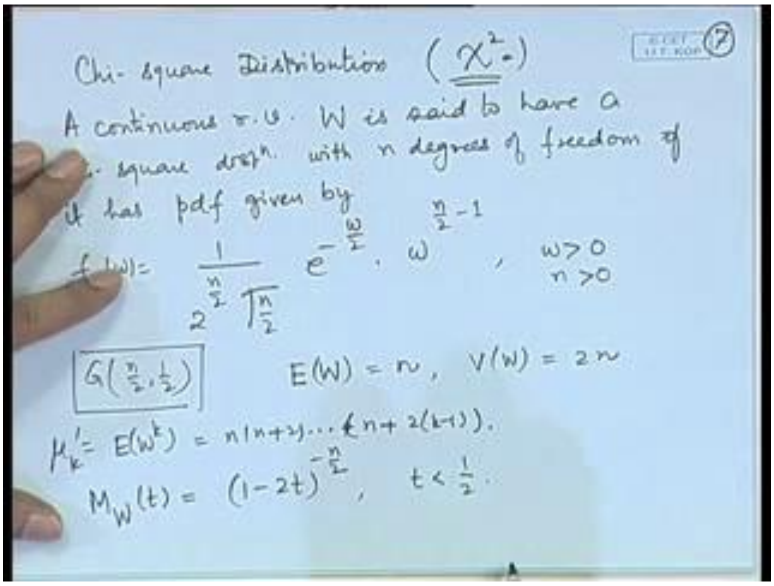
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 46
Chi-Square Distribution

Now, we discuss another sampling distribution which is known as Chi-Square Distribution.

(Refer Slide Time: 00:23)

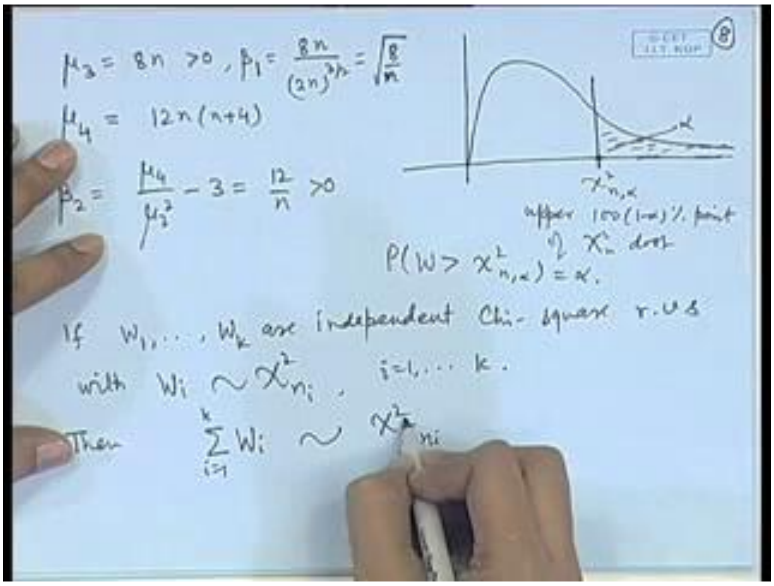


And it is used as Greek letter chi, so chi square distribution. So, a continuous random variable say W is said to have a chi square distribution with n degrees of freedom if it has probability density function given by say f w equal to 1 by 2 to the power n by gamma n by 2 e to the power minus w by 2, w to the power n by 2 minus 1; where w is positive and of course n has to be positive. If we see carefully it is actually nothing but a gamma distribution with parameters n by 2 and 1 by 2. So, this is only a special case of gamma distribution.

So, why we are calling it as a sampling distribution? So, we will show that this distribution arises in sampling from a particular population; that means we have certain characteristic for which this will be the distribution. Before doing those things let us look at the usual characteristics like mean variance and other things. Since, it is a gamma distribution we already know the mean it will be n by 2 by 1 by 2 that is equal to n. So, the term which we are calling as degrees of freedom is actually the mean of the chi square distribution. Similarly, if we look at the variance in gamma r lambda distribution the variance was r by lambda square. So, it becomes n by 2 divided by 1 by 2 square that is equal to twice n; so that is two times degrees of freedom.

