Probability Methods in Civil Engineering

Prof. Dr.Rajib Maity

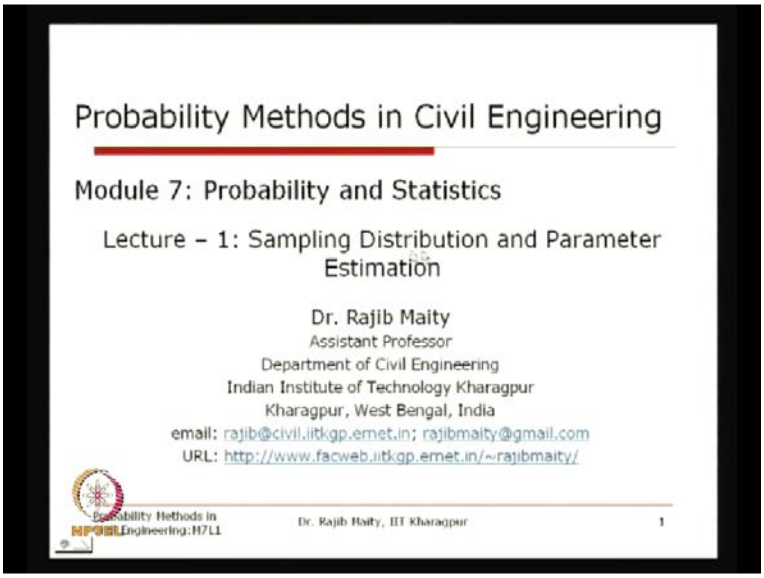
Indian Institute of Technology, Kharagpur

Lecture No. # 35

Sampling Distribution and Parameter Estimation

Hello and welcome to this lecture today. We are starting a new module.

(Refer Slide Time: 00:24)



This module is on probability and statistics. In last couple of modules, we have seen, we have discussed about different theories related to probability. Basicallythose, whatever we have discussed, is associated with the population.

Now, when we talk about the statistics, this statistics is related to the to few samples taken from this population.More precisely speaking, it is some of this random sampling. So, the the sample that we take that should be the random sample for from the population.

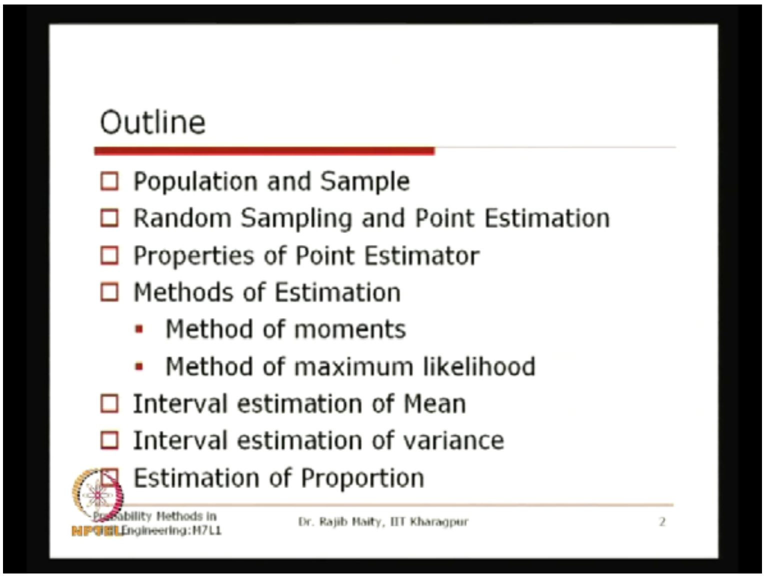
So, now, as we are using the word the population sample, so, at the starting of this lecture, we should know what does this actually means.You know that when we are talking about some of this parameters for those distribution, that we have seen in the last lecture in this last module, that there are certain parameters are there which, is associated for some standard probability distribution, that we have seen.

So, basically we have to estimate those parameters. Those parameters when we estimate, we should estimate, we should have some sample data.From there, we should estimate those those parameters.Now again, the samples means, when we are taking the random samples from the population, now once we change the sample; obviously, that that estimation of those parameters may also change.Obviously, it will change because two samples will never be exactly identical.So, those those estimations also we will have some some distribution. So, those are basically called the sampling distribution. So, basically our first lecture that we are going to take in this module is is on this sampling distributions and the parameter estimation.

So, in this parameter estimation that what we will do we will do is based on this, some of this sampling distribution.You know that, whenever we talk about the mean or standard deviation or this type of parameters and which is estimated from the samples, those are also follow some distribution. So, we will see what that distribution refers to and then we can go for its estimation.Now, in the in the estimation theory also, we can do this estimation in terms of for two different types.One is that, that it is the point estimation.The point estimation means just one value. I can estimate from whatever information that is available to me through the random samples.Or, what I can do is that, I can look for some of this interval of this of this estimation.

So that, this is the interval, this is the lower limit,and this is the upper limit of this particular parameter. So, that it will be more logical to say or in the in the statistical sense, that instead of giving a single value as the estimate, sometimes it is preferable to give some of this interval.So, the all these things we will discuss in this lecture.

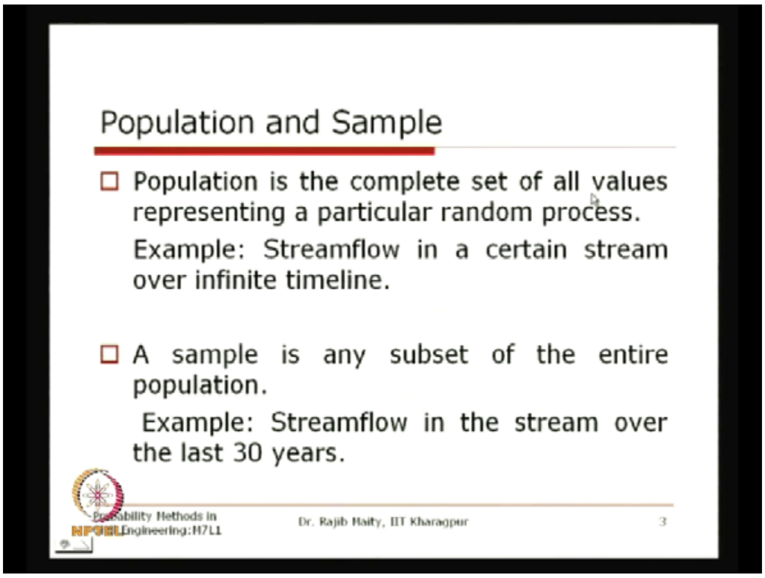
(Refer Slide Time: 03:30)



So, our outline of today's lecture is that, first we will discuss about the population and sample, then the random sampling and the point estimation.As I was mentioning, thus the single value, we will just pick up and we will just estimate from the sample.Then, properties of this point estimation and there are two methods for this estimation.One is that called Method of moments and other one is Method of maximum likelihood.Then, the interval estimation of this mean, and theninterval estimation of the variance, and then estimation of the proportion.

So, all these things will go, basically this outlines should not be a should not be a dead end here. So, it should continue and we should have this hypothesis testing also.Because, these are all very well related to whatever we are going to discuss. So, may be one after another, we will take all these topics. So, to discuss about our the statistics, which is related to each and every point, whatever we will be discussing. We will see that what and where these things are useful to handle in some of the civil engineering problems. So, that, we will see. So to start with, we will start that about the concept of this population and sample.

(Refer Slide Time: 04:52)



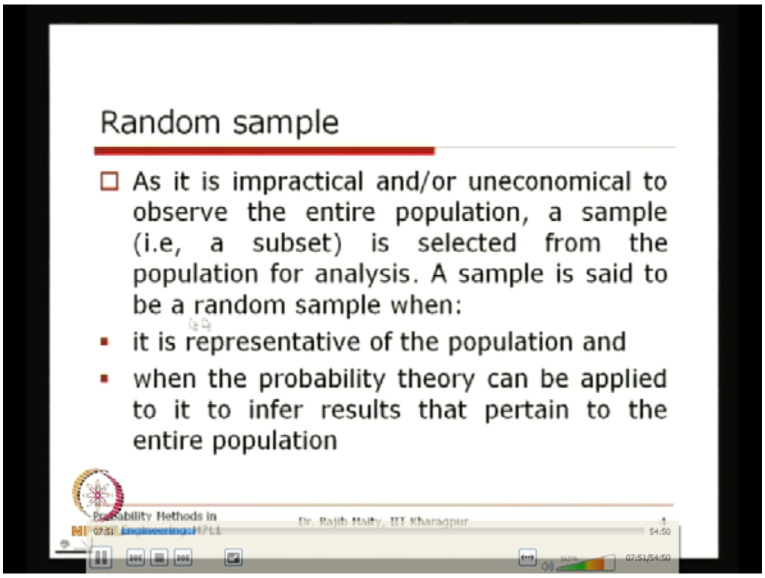
So, the population is the complete set of all values representing a particular random process. So, when we talk about this, all values which represents a particular random random process; that means, this all is very very I should say, that is including of, whatever the possible range is there, that should include everything.

So, what I can give the example is that, the stream flow in a certain stream over the infinite timeline. So, if we are collecting the time series of the of the stream flow value, so, there should not be a finite time over which we are looking the data. So, it should be on this infinite timeline of whatever has happened and what is going to happen, everything. So, that is what is our our population is.When we talk about a sample, that sample is any subset of the entire population. Its corresponding example for this stream flow that we have used here is, the stream flow in the stream over the last 30 years. So, if I just take over the last 30 years, then that particulardata set that is available to us is the sample of this population.

Now, you can see that whatever the probability theory that we have discussed earlier, is basically related to its itsentire entire possible values.That is, all possible values of that of that random process, but in reality, we never get, we never have this entire or whatever whateverall possible values, that is possible that we may not have.

