INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR

NPTEL
National Programme
on
Technology Enhanced Learning

Probability Methods in Civil Engineering

Prof. Rajib Maity

Department of Civil Engineering
IIT Kharagpur

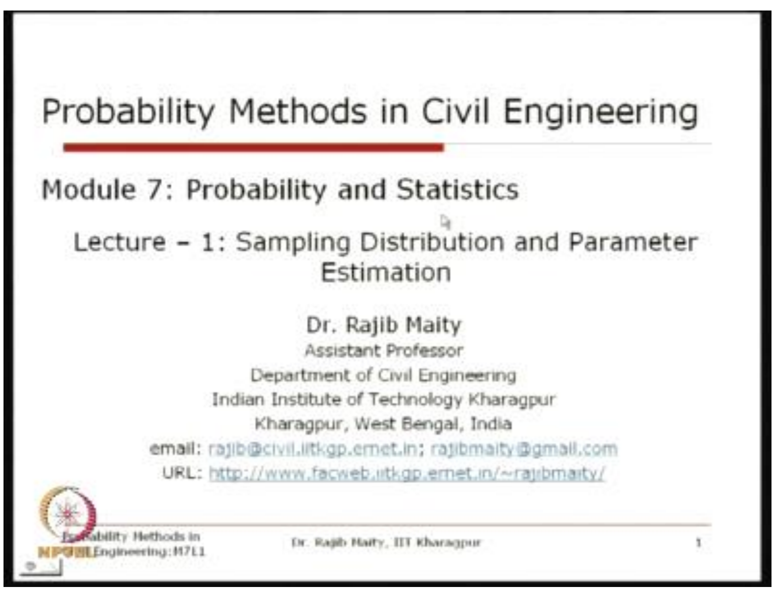
Lecture – 21

Topic

Sampling Distribution and Parameter Estimation

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

(Refer Slide Time: 00:24)



And the module is on probability and statistics in last couple of modules we have seen we have discussed about different theories related to probability basically those whatever we have discussed is associated with the population. Now when we talk about the statistics the is statistics is related to the few samples taken from this population more precisely speaking it is some of this random sampling so the sample that we take that should be the random sample for from the population.

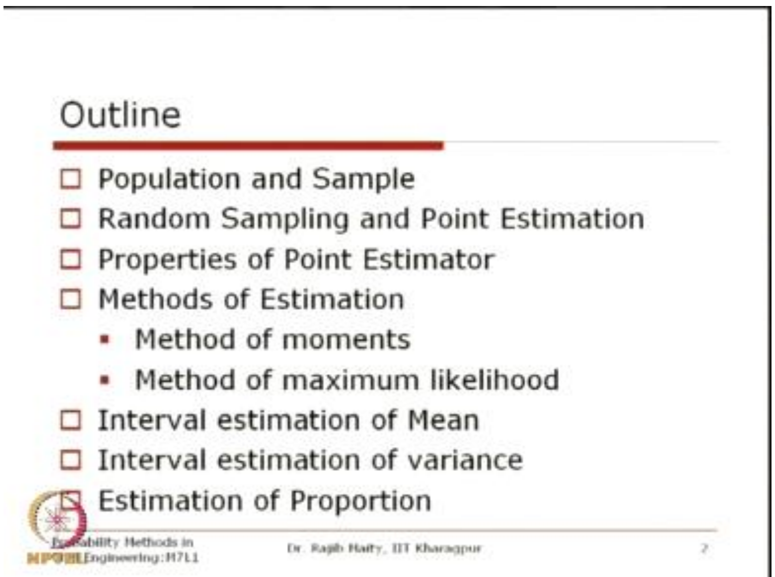
So now as we are using the word the population sample so at the starting of this lecture we should know what does this actually means and you know that when we are talking about some of this parameters for those distribution that we have seen in the last lecture in this last module that there are certain parameters are there which is associated for some standard probability distribution that we have seen. So basically we have to estimate those parameters and those parameters when we estimate we should estimate we should have some sample data and from there we should estimate 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 estimation of those parameters may also change obviously it will change because two samples will never be exactly identical. So those estimations also we will have some distribution. So those are basically called the sampling distribution and so basically our first lecture that we are going to take in this module 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 can 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 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.

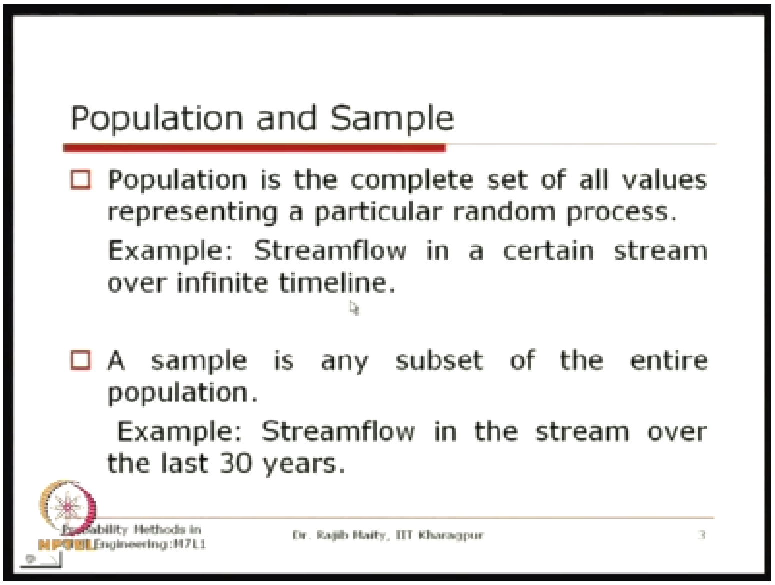
(Refer Slide Time: 03:30)



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 and then the interval estimation of this mean and then interval 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 the 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: 04:53)



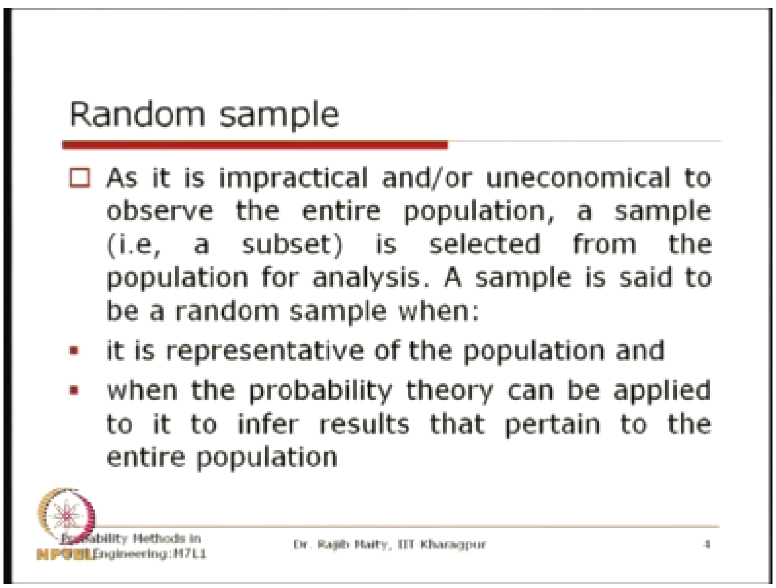
So the population is the complete set of all values representing a particular random process so, when we talk about this all values which represents a particular random process that means this all is very I should say that is including of whatever the possible range is there that should include everything. 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 population is and when we talk about a sample that sample is any subset of the entire population and it is 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 particular dataset 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 it is 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 the all possible values that is possible that we may not have. So we have to rely on some sample and somehow we have to relate whatever we can estimate from the sample and we have to relate it to it is population.

So this is basically the role of this statistics and with this we will just see how we can estimate the parameter show we can estimate the required information from the samples and 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.