We may write a general term like mu k prime that is equal to expectation of w to the power k that is n into n plus 2 and so on up to n plus 2 into k minus 1. We may look at the moment generating function that is equal to 1 minus 2 t to the power minus n by 2; this is valid for t less than half.

(Refer Slide Time: 03:12)



In particular we may look at the third central moment; that is 8 n. So obviously, it is a positively skewed distribution. Since gamma distribution is positively skewed, so of course, depending upon different values of n you will have different shapes for the chi square variable. But of course if we look at the measure of symmetry that is eta one that is 8 n divided by 2 n to the power 3 by 2, so it is becoming root 8 by n and as n becomes large this is approximately 0.

Similarly, if you look at mu 4 that is 12 n into n plus 4, and the measure of the kurtosis is mu 4 by mu 2 square minus 3 which is equal to 12 by n. So it is positive, so that means the peak is higher than the normal, but as n becomes large this is approximately normal. In fact, as n becomes large this is tending towards normality. That fact you can see from here also because it is 1 minus 2 t to the power minus n by 2. Here if I take limit as n tends to infinity, so we can after certain adjustments show that this will tend to the moment generating function of a normal variable. We will come to that later after representation of chi square is known.

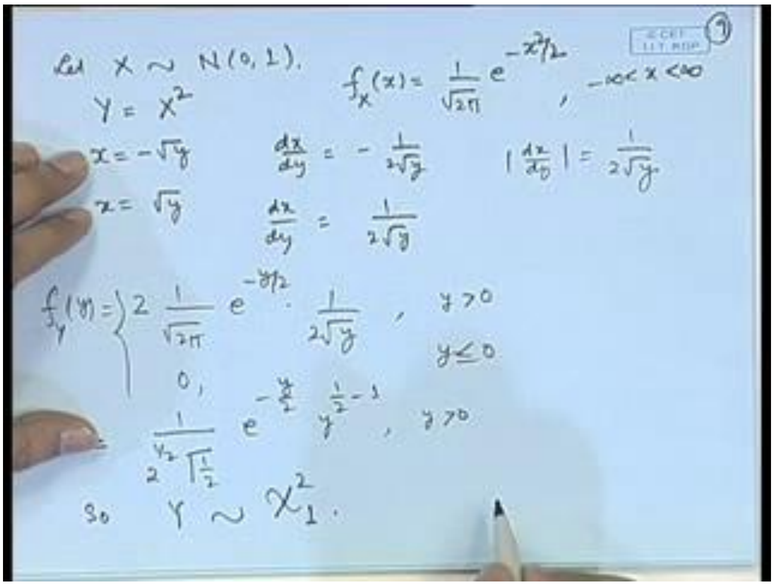
Now, depending upon the different values of n the shape of this will be different. And since it is a special case of gamma distribution the tables of gamma distribution can be used to determine the probabilities. However, tables of chi square distribution are available for a specific probability. So, if this probability is say alpha then the point on the axis is called chi square n alpha; that is upper 100 alpha 1 minus alpha percent point of chi square n distribution. That means probability of w greater than chi square n alpha is equal to alpha.

Now, we see that why this is a particular case of a sampling distribution. So, we will try to derive. Since it is a special case of gamma distribution we have already seen that in the gamma distribution if the scale parameter is kept fixed a certain additive property is satisfied. Therefore, if say w 1, w 2, w k are independent chi square random variables with say w i following chi square n i for i is equal to 1 to k. Then say sigma of w i i is equal to 1 to k that will follow chi square sigma of n i.