So, we have to rely on some some sample and somehow we have to relate whatever we can estimate from the sample.We have to relate it to to its population. So, this is basically the role of this statistics and with this, we will just see how we can estimate the parameters how we can estimate the required information from the samples.Then we will infer something about its population.Because as I told that, having the entire population is not possible in for almost all the cases.

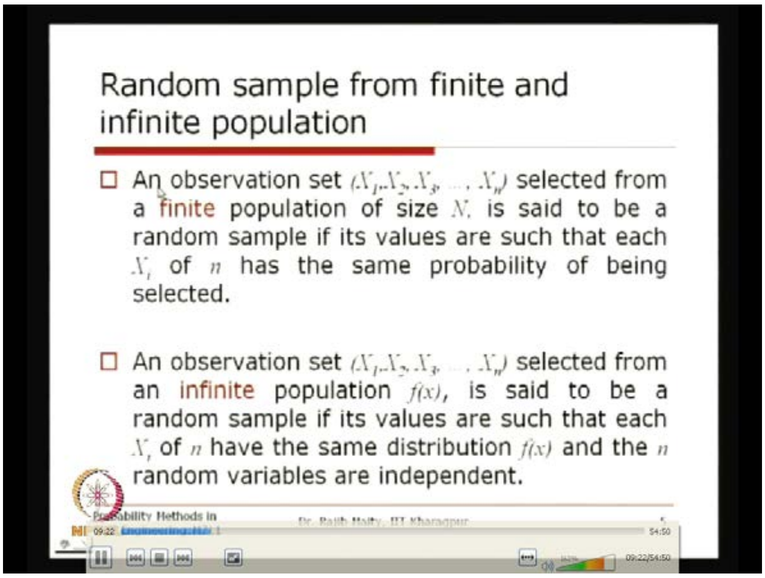
(Refer Slide Time: 07:19)



So in now, in sample, I was mentioning that the random samples, when should we call that a sample is a random is a random sample. So, as it is impractical or and or uneconomical to observe the entire population that I was mentioning. So, a sample that is a subset is selected from the population for analysis.A sample is said to be random sample, when these two conditions are satisfying.One is that, it is the representative of the population. So, whatever the possible cases that we can see in the population, the sample should have have that representation in it.Thesecond one is that, when the probability theory can be applied to it, to infer the results that pertain to the entire population.

So, the probability theory as I was telling, it is basically relates to the population.Now, if we can apply the probability theory to the sample to infer the results that pertains to the population, then we can say that, whatever the sample we are taking, that is the random sample. So, these are just the concept and with this concept, what we can do is that, we can use whatever we have seen from the probability theory, we can apply to the to this random samples.

(Refer Slide Time: 08:48)



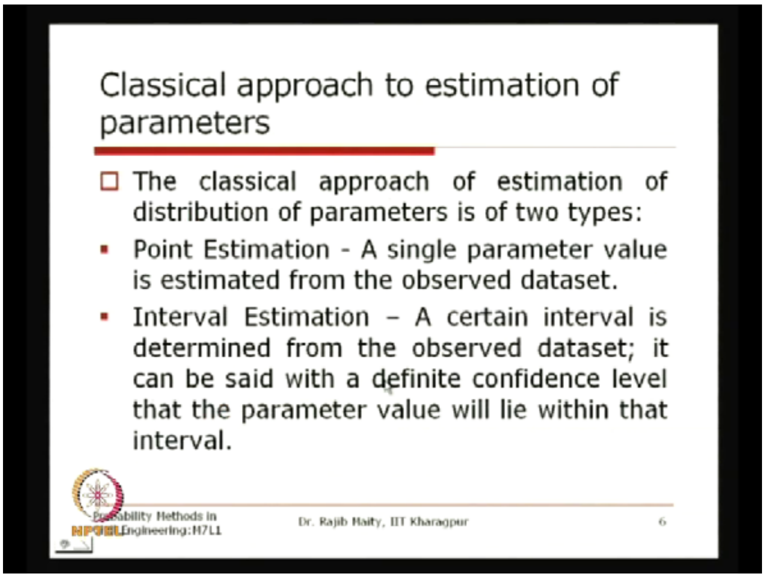
Now, the random sample for finite and infinite populations. So, you know that there are certain cases where the population can be finite, or in other cases it can be infinite, or sometimes it can be count ably infinite.Even though I know that; obviously, there is some maximum number is there. But we can sometimes; we can also say that that is count ably infinite. So, in such cases, what should be the random sample?

So, first we will take the case of this finite population.An observation set x 1, x 2, x 3 up to x n is selected from a finite population of size n. So, that entire size of the population is n is said to be a random sample, if its values are such that, each x I of n has the same probability of being selected.Now, so as we can see that there is this, the size of this population is finite that its maximum number is n. So, this x 1, x 2, x 3,so, we can say that it is a random sample, if each of this item have the, so this, each of this element of this set have the equal probability of being selected from the population.

The second one, if the population is infinite, an observation set again that x 1, x 2, x 3, x n is selected from an infinite population f x.So obviously, when we talk about this infinite series, we generally denote it through its probability density function.Here, it is that f x is said to be a random sample.If its values are such that, each x I of n have the same distribution f x and the n random variables are independent.

So each observation, now the first observation,second observation,third up to n, these are also; we can say that these are also a random variables. So, these variables should have the same distribution, which is f x and which is same to the distribution of the population.And all these n elements of this set are independent to each other, then we can say that this set is one random sample.

(Refer Slide Time: 11:17)



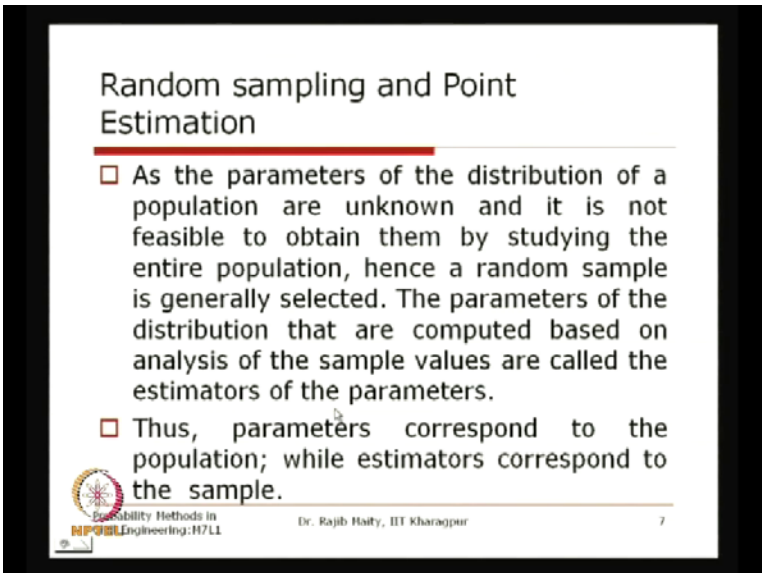
The classical approach to estimation of this parameter. So, the classical approach of the estimation of this of the distribution parameters is of two types.One is the point estimation and other one is the interval estimation. So, this point estimation means a single parameter value is estimated from the observed dataset that is from the sample. So, from the sample a single parameter value is estimated.

Now,we will we will discuss how to estimate that parameter.That we will see so, but, here I want to trace the word single. So, when the single parameter is estimated, that is known as the point estimation.On the other hand, it will be the interval estimation, when a certain interval is determined from the observed dataset.It can be said with a definite confidence level that the parameter value will lie within that interval.

So, there are now, might be many new things here.One is that it is the interval. So, this word, the certain interval when we are predicting, then it is the interval estimation.Now, when we say that this is the interval estimation, then certain confidence level comes. So that, with this much confidence or with this much confidence and this confidence are also in this in the statistical sense. So, sometimees the general values that we use for this confidence levels are the 90 percent confidence,95 percent confidence, 99 percent confidence. So, this confidence level can vary from the 0 to 1 basically. So, means 0 percent to 100 percent. So…So, so whenever we say that it is an interval estimate, we this interval estimations are always associated with some confidence level.We will see all these things, how we can how we candeclare that, this is the confidence for this this interval. But, one thing is should be cleared here now is that, this is this confidence level.If I increase the confidence level, so that means, as you can read from this sentence, that it can be said with a definite confidence level, that the parameter value will lie within that interval.

Now, suppose that I take two cases; one is that 95 percent confidence level, other one is the 99 percent confidence level.Then the, obviously, the 99 percent confidence level is more. So that, thethat the chance of the exact parameter value will lie within that interval.In case of the 99 percent confidence, should be more so; obviously, the interval estimate that we are doing should be more should be more wider in case of 99 percent confidence interval, compared to what we can estimate in case of the 95 percent confidence interval. So, more the confidence level, wider is the interval. So, that is what you can we we can see at least at this test.We will discuss all these things.

