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

Lecture - 61
Interval Estimation – I

So, far this in this course on Statistical Inference we have concentrated on 2 main topics; one was the point estimation where for estimating a parametric function g theta we propose a point estimator that is we call it t x. And we have discussed methods of obtaining various types of estimators and how to evaluate that is which estimator should be used so what are the criteria like and biasness, consistency in minimum variance and bias estimation minimum risk and so on.

Another area that we have covered so far is the problem of testing of hypothesis; that means, we want to make certain assertion about a probability distribution and then we want to test whether our assertion is tenable in the light of the data that we are available towards or not.

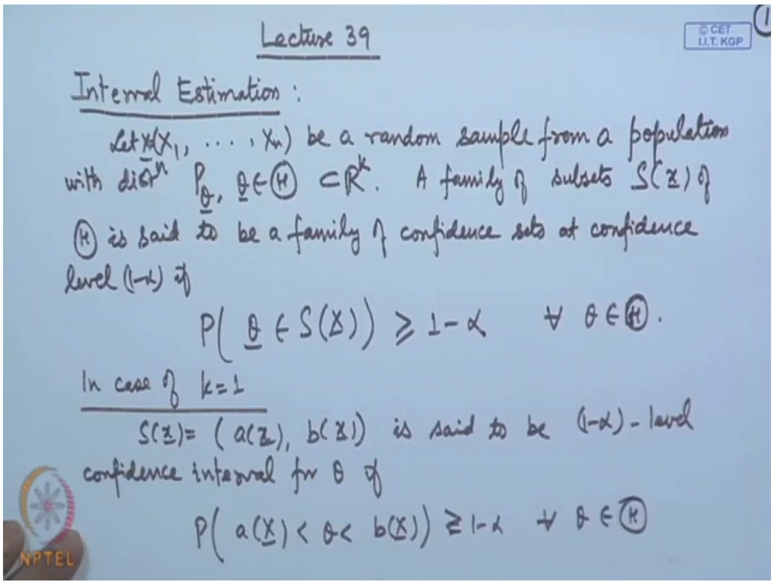
A third and equally important area of statistical inference is the Interval Estimation. In the point estimation we propose a value. So, for example, we say suppose we are considering estimation of the average income of a population or average per capita expenditure of a population, then we may consider sample and based on that we calculate a value. And then we say for example, the statement like this per month per capita expenditure on say health is 200 rupees.

However, there are some statisticians and also from practical point of view sometimes it is more useful if in place of a value we give a range of values. So, rather than saying that the percentage of cure by a certain medicine is 75 percent, it is better to give a statement that it is more than 70 percent, which may give more faith in people to use that medicine. If we say it is more than 70 percent rather than giving exact an value if we say it is more than 70 percent.

Similarly, if we want to say per capita expenditure as I mentioned just now for a point estimation problem in place of saying say 200 rupees we may say it is between 175 rupees to 225. Now when we propose an interval so first of all this is called an interval estimator. Now secondly, there is a probability associated with this thing or you can say a probability statement associated with this because how often this interval will actually include the true value and how often it will not improve the true value, that is known as the confidence level.

So, in the problem of interval estimation we propose confidence intervals. I will start with the general definition of a confident set and then we will move over to confidence intervals when we deal with the real valued parameters.

(Refer Slide Time: 03:38)

