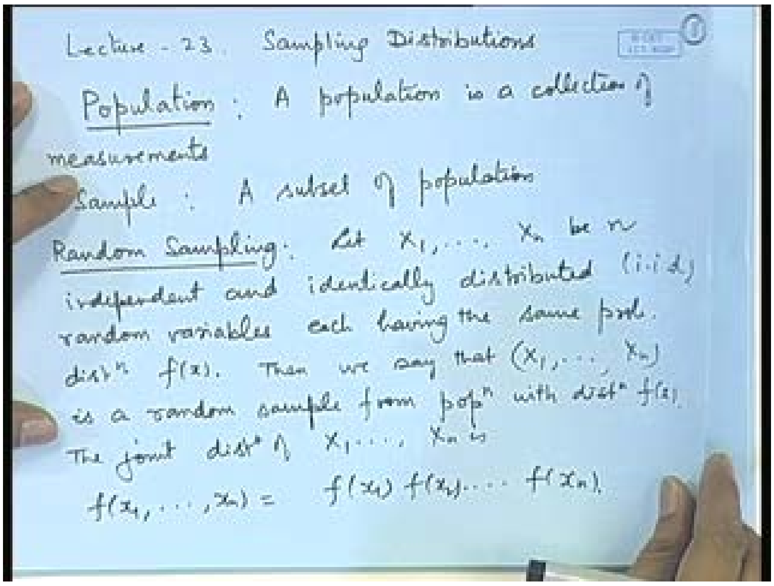
Probability and Statistics
Prof. Dr. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Module No. #01
Lecture No. #23
Sampling Distributions-I

(Refer Slide Time: 00:24)

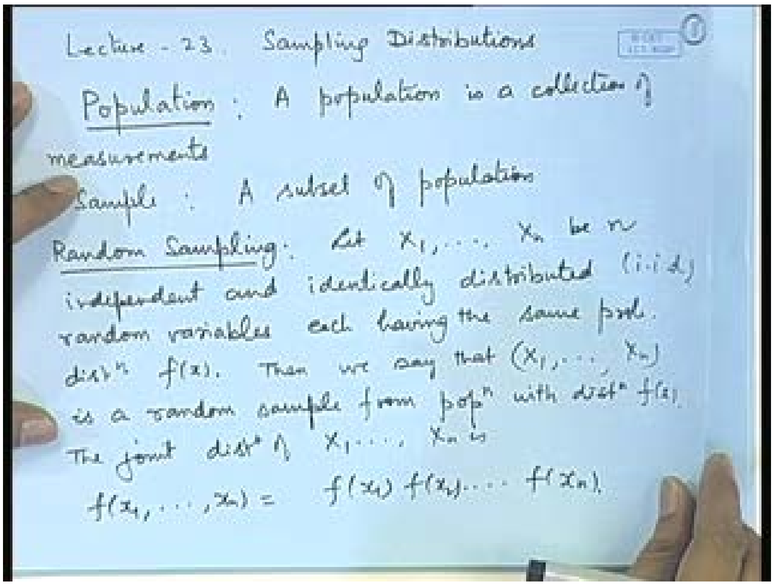


So, today we will introduce sampling distributions. So, first of all, we introduce, what we mean by a sample? What is a population? So, we introduce the term population. So, a population is a collection of measurements on certain characteristic. For example, if we are studying heights of people, then the measurements of the heights of our desired target population, that will be the statistical population. If we are interested in the lives of the people or longevity of the people, then, if we consider say, the total life, total age at death of a set of people, then that is our target population.

If we are interested in, say the number of smokers in a population then the characteristic of recording, that is whether a person is a smoker or not a smoker for a certain set that, is our target population. So, statistical population is a collection of measurements whether it is numerical or a qualitative measurements. A sample is a a subset of population. So, since it may not be possible to have the complete enumeration of the population, in various studies it is enough if we consider a certain sample of the population. So, a general random sample which we consider in statistics, is taken in such a way that the probability of selecting each observation is same; however, this is the methods of doing sampling, it is a part of another topic called sampling theory or sampling techniques.

In this particular course, we are assuming that we already have a random sample and then we proceed with that. So, what is a random sample in the context of distribution theory? So, we say that let x 1, x 2, x n be n independent and identically distributed; that is i i d random variables, each having the same probability distribution f x. Then, we say that x 1, x 2, x n is a random sample from population with distribution f x. And the joint distribution of x 1, x 2, x n is defined as f of x 1, x 2, x n is equal to product of f of x 1, f of x 2, f of x n.