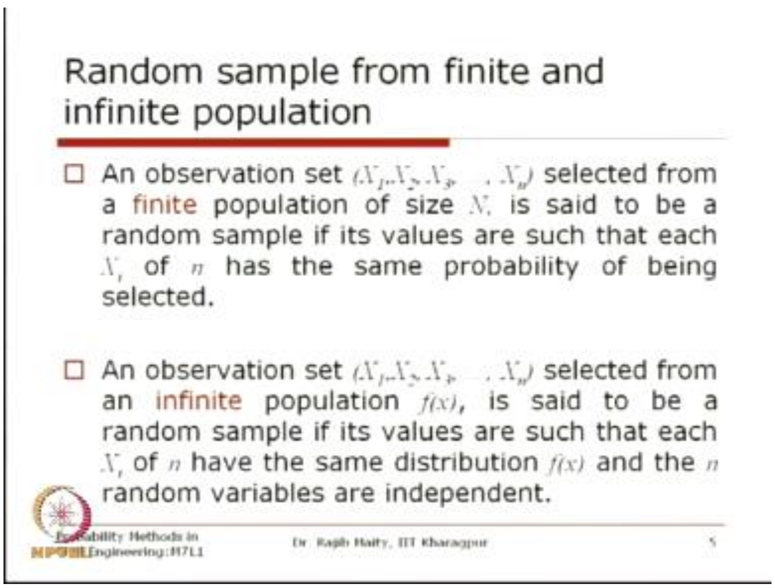
(Refer Slide Time: 07:19)



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 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 now again this what is important that it is should be the representation of the population so what whatever the possible cases that we can see in the population the sample should have that representation in it and second 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: 08:51)



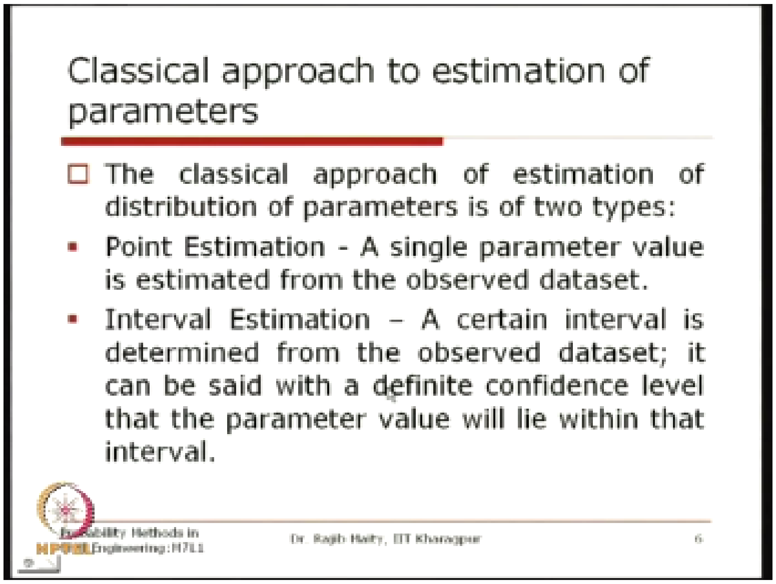
Now the random sample for finite and infinite populations. So you know that there are certain cases where the population can be finite or in other cases it can be infinite or sometimes it can be count ably infinite. So in such cases what should be the random sample? So first we will take the case of this finite population and observation set x1, x2, x3……Xn 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 xi 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 x1, x2, x3 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.

And the second one if the population is infinite and observation set again that x1, x2, x3, xn is selected from an infinite population fx. So obviously when we talk about this infinite series we generally denote it through its probability density function here it is that fx is said to be a random sample if its values are such that each xi(n) have the same distribution fx 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 fx 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 data set 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 data set it can be said with a definite confidence level that the parameter value will lie within that interval.

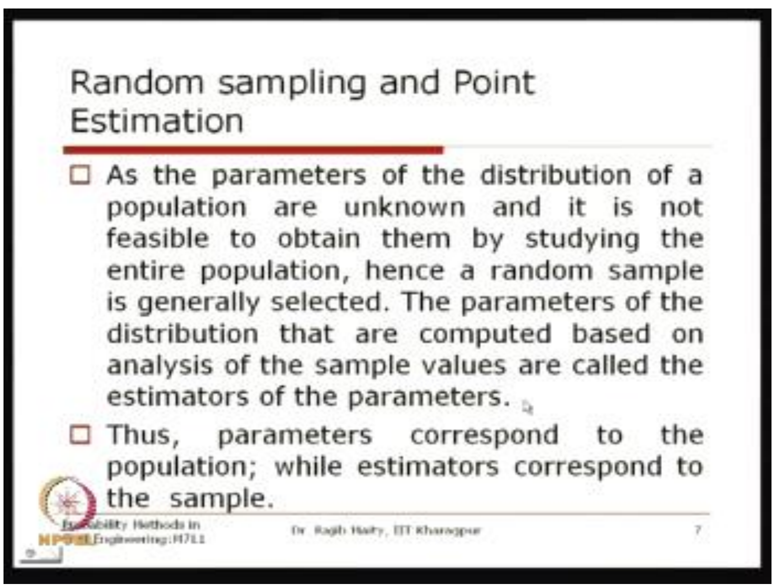
So there are now might be many new things here one is that it is the interval. So this word the certain interval when we are predicting then it is the interval estimation. Now when we say that this is the interval estimation then certain confidence level comes. So that with this much confidence or with this much confidence and this confidence are also in this in the statistical sense.

So sometimes the general values that we use for this confidence levels are the 90 % confidence, 95 % confidence, 99% confidence. So this confidence level can vary from the 0 to 1 basically. So means 0 % to 100 %. So whenever we say that it is an interval estimate we this interval estimations are always associated with some confidence level and we will see all these things how we can how we can declare that this is the confidence for 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% confidence level other one is the 99% confidence level then obviously the 99% confidence level is more.

So that the that the chance of the exact parameter value will lie within that interval in case of the 99 % confidence should be more so obviously the interval estimate that we are doing should be more should be more wider in case of 99 % confidence interval compared to what we can estimate in case of the 95 % confidence interval. So more the confidence level wider is the interval. So that is what you can we can see at least at this test we will discuss all these things.

(Refer Slide Time: 14:19)



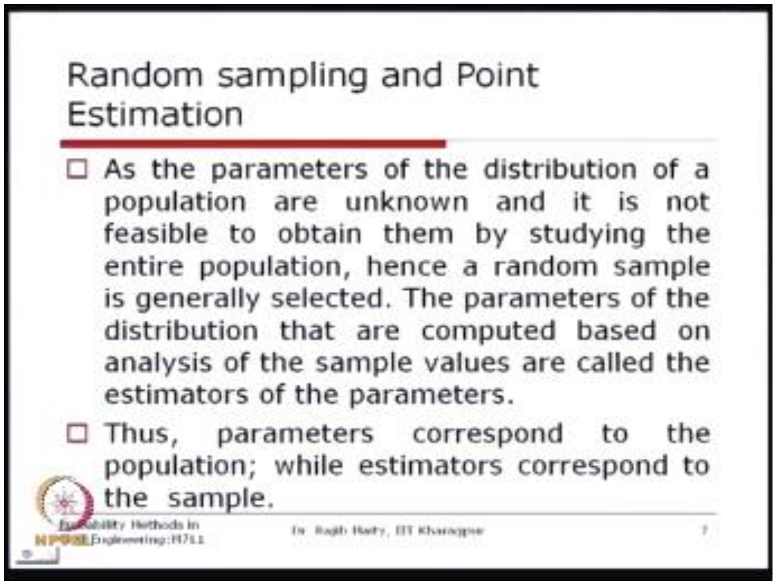
Now the random sampling and point estimation so first we will take that point estimation out of these two estimates. So as the parameters of the distribution of a population are unknown and it is not feasible to obtain them by studying the entire population hence the random sample is generally selected,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