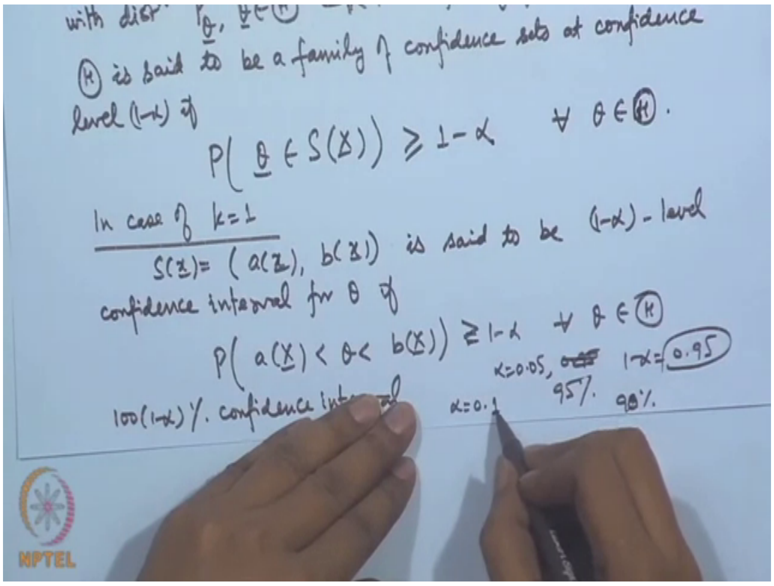


So, as before let X 1 X 2 X n, so, this is a random sample X n be a random sample from a population with probability distribution say P theta; theta belonging to say a script theta. Now in general theta can be a vector so let me assume in the beginning it is k dimensional and then specific retention will be paid to the case of k is equal to 1.

A family of subsets S x of theta is said to be a family of confidence sets at confidence level 1 minus alpha if probability of theta belonging to S X that is greater than or equal to 1 minus alpha for all theta belonging to script theta. So, you can see here that this interval which is of course a random interval S X is a subset of theta. So, a random interval it includes theta the probability of this statement is greater than or equal to 1 minus alpha, then we say S X is a family of confidence sets at confidence level 1 minus alpha.

Now, in the case of if I consider 1-dimensional k that is real parameter then we may consider this set to be an interval. So, we may say S x is equal to an interval of the form say a x to b x this is said to be a 1 minus alpha level confidence interval for theta if probability of a x less than theta less than b x is greater than or equal to 1 minus alpha for all theta.

(Refer Slide Time: 06:30)

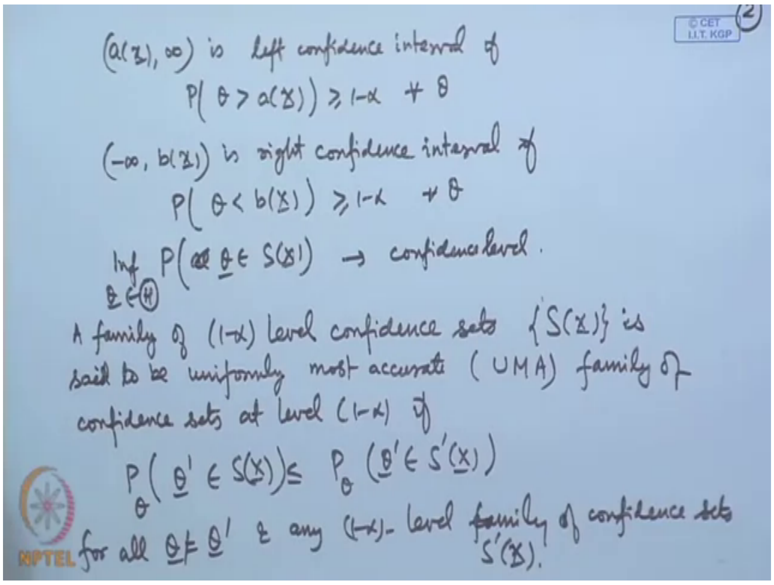


It is also written as sometimes 100 1 minus alpha percent confidence interval. So, in place of saying 1 minus alpha we also say 100 1 minus alpha percent. So, for example, 0.95 suppose I take alpha is equal to 0.05 then 1 minus alpha is equal to 0.95. So, this is the level and we may say 95 percent confidence interval or if I choose say alpha is equal to say alpha is equal to 0.1 then we will say 99 sorry 90 percent confidence interval because here 1 minus alpha will become 0.9.

