look at the first term and the first term will be X 1 by root n plus X 2 by root n plus X n by root n. So, that is root n X bar. So, Y 1 is root n X bar.

So, from here we get sigma Y i square i is equal to 2, 2 to n is equal to sigma Y i square minus Y 1 square which is same as now sigma X i square this is i is equal to 1 to n, sigma i is equal to I to n minus n X bar square, that is equal to sigma X i minus X bar whole square, which is the term which appears in the s square term that is the sample variance. Therefore, this new transformation is giving me Y 1 as well as sigma, that is which is X bar which is a term or X bar and which is another term which is a term of S square. So, our desired objective was to get the distributions of X bar and S square and this particular transformation helps us in at least representing these 2 terms in terms of transformed variables.

Now, let us look at the distribution so. Firstly, we write down the joint density function of X 1, X 2, X n. So, each of these X i's are normal mu sigma square variables. So, that distribution we write as 1 by sigma root 2 pi to the power n, e to the power minus 1 by 2 sigma square sigma X i minus mu whole square. Now this we expand we can write as 1 by sigma root 2 pi to the power n, e to the power minus 1 by 2 sigma square and now we get sigma of X i square minus 2 mu sigma X i plus n mu square.

When we consider the transformation Y is equal to B X and B is an orthogonal matrix, then we know that the determinant of an orthogonal matrix is either plus 1 or minus 1. So, Jacobian of the transformation that will be having absolute value 1; so this will be used for calculation of the transform density.

(Refer Slide Time: 08:09)



So, if I consider the joint density of Y 1, Y 2 Y n then it is obtained as. So, in the density of X 1, X 2, X n let us substitute the transformed values in terms of Y i's. So, sigma X i square will become sigma Y i square, and this sigma X i is nothing, but n X bar. Now X bar is Y 1 by root n. So, this also we can substitute and multiplied by the Jacobian of the transformation that is unity. So, we get the transformed density as 1 by sigma root 2 pi to the power n, e to the power minus 1 by 2 sigma square, sigma Y i square minus 2 mu root n Y 1 plus n mu square this is i is equal to 1 to n.

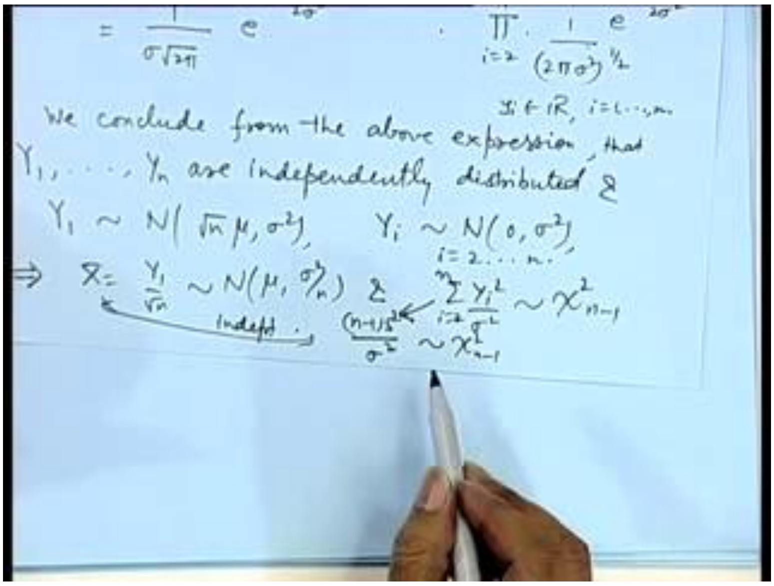
Now, in this particular exponent if I consider this sigma Y i square I take I write this as Y 1 square plus sigma Y i square from 2 to n. Now this Y 1 square minus 2 mu root n Y

1, n plus n mu square becomes a perfect square. So, we can represent it as 1 by sigma root 2 pi, e to the power minus 1 by 2 sigma square Y 1 minus root n mu square and the other terms we write separately as product i is equal to 2 to n, 1 by 2 pi sigma square to the power half, e to the power minus Y i square by 2 sigma square. The range of the transformed variables each of these Y i’s are also from minus infinity to plus infinity.