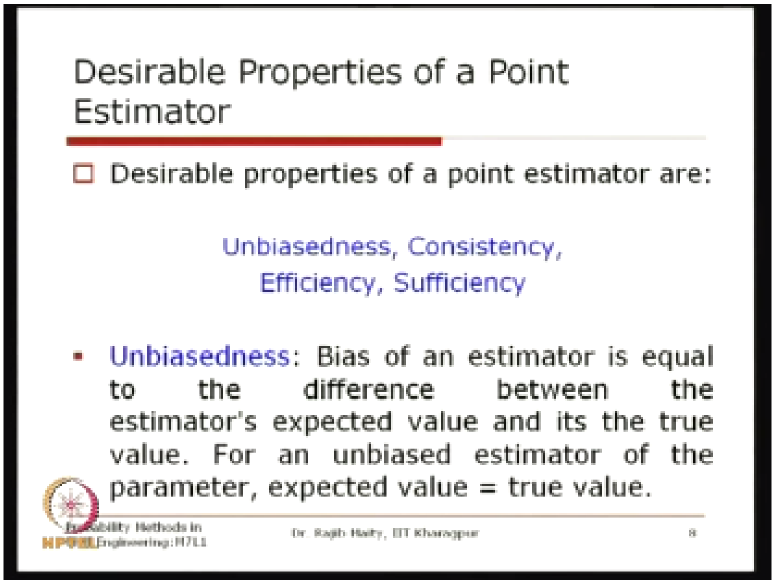
So that so this is what when we are talking about the parameters of a distribution or of a of a particular probability distribution say so that is generally corresponds to the population and when we talk about the estimator that is correspond to the sample. So if I take the example of this normal distribution that we have that we have seen earlier, is that there are two parameters now.

One is the μ and the other one is the σ^2, μ is the mean and the σ^2 sigma square is the variance. So that μ and σ^2 these are the properties these are the these are related these are asssociated with the population. Now how we will estimate that mean that is μ 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: 16:09)



(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 and 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.

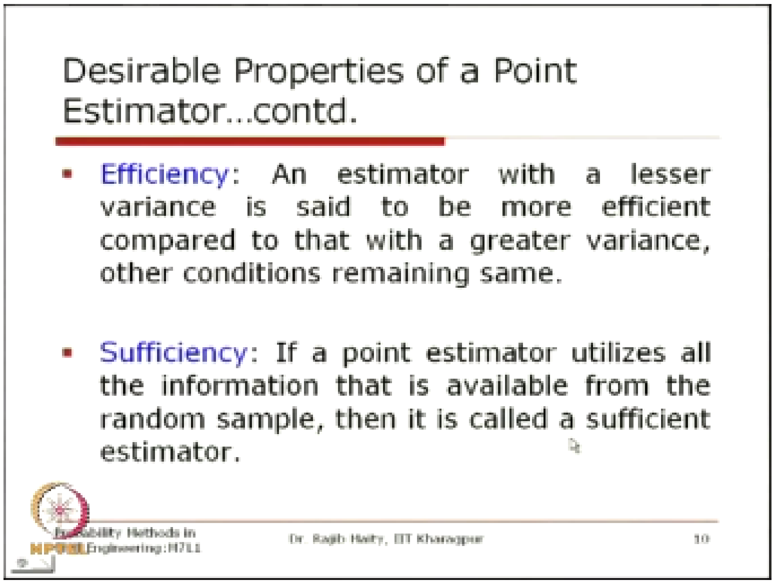
Or atleast the more maximum number of this properties should be satisfied by those estimators. So these four properties are the unbiasedness, consistency, efficiency and sufficiency. So we will take one by one the unbiasedness is the bias of an estimator is equal to the difference between the estimators 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 μ. 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 value.

Now the difference between 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 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 population parameter so if that is the case that is basically the check for this unbiasedness.

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 ∞ 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 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.

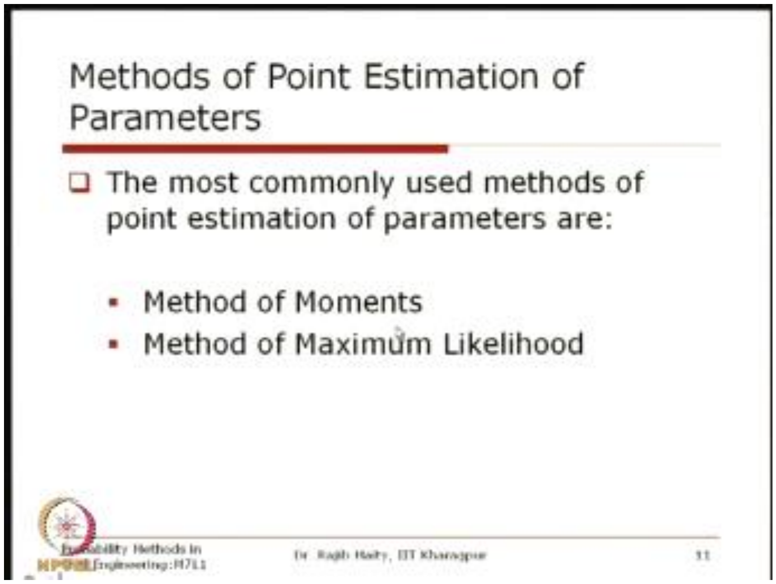
(Refer Slide Time: 19:24)



Third one is the efficiency an estimator with lesser variance is said to be more efficient compared to that with a greater variance keeping 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 estimator having the lesser variance is more efficient in such cases.

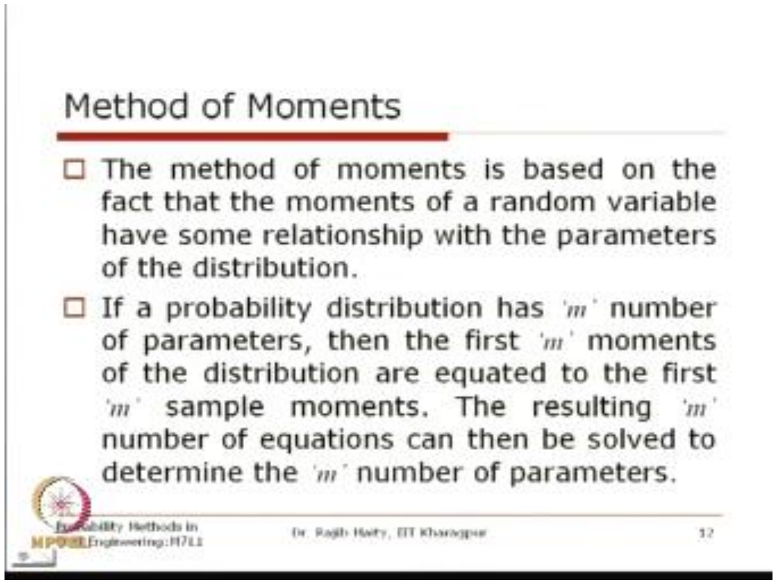
Sufficiency if a point estimator utilizes all the information that is available from the random sample then it is called a sufficient estimator.

(Refer Slide Time: 20:32)



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 so first we will take this method of moments.