Now when I say this interval a x to b x it may also happen that one side is taken to be like b x could be plus infinity or a x could be minus infinity then it becomes a one-sided confidence interval. So, sometimes like I mentioned the statement that the percentage of cure by a certain medicine is for a certain disease is at least 70 percent.

So, here it is one sided interval, that is we are saying P is greater than or equal to 0.7 or we may say that then rate of failure is at most 1 percent or at most 10 percent for a certain manufacturing process. So, in that case we are putting an upper bound. So, let me formally will give this one-sided confidence intervals also.

(Refer Slide Time: 08:15)



So, a x to infinity is you can say left confidence interval if probability of theta greater than a x is greater than or equal to 1 minus alpha. Similarly minus infinity to b x is right confidence interval, if probability of theta less than b x is greater than or equal to 1 minus alpha. One point which I will like to repeat here when I say this probability statement this is statement is a probability statement involving the random variable which is in the form of S X here because X 1 X 2 X n are random variable. So, this S X is actually a random variable.

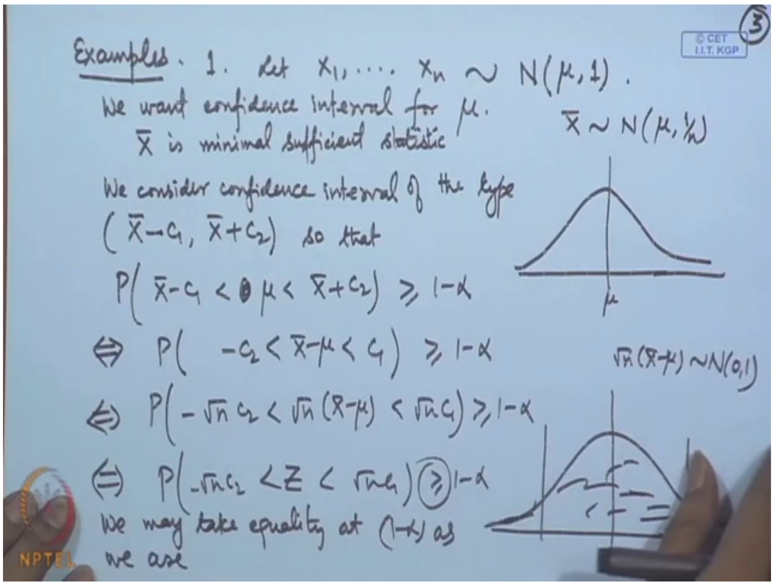
So, one should not understand that we are considering some probability statement for parameter theta; theta is actually fixed. So, in all of these statements the probability statement is valid because of the distribution of X 1 X 2 X n is involved here.

We gave a, now definitely when we are having an interval and certainly we would like to have some sort of understanding for example, if I say 95 percent confidence interval. Now certainly if there is an error interval which will give me 99 percent confidence then I will considered it to be better if the length is likely to be the same. On the other hand if I fix a confidence level say I say 90 percent or 95 percent at that level I would like to choose that interval which has the shortest length. Otherwise you can always say like minus infinity to infinity it includes all the parameter values if probability 1.

So, this statement will be Wakeways statement or you can say it is not going to be useful in the inference point of view. So, we need to make a statement which will be in the valid as well as which will be more precise or accurate, that brings us to the problem of finding out shortest length confidence intervals or if we fix the width then the maximum confidence level ok. So, we if we put say infimum of say probability ax or say theta belonging to S X.

If I consider this is called the confidence level a family of 1 minus alpha level confidence sets S x is said to be uniformly most accurate. So, we call it UMA; UMA family of confidence sets at level 1 minus alpha if probability of theta prime belongs to S X is less than or equal to S prime X. For all theta not equal to theta prime and any 1 minus alpha level family of confidence sets S prime X. So, S X will be said to be uniformly most accurate.