Now, you see this representation we are able to express the joint density of Y 1, Y 2, Y n as product of the certain functions where each function is strictly dependent only on each Y i. So, these are n functions. So, this is a function which is dependent upon Y 1 alone and here we have n minus 1 function each function is dependent upon Y 2, Y 3, Y n respectively. So, if we integrate with respect to Y i’s we will get individual terms; that means, Y 1, Y 2, Y n are independent and you are also able to say that these Y i’s are independently normally distributed, because of the form of the density. So, we conclude that we conclude from the above expression that Y 1, Y 2, Y n are independently distributed and Y 1 follows normal root n mu and sigma square and remaining Y i's follow normal 0 sigma square for i is equal to 2 to n.

So, this implies that since Y 1 is root n X bar, that is X bar that is equal to Y 1 by root n that will follow normal mu and sigma square by n and the sum of the squares of this Y i's divided by sigma square that is sigma Y i square by sigma square i is equal to 2 to n, that will follow chi square on n minus 1 degrees of freedom.

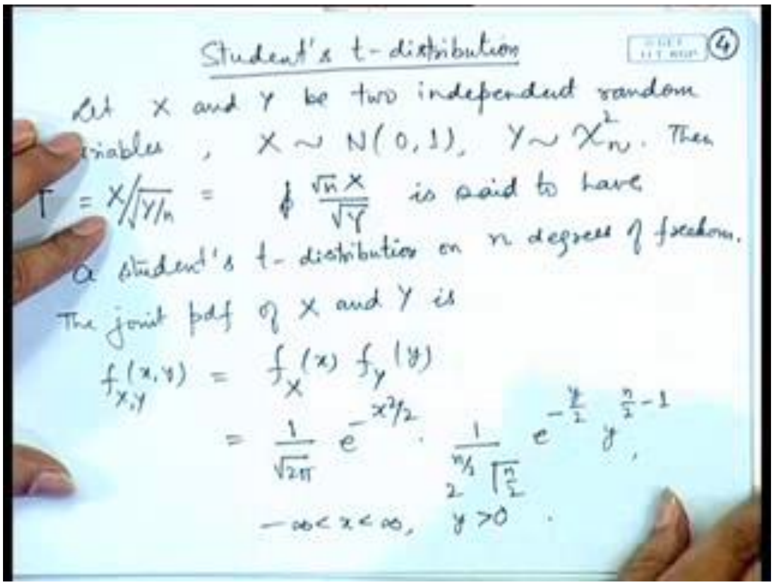
(Refer Slide Time: 12:28)



But this term is nothing, but n minus 1 S square by sigma square. So, that is following chi square on n minus 1 degrees of freedom and further these 2 are independent.

So, the result which we had proved using moment generating function, we have proved using transformations of the variables also. This Helmert's orthogonal transformation is quite useful, and that actually suggests that a procedure for obtaining the distributions of the sums of random variables and the squares of random variables. So, many times this statement is quite useful.

(Refer Slide Time: 13:21)



Now, we move over to another sampling distribution called t distribution. So, we call it students t-distribution, there is a story behind that why it is called a student's t-distribution, basically it was discovered by W.S Gosset a statistician in England; however, he worked in a brewery and therefore, it was not permitted for him to give his affiliation as working in a brewery. So, he used a pseudo name student and therefore, the distribution became famous as students t-distribution.

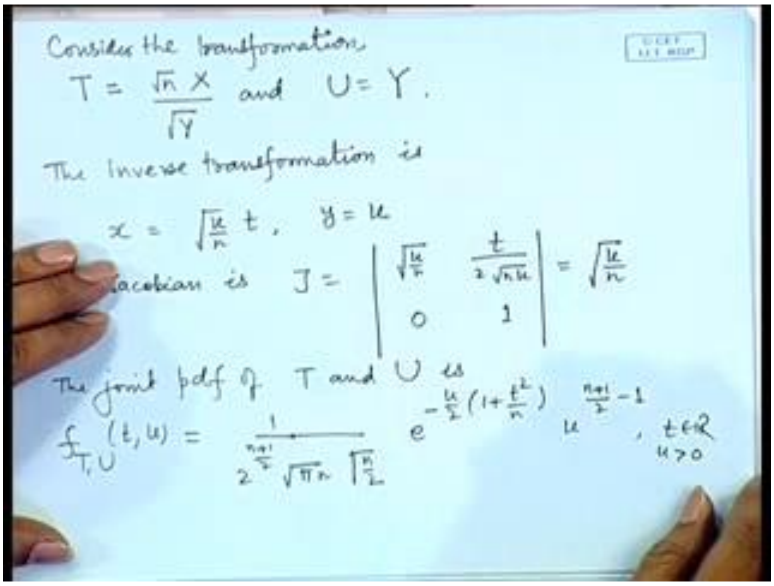
So, if I have let X and Y be 2 independent random variables, let us assume that X follows normal 0, 1 distribution and Y follows a chi square distribution on n degrees of freedom. If I consider the ratio X divided by Y by n and square root. So, let me write it as square root n X by square root of Y, let me call it say T, then this T is said to have a student's t-distribution on n degrees of freedom.