(Refer Slide Time: 04:15)

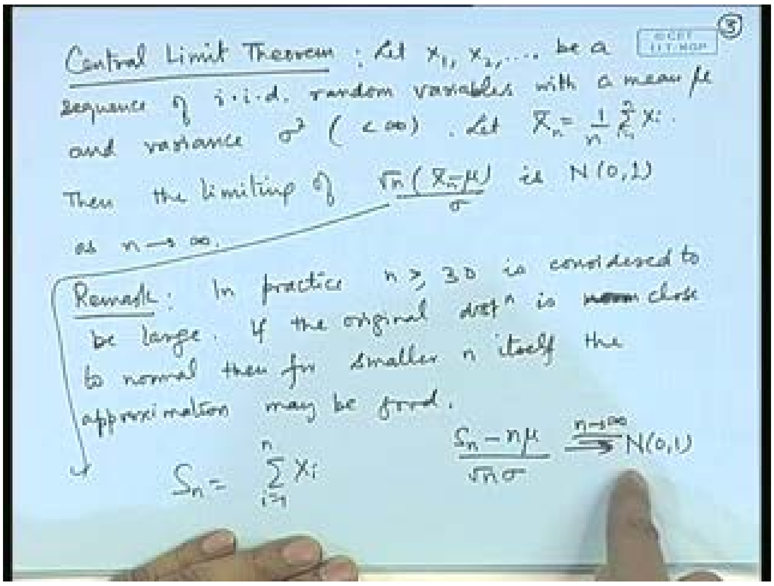


Any characteristic of the sample, we call it as a statistic. So, a function of random sample, let us say T; that is T of x 1, x 2, x n. This called a statistic. For example, we may consider x bar, that is 1 by n sigma x i, that is the sample mean. We may consider sample sum of squares from the deviation from the mean. We may consider 1 by n minus 1 of this, which we usually note by s square; that is sample variance. We may be interested in say sample median. We have already talked about ordered statistics. So, that is also a statistics. Sample median, we may define to be the x of n plus 1 by 2; that is n plus 1 by 2, a third order statistics if n is odd; that means, the middle order statistics or if n is even, then, we may take the mean of the middle two; that is x n by 2 plus x n by 2 plus 1 by 2.

We may consider say sample range; that is the difference between the largest and the smallest. So, these are Examples of certain statistics. And when we are dealing with a sample, we are interested in the characteristics and therefore, we will be interested in their distributions. So, the distributions of the statistics they are known as sampling distribution. So, we may formally define, a sampling distribution is the probability distribution of a statistic is called a sampling distribution. Now, as such here the distribution of x 1, x 2, x n is known. So, the joint distribution of the sample is known to us.

So, if we consider any function of that T of x 1, x 2, x n; the derivation of the distribution relates to the techniques which we have defined in the previous lecture, that is for transformation of random vectors; that means, we may consider say one variable as T of x 1, x 2, x n and we may define some other variables say u 1, u 2, u n minus 1. So, that we have a n to n transformation and we may determine using any techniques for determining the sampling distribution.

(Refer Slide Time: 07:34)



However, there are some particular characteristics such as sample mean or the sample variance which plays very important role and here we will consider the distributions of that. So, one of the first results which is related to the distribution of the sample mean is a quite important result, in the sense that, it applies to a very large number of situations, it is known as Central Limit Theorem. So, let x 1, x 2, and so on be a sequence of independent and identically distributed random variables. So, basically what we are saying is that, we are taking a sample with a large size. So, i i d random variables with a mean mu and variance sigma square, we assume it to be finite. So, if we are assuming that say x n bar is the mean of the n observations. Then, the limiting distribution of root n x bar minus mu by sigma is normal 0 1 as n tends to infinity.

That means, the standardized sample mean has a limiting standard normal distribution. Now, if we carefully look at the conditions of a theorem, this is pretty general. We are not making any assumption on the form of the distribution of x i's. All that we are assuming that the mean is given and the variance is given. In that case, the limiting distribution of the sample mean after a certain change of location and scale is standard normal provided the sample size is large. In fact, this is the result which places the normal distribution in the center of statistical theory.

What happens that in practice, when we are doing, when we are taking observations or measurements on certain thing, we are usually not taking one observation. For example, we may be measuring length of certain article. Suppose, if it is a physical experiment. So, we in place of taking one measurement, there is some measuring device and we take say thirty measurements and we take average of those measurements to say that, this is the actual estimate of the length of that Equipment.

So, in that case basically, what we are using is the x n bar, rather than individual x i. The same thing is used at various places. For example, if we are looking at the average crop per field, then we are not taking individual crops rather than we are taking a sample of the fields and then we take the average; that means, the crops of the individual fields and then we take the average of that.