(Refer Slide Time: 13:31)



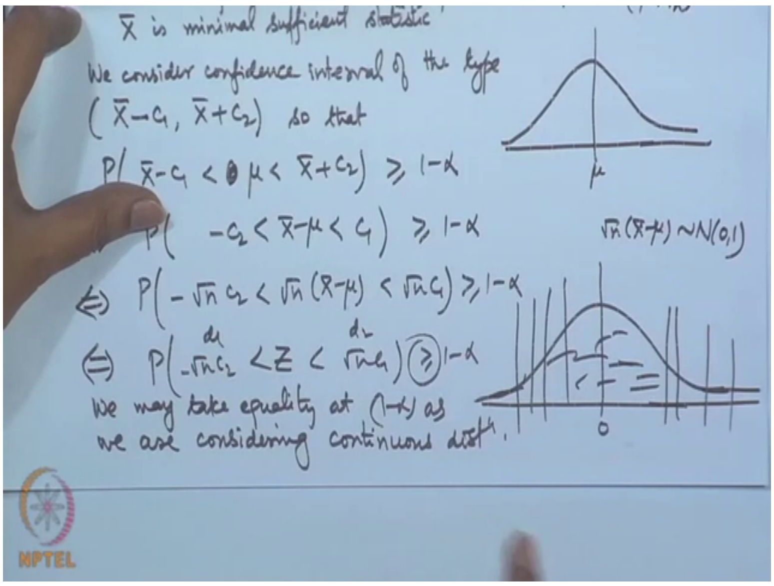
Let me give some example of a confidence interval. Let us consider a normal population let X 1 X 2 X n be a random sample from a normal distribution with mean mu and I am considering for convenience variance to be unity. That means, variance is known so, I have converted to unity, we want the confidence interval for say mu. We want confidence interval for mu. Now I am looking at from the practical point of view in this problem X bar is a minimal sufficient statistic; is not it? X bar is a minimal sufficient it is also complete of course.

So, we will actually we have seen when we were doing the estimation. So, X bar was actually given umue mle mini max estimator etcetera. Also when we looked at the test for mu in this particular problem the test was based on X, bar the uniformly most powerful test ump unbiased test or the likelihood ratio test they were all based on x bar in fact, the invariant test also. So, when we were setting up a confidence interval it is natural that we will restrict attention to the minimal sufficient a statistic X bar. And also what is the distribution of X bar? X bar follows normal mu 1 by n.

Which is actually a symmetric distribution. So, naturally it suggests that we take an interval around mu which is based on X bar. So, we consider confidence interval of the type say X bar minus sum C 1 to x bar plus sum C 2 and probability of this interval including mu should be at least 1 minus alpha. Now this statement is equivalent to probability of a X bar minus mu is less than C 1 greater than minus C 2 that is greater than or equal to 1 minus alpha. Now we may also put from here root n X bar minus mu that follows normal 0 1.

So, if we consider the probability density function of standard normal distribution then I am choosing two points such that the probability in between is equal to 1 minus alpha. So, we can write it as minus root n C 2 less than root n x bar minus mu less than root n C 1 greater than or equal to 1 minus alpha. So, probability of Z lying between these two values minus root n C 1 to root n C 1 and you can see actually we are dealing with the continuous distribution.