Now, this degrees of freedom terminology is coming from the chi square distribution, where we express that what is the meaning of the terms degrees of freedom, a chi square distribution on n degrees of freedom was represented as the sum of squares of n independent standard normal random variables. So, in the definition of t distribution I am using the degrees of freedom of chi square, and therefore, this t distribution is said to have set to be on n degrees of freedom.

Now, since here X and Y are independently distributed random variables, the derivation of the density of t is an exercise of deriving distribution of a function of random variables. So, we can write down the joint distribution of X and Y and create a transformation, in which one of the variables will be defined by t, and some other variable and we derive the distribution.

So, let us do it in the following way. Firstly, we look at the joint probability density function of X and Y. So, it is equal to the product of the individual distributions of X and Y. Now X is normal 0, 1 so the density is 1 by root 2 pi, e to the power minus X square by 2, and the density of Y is chi square n that is 1 by 2 to the power n by 2, gamma n by 2, e to the power minus Y by 2, Y to the power n by 2 minus 1. Here the range of the X variable is from minus infinity to infinity and range of Y variable is positive.

(Refer Slide Time: 17:01)



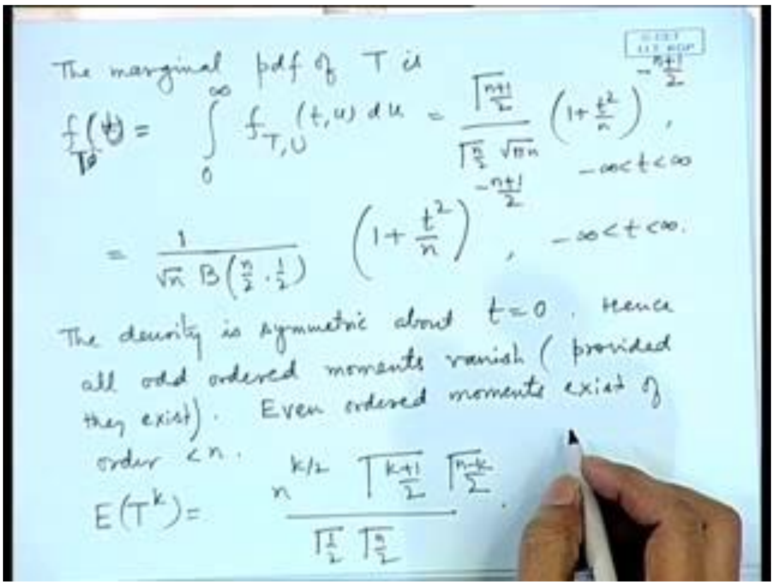
So, now we are considering the transformation T is equal to root n X by root Y. So, let us consider this transformation. So, the second variable we can consider as say V is equal to or say U is equal to Y, because we have to consider a one to one transformation or at least the number of variables should be same, so that we can find the joint density and then we can integrate out the not desired variable.

So, the inverse transformation here will be. So, x is equal to root u by n t and y is equal to u. So, we consider the Jacobian del x by del t that is root u by n, del x by del u that is t by 2 root n u, del y by del t that is 0 and del Y by del u that is 1. So, it is equal to root u by n. So, if we substitute this in the joint density of X Y and multiplied by Jacobian, we get the joint probability density function of T and U f T, U.

So, let us substitute the values here 1 by root 2 pi and all this thing is constant, so we combine it together, it becomes 1 by 2 to the power n plus 1 by 2, root pi n, gamma n by 2, e to the power minus u by 2, 1 plus t square by n, u to the power n plus 1 by 2 minus 1. So, this is after combining the coefficients and another thing you observe here that if we are making this particular transformation, the range of T remains from minus infinity to infinity and U is the chi square variable. So, U is positive. So, t belongs to r and u is positive.