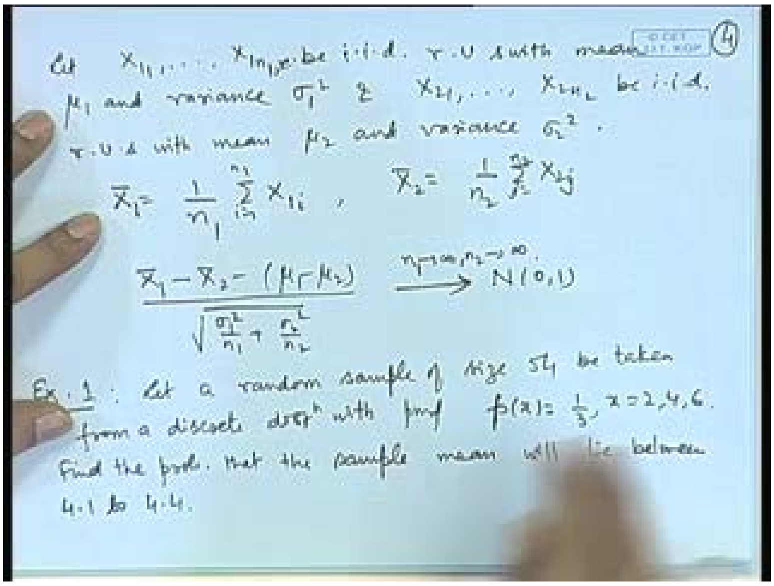
So, likewise in large number of practical situations, we are interested or we are actually using the mean rather than the individual observations. And therefore, the distribution of the sample mean is what should be used. And this particular result, which is known as a central limit theorem, it says under very pretty general conditions, that the distribution is actually normal. Another thing that we should notice here is that, here we have assumed that the distributions; that means, the random variable x 1, x 2, Etcetera are from the same population; that means, it is the sample from the same population. In fact, this central limit theorem has been further generalized; that means, we may lose the condition of say identically distributed or we may lose the condition of independence also and even then, under certain conditions the central limit theorem holds. However that is part of another study. Right now, we are concerned with this, sampling distribution, in which case, we take x i's to be independent and identically distributed random variables.

Now, one question may arise that how large n should be such that this approximation is good. So, in practice n greater than or equal to 30 is considered to be large. If the original distribution is normal or it is close to normal, then for smaller n itself the approximation may be good. One more point, earlier we have seen that binomial distribution was approximated to normal or the poisson distribution was approximated to normal. So, that is actually a, special case of central limit theorem. Because, what is a binomial random variable? It is the sum of successes is individual trial. So, if you are taking x 1, x 2, x n, so basically, it becomes the distribution of the sample sum. So, actually you can write a equivalent form also.

Suppose, I define S n equal sigma x i, i is equal to 1 to n. Then, an Equivalent form is that, if we write S n minus n mu by root n sigma. Then, this will be converging to a standard normal random variable as n tends to infinity. So, the binomial approximation to normal is actually a special case of the central limit theorem. Similarly, the poisson distributions approximation to the normal is also a special case here, because a poisson random variable is the number of arrivals. So, if we are looking at the arrivals in the individual instants, how is were small that instant we may choose? Then, x is denoting the number of arrivals in the full length of the time which is becoming a sum and therefore, the sum of the observations must follow approximately normal distribution.

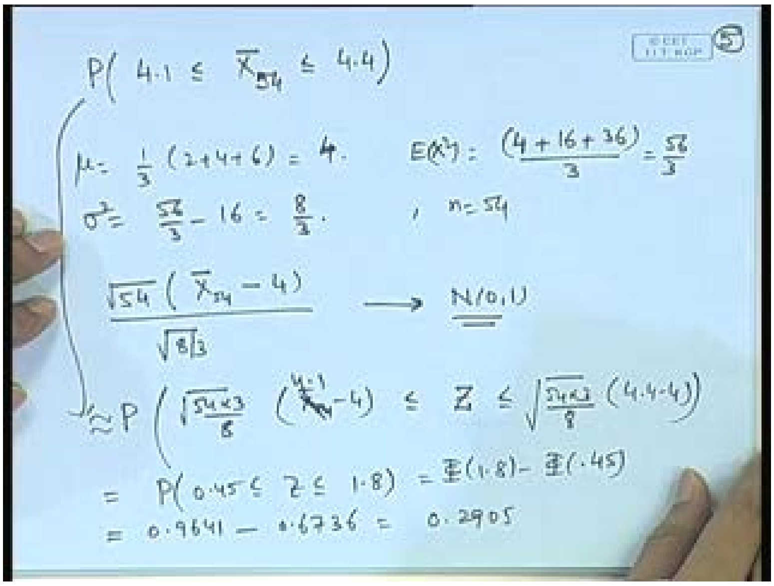
Now, in case of one sample we have straight forwardly for sample mean. Suppose, we have two samples, then the second sample mean may also have normal. And therefore, if we use the linearity property of the normal distributions, then the differences etcetera may also follow a certain central limit theorem.