The proof is extremely simple because, if we apply the property that the moment generating function of the sum is the product of the individual moment generating functions if the random variables are independent then the distribution of the mgf of u will be product of the mgf’s of w i which will be 1 minus 2 t to the power minus n i by 2. So, if we multiply out this will become sigma n i by 2. So, the distribution of the sum of the chi squares is again a chi square and the degrees of freedom are added.

Next we look at the following result.

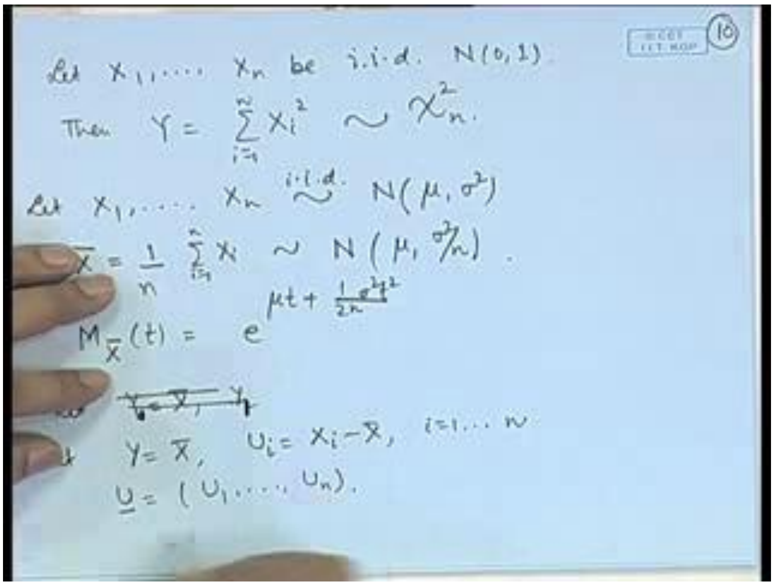
(Refer Slide Time: 07:25)



Let x follow normal 0, 1; let us define say y is equal to x square; we want the distribution of y. So, we look at the inverse transformation it is a 2 to 1 transformation the joined the density of x is given to be 1 by root 2 pi e to the power minus x square by 2 where x lies between minus infinity to infinity; x is equal to minus root y and x is equal to plus root y are two inverse images for any y positive. So, if you look at d x by d y term that is minus 1 by 2 root y or plus 1 by 2 root y. So, when we take absolute value of d x by d y in both the reasons it is 1 by 2 root y.

So, the density function of y is obtained as 1 by root 2 pi e to the power minus y by 2, 1 by 2 root y and second time again the same term will come so we will write two times, this is for y positive and 0 for y less than or equal to 0. So, this is equal to 1 by 2 to the power half root pi we can write as gamma half e to the power minus y by 2 y to the power minus 1 by 2 which we can write as 1 by 2 minus 1 for y positive. So, if we look at the density of a general chi square distribution here if you substitute n is equal to 1 then we get this density function. This proves that the square of a standard normal variable is a chi square variable with one degree of freedom; that is y follows chi square on one degree of freedom.

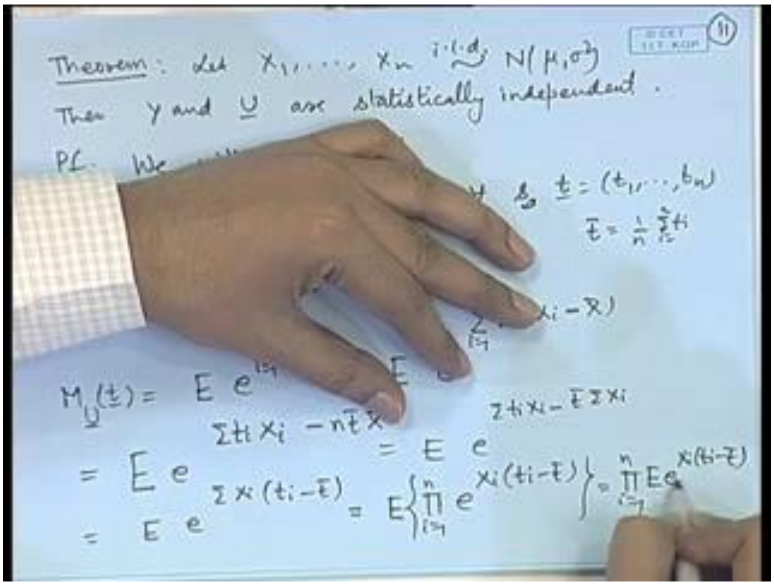
(Refer Slide Time: 09:40)