(Refer Slide Time: 21:09)



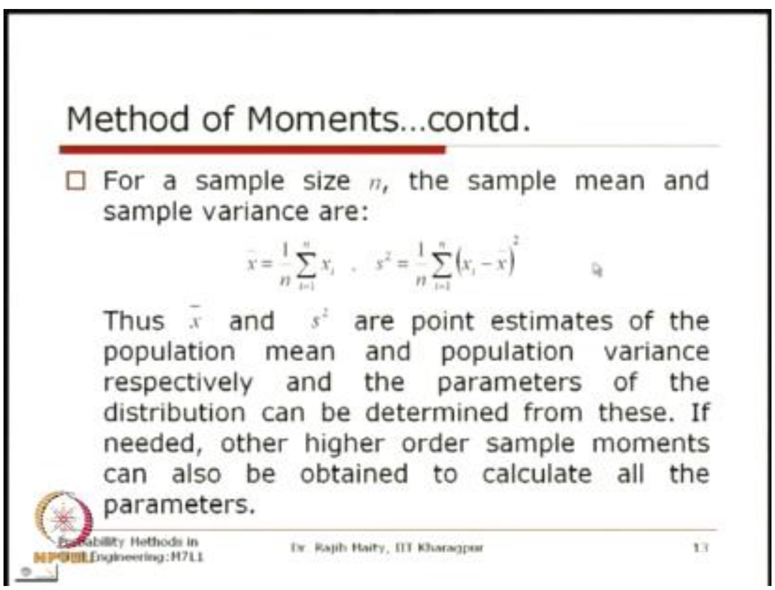
The method of moment is based on the fact that the moments of a random variable have some relationship with the parameters of the distribution so whatever the sample that we are having we will calculate what are the moments of that random variable and that should have some relationship with the parameters of the distribution and 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 0 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 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 λ 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 get that what is the estimate for this λ we take another example that is say that normal distribution or γ distribution both are having two parameters.

So for normal it is μ and σ^2 and for γ you know that it is α and β 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 parameters that is first one should be equated to the mu in case of normal distribution and second one should be equated to the σ^2 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: 23:54)



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 q which is satisfying all the properties that we can that we have discussed that is four requirements for the estimator.

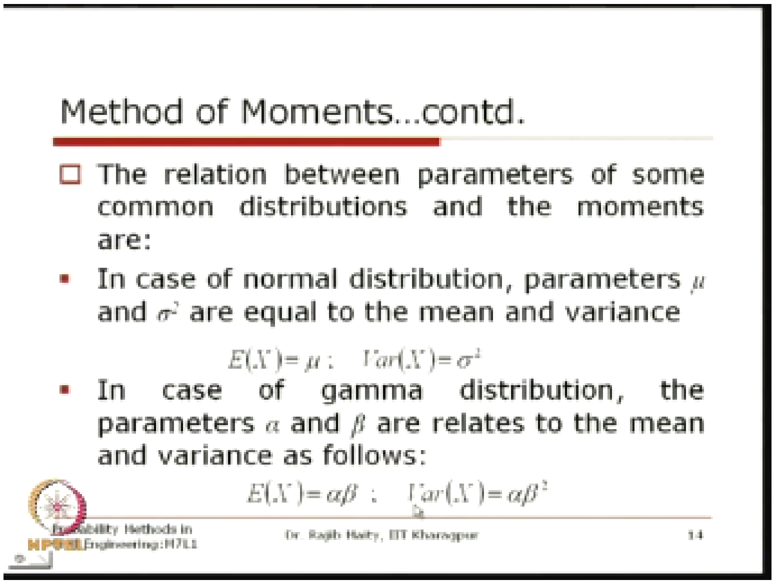
Now this is s^2 which is the sample variance which is 1 by n summation of x I - x bar again this x bar is that mean estimated through this equation and this is square summing up from this I - I = 1 to n now this is the point estimator for the variance now this one it can we can show that if we use this 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 – 1 so if we use that n – 1 so in many text what you will see that this estimator for this variance is that a 1 by n - 1 so that -1 is basically to make the estimator unbiased and you know that this – 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- 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 – 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 - μ that square and if you if we use this quantity then there is no need to make that n - 1 then that1 - 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 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, we can go to the sample estimate of the catharsis and all.

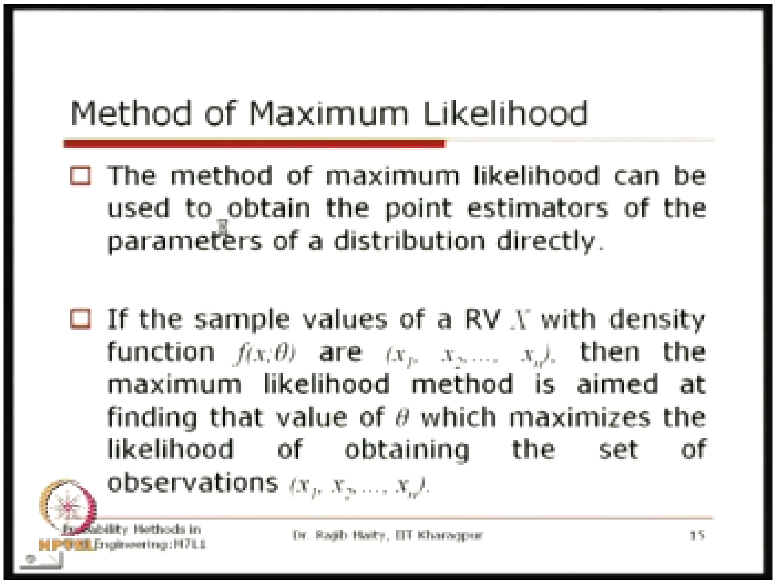
(Refer Slide Time: 27:10)



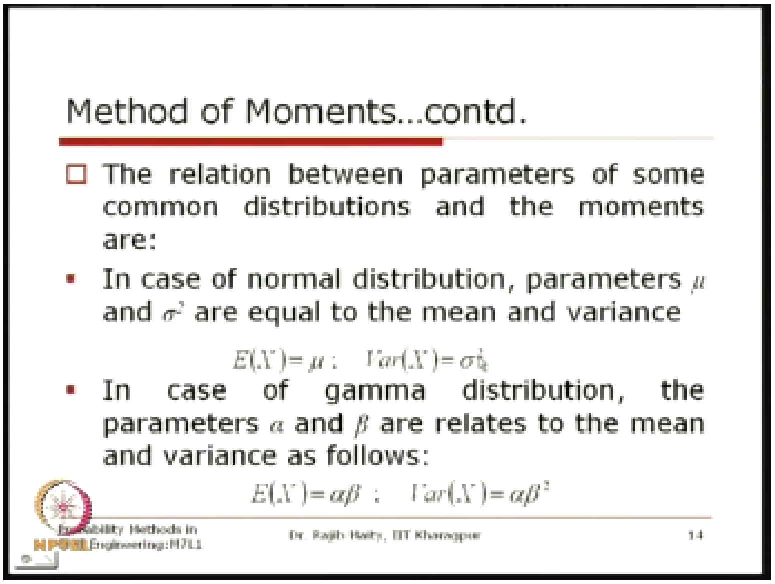
So now, the relation between the parameter of some common distribution and the moments are say for example, here we are taking the normal distribution first there are two parameters that you know, one is that μ and other one is the σ^2 are equal to the mean and variance, like this. So, expectation of the X is equals to mean and the variance is equal to σ^2. So, what you can see is that, directly we can use 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 α and β that relates to the mean and variance as follows that you know, that expectation of X in case of the gamma distribution is the αβ and the variance is $\alpha\beta^2$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.