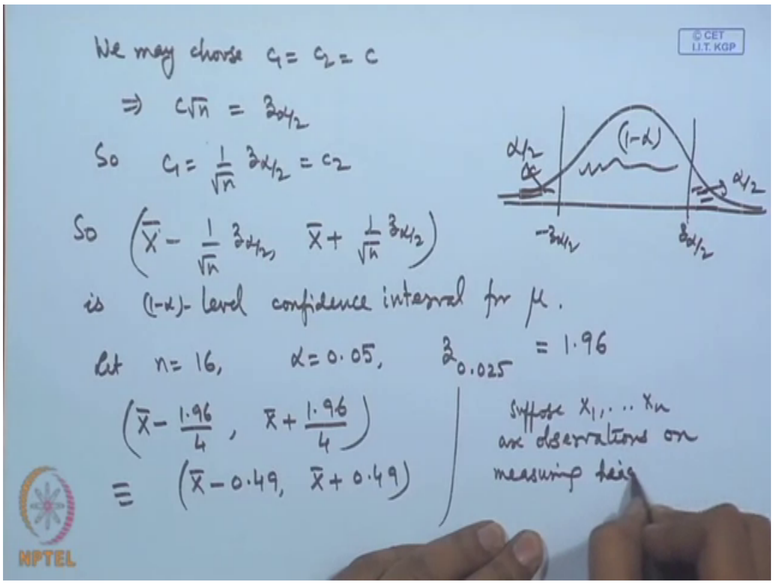
(Refer Slide Time: 17:44)



So, we can achieve actually equality here, we may take equality at 1 minus alpha as we are considering continuous distribution, another point you can see when I am putting two points here. So, let me call d 1 and d 2, then there can be many points where the probability can be equal to 1 minus alpha like if I shift on the side or if I shift on the side I can consider so many intervals here.

But once again let us bring down the problem to the finding out shortest length. Now because of symmetry of the standard normal distribution we can consider a symmetric interval and if I choose the symmetric interval it will be the shortest because the probability in the middle is the highest for a normal distribution.

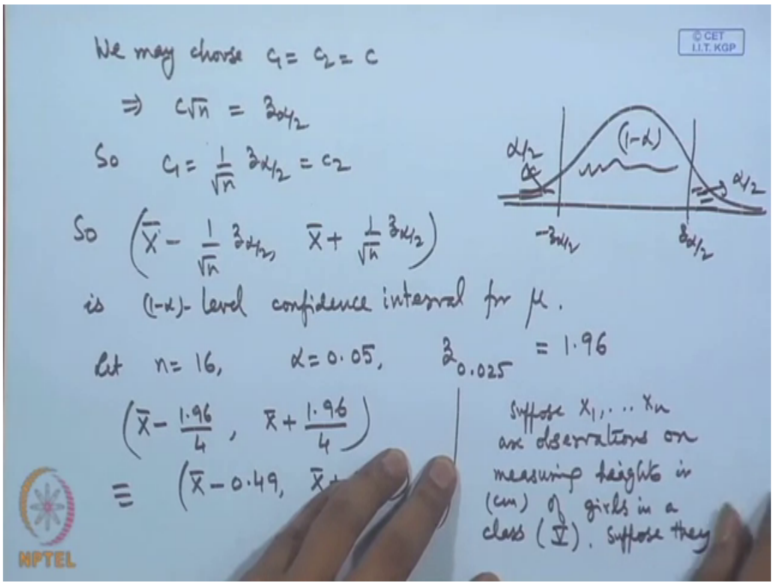
(Refer Slide Time: 18:37)



So, we may choose say C 1 is equal to C 2 is equal to say C then this implies C root n that should be equal to z alpha by 2 that is a point on the this is z alpha by 2; that means, this probability is alpha by 2. So, minus z alpha by 2 is the probability where here at this probability is alpha by 2. So, in between probability is 1 minus alpha. So, we are getting here that C 1 is equal to 1 by root n z alpha by 2 that is equal to C 2. So, X bar minus 1 by root n z alpha by 2 to X bar plus 1 by root n z alpha by 2 this is 1 minus alpha level confidence interval for mu.