So, let us consider say x 1, x 2, x n independent and identically distributed say standard normal random variables, then y is equal to sigma x i square i is equal to 1 to n this will follow chi square n. Since we have already proved each of the xi square that will be chi square 1, if x 1, x 2, x n are independent then x 1 square x 2 square x n square also independent therefore the distribution of the sum will be the chi squares added, and since chi squares are satisfying an additive property this becomes chi square n distribution. Now you see if x 1, x 2, x n is random sample from a standard normal variable then sigma x i square is a statistic, and therefore chi square becomes a sampling distribution. We will consider a further elaborate description of chi square in the next section.

So, now let us consider say x 1, x 2, x n be a random sample from say normal mu sigma square, so in place of normal 0 1 now let us consider normal mu sigma square. So, if we define x bar as the mean then by the linearity property this will follow normal mu sigma square by n, therefore the moment generating function of x bar will be e to the power mu t plus half sigma square t square by n. So, we prove the following result; let us denote by say y 1 is equal to x bar y 2 and so let us put say y is equal to x bar and or let me change the notation, let say y is equal to x bar and say u i is equal to x i minus x bar for i is equal to 1 to n. Let us use the vector notation u for u 1, u 2, u n; then we have the following theorem.

(Refer Slide Time: 12:38)



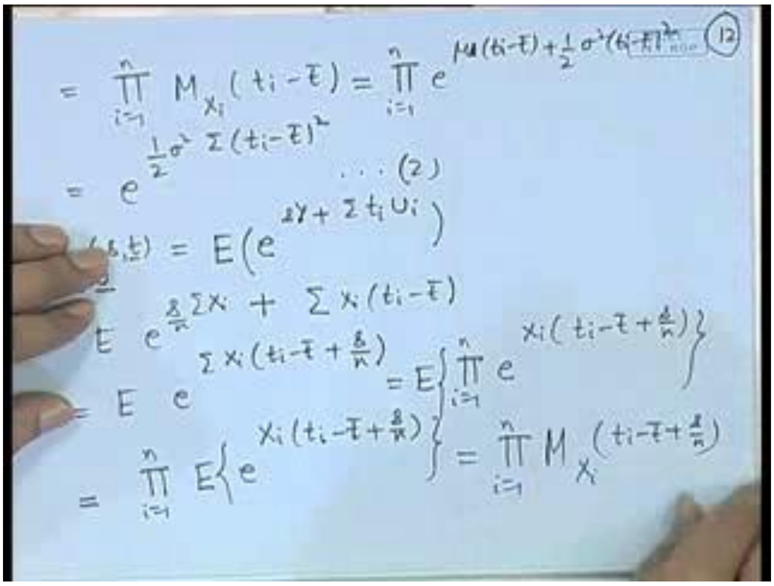
Let x 1, x 2, x n be a random sample from normal mu sigma square distribution then y and u are statistically independent. That means, x bar is independent of x 1 minus x bar x 2 minus x bar x n minus x bar.

Now, to prove this result we will use a moment generating function approach. We will show that the joint mgf of y and u at the point say s and t is equal to the moment generating function of y at s into the moment generating function at t; where t is equal to t 1, t 2 t n, so this is true for all s and all t. So, we need to evaluate the moment generating function of y u and the individual moment generating functions. So, already the moment generating function of y that is x bar is given to us. So, M y s is e to the power mu s plus half sigma square s square by n.