(Refer Slide Time: 28:19)

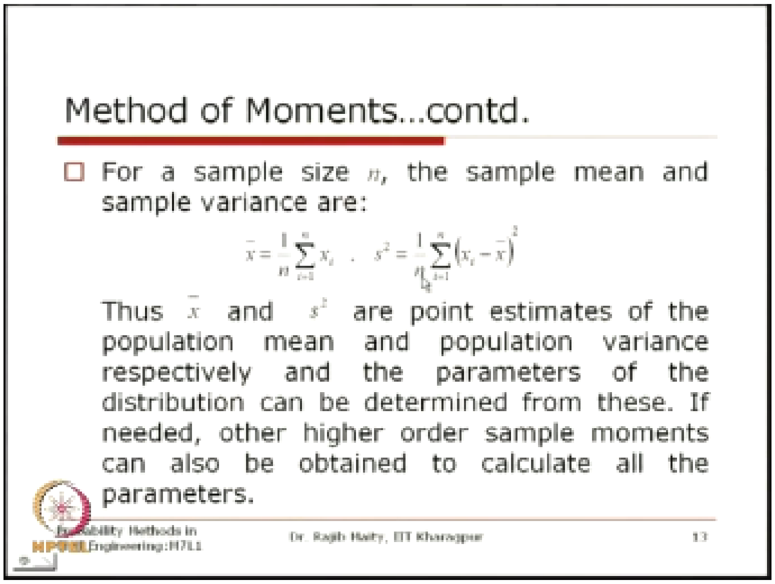


(Refer Slide Time: 28:22)



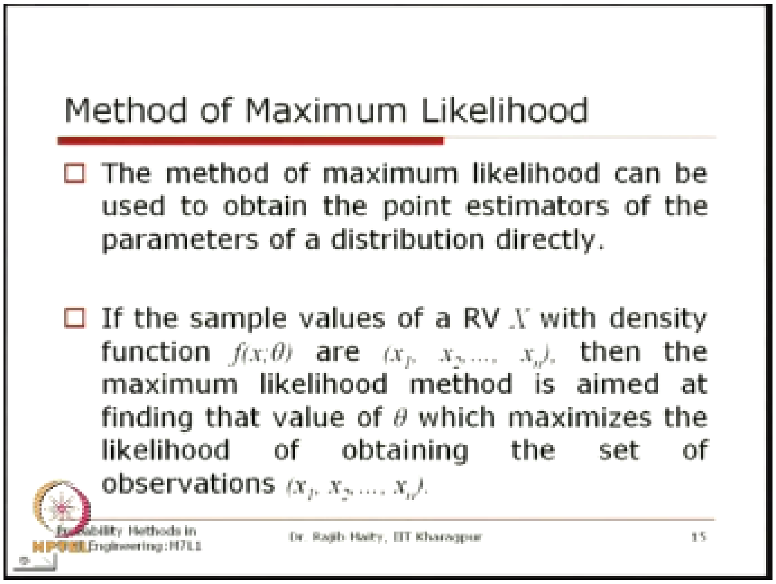
So okay, fine so as to conclude this one is that now you are having basically, this you can estimate from this sample this also you can estimate from the sample. Now, if you equate this one with these two parameters, that is what we are saying, so we are having that two unknown shear α and β and there are two equations. So, these two can be can be solved to get what is the estimate of this α and β and here it is straight forward because, we are getting this μ is the expectation which is directly should be the equal to this $\bar{x}$ whatever you have seen and the variance should be that σ^2, what you have seen in this slide.

(Refer Slide Time: 29:00)



Obviously, if you are using this $\bar{x}$ that is which is also estimated from this sample, this instead of 1/n it should be n-1.

(Refer Slide Time: 29:09)

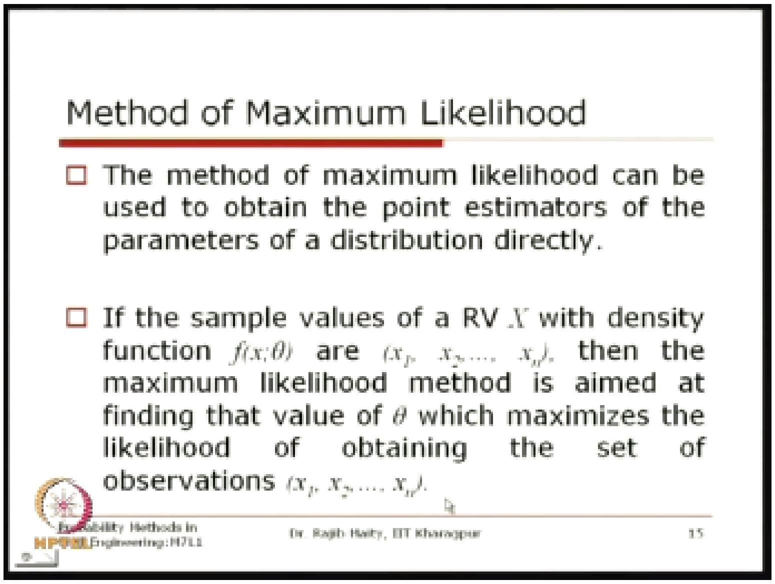


Now, the second method is that method of maximum likelihood. The method of maximum likelihood can be used to obtain the point estimator of the parameters of a distribution directly. So, there are some shortcomings of this method of moments, which we generally suppose that sometimes the estimator that is using the method of moments, what we get is that sometimes the solutions are once it is solved, we get that 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 more effective. Because, here directly the distribution we are using and where we are developing a likelihood function and that likelihood function is maximized to estimate that 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.

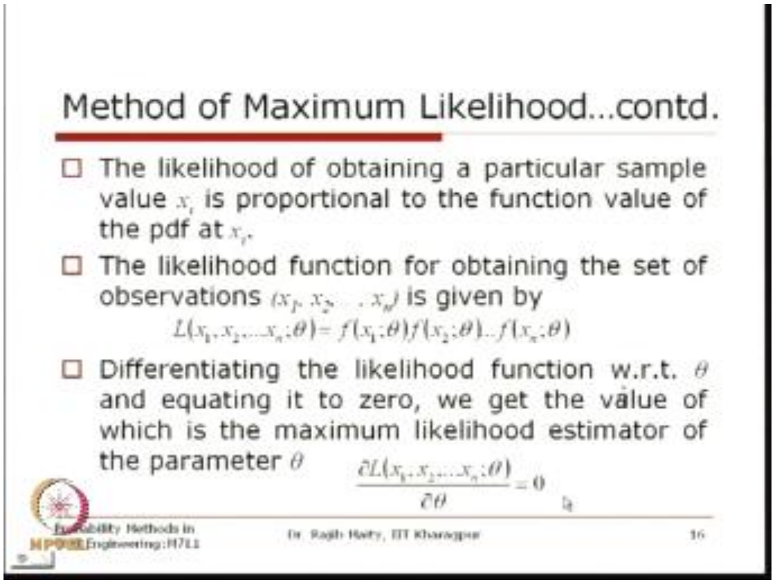
Now, as many parameters are there, so we have to maximize all those parameters. So that, we will get those many equation to have to solve them to get those estimate.

(Refer Slide Time: 30:31)



So, suppose that if a sample value of a random variable X with density function fx and the parameter is θ here are this x_1, x_2, x_n then, the maximum likelihood method is aimed at finding that value of θ which maximizes the likelihood of obtaining the set of observations x_1, x_2, x_n.