To get the density of T we integrate this joint density with respect to U. So, if you integrate with respect to U from 0 to infinity you observe this term, it is e to the power minus u into something, and then u to the power some power, which is of the nature of a gamma integral or a gamma function.

(Refer Slide Time: 20:20)



So, it can be easily evaluated and we get the marginal probability density of T is let me call it f U that is integral f T, U d u from this is f T. So, I have written wrongly this is f T from 0 to infinity.

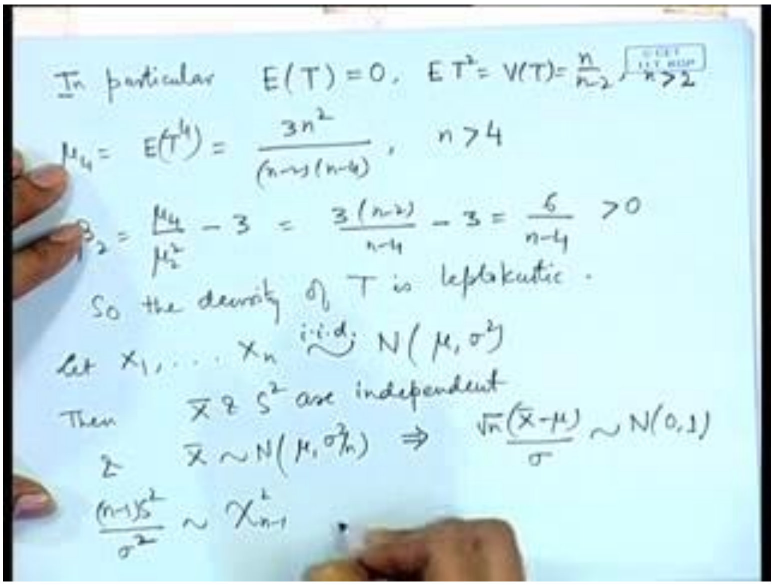
So, here the order of the gamma function is n plus 1 by 2. So, in the numerator I will get gamma n plus 1 by 2 divided by u as a multiple half, 1 plus t square by n. So, in the denominator I will get half, 1 plus t square by n to the power n plus 1 by 2. Now there is a 2 to the power n plus 1 by 2 term that will cancel out, so we are left with this density as gamma n plus 1 by 2 divided by gamma n by 2 root pi n, 1 plus t square by n to the power minus n plus 1 by 2 and the range of the variable is from minus infinity to infinity so this is the density of the t distribution on n degrees of freedom.

This particular coefficient we can write in a slightly different way also, because root pi if we observe it is gamma half. So, we can utilize the beta function notation and it becomes 1 by root n beta, n by 2, 1 by 2 and 1 plus t square by n to the power minus n plus 1 by 2.

Obviously if you look at his one the density is a symmetric function in t around 0, because if we replace t by minus t you get the same function. Another thing you observe that as t becomes large this will go towards 0, because the 1 plus t square by n term is in the denominator; also you observe that higher the power of n higher the value of n the conversions to 0 will be faster. So, basically that determines the shape of the t distribution, so we look at this thing.

The density is symmetric about t is equal to 0, hence all odd ordered moments vanish provided they exist, even ordered moments can be calculated. Now if you evaluate e to the power 2 k integral of this term, then you can reduce it to a gamma function and do the calculation. So, the even ordered moments exist of order less than n. So, we have expectation of T to the power k for the even ordered moment as n to the power k by 2, gamma k plus 1 by 2, gamma n minus k by 2 divided by gamma half, gamma n by 2.

(Refer Slide Time: 24:17)



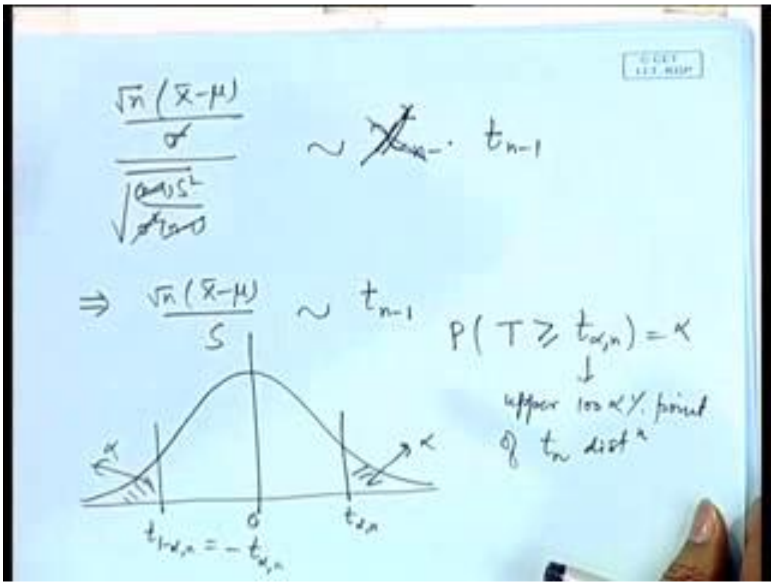
So, in particular expectation of T is 0, expectation of T square is that is variance of T that is n by n minus 2 which is existing for n greater than 2. You observe this is somewhat peculiar number n by n minus 2, as n becomes large this becomes close to 1.