(Refer Slide Time: 14:41)



Let me give a generalization of this one. So, let say x 1 1, x 1 2, x 1 n 1 be Etcetera. So, let me take n only be i i d random variables, with say mean mu 1 and variance sigma 1 square and say x 2 1, x 2 2, x 2 n 2, be i i d random variables with mean mu 2 and variance sigma 2 square. So, you consider the random variables say x 1 bar which is actually the mean of the first sample. And x 2 bar is equal to say mean of the second sample. Let me put j here. And construct the random variable x 1 bar minus x 2 bar minus mu 1 minus mu 2 divided by square root sigma 1 square by n 1 plus sigma 2 square by n 2. Then, this converges as n tends to infinity to a normal 0 1; that is here n 1 tending to infinity and n 2 tending to infinity. So, this result is quite useful, the original central limit theorem and this results. To solve variety of probability problems where original probability distribution of the sum may be quite complicated and but using this, we can derive the probabilities. Let me give some example here. Let a random sample of size say 54 be taken from a discrete distribution with probability mass function say p x is equal to 1 by 3 for x equal to 246. Find the probability that, the sample mean will lie between say 4.1 to 4.4.

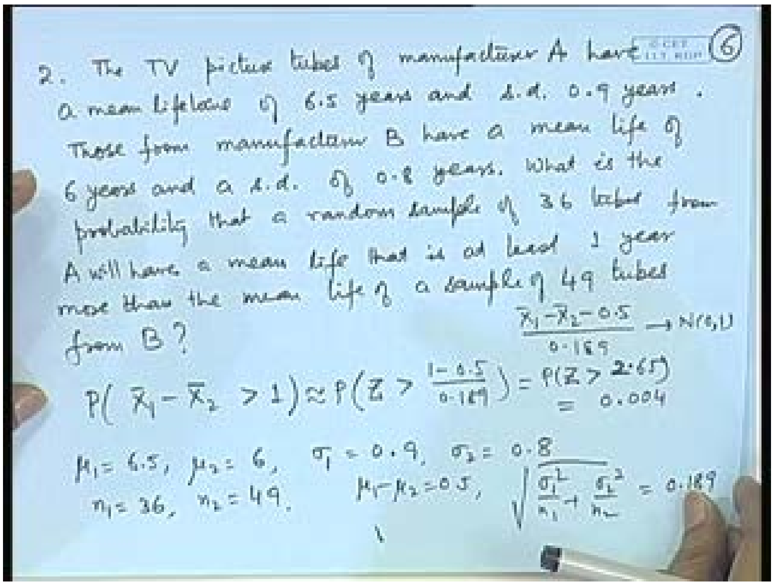
(Refer Slide Time: 17:46)



So, basically we are interested to get the probability of 4.1 less than or equal to x n. So, here x n is X 54 bar less than or equal to 4.4. Now, this is a discrete uniform distribution centered at 24 and 6. If we look at the mean of this one, mu is 1 by 3, 2 plus 4 plus 6 that is equal to 4. And if we look at the variance, so, we can check say expectation x square that is equal to 4 plus 16 plus 36 by 3, that is 56 by 3. So, variance of x that is equal to 56 by 3 minus 16, that is 8 by 3. And n is here 54. So, we have the distribution of root 54 into X 54 bar minus 4 divided by root 8 by 3. This will be approximately normal 0 1. So, if we use this property here, the probability of X 54 bar lying between 4.1 to 4.4 is approximately same as, so, root of we may take it to the numerator. So, it becomes root of 54 into 3 by 8 X 54 bar minus 4. So, this is the 4.1 minus 4 less than or equal to z; this is approximately root 54 by into 3 by 8 into 4.4 minus 4.

So, if we simplify these terms, it is probability of z lying between 0.45 to 1.8 which is approximately. So, phi of 1.8 minus phi of 0.45. So, from the tables of the normal distribution, these values are 0.9641 minus 0.6736; that is equal to 0.2905; that is approximately 30 percent of the time the sample mean will lie between 4.1 to 4.14. Here, we notice that the original distribution is uniform. So, the distribution of X 54 bar will be very complicated. We have seen earlier, that the sum of two independent continuous uniform distributions is triangular distribution. If we take three of the independent continuous uniform distributions, the form is some sort of parabolic in nature. So, if we take 54 such observations and try to find out the actual distribution that is very complicated. And here using the central limit theorem, easily we are getting an approximate value for this. And 54 is in fact, a large sample size and therefore, this approximation will be almost quite good.