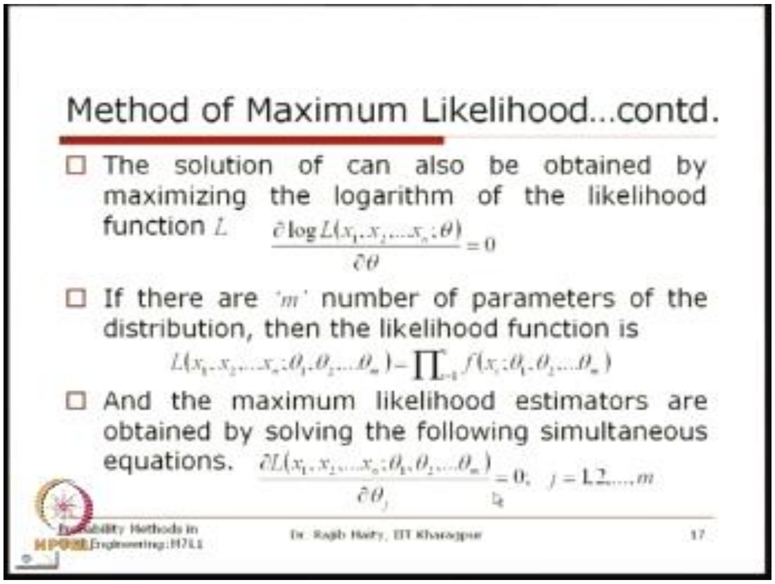
(Refer Slide Time: 30:54)



So, what we have to do is that the likelihood of the, of obtaining a particular sample x_i is 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 θ now and equating it to 0, so, this is basically we are maximizing the likelihood function we are finding where this likelihood function will be maximized. We get the value that is θ value of that estimate, that is θ hat, we can just set which is the maximum likelihood estimator of this parameter θ that is this one will equate to 0.

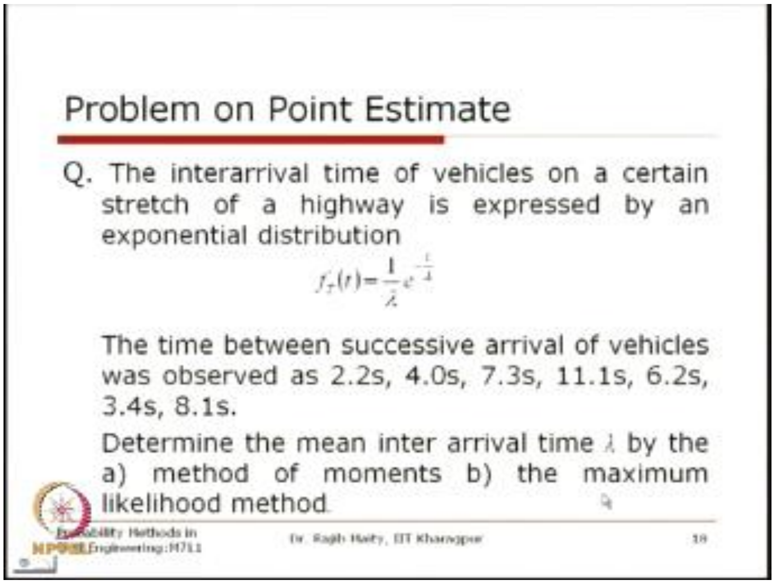
(Refer Slide Time: 32:11)



And then we will get that estimate of that θ hat. 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 θ_1, θ_2 up to θ_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 θ_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: 33:21)

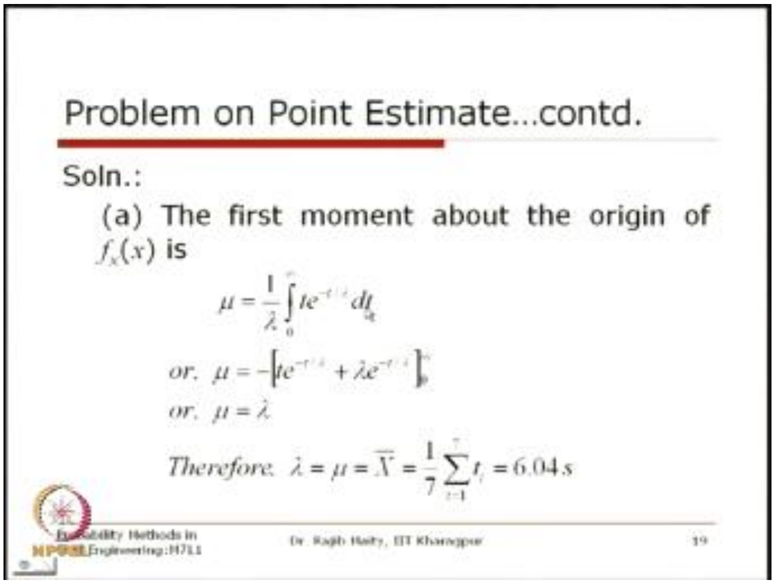


Now, we will take one example using both the methods that we have seen just now both that method of method of moment and method of maximum likelihood. So, and here we have taken that example of this exponential distribution and you know this exponential distribution here, the example is on this interarrival time of this vehicle on a certain stretch of a highway is expressed by an exponential distribution, where this $f_T=1/\lambda\ e^{-t/\lambda}$.

Now, means some places or even in this lecture also, earlier we have, we might have used some other form, that is $\lambda\ e^{-\lambda t}$ so, it does not matter so there just parameter is taken that $1/\lambda$ 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 other form also that is $\lambda\ e^{-\lambda t}$ then also we can we can get the same result that we see now.

So and obviously here this $t \geq 0$. Now, there are some samples has been taken, that is the time between the successive arrival of the vehicle was observed as 2.2 seconds, 4 seconds, 7.3 seconds, 11.1seconds, 6.2 second, 3.4 seconds and 8.1 seconds. So, this is a sample that we have collected. Now determine the mean, interarrival time that is λ 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 λ for this distribution by two methods, one is the method of moments and other one is the method of maximum likelihood.

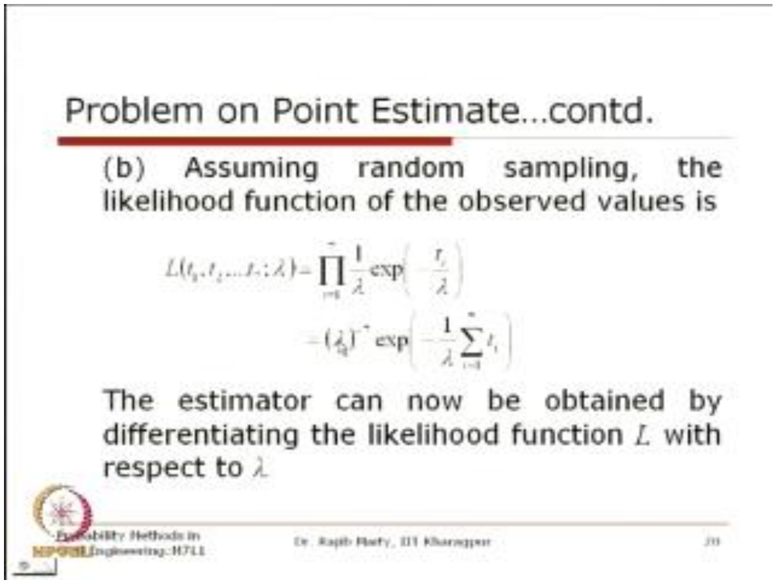
(Refer Slide Time: 35:12)



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 and taking this integration over the entire support of this distribution and here this exponential distribution having the support 0 to ∞.

So, we will take this integration and if we solve this one, we can see that this μ=λ here so this is μ so the λ=μ which is, we can obtained from this sample of this $\overline{x}$, which is the estimator for this mean. So, this 1/7 ∑ of all this t_i, so, it is the arithmetic mean. So, 6.04 second is the estimate for this λ. Now, if we use the other form of this exponential distribution that is that is $\lambda e^{-\lambda t}$ 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 parameter is used here.