.(Refer Slide Time: 14:18)



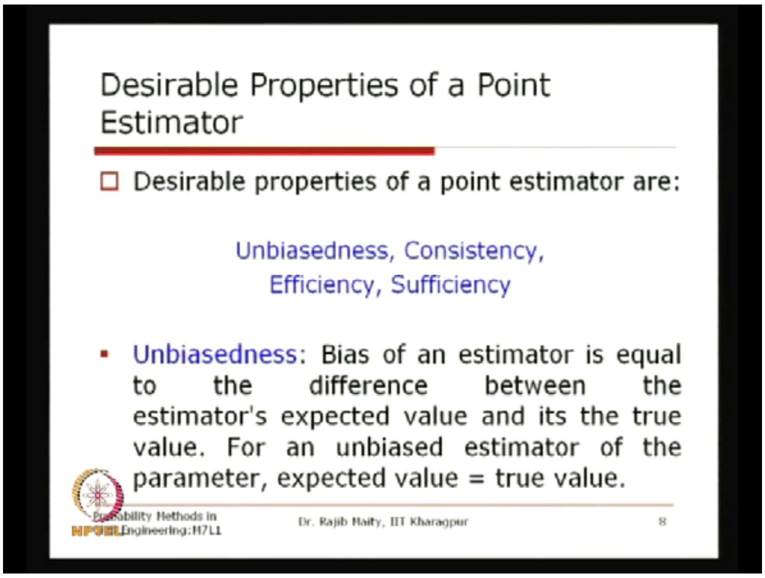
Now, the random sampling and point estimation. So, first we will take that point estimation out of these two estimates. So, as the parameters of the distribution of a population are unknown and it is not feasible to obtain them by studying the entire population, hence, the random sample is generally selected.The parameters of the distribution that are computed based on the analysis of sample values are called the estimator of the parameters.

So, whatever the quantity that we try to estimate as the estimate of that parameter is known as the estimator. So, there are two things, one is the estimation and the other one is the estimators. So, through what through what function I can say, through what function I am trying to estimate that parameter that is known as the estimator.

Thus parameters correspond to the population; while estimators corresponds to the sample. So that, so this is what, when we are talking about the parameters of a distribution or of a of a particular probability distribution, say so, that is generally corresponds to the population.When we talk about the estimator that is correspond to the sample. So, if I take the example of this normal distribution that we have that we have seen earlier, is that there are two parameters now.One is the mu and the other one is the sigma square, mu is the mean and the sigma square is the variance.

So that, mu and sigma square, these are the properties,these are the these are related, these are associated with the population.Now, how we will estimate that mean that is mu and how we will estimate that variance from a sample. So, those are the estimator and those are associated with the sample. That we will see now.

(Refer Slide Time: 16:11)



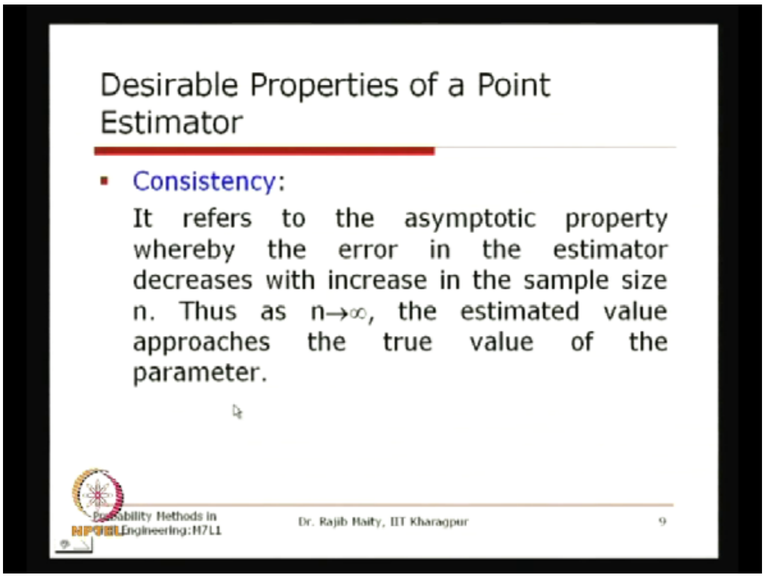
So, when we say that there is a point estimator and that point estimator is, you know that should have some properties.This properties, there are four different properties that a point estimator should have before we can use that one.Because, you see that there could be the many functions that we can use to estimate that particular parameter. But, so, we have to take that particular function, which is a satisfying all these all these properties or at least the the more maximum number of this properties should be satisfied by those estimators.

That these four properties are the unbiasedness, consistency, efficiency and sufficiency. So, we will take one by one.The unbiasedness is the bias of an estimator is equal to the difference between the estimator's expected value and its true value.For an unbiased estimator of the parameter, expected value should be equal to the true value.

Now, when we are talking about this true value; that means, it is the value for the population that is mu. Now, whatever the estimator, from the estimator, which is, we are getting which we should, for which we should use the sample that is available with us.And I, as I told that this estimators also have some distribution. So, as these estimators are also having some distributions so, we can also calculate, whatever the properties that we knew from our earlier lectures and modules is that, so one is that the expected value. So, if I take the expectation of that estimator itself, so, that will reach to one one value.

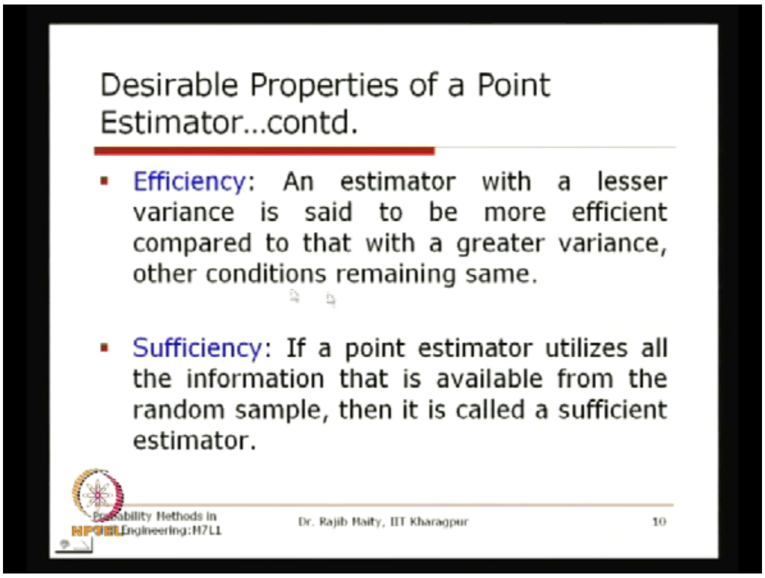
Now, the difference between the the this expected value of the estimator and its true value should be obviously, the desirable thing is that, it should be as minimum as possible. So, and that is the bias. So, the estimators should be unbiased and this is the property that is written here.At least, when when we see that if the n tends to infinity, means n, when the n is the number of sample that we take. So, when the n tends to infinity, then if that expectation of the estimator should be equal to its the populationparameter. So, if that is the case that is basically the check for this unbiasedness.

(Refer Slide Time: 18:42)



Second is the consistency.It refers to the asymptotic property, whereby, the error in the estimator decreases with the increase in the sample size of n.Thus as n tends to infinity, estimated value approaches to the true value of the parameter. So, this is the consistency.That is, as we are increasing, the increasing the number of element in the sample, the estimators should be such that that, the value of the estimators should be should approach to the to the true value of the parameter, that is the parameter of the population.

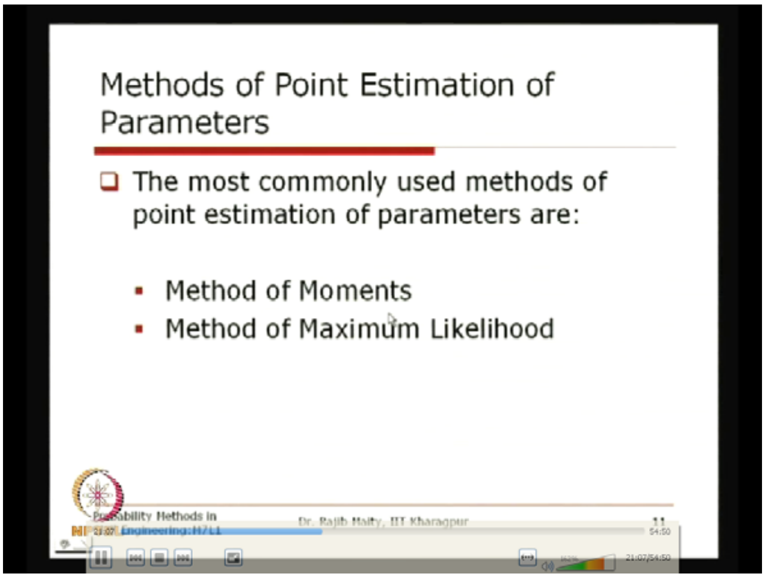
(Refer Slide Time: 19:24)



Third one is the efficiency.An estimator with lesser variance is said to be more efficient compared to that with a greater variance keeping,keeping other conditions same. So here, you can see that suppose, we got two estimators and both the estimators can satisfy the first two properties; one is the both are unbiased and both are what is called, that consistent. So, if both, are both are satisfying the first two condition, then we have to we have to select the estimator, which is having the less amount of the variance.

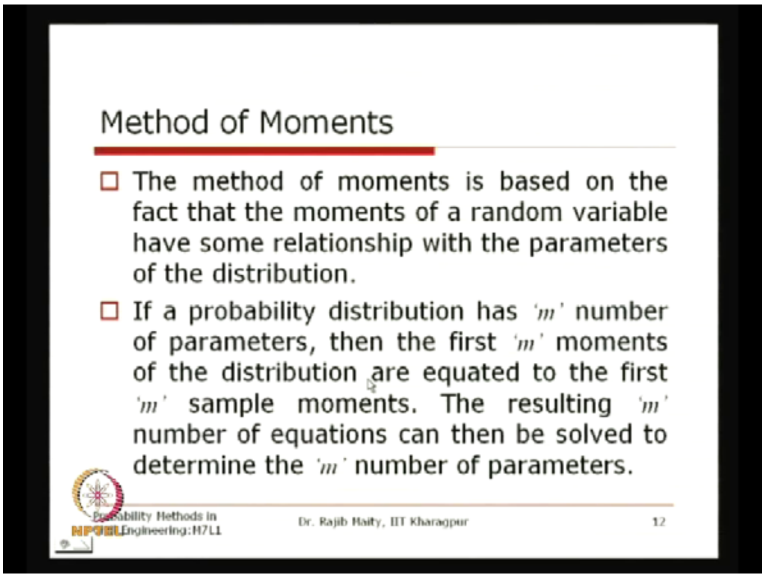
So, this is basically related thing, which one is more efficient more efficient estimator, the the estimator having the lesser variance is is more efficient in such cases. Sufficiency, if a point estimator utilizes all the information that is available from the random sample, then it is called a sufficient estimator.