(Refer Slide Time: 21:08)

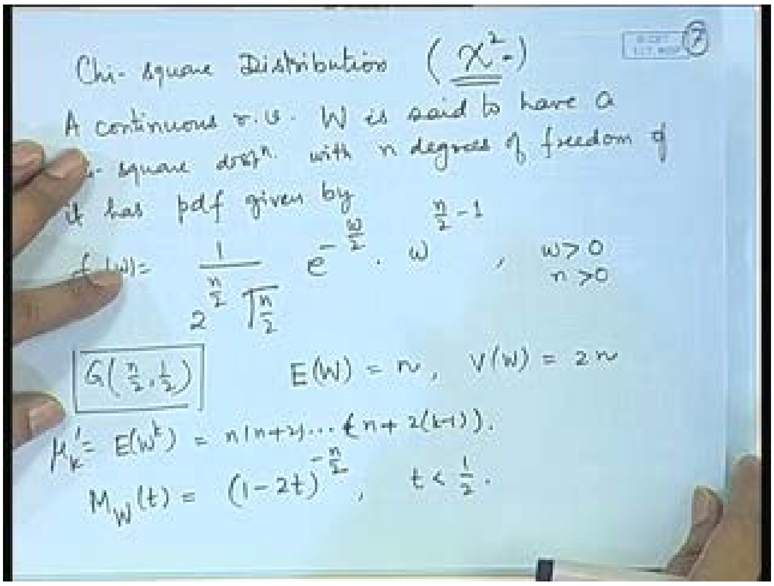


Let us take another example. The TV picture tubes of say manufacturer A have a mean life time of 6.5 years and standard deviation say 0.9 years. Those from manufacturer B have a mean life of 6 years and a standard deviation of 0.8 years. What is the probability that a random sample of say 36 tubes from A will have a mean life, that is at least one year more than the mean life of a sample of 49 tubes from B.

So, here we will apply the extended version of the central limit theorem, because we are dealing with the two samples. So, we can consider that x 1 bar minus x 2 bar minus mu 1 minus mu 2 divided by sigma 1 square by n 1 plus sigma 2 square by n 2 will be approximately standard normal distribution. So, here we see that, we are supposed to find out the probability of x 1 bar minus x 2 bar greater than 1. Now, we look at the parameters here mu 1 is 6.5 mu 2 is 4 sigma 1 is 0.9, sigma 2 is 0.8, n 1 is equal to 36 and n 2 is equal to 49. So, if we calculate say mu 1 minus mu 2; that is 0.5. And square root of sigma 1 square by n 1 plus sigma 2 square by n 2 that is equal to 0.189.

So, x 1 bar minus x 2 bar minus 0.5 divided by 0.189. It is approximately normal distribution. So, if we are to calculate this probability, we can approximate it by probability z greater than 1 minus 0.5 divided by 0.189; that is equal to probability z greater than plus 2.65. If we see from the tables of the normal distribution, this probability is only be 0.004. So, the probability that a random sample of size 36 from A will have a mean life at least one year more than the mean life of another sample of 49 from B is extremely small. So, here you see that, actually the mean life difference is only 0.5 here. So, we are expecting in the sample means to have difference of 1. So, the probability of that is going to be very small now. Another point, that we notice here is that if the original distribution itself is normal, then square root and x bar minus mu by sigma has exactly a standard normal distribution. So, approximation is exact when we have normal distribution. Central limit theorems are widely used in statistical theory.

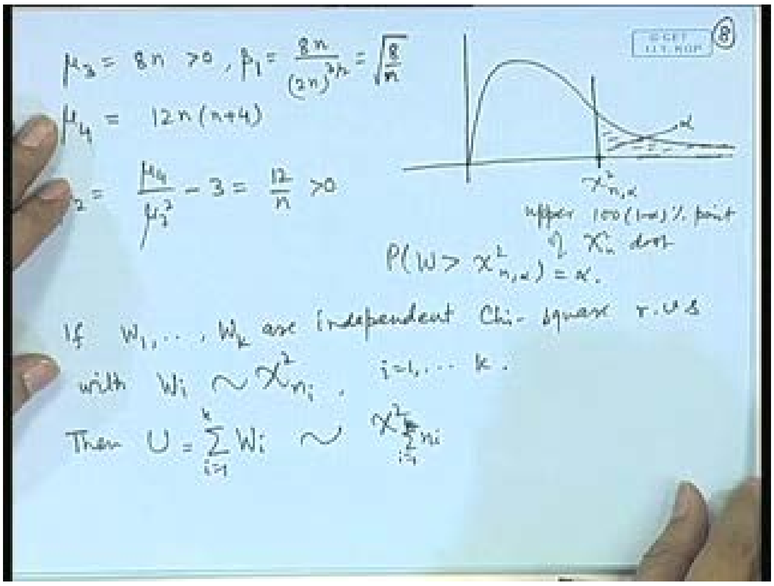
(Refer Slide Time: 25:59)