Now, we calculate the moment generating function of u also; that is expectation of e to the power sigma t i u i that is equal to expectation of e to the power sigma t i x i minus x bar. At this stage I will introduce some notation, so this becomes expectation t I; E to the power sigma t i x i. Now the second term here is minus x bar into sigma t I, so we use a notation say t bar as the mean of t i's. So, sigma t i becomes n t bar so n t bar x bar.

Now once again, since x bar is 1 by n sigma x i n x bar becomes sigma x i. So, we can again use it here, so this becomes expectation of E to the power sigma t i x i minus t bar sigma x i which is equal to expectation of E to the power sigma x i into t i minus t bar. Since the random variables are independent x 1, x 2, x n are independent random variables expectation of a product becomes the product of the expectations. However, we notice that these expectations are nothing but the moment generating functions of x i at the point t i minus t bar.

(Refer Slide Time: 16:10)



So, this is equal to the product of the moment generating functions of x i at t i minus t bar. Since x i's are normal we know this values so we substitute it here; this is product of i is equal to i to n e to the power mu i t i minus t bar plus half; this is not mu i this is only mu sigma square t i minus t bar whole square.

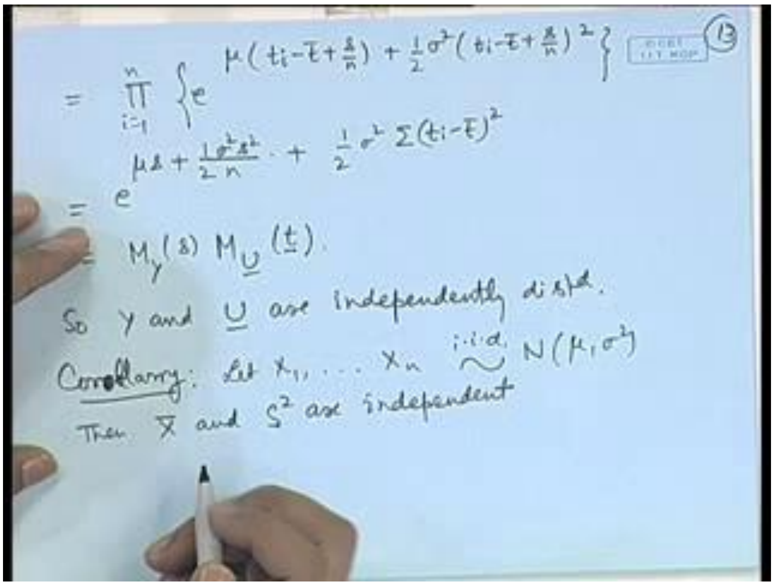
So, now we apply this product here. So, the first term vanishes and the second term becomes e to the power half sigma square sigma t i minus t bar whole square. So, we have calculated the right hand side here that is m y s is calculated here m u t is also now calculated. Now we calculate the joint mgf of y and u at a point s and t.

So, M y u at s t; so by definition of the joint mgf it is equal to expectation of E to the power s y plus sigma t i u i. The second term has already been simplified that is e to the power sigma t i u i has already been simplified as e to the power sigma x i t i minus t bar. So, we will use that here it becomes expectation of e to the power s sigma x i by n that is y is the mean so sigma x i by n plus the second term sigma t i u i we are using the simplification that we did just now that is sigma t i u i is equal to sigma x i t i minus t bar.

We again combine these terms here; so this becomes expectation of e to the power sigma x i t i minus t bar plus s by n. So, this we express as product i is equal to 1 to n e to the power x i t i minus t bar plus s by n. And once again since x i's are independent random variables expectation of the product becomes the product of the expectations; product i is equal to 1 to n expectation of e to the power x i t i minus t bar plus s by n.

So, notice here that now this has become mgf of x i at the point t i minus t bar plus s by n. So, it product of the mgf's at the point t i minus t bar plus s by n.