(Refer Slide Time: 20:33)



Now, now we will discuss this thing through, there are two commonly used method of this point estimator. This will apply to one of the known distribution and can easily be handled through the, this hand calculation problem, that we will see. So, basically the two commonly used method for this point estimation is are the point estimation of the parameter are the method of moments and other one is method of maximum likelihood.

(Refer Slide Time: 21:10)



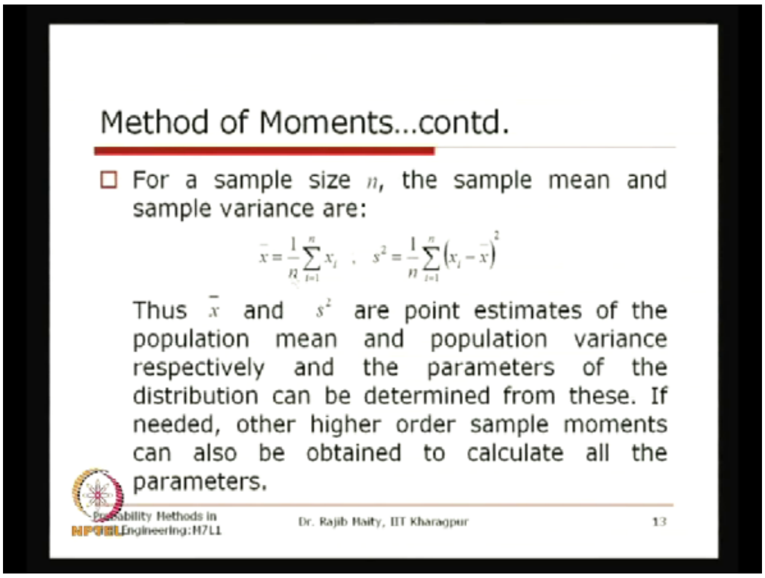
So, first we will take this method of moments.The method of moment is based on the fact, that the moments of a random variable have some relationship with the parameters of the distribution. So, whatever the sample that we are having, we will calculate what are the moments of thatrandom variable and that should have some relationship with the parameters of the distribution.We have discussed this in the first moment, second moment with respect to the origin.

You know the first moment with respect to the origin is the mean, and second moment with respect to the origin, we generally we can take, but, we generally do not take from the second moment onwards.We take it with respect to the mean and that gives the expression for the spread of the distribution.All these things, we have discussed earlier and we have also discussed that the first moment with respect to the mean is zero. So, that concept we have discussed earlier. So, here we will use that concept to use in this method of moments.

So, if a probability distribution has m number of parameters, then the first m moments of the distribution are equated to the first m sample moments.The resulting m number of equations can then be solved to determine the m number of parameters. So, we will take two examples.One is that; say for example, that exponential distribution. So, exponential distribution is having one parameter, that you know, the lambda, one parameter is having. So, the first moments, the first moment of the sample that you are having should be equated to the first sample moment.Whatever the sample that you are having, we can also calculate what is this first moment, with respect to origin,of course, and that we can equate to to get that what is the estimate for this lambda.

We take another example.That is, say that normal distribution or gamma distribution, both are having two parameters. So,for normal it is mu and sigma square and for gamma you know that it is alpha and beta. So that, two parameters are there, then the first two moments, first moment and the second moment should be equated to the first two moments of the sample and this should be equated to those those parameters.That is, first one should be equated to the mu in case of normal distribution, and second one should be equated to the sigma square in case of normal distribution. So, there are two unknowns now and there are two equations also.That we can solve, to get what are the estimate of those parameters.

(Refer Slide Time: 23:52)



Now, for a sample size n, the sample mean and sample variance that what we use as their estimator, is that x bar is 1 by n summation of all the all the sample that we are having.Here, the sample size is n as it is mentioned here. So, we will sum up all the sample and divide by n, that is, basically the arithmetic mean.And this is, we can show that this is the estimator for the sample mean, which is satisfying all the properties that we can, that we have discussed.That is, four requirements for the estimator.

Now this this is s square, which is the sample variance, which is 1 by n summation of x I minus x bar.Again, this x bar is that mean estimated through this equation and this is square summing up from this I minus I equals to 1 to n.Now, this is the point estimator for the variance.

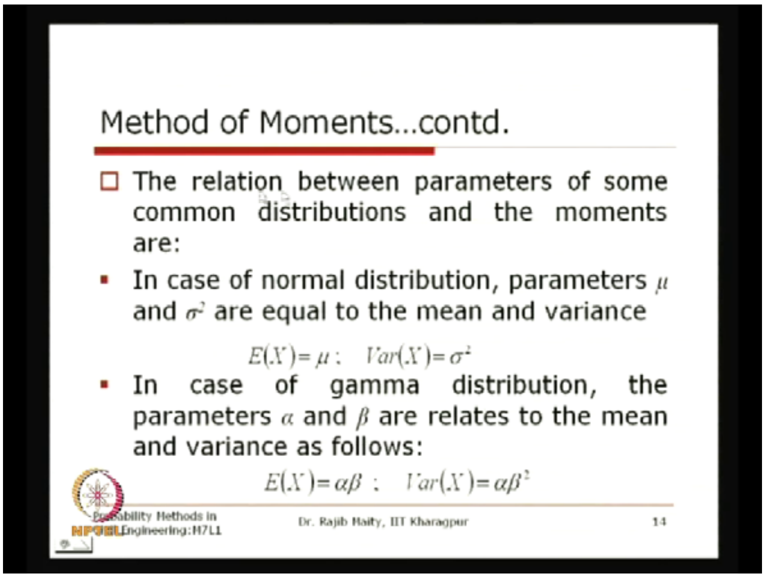
Now, this one, it can, we can show thatif we use this one, this will not be not be unbiased.Now, to make this one unbiased, that is, as I was telling that if the n tends to infinity, it should reach to the actual value.Then, what we have to use?We have to use that 1 by n minus 1. So, if we use that n minus 1, so, in many text what you will see that this estimator for this variance is that a 1 by n minus 1. So, that minus 1 is basically to make the estimator unbiased.And you know that, this minus 1, that we are talking about is basically a, we can say that it is a, it is that degrees of freedom. So, one degree of freedom is lost here, that is why we have to, we need to get that n minus 1.Why it is lost is that, we are using this mean which is also estimated from the sample itself.

So, that is why. So, one degrees of freedom is lost here, and so, we have to write that n minus 1.Now, somehow if you know what is the population mean and if you can replace this x bar with respect to the population mean, that is x I minus mu, that square and if you if we use this quantity, then there is no need to make that n minus 1, then that 1 minus n is sufficient.

So, thus this x bar and s square are the point estimates of the population mean and the population variance. So, these are the point estimate for those things, which is obtained from the sample and the parameters of the distribution can be determined from these. So, these are the first two,that mean the population are are the estimate for the population mean and population variance. So if needed, other higher order sample moments can also be obtained, to calculate all the parameters.

So that means, that we can go to the skewness.With sample estimate of the skewness, we can go to the sample estimate of the catharsis and all.

(Refer Slide Time: 27:10)



So now, the relation between the parameter of some common distribution and the moments are say for example, here we are taking the normal distribution first.There are two parameters that you know, one is that mu and other one is the sigma square are equal to the mean and variance, like this. So, expectation of x is equals to mean and the variance is equal to sigma square.

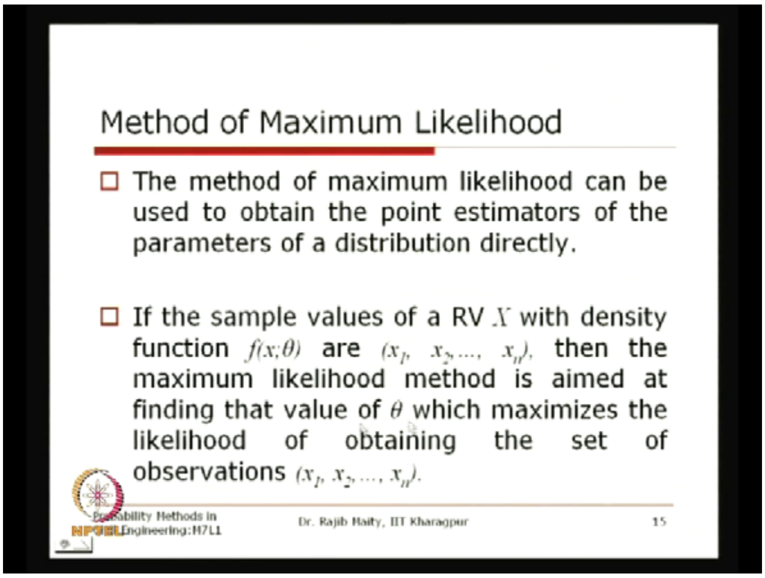
So, what you can see is that, directly we can use whatever, we, whatever the estimate that we have done, that we can put here to get what is this population mean and the population variance.

Now, in case of the gamma distribution, now the parameters are the alpha and beta, that relates to the mean and variance as as follows.That you know, that expectation of a x in case of the gamma distribution is the alpha beta and the and the variance is alpha beta square.

So, all this things we have discussed in the earlier modules.You can refer to that those lectures.Here, we are just using that what we have seen in the earlier modules.