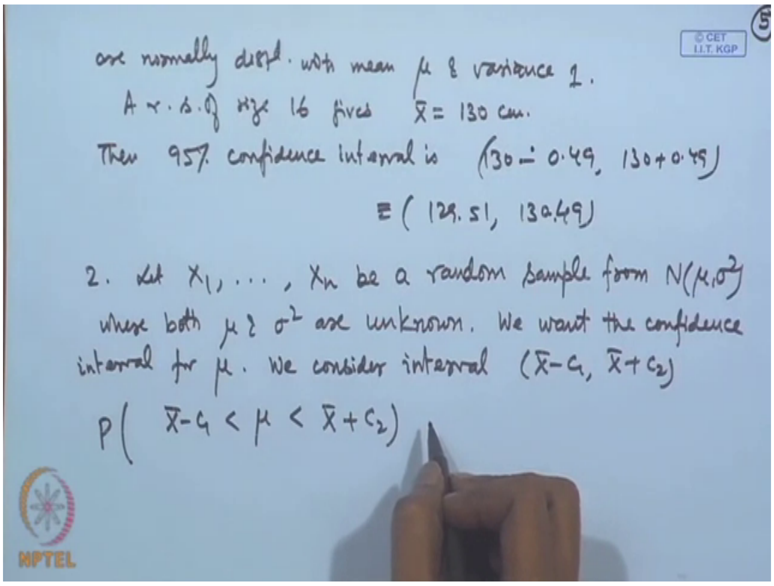
Let me just take a practical example, suppose I take n is equal to 16 say and alpha is equal to say 0.05. So, then I have to see the value of z 0.025 from the tables of standard normal distribution this value is 1.96. So, X bar minus 1.96 by 4 to X bar plus 1.96 by 4 that is X bar minus 0.49 to x bar plus 0.49.

(Refer Slide Time: 21:02)



So, if I in a practical problem suppose we are considering suppose X 1 X 2 X n are observations on measuring say heights in say centimeters of girls in a class suppose class is 5 ok.

(Refer Slide Time: 21:35)



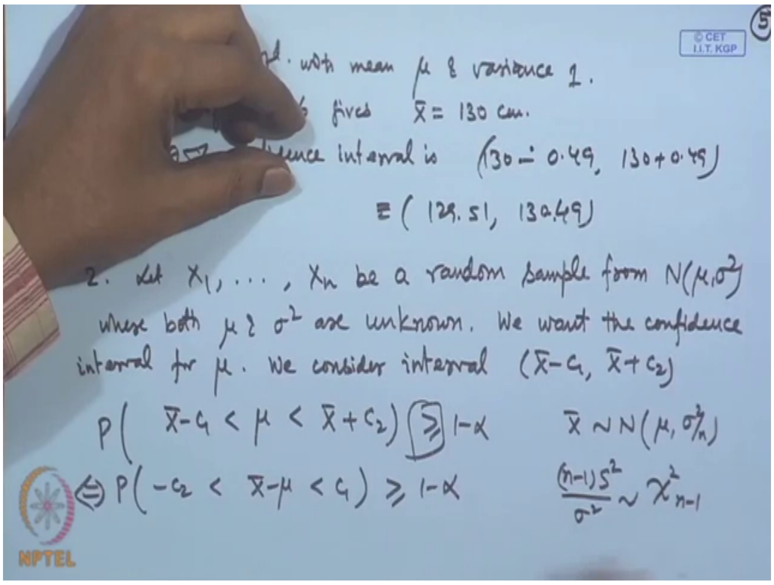
And the heights in centimeters are suppose they are normally distributed with mean mu and variance 1 ok. A random sample of size 16 gives say X bar is equal to the observed value of the X bar suppose it turns out to be. So, in class 5 we may assume something like say 48.

So, something like say 130 centimeters say then 95 percent confidence interval is 130 point 130 minus 0.49 to 130 plus 0.49, that is 129.51 to 130.49. That means, we are saying with 95 percent confidence that the average height of the girls in a class 5 will be between 129.51 centimeter to 130.49 centimeters.