Mu 4 is expectation of T to the power 4, that is 3 n square by n minus 2 into n minus 4, which is valid for n greater than 4. So, we can calculate the measure of kurtosis that is beta 2, that is mu 4 by mu 2 square minus 3. So, you look at this term we divide by n square, so this cancels out and we get 3 into n minus 2 by n minus 4 minus 3, which is simply 6 by n minus 4 so; obviously, this is positive, because we are considering n to be greater than 4, but you can observe here that if n becomes sufficiently large, then this number becomes small and that means, the kurtosis moves towards normality as n becomes large; in general it is leptokurtic density of T is Leptokurtic.

Now, let us consider this distribution as a sampling distribution, because right now I have given a distributional theoretic representation of this t variable, because we are writing it as only ratio of 2 variables in a particular form, but we can make use of the fact that chi square is itself is a sampling distribution, so whether we can represent t also in the same fashion. So, let us see.

If I consider a random sample from normal mu sigma square distribution, then X bar and S square are independent that we proved and what are the distributions? X bar follows normal mu sigma square by n, this means that I can consider root n X bar minus mu by sigma and this will follow normal 0, 1 n minus 1 s square by sigma square follows chi square on n minus 1 degrees of freedom, and this variable and this variable is independent.

(Refer Slide Time: 27:30)



So, if I make use of this definition of the 2 variables of the t variable, then I can write root n X bar minus mu by sigma divided by root n minus 1 S square by sigma square into n minus 1, this must follow chi square distribution on n minus 1 sorry this must follow t distribution on n minus 1 degrees of freedom. Now if you simplify this term here n minus 1 cancels out sigma cancels out. So, we are left with root n X bar minus mu by S, this follows t distribution on n minus 1 degrees of freedom therefore, t distribution is a sampling distribution.

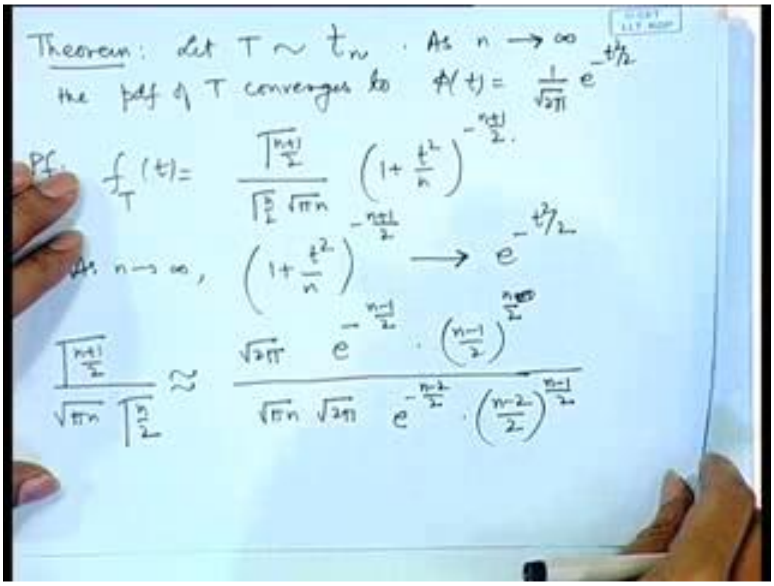
Another interesting thing is you can observe when I considered root n X bar minus mu by sigma this is standard normal, and here root n X bar minus mu by S is there; that means, sigma is replaced by S, later on we will see in the inference portion that S is actually an estimate for sigma. So, when sigma is not known we have to work with S and the distribution of that is known. In fact, in the context of this only this distribution was derived.