So, fine. So, as to conclude this one is that, now you are having, basically, this you can estimate from this sample.This also you can estimate from the sample.Now, if you equate this one with these two parameters, that is, what we are saying, so, we are having that two unknown shear alpha and beta and there are two equations. So, these two can be can be solved to get what is the estimate of this alpha and beta.And here, it is straightforward because, we are getting this mu, is the expectation which is directly should be the equal to this x bar.Whatever you have seen and the variance, should be that sigma square, what you have seen in this slide. Obviously, if you are using this x bar, that is, which is also estimated from this sample, this instead of 1 by n, it should be n minus 1.

(Refer Slide Time: 29:12)

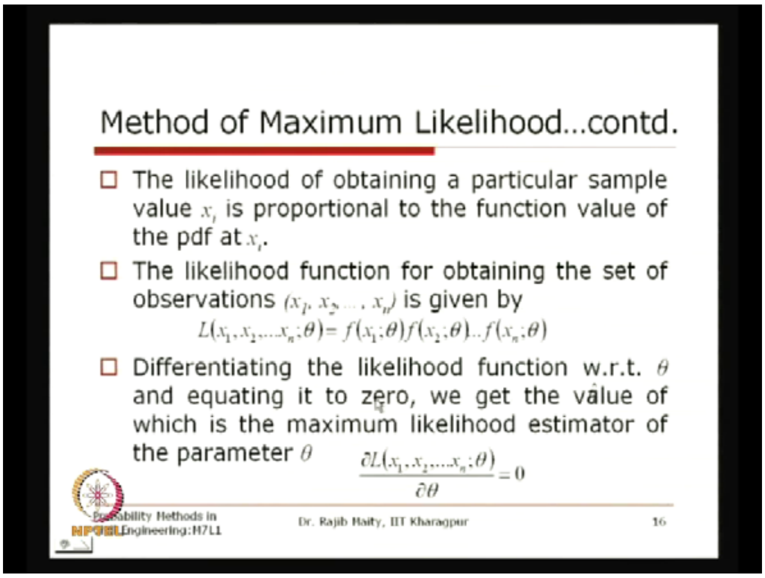


Now, the second method is that method of maximum likelihood.The method of maximum likelihood, can be used to obtain the point estimator of the of the parameters of a distribution directly.

So, there are some shortcomings of this method of moments, which we generally suppose that, sometimes the estimator that is using the method of moments, what we get is that, sometimes the solutions are once it is solved, we get that the, it is not within the within the range of these parameters. So, sometimes this kind of things are observed. So, that is the criticism over the method of moments. So there, this method of likelihood has found to be moremore effective.Because, here directly the distribution we are using and we are we are developing a likelihood function and that likelihood function is maximized to estimate thatthat particular parameter.Now, you know that how to maximize that likelihood function.First of all, we will see what is the likelihood function and then we will maximize that one.Now, as many parameters are there, so, we have to maximize all those parameters. So that, we will get those many equation to have to solve them to get those estimate.

So, suppose that if a sample value of a random variable x with density function f x and the parameter is theta here are this x 1 x 2 x n then, the maximum likelihood method is aimed at finding that value of theta, which maximizes the likelihood of obtaining the set of observations x 1 x 2 x n.

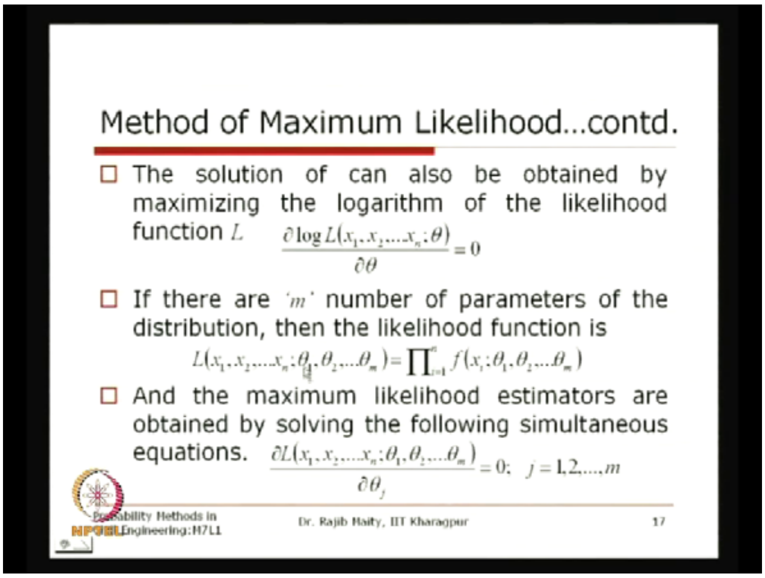
(Refer Slide Time: 30:54)



So, what we have to do is that the likelihood of the of obtaining a particular sample x Iis proportional to the function value of the pdf at x i. So, so this x I means from this x 1 x 2 x 3, we have to calculate what is that likelihood function. So, the likelihood function for obtaining the set of this observation x 1 x 2 x n is given by these values of this distribution at each point.That is, what is the value of that function at x 1, at x 2, at x 3 and all and their multiplication.

So, this differentiating the likelihood function with respect to theta now and equating it to 0, so, this is basically we are we are maximizing the likelihood function.We are finding where this likelihood function will be maximized.We get the value, that is theta value of of that estimate, that is theta hat, we can just set which is the maximum likelihood estimator of this parameter theta, that is this 1 will equate to 0 and then we will get that estimate of that theta hat.

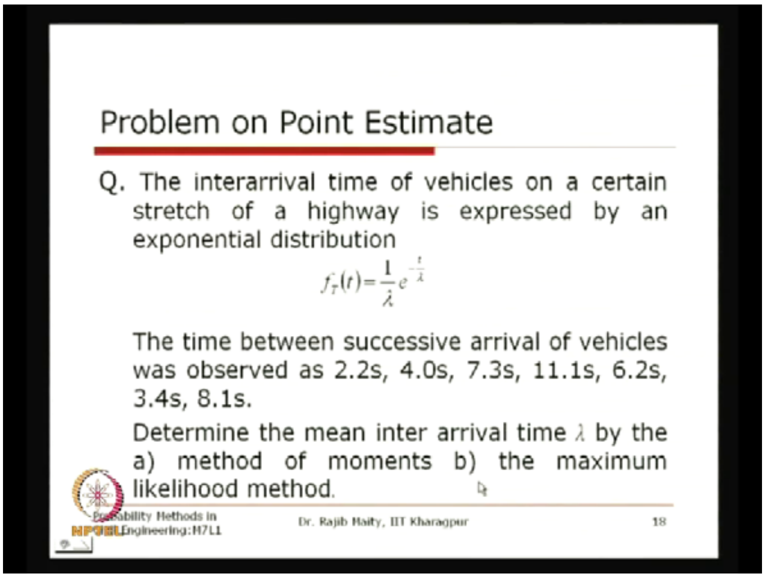
(Refer Slide Time: 32:11)



The solution can also be obtained by maximizing the logarithm of this likelihood function. So, if we take the log also, sometimes this it will be more that. So, far as mathematical calculation is concerned, it a may become easier that will take the log of this likelihood function and will differentiate with respect to the parameter, can if there are m numbers of parameters of the distribution, then the likelihood function is like this. So, there are theta 1 theta 2 up to theta m, these are the parameters of the distribution and this is you know, that this is a multiplication sign of this at each sample point x i.

The maximum likelihood estimators are obtained by solving the following simultaneous equations. So, we will, we have to take the differentiation with respect to each parameter theta j, j can vary from 1 to up to m.And we can take this; we can take this parcel derivatives equated to 0. So, we will get m equations to solve them, to get that estimate of this m parameters.

(Refer Slide Time: 32:21)



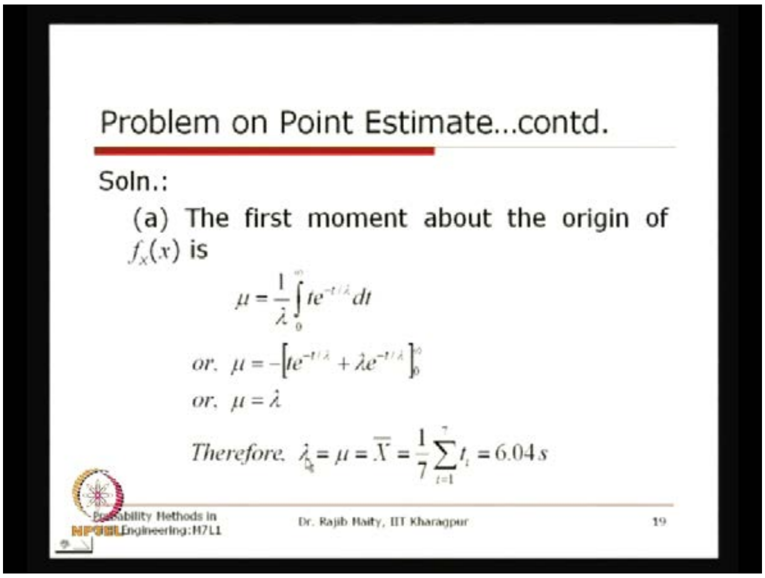
Now, we will take one example using both the methods that we have seen just now.Both, that thatmethod of method of moment and method of maximum likelihood.

So, and here we have taken that example of this exponential distribution.You know this exponential distribution here, the example is on this interarrival time of this vehicle on a certain stretch of a highway is expressed by an exponential distribution, where this f t is equals to 1 by lambda e power minus t by lambda.

Now, means some places or even in this lecture also, earlier we have, we might have used some other form, that is, lambda e power minus lambda t. So, it does not matter. So, there just parameter is taken that 1 by lambda here. So, here also. So, the form is not changing, only thing the parameter represents in a different way. So, if we use that, the the other form also that is lambda e power minus lambda t, then also we can we can get the same result that we see now.