Now, we discuss another sampling distribution which is known as chi square distribution. And it is used as chi, 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 2 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 two 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. So, 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 a degrees of freedom is actually the mean of the chi square distribution. Similarly, if we look at the variance in gamma or 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: 28:48)



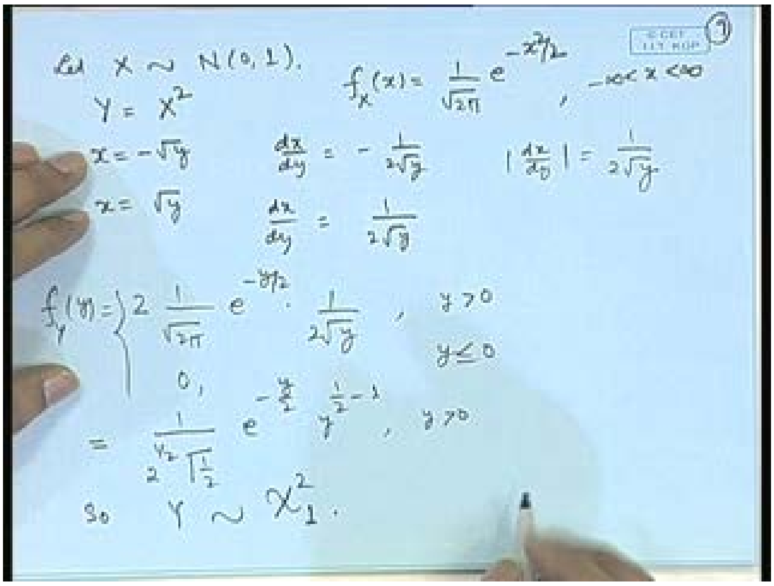
In particular, we may look at the third central moment that is 8 n. So, obviously, it is a positively skewed distribution, since gamma distribution is positively skewed. So, of course, depending upon different values of n, you will have different shapes for the chi square variable. So, mu 4, but of course, if we look at the major of symmetry, that is beta 1 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 we look at mu 4, that is 12 n into n plus 4 and the major 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.

So, in fact as n becomes large, this is tending towards normality. That fact, we 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 specific probabilities. So, if this probability 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. So, 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 we consider, 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 1, 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 y, 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.

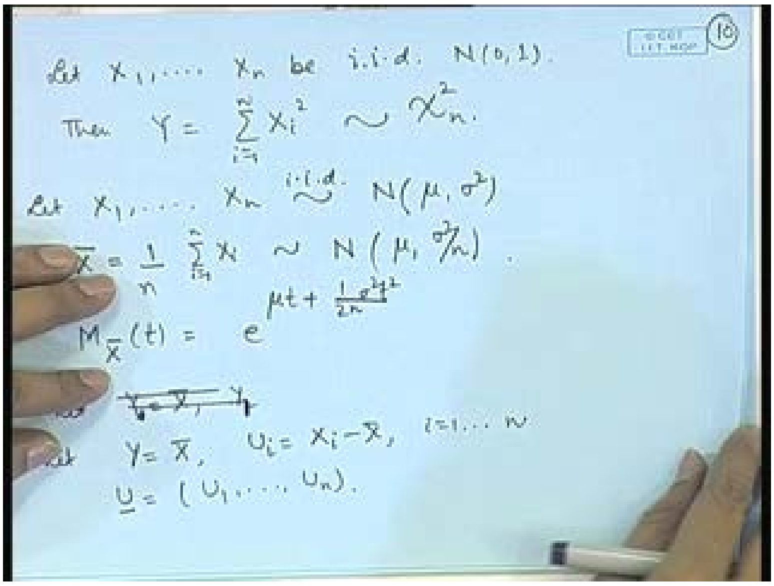
(Refer Slide Time: 33:03)



Next, we look at the following result. 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 joint, 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 we look at dx by dy 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 we substitute n is equal to 1, then we get these 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.