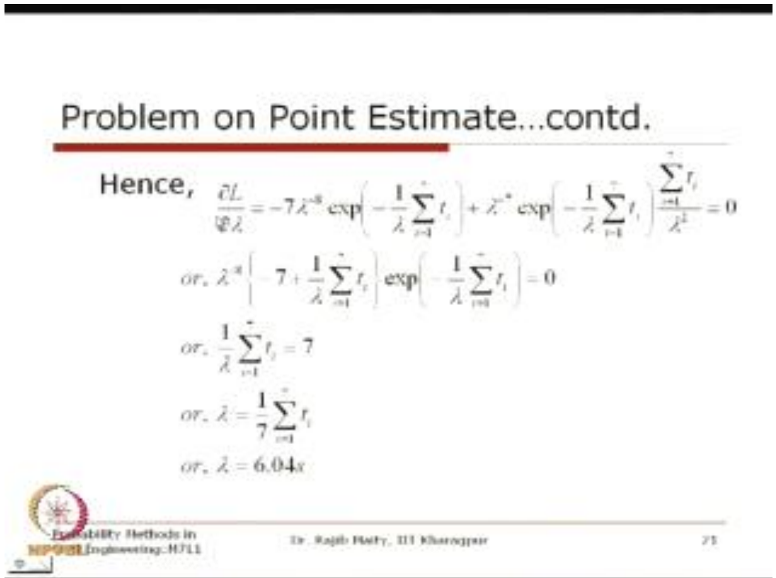
(Refer Slide Time: 36:26)



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$1/ \lambda e^{-t_i}/\lambda$. So at all this observation we are getting this value and we are multiplying them with each other.

So this λ^{-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 λ.

(Refer Slide Time: 37:13)

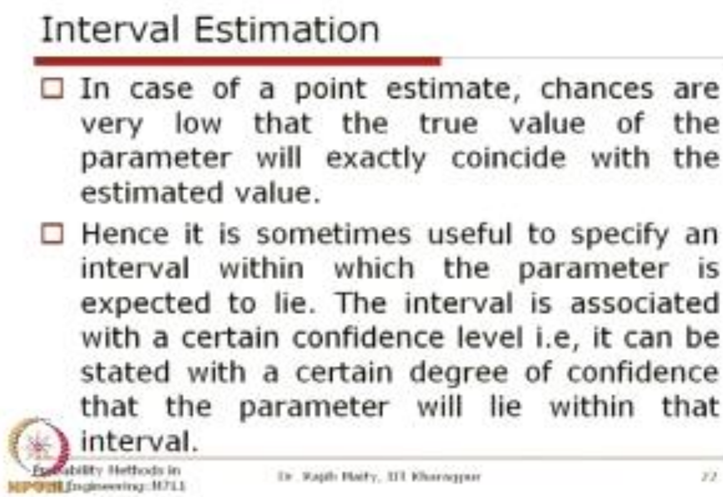


So if we do that that if we do this derivative we take with respect to the λ then we will get this form and if we just equate it to the 0, then after solving this form, you will get that again the lambda is equals to 6.04 second. So for this one this example that we have shown for both the method, whatever the parameter that we have that we have estimated are same for the λ 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. 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 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 λ value that you have seen that 6.04 second. So only a single value that you have seen in the point estimate. Now the interval estimate we generally look for a interval, in which, with some confidence that actual value of the parameter should like.

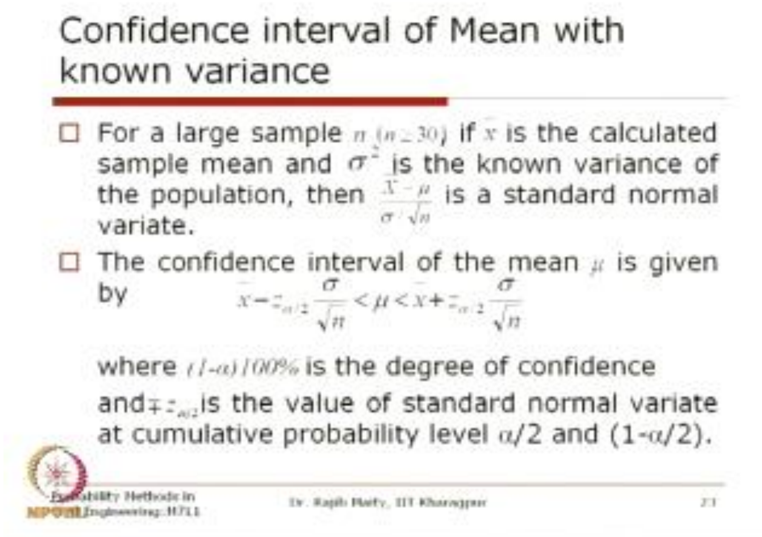
(Refer Slide Time: 38:30)



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 and 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, 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.

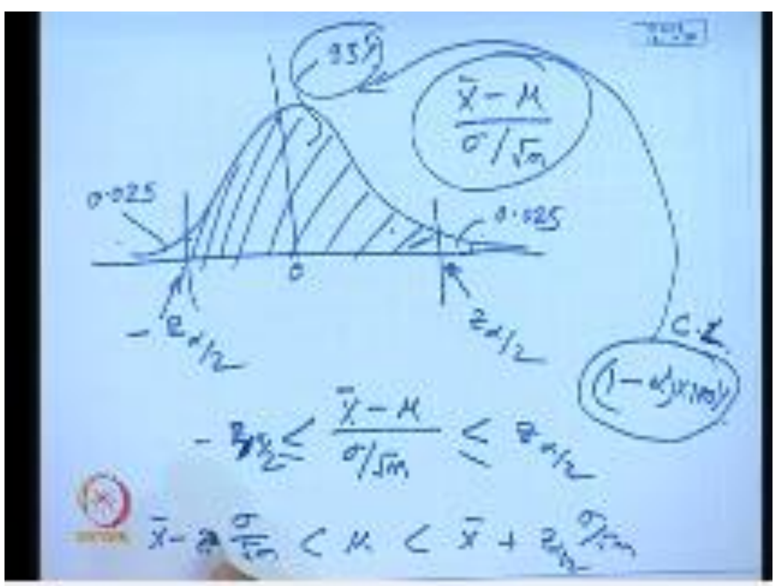
So for a large sample n, so large sample is again, this is a subjective word. So generally we can we have seen that, if $n \geq 30$, then you can say that in case of mean only, so not for all in general case, in case of mean, if the sample size is greater than 30, then we can say that it is a large sample. So, in such case, if x bar is calculated sample mean and the σ^2 is the known variance of the population.

So, σ^2 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 a normal distribution with mean equals to μ, which is the population mean and the variance of this x bar, that is, this sample mean, variance of the x bar is σ^2/n or the standard deviation of this x bar is $\sigma/\sqrt{n}$.

So this 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 μ and its standard deviation is $\sigma/\sqrt{n}$. 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 $x/\ \mu\ /\sigma^2\sqrt{n}$ 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 μ is given by is this that is, x bar. Basically we are just equating it with two side of this standard normal distribution. Now if you if you see this one here.

(Refer Slide Time: 42:36)



So if this is your standard normal distribution then basically that that quantity, that is x /- μ /σ /$\sqrt{n}$, that should so this is your, this is the standard normal distribution. So this 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.