So, and obviously, here this t is greater than equal to 0.Now, there are some samples has been taken, that isthe time between the successive arrival of the vehicle was observed as 2.2 seconds, 4 seconds, 7.3 seconds, 1.1 seconds, 6.2 second, 3.4 seconds and 8.1 seconds. So, this is a sample that we have collected.Now, determine the mean, inter arrival time that is lambda by, so, or I, if I do not even want to mention this; what is this, I just want to estimate, what is the parameter lambda for this distribution by two methods.One is the method of moments and other one is the method of maximum likelihood.

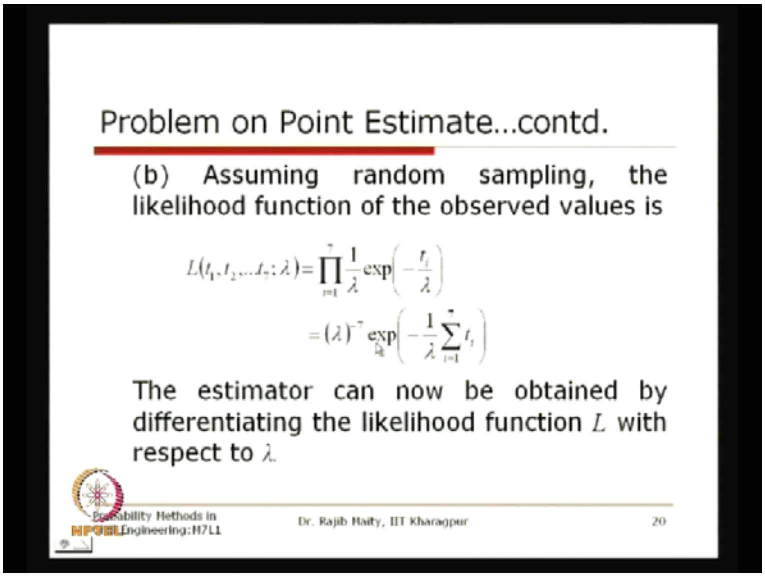
(Refer Slide Time: 35:12)



So, first the method of moment.When we have said that, we should take the moment with respect to the origin and that we should equate with this mean. So, if we take this moment, you know that this is a first moment with respect to mean that we have discussed in the earlier classes, is that that t multiplied by that distribution.Taking this integration over the entire support of this distribution and here this exponential distribution having the support 0 to infinity. So, we will take this integration and if we solve this one, we can see that this mu is equals to lambda here. So, this is mu. So, the lambda is equals to mu, which is, we can obtained from this sample of this x bar, which is the estimator for this mean. So, this 1 by 7 summation of all this t I, so, it is the arithmetic mean. So, 6.04 second is the estimate for this lambda.

Now, if we use the other form of this exponential distribution that is that is lambda e power minus lambda t then, you can see that this mu will become that 1 by lambda. So, there the lambda will be equals to 1 by x bar and this will be 1 by 6.04 second. So, it depends on what way parameters is used here.

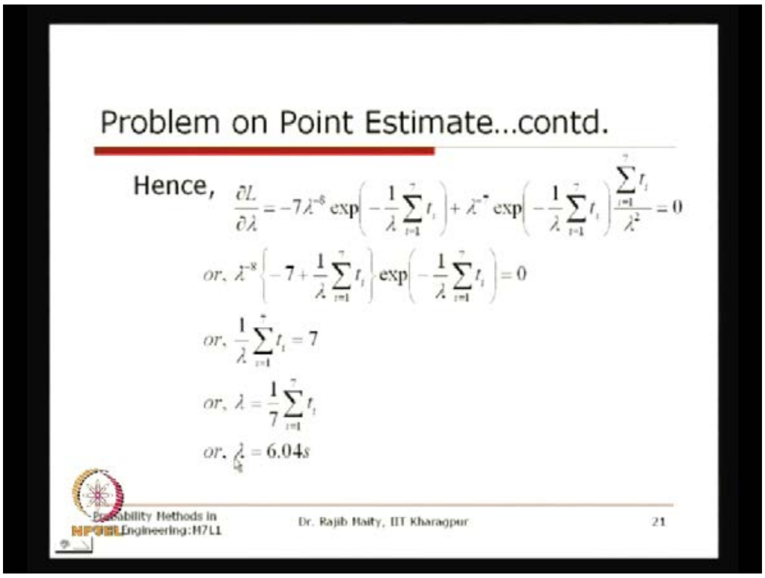
(Refer Slide Time: 36:26)



And, the other one that is use of this maximum likelihood maximum likelihood function.That is, so, assuming the random sampling, the likelihood function of the observed value is that is t 1, t 2, t 3 up to t 7. So, all this sample will take and this will get what is the likelihood function first.

So, 1 by lambda e power. So, e power minus t I by lambda. So, at all this observation, we are getting this value and we are multiplying them with each other. So, this lambda power minus 7 exponential of this value, we will get and the estimator can now be obtained by differentiating the likelihood function l with respect to the parameter that is lambda.

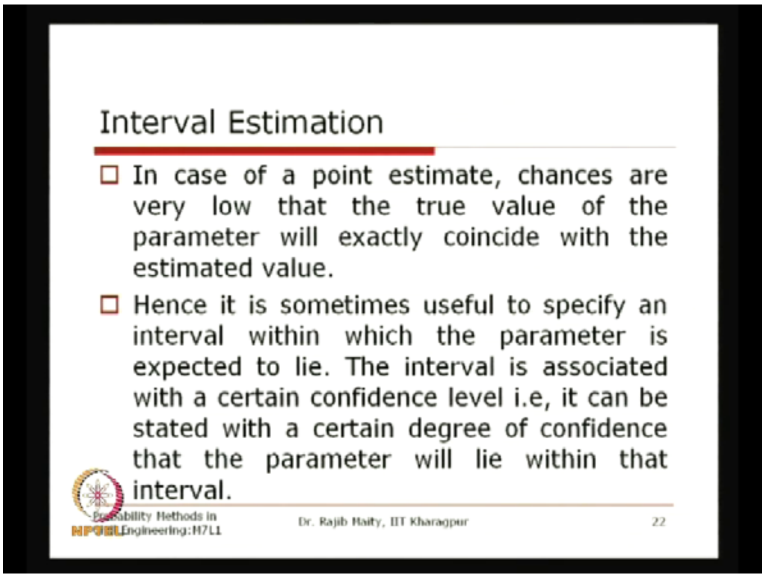
(Refer Slide Time: 37:12)



So, if we do that that, if we do this derivative, we take with respect to the lambda, then we will get this form.If we just equate it to the 0, then after solving this form, you will get that again the lambda is equals to 6.04 second.

So, for this one, this example, that we have shown for both the method, whatever the parameter that we have that we have estimated are same for the lambda is equals to 6.04 second for both the methods; that is method of moment and method of maximum likelihood. But sometimes, in some problems, in some distribution this two estimate may not be same.

(Refer Slide Time: 37:56)



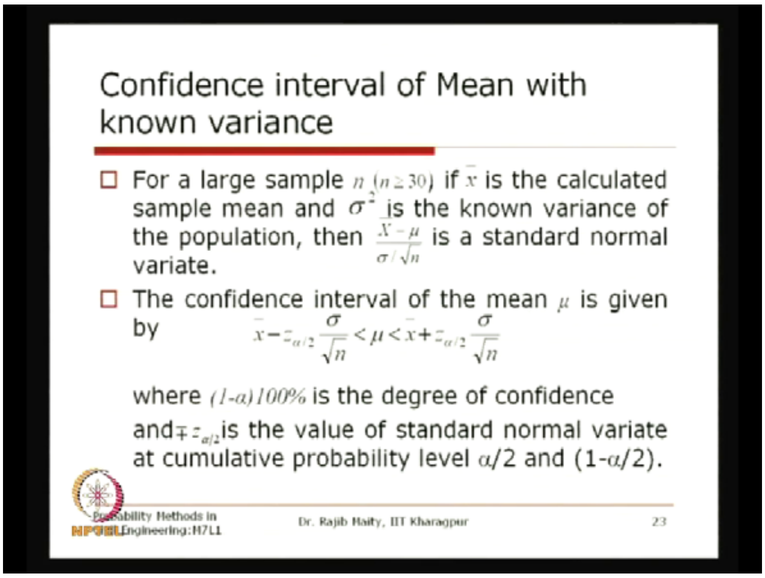
So, next we will take that interval estimation. So, as I was telling that in this point estimation, we generally get a single value.That is, what we have seen in in the previous example also, in both the methods of method of moments or method of maximum likelihood, we get the single value.That is, the lambda value that you have seen that 6.04 second. So, only a single value that you have seen in the point estimate.

Now, the the interval estimate, we generally look for a interval, in which, with some confidence that actual value of the parameter should lie. So, this is what this interval estimation.

So, in case of point estimate, the chances are very low that the true value of the parameter will exactly coincide with the estimated value. So, as the sample is finite, always there will be some error. So, hence sometimes, it is desirable or it is useful to specify an interval within which the parameter is expected to lie.

The interval is associated with certain confidence level that is, it can be stated with certain degree of confidence that the parameter will lie within that interval. So, this is what we have, we are discussing few minutes ago, that always this estimate whenever we say that this is the this is the interval, so, that interval must be associated with some of some confidence level, in statistical sense.