One interpret it in a frequency sense that if we conduct or we collect the sample 100 times then 95 percent of the times the sample will include the true parametric value and 5 percent of the time it is not like it may not include the 2 parameter value. Let me consider some more examples here in this previous case I have considered variance to be known. So, I took it to be 1, but suppose the variance is unknown, if the variance is unknown then the distribution of X bar will involve sigma X square and in that case we need to do little bit of modification.

So, let me take another example let X 1 X 2 X n be a random sample from normal mu sigma a square where both mu and sigma square are unknown and as before we want the confidence interval for mu. Now as I argued in the previous case X bar is some sort of you can say good estimator for this problem for mu. So, once again we will write the same thing that we consider interval of the type interval X bar minus C 1 to X bar plus C 2 that is probability of X bar minus C 1 less than mu less than the X bar plus C 2 should be greater than or equal to 1 minus alpha.

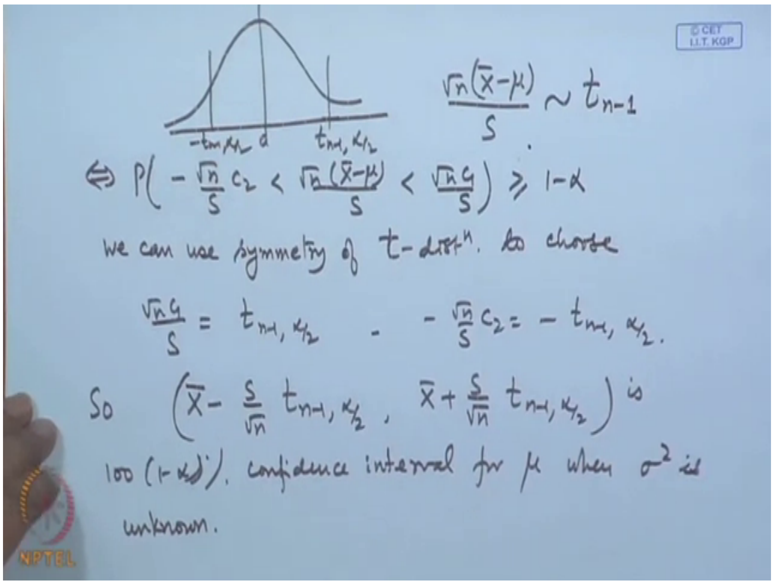
(Refer Slide Time: 25:13)



And of course and since we will be dealing with the continuous distributions I will actually find out the solution corresponding to equal to 1 minus alpha. So, probability of now this is statement you can modify minus C 2 less than X bar minus mu less than C 1 is greater than or equal to 1 minus alpha.

Now in this case if I am again considering X bar the distribution of X bar that is normal mu sigma a square by n. So, if, I write in the state forward fashion this will depend upon; the interval will depend upon sigma here. So, we make use of the t statistic here if you remember n minus 1 S square by sigma a square follows xi square distribution on n minus 1 degrees of freedom.

(Refer Slide Time: 26:10)