Now, regarding the probability points of t distribution, so the t distribution is a symmetric distribution about 0. So, if this point I call t alpha n, then the probability beyond this must be alpha; that means, probability of T greater than or equal to t alpha n is equal to alpha; that means, t alpha n is upper 100 alpha percent point of t distribution on n degrees of freedom. Because of the symmetry if you consider t 1 minus alpha n then that will be equal to t minus t alpha n; that means, if this value is alpha, then this point is t 1 minus alpha n by this definition, but because of the symmetry this will be equal to minus t alpha n.

Now, few things that we observe let us recollect that; when I wrote the density I said it is symmetric about 0 the odd ordered moments vanish even ordered moments can be found, mean is 0, the variance approaches 1 as n becomes large, the peakedness approaches normal p as n becomes large. So, these things and another thing I said that if you replace sigma by S then you have a t distribution, here you have a normal distribution.

This shows some sort of close similarity between t distribution and n t distribution and a standard normal distribution. Actually it is true; in fact we can prove that as n becomes large, the t distribution can be approximated by a standard normal distribution. So, we prove the following result.

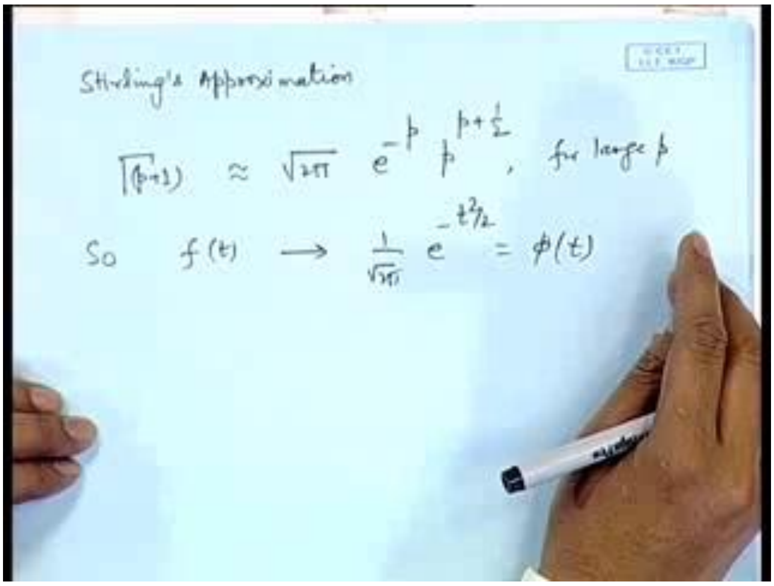
(Refer Slide Time: 30:55)



So, we consider let T be a t random variable on n degrees of freedom, then as n becomes large the pdf of T converges to phi t. Phi t is the probability density function of a standard normal random variable. So, to prove this let us write down the density function of t as derived that is gamma n plus 1 by 2 divided by gamma n by 2 root pi n and 1 plus t square by n to the power minus n plus 1 by 2. Now you observe this term as n becomes large or n tends to infinity, this converges to e to the power minus t square by 2.

So, as n tends to infinity, 1 plus t square by n to the power minus n plus 1 by 2 converges to e to the power minus t square by 2. So, let us look at the remaining terms we must actually prove that this remaining term converges to 1 by root 2 pi. So, if we look at this term n plus 1 by 2, gamma root pi n gamma n by 2.

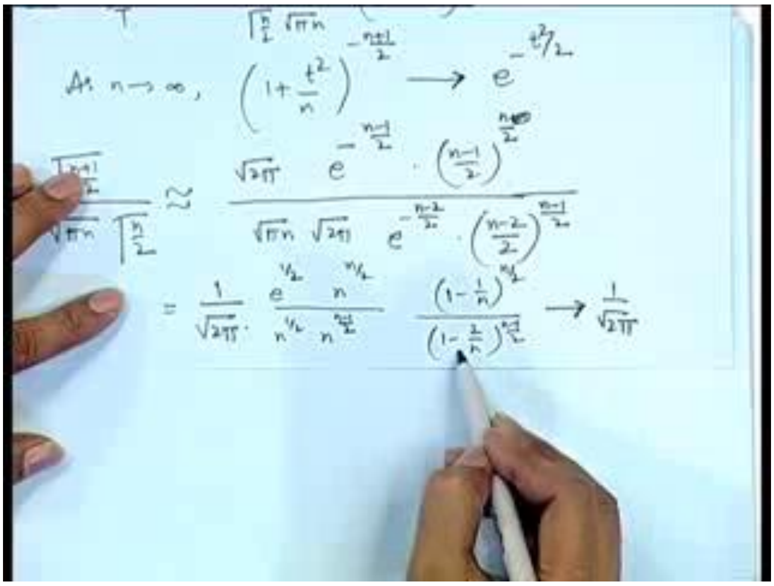
(Refer Slide Time: 32:48)