(Refer Slide Time: 19:32)



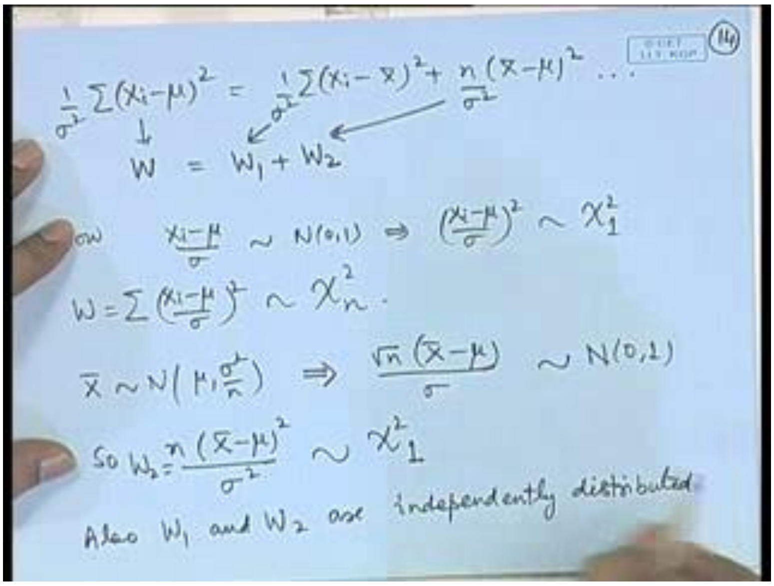
So once again making use of the fact that the x i as a normal mu sigma square distribution, we know the form of the mgf, so we substitute it here; that is equal to product i is equal to 1 to n e to the power mu into t i minus t bar plus s by n plus half sigma square t i minus t bar plus s by n whole square.

Now, this if we apply this product here e to the power sigma mu t i minus t bar becomes 0, on the second term you will get e to the power mu s. Here if we expand one term is t i minus t bar whole square and another term is s square by n. So, this we write here as half sigma square s square by n plus 1 by 2 sigma square sigma t i minus t bar whole square. The second term actually vanishes because cross product term will give s by n into t i minus t bar, so sigma of that will be 0. So, if we utilize the relations one, so the first term here is nothing, but the mgf of y at the point s and the second term from the equation number two is M u t. So, we have proved that the joint mgf of y and u is equal to the product of the mgf's of y and u respectively. So, y and u are independent.

So, as a consequence we have the following corollary; that is let x 1, x 2, x n be a random sample from normal mu sigma square distribution then x bar and s square are independent. That is in a random sampling from a normal distribution the sample mean and the sample variances are independently distributed. So, the proof follow the immediately if we notice that s square is a function of u. So since x bar and u are independent, therefore x bar and x square are independent.

Now, we will show that this is helpful to derive the distribution of s square.

(Refer Slide Time: 22:28)