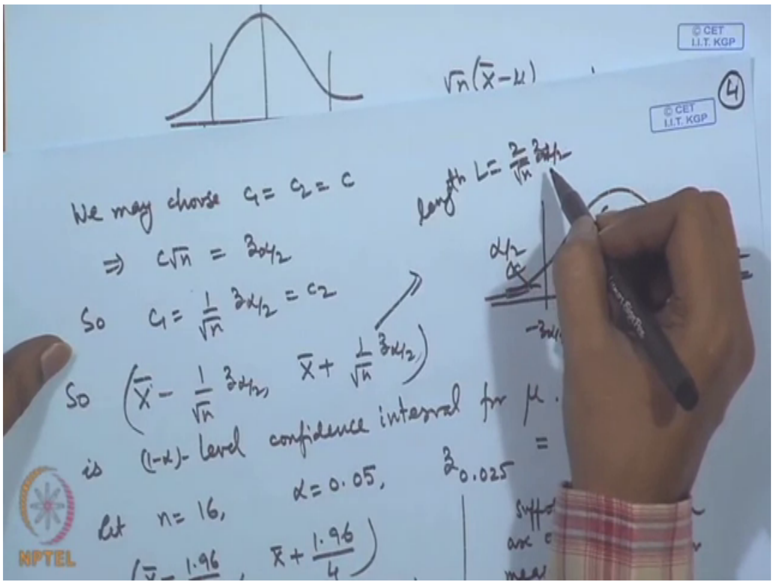


And as a consequence we had seen that X bar minus mu root n by S this follows t distribution on n minus 1 degrees of freedom. So, from this statement then we write probability of minus root n by S C 2 less than root n X bar minus mu by S less than root n C 1 by S that is greater than or equal to 1 minus alpha. Since this is again a symmetric distribution this t distribution as you know it is a symmetric distribution around 0. So, we may choose the point as before as t n minus 1 alpha by 2 and minus t and minus 1 alpha by 2.

So, we can use symmetry of t distribution to choose root n C n by s is equal to t n minus 1 alpha by 2 and minus root n by S C 2 is equal to minus t n minus 1 alpha by 2.

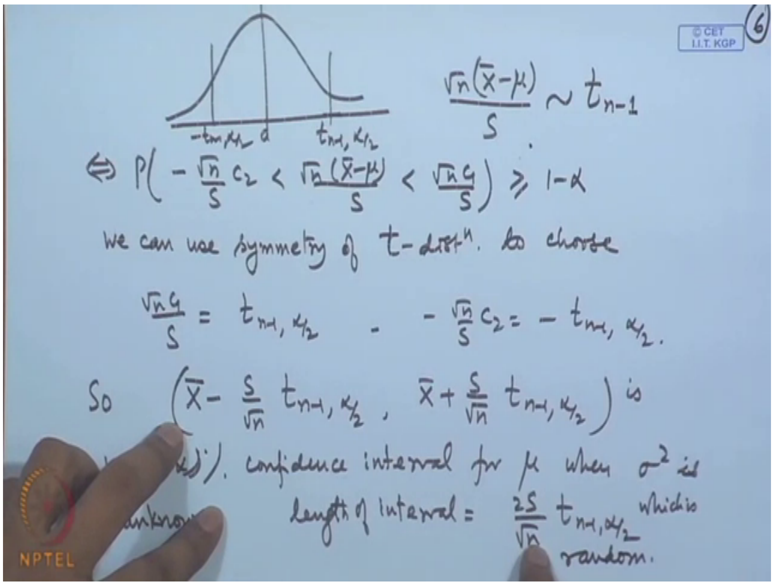
So, from here the confidence interval comes out to be X bar minus S by root n t n minus 1 alpha by 2 to X bar plus s y root n t n minus alpha by 2. So, this is 100 1 minus alpha percent confidence interval for mu when sigma a square is unknown. When sigma a square was known then here sigma would have come here as I solve the problem with sigma is equal to 1. Also let us note down some feature here.

(Refer Slide Time: 28:27)



For example what is the length of the interval, in this case the length of the interval that is equal to 2 by root in z alpha by 2 which is actually free from the parameters and it is free from the random variable.

(Refer Slide Time: 28:48)



In this case if you see the length is length of interval that is equal to 2 S by root n t n minus 1 alpha by 2 which is random, but of course, you can see that has n becomes large this will become a small the length of the interval will become a small actually it will go to 0. Another point you can see that as n becomes largest value will converge to z value, that is z alpha by 2 and S will be some number which will actually converts to sigma by using strong law of large numbers or consistency of s for sigma.

So, this will converge to; so if n is large actually this is approximately the same as the one for the known sigma case. And in both the cases that I have given here the length of the intervals as you can see as n becomes large the length of the interval become a small. So, the interval become sharper for larger sample sizes. So, which again suggests that the more than area or you can say more observations are there then the better inference will be drawn because the interval will become more crisp.