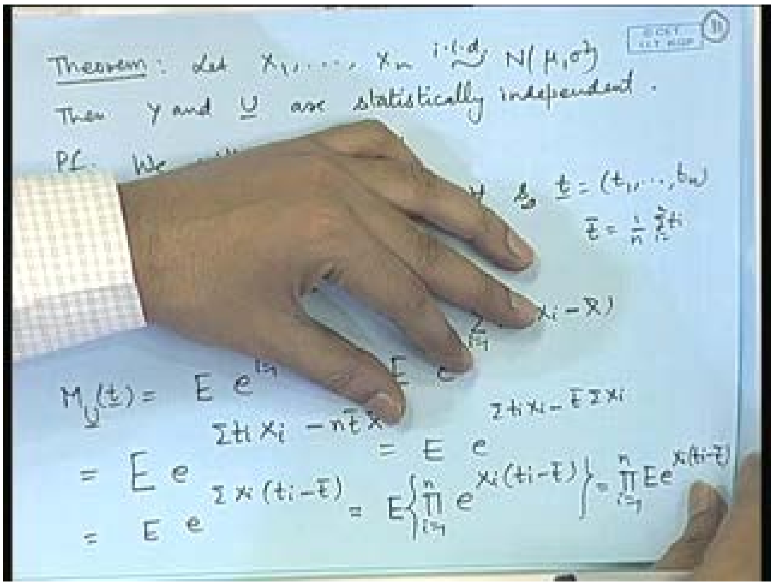
(Refer Slide Time: 35:16)



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 x i square that will be chi square 1, if x 1, x 2 and x n are independent, then x 1 square, x 2 square, x n square are 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, and x n is a 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.

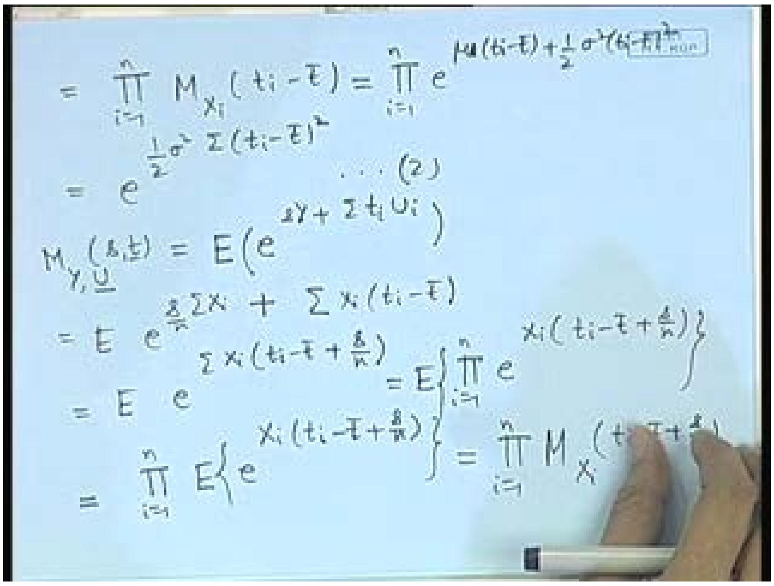
(Refer Slide Time: 38:15)



Then, we have the following theorem. 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 of U 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 mu 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, this expectations are nothing but the moment generating functions of x i at the point t i minus t bar.

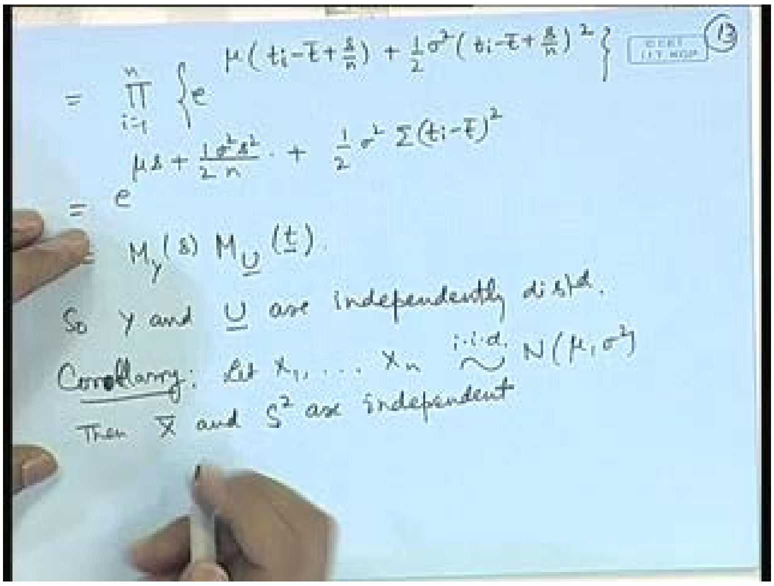
(Refer Slide Time: 41:46)



So, this is equal to 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 1 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 the 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 is product of the MGF's at the point t i minus t bar plus s by n. So, once again making use of the fact that the x i have a normal mu sigma square distribution, we know the form of the MGF. So, we substitute it here.