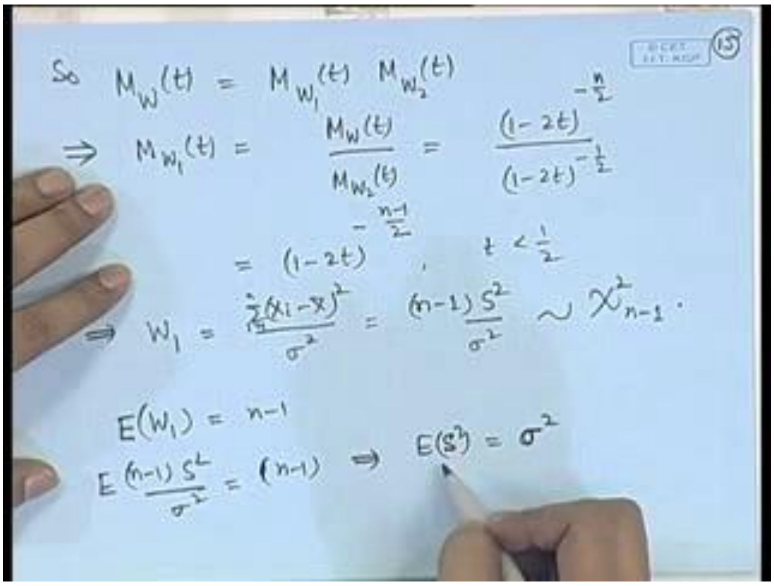


Now we look at the following quantities: consider say sigma of x i minus mu square, so here if we add and subtract x bar this becomes sigma x i minus x bar whole square plus n times x bar minus mu square. So, if we divide it by a sigma square here then we have this relationship. So, let us name these variables as say w is equal to w 1 plus w 2 say. So, this is w variable this is w 1 variable this is w 2 variable.

So, now if i have x i's following normal mu sigma square then x i minus mu by sigma follows normal 0, 1. This implies that x i minus mu by sigma x square follows chi square 1, and therefore the sum of these that is sigma x i minus mu by sigma whole square that is w this will follow chi square distribution on n degrees of freedom.

Further, the distribution of x bar is normal mu sigma square by n. So, from here we conclude that x bar minus mu root n by sigma this will follow normal 0 1 distribution. So, if I take the square n x bar minus mu by sigma whole square, so this will follow chi square distribution on one degree of freedom; this is w 2 variables. Also w 1 and w 2 are independent, because w 1 is a function of s square and w 2 is a function of x bar we have already proved that x bar and x square are independent. So, here we have written w 1 a chi square n variable as a sum of two independent random variables of which one of them is already a chi square.

(Refer Slide Time: 25:09)



So now if we use the moment generating function property that is M w t will be equal to M w 1 t into M w 2 t. So, this means that M w 1 t is the ratio of M w t divided by M w 2 t. Since some mgf of a chi square variable is known that is 1 minus 2 t to the power minus n by 2 divided by 1 minus 2 t to the power minus 1 by 2 because w 2 is a chi square 1. So, this becomes 1 minus 2 t to the power minus n minus 1 by 2 for t less than half. That means, w 1 that is sigma x i minus x bar whole square by sigma square which we can also write as n minus 1 s square by sigma square this follows a chi square distribution on n minus 1 degrees of freedom.

So, this means that chi square is a distribution of the sample variance after a certain scaling. So, this shows that chi square is a sampling distribution. Either we consider a standard normal random variable, so sum of a squares of n independent random variables normal random variables is chi square on n degrees of freedom or if we are considering arbitrary normal random variables then if we consider the scaled distribution of the sample sum of a squares from the deviation that is n minus 1 s square by sigma square then that is chi square on n minus 1 degrees of freedom.

Here we want to clarify one question that, although this is sum of n variables in fact each of x i minus x bar is a normal random variable, in fact x i minus x bar will follow a normal distribution with mean 0 because x i as mean 0 and x bar as mean 0 and variance will become sigma square 1 plus 1 by n. So, it is a sum of n squares of random normal random variables, but this is not independent because sigma x i minus x bar is 0, so only n minus 1 of these are independent that is why the degrees of freedom are your n minus 1.

So, in some sense these degrees of freedom can be related to the fact that a general chi square random variable is sum of squares of n independent squares of standard normal variables. So, when we consider any other then it need do not be. So, we have established here chi square as a sampling distribution. And in particular if we are interested to find out certain statement about s square then we can answer that.

For example, if we look at expectation of w 1 then it is equal to n minus 1. So, expectation of n minus 1 s square by sigma square follows so that is equal to n minus 1. So, this means expectation of s square is equal to sigma square. So, this means that on the average that is s square that is the sigma x i bar square divided by n minus 1 is unbiased for sigma square not divided by n, and that is why in particular we consider sample variance as where the divisor is 1 by n minus 1 not 1 by n because this is coming as an unbiased estimator for sigma square. In the inference course we will deal in detail about the criteria of un-biasness form.

In the next lectures we will take up other sampling distribution such as t and if f distributions. So today we will stop at this point.