(Refer Slide Time: 39:27)



Well. So, the confidence interval of the mean with known variance. So, so whatever the estimators that we have seen for this mean, and as we are taking it from the sample, so, that will also have some sampling distribution; it should have.And if we somehow know what is the variance and known variance means, we know the population variance.If we know that, one then, how we can get that confidence interval for the mean. So, for a large sample n, so, large sample is again, this is a subjective word. So, generally we can, we have seen that, if n is greater than equal to 30, then you can say that in case of mean only, so, not for all in general case, in case of mean, if the sample size is greater than 30, then we can say that it is a large sample. So, in such case, if x bar is calculated sample mean and the sigma square is the known variance of the population. So, sigma square I know which is exactly which is actual value of this population.How we know, that is the second issue, but, we know this sample population we know the variance of the population.Only thing we are interested to know, what is the interval for this sample mean.

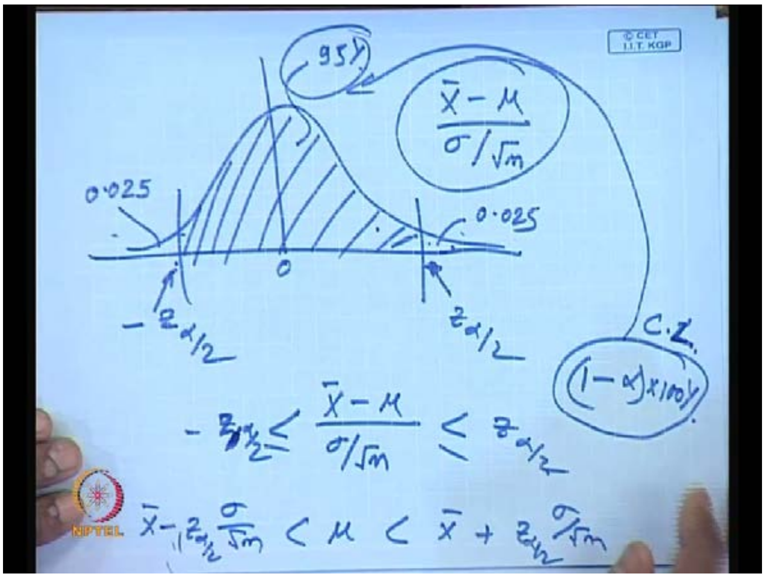
Then, it is, it can be shown that this x bar is is a normal distribution with mean equals to mu, which is the population mean and the variance of this x bar, that is, this sample mean, variance of the x bar is sigma square by n or the standard deviation of this x bar is sigma by square root n.

So, this what I want to repeat once again, that this sigma square is the variance of the population for the random variable x.Now, what we are talking about is this x bar.This x bar is again another random variable, which is normally distributed having the same mean of the population, which is mu and its standard deviation is sigma by square root n.

Now, you know from the, from our earlier lecture, that this if we just take this quantity now, that this, what is the random variable minus its mean divided by its standard deviation is a is a is a standard normal distribution. So, that is why this x bar minus mu by sigma square root n is a standard normal variate.

So, now for once we know that this is the quantity and this follow a standard normal distribution.Now, we can calculate whatever the confidence interval that we are looking for. So, the confidence interval of the mean mu is given by is this, that is, x bar.Basically, we are just equating it with two side of this standard normal distribution.

(Refer Slide Time: 42:36)



Now, if you if you see this one here.So, if this is your standard normal distribution, then basically that that quantity, that is x bar minus mu by sigma by square root n, that should, so, this is your, this is the standard normal distribution. So, this the distribution that I have drawn is a standard normal distribution. So, this should lie between these two values here, should in such a way that this area should be your, that whatever the confidence limit that you wish to specify.

Now, so, so this is the confidence level of that estimated that it should have.Now, if I just say that this this confidence level is, say at some some level, say that 95 percent confidence level. So, whatever is remaining here is your 0.025 and whatever is remaining here is again 0.025.

So, we have to find out these two values.Suppose that, if I just put that one, z alpha by 2 and and this is say that z alpha by 2; obviously, this will be the negative side, this is 0 for the standard 1. So, this one, so, this quantity should lie between this z alpha by 2 and this z alpha by 2, where this alpha by 2 are basically the cumulative probability up to that point.

So...So, this x bar minus mu divided by sigma by square root n should lie between this two, z alpha by 2 and this minus z alpha by, sorry, z alpha by 2. So, z is the is the variate, is the is the reduced variate for this standard normal distribution.And, this alpha by 2 is this part where we are. So, the confidence, if I just want to relate to this confidence level is that, it will be that 1 minus alpha multiplied by 100 confidence level. So, this is the confidence level that we can say. So, that if this area, this white area is the alpha by 2, then this 95, that is here in case of 0.25 here, can be relate to this one.

So, this is that confidence interval that we are, that is the confidence level that we are talking about for this, once we can put this limits as this z alpha by 2.Now, if I just do some arithmetic change, then it will come like this; that is, mu should be equals to here that x bar plus z alpha by 2 sigma by square root n.Here it will be, x bar minus z alpha by 2 sigma by square root n.

So, you remember, that even this alpha by 2 when we are talking about. So, so this alpha by 2, it is automatically, when it is coming to this negative side, it have this negative value. So, do not confuse that this will have this have the negative value and this again another negative sign is here. So, this will not be the plus. So, this is a single value that we are using here, which should havethat this particular quantity.From the symmetry, both the values will be same numerically, only this one will be positive and this one will be negative.

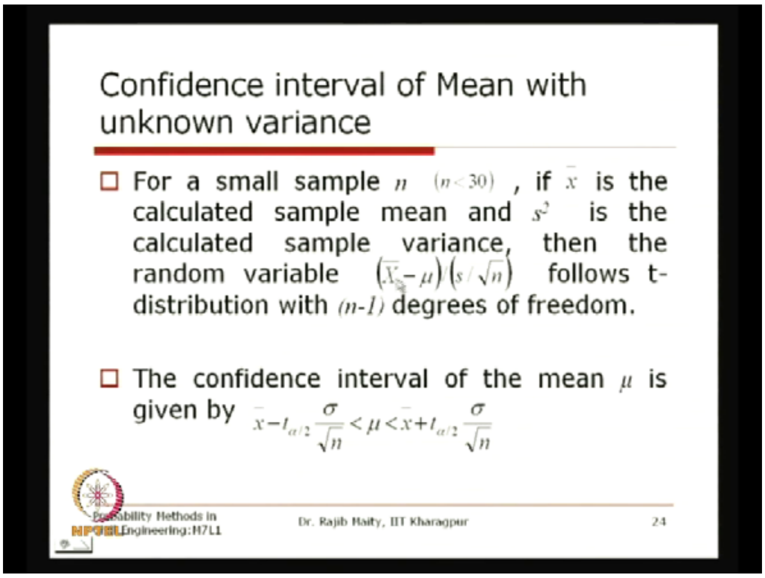
So, this is what is mention here, that is x bar minus z alpha by 2 sigma by square root n minus mu less than x bar plus z alpha by 2 sigma by square root n. So, you know that from the continuous distribution less than and less than equals to are same.

Where, this 1 minus alpha into 100 percent, this is the quantity, which is the degree of confidence.And here, this minus plus z alpha by 2 is the value of the standard normal variate, at the cumulative probability level alpha by 2 and 1 minus alpha by 2.

So, when you are taking this minus sign minus z alpha by 2, this is is the probability level at this alpha by 2. So, if it is 95, if you put this alpha equals to 0.95 then; obviously, 1 minus, so, sorry, this is full is equals to 0.95, then alpha will become 0.05 and this alpha by 2 will become 0.025.

Now, at this 0.025, this z alpha by 2 will be, say if it is 0.95, you know that this value will be 1.96. So, this will be x bar minus 1.96 multiplied by sigma by square root n and here also, plus 1.96 sigma by square root n this will come.

(Refer Slide Time: 47:36)



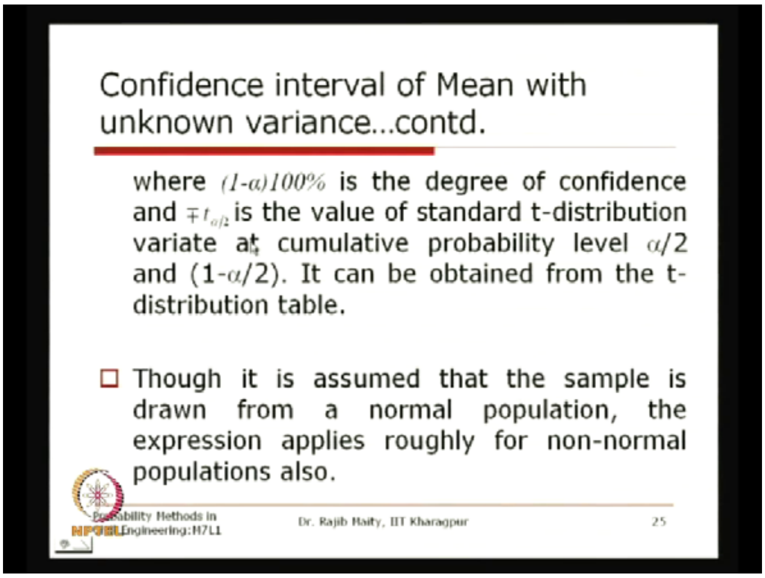
Now, as we are discussing that if the sample is more, if the sample is large, if it is more than 30, now in other case, if the sample size is small, say if it is less than 30 and if the x bar is the calculated sample mean and this, then this s square is this calculated sample variance, then the random variable x bar minus mu divided by s square root of n, follow a t distribution with n minus 1 degrees of freedom.

Here you see, this is that the variance, we do not know that is the unknown variance. So, this one also we have to calculate from the from the sample itself. So, this also, again when you take that this x bar minus this population mean divided by this sample variance divided by square root n.

So, this one instead of following this standard normal distribution, it will follow the t distribution with n minus 1 degrees of freedom. So, this t distributions and all we have discussed earlier.Only thing is that, if the sample size is small, this quantity will follow the t distribution.If the sample size is more and thisthis variance is known, then this will follow a normal distribution.