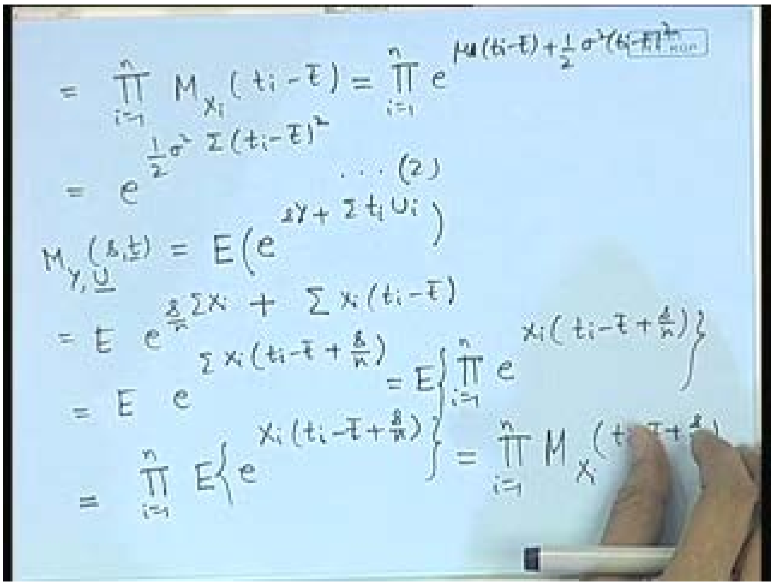
(Refer Slide Time: 45:08)



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. So, now this is, if we apply this product here, e to the power sigma mu t i minus t bar becomes 0. And the second term, we 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 variance 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 S square are independent.

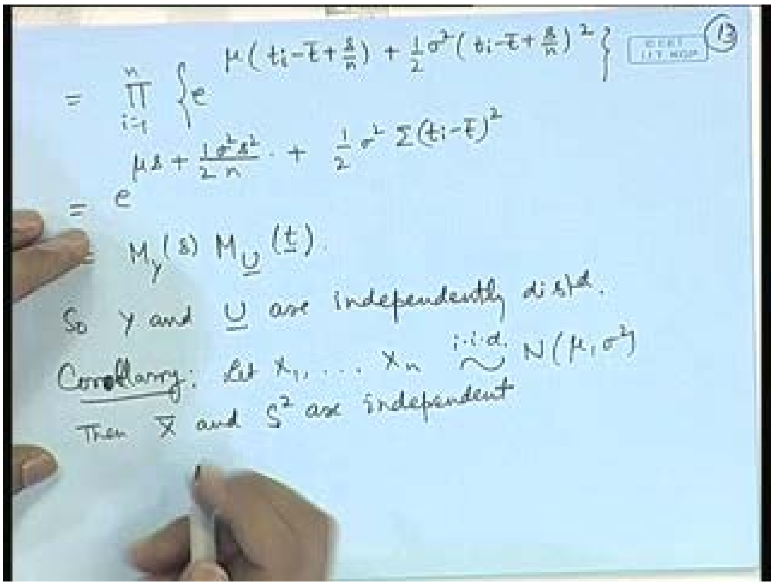
(Refer Slide Time: 48:06)



Now, we will show that, this is helpful to derive the distribution of S square. So, 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 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. So, This implies at x i minus mu by sigma 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 variable. 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 s 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: 50:44)



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, the MGF of a chi square variable is known, there 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 is 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 one degree of freedom.

So, this means that chi square is a distribution of the sample variance after certain scaling. So, this shows that chi square is a sampling distribution. Either, we consider a standard normal random variable, so, sum of squares of n independent random variables normal random variables is chi square on degrees of freedom or if we are considering arbitrary normal random variables, then if we consider the scaled distribution of the sample sum of squares from the deviation, that is n minus 1 S square by sigma square, then that is chi square on n minus one 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 has mean 0 and x bar has 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 dependent; that is why, the degrees of freedom are here 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, if when we consider any other, then it need 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 minus x square bar square divided by n minus one 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, 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 unbiasness for…

In the next lectures, we will take up other sampling distribution such as t and f distributions. So, today we will stop here.