Now, there is a formula called Sterlings approximation. So, Sterlings approximation is let me write it here, that gamma p plus 1 can be approximated by root 2 pi, e to the power minus p, p to the power p plus half for large p; that means, for large p gamma function can be approximated by an exponential and binomial type of term. So, is a mathematical formula we can use it here. So, I am saying n is large then I can represent these things as root 2 pi, e to the power minus n minus 1 by 2, n minus 1 by 2 to the power n plus 1 by 2, no n minus 1 by 2 by to the power n by 2 and in the denominator you have root pi n, root 2 pi, e to the power minus n minus 2 by 2, n minus 2 by 2 to the power n minus 1 by 2.

So, we can do some simplification this root 2 pi etcetera will cancel out, and here you have 1 by 2 to the power n by 2, and 1 by 2 to the power n minus 1 by 2. So, 1 root 2 will come here.

(Refer Slide Time: 34:22)

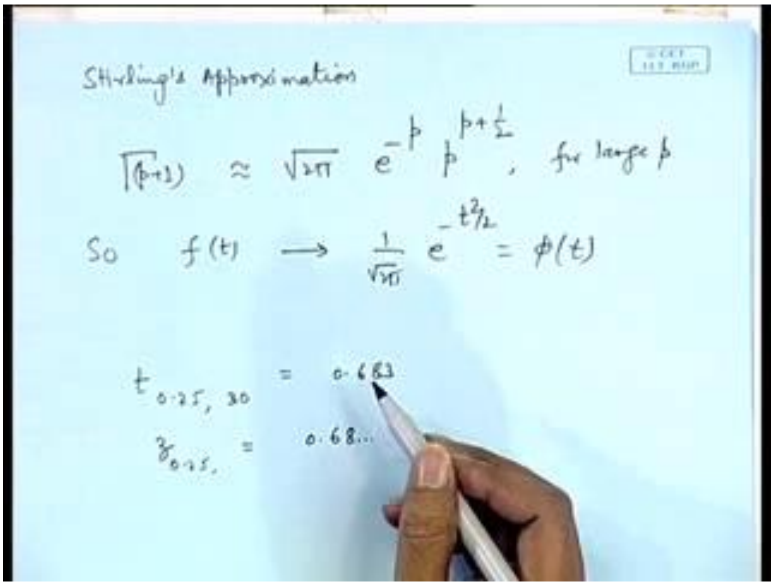


So, this is giving rise to 1 by root 2 pi and then we get this e to the power half and here I can take common n in the numerator and denominator. So, the those terms are getting canceled out that is n to the power n by 2, in the denominator I have n to the power half here and n to the power n minus 1 by 2. So, these all terms get canceled out and you are left with 1 minus 1 by n to the power n by 2, divided by 1 minus 2 by n to the power n minus 1 by 2.

So, if I take the limit as n tends to infinity, this goes to e to the power half and this goes to e therefore, the limit is simply 1 by root 2 pi, because this e to the power half e to the power half and e, they get canceled out. So, this proves that this f t converges to; f t converges to 1 by root 2 pi, e to the power minus t square by 2 that is the density function of a standard normal variance.

The question arises that for what sufficiently large value of n is this approximation good? The answer is that for n greater than or equal to 30, the approximation is extremely good and the tables of t distribution most of the times they show. So, if we look at a standard table of t distribution, unfortunately this cannot be seen here, but I will just write here the t value say at 0.25 and 30 is 0.683.

(Refer Slide Time: 36:13)



If I look at the same value for normal distribution then the point where the probability is 0.25 above the given point, it is 0.67 that is the if I consider that as z point, z 0.5 is equal to 0.68 and since this tables is given only up to 2 places, so I cannot predict here, but it is pretty close as you can see from there.

In fact, the tables are not given beyond 30 in most of the cases, because the approximation is extremely good. In fact, at 120 the value is almost equal.