So, here in this case, this this confidence interval will be x bar minus t alpha by 2 sigma by square root n and x bar plus t alpha by 2 sigma by square root n.So, this is a t distribution with degrees of freedom n minus 1.And, you can see even that from standard text book about this t distribution, when this n n goes beyond 30, then the value of this t alpha by 2 and the z alpha by 2 are essentially same.

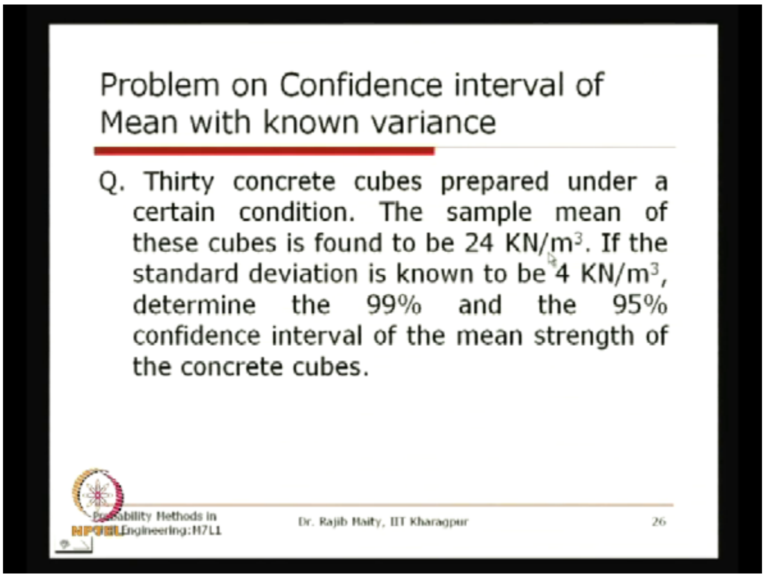
(Refer Slide Time: 49:30)



So,where this again, the 1 minus alpha into 100 percent is the degree of confidence and minus t alpha by 2, is the value of the standard t distribution variate at cumulative probability again, that alpha by 2 and this 1 minus alpha by 2. So, basically the difference between the t and the and the standard normal distribution is at this lower end level, where you will get a a wider estimate of the interval because,the when the sample size is less, as there are uncertainty is more.

Though it is assumed that the sample is drawn from a normal population, the expression applies roughly for non normal population also. So, basically when the population is normal distributions, these are very well acceptable method. But, even though it is non normal also, this method we can apply.

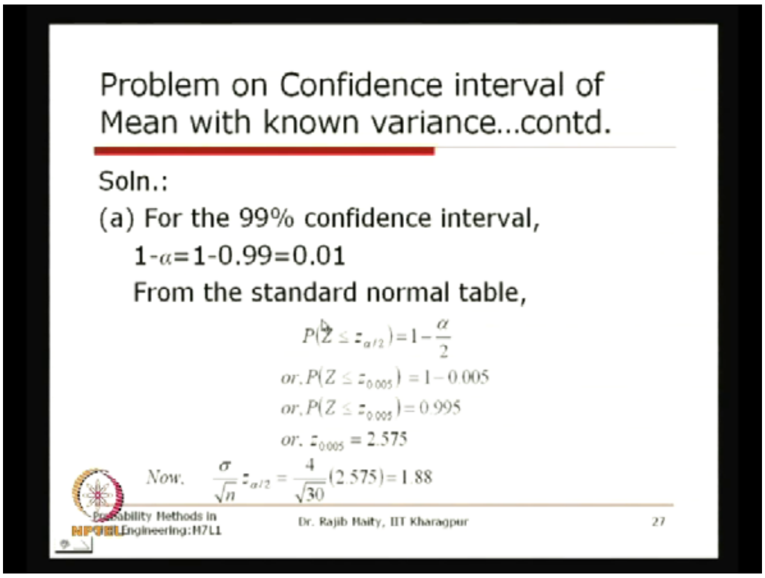
(Refer Slide Time: 50:26)



So, we will take one example of this, whatever we have seen, is that 30 concrete cubes prepared under certain condition.The sample mean of these cubes is found to be 24 kilo Newton per meter cube, if the standard deviation is known to be 4 kilo Newton per meter cube, determine the 99 percent and 95 percent confidence interval of the mean strength of the concrete cube.

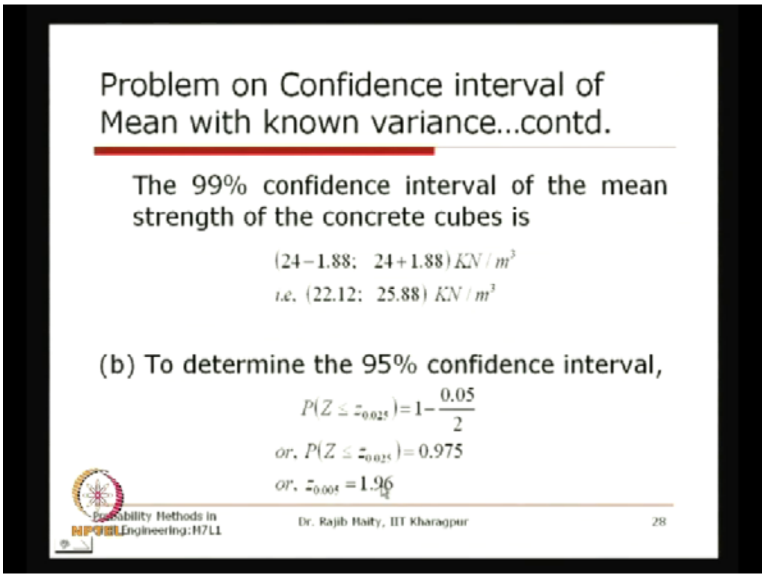
So, this 4 kilo Newton per meter cube is known; it is from the population.And this one, when we say that this one is obtained from this sample.

(Refer Slide Time: 51:10)



So, if we want to solve this one, then you know, that we first of all, we will get that what should be the quintile value for this z alpha by 2 and this is for this 99% confidence interval from the standard normal table.You can see that it is 2.575, which is z alpha by 2 value.So, this sigma by square root n multiplied by z alpha by 2 is equals to 1.88, whatever the data is available.

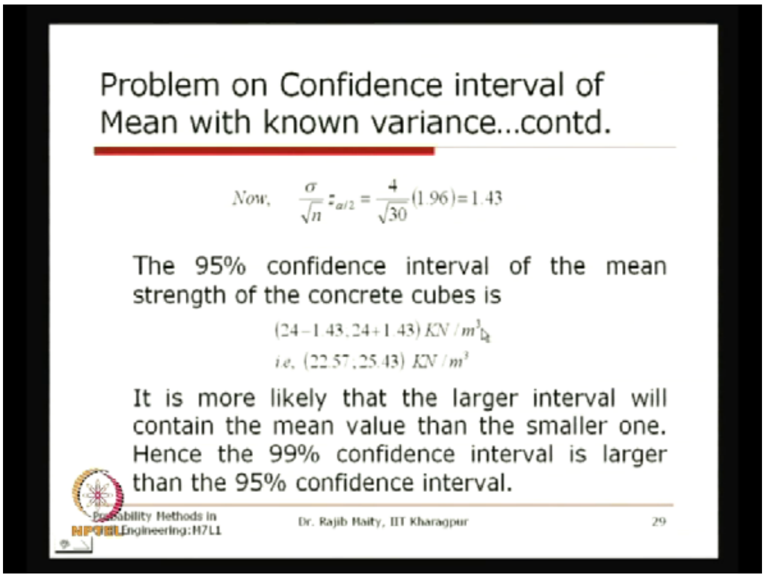
(Refer Slide Time: 51:35)



So, the 99 percent confidence interval will be the mean minus that quantity 1.88 and mean plus 1.88. So, the confidence interval is 22.12 and 25.88 kilo Newton per meter square. Sorry.

To determine the 95 percent confidence interval, again we have to find out the z alpha by 2 and this is your 1.96, earlier it was 2.575. So, it is now 1.96.

(Refer Slide Time: 52:03)



If we calculate this one, it will become 1.43 and the 95 percent confidence interval of the mean strength of this concrete will be 24 minus 1.43 and 24 plus 1.43. So 22.57 and 25.43 kilo Newton per meter square.Sorry.This is or this is the density sorry sorry this is the density. So, this it is kilo meter per meter cube.

But, what we should observe here is that there are two confidence interval.We have determined one is the 99 percent confidence interval and other one is the 95 percent confidence confidence interval.

So, the 95 percent confidence interval is 22.57 to 25.43 whereas, the 99 percent confidence interval is 22.12 and 25.88. So, you can see that this 99 percent confidence interval is wider, because it is more likely that the larger interval will contain the mean value than the smaller one. So, hence the 99 percent confidence interval is larger, when the 95 percent confidence interval whereas compare to this 95 percent confidence interval.

So, this example, that we have used, it is we we we have taken that a variance is known. So, once we we have decided the variance is known, we have used the standard normal distribution. But, in the other case, we have seen that sometimes the sample size is less and we have to we have to use, in that case, we have to use that, what is the t distribution.From the t distribution interval, we have to use that one.

So, maybe we will take up the same example, but, in that time we will just declare that whatever the distribution, whatever the standard deviation that we got, it is not from the population, but, from the sample. So, that example we will basically start from the next lecture and after that we will take what should be the, that estimation for the other parameters like the variance proportion and all.And, we will also relate, we will also see about this test of hypothesis which is obviously, essential when we are comparing the mean of from the two different samples or the variance or proportion from. So, when it is related to two different samples, then we have to go for those testing. So, we will start from this point in the next lecture.Thank you.