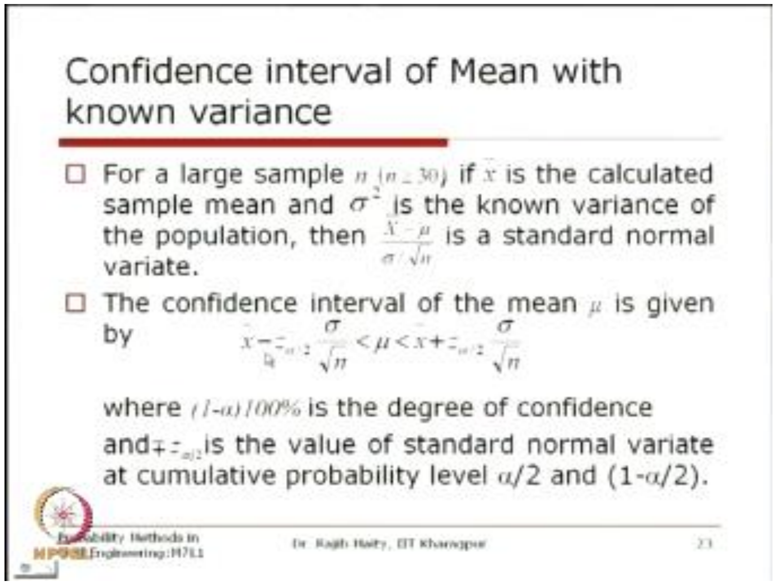
Now so this is the confidence level of that estimated that it should have now if I just say that this confidence level is, say at some level say that 95 percent confidence level. So, whatever is remaining here is your 0.025 and whatever is remaining here is again 0.025.So we have to find out these two values suppose that if I just put that one, z α/2 and this is say that z α/2 obviously, this will be the negative side, this is 0for the standard 1.

So, this one so this quantity should lie between this z α/2 and this z α/2 where this α/2 are basically the cumulative probability up to that point. So this x /- μ /σ/$\sqrt{n}$ should lie between this two, z α/2 and this - z α/2 sorry, z α/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 –α x100 confidence level.

So this is the confidence level that we can say, so that if this area, this white area is the α/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 limit as this z α by 2.Now,if I just do some arithmetic change, then it will come like this; that is, μ should be equals to here that x bar + z α by 2 sigma by square root here it will be, x bar - z α by 2 sigma by square root n .So, you remember, that even this α/2 when we are talking about. So, so this α/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 +. So, this is a single value that we are using here, which should have that 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.

(Refer Slide Time: 46:17)



So, this is what is mention here, that is x bar - z α by 2 sigma by square rooting - μ less than x bar + z α by2 sigma by square root n. So, you know that from the continuous distribution less than and less than equals to are same.

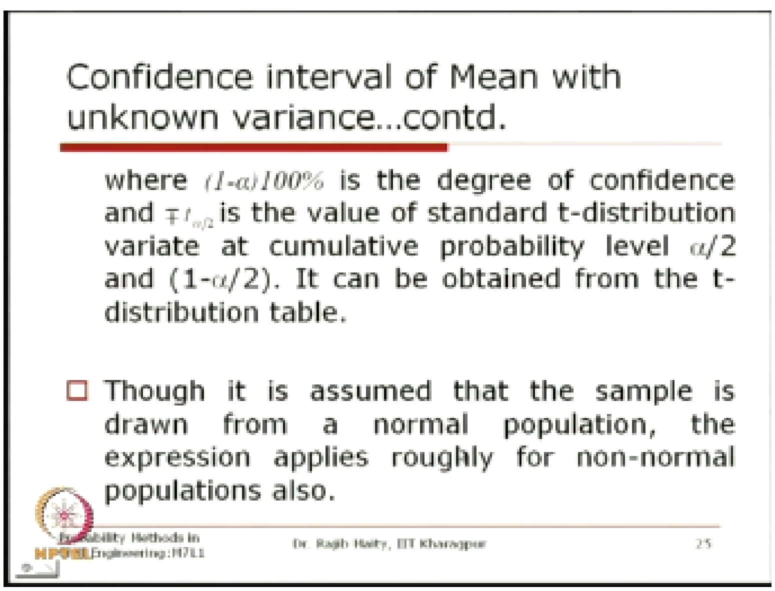
Where, this 1 - α into 100 percent, this is the quantity, which is the degree of confidence. And here, this - + zα /2 is the value of the standard normal variate, at the cumulative probability level α by 2 and 1 - α by 2.So, when you are taking this - sign -z α by 2, this is the probability level at this α by 2. So, if it is 95, if you put this α equals to 0.95 then; obviously,1 -, so, sorry, this is full is equals to 0.95, then α will become 0.05 and this α / 2 will become 0.025.Now, at this 0.025, this z α/2 will be, say if it is 0.95, you know that this value will be 1.96. So, this will be x bar-1.96 multiplied by sigma by square root n and here also, + 1.96 sigma by square root n this will come.

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 bar - μ divided by s square root of n, follow a t distribution with n - 1degrees 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 -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 - 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 that distribution. If the sample size is more and this variance is known, then this will follow a normal distribution.

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

(Refer Slide Time: 49:30)

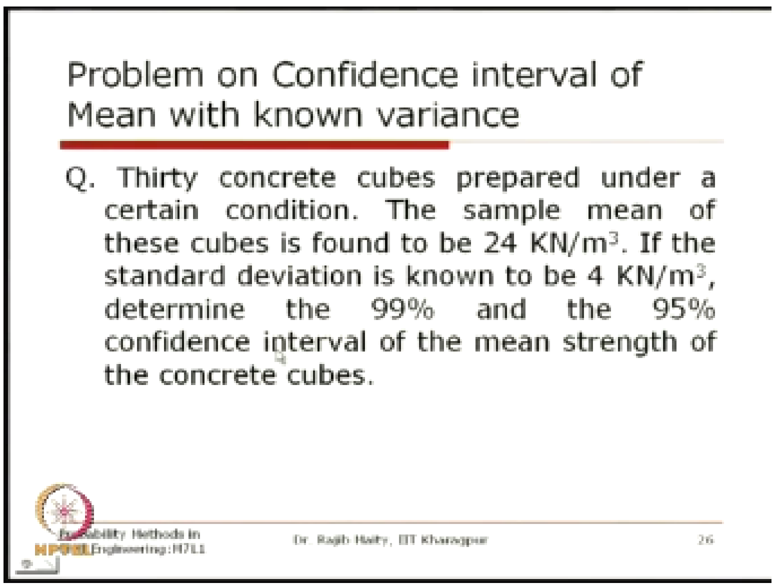


So, where this again, the 1 - α into100 percent is the degree of confidence and- t α by 2, is the value of the standard t distribution variate at cumulative probability again, that α by 2 and this 1 -

α / 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 wider estimate of the interval because, the when the sample size is less, as there are uncertainty is more.

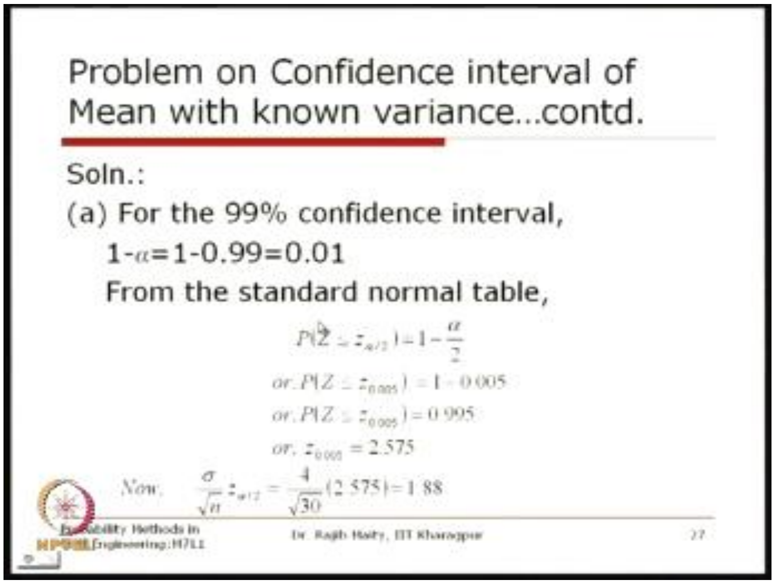
Though it is assumed that the sample is drawn from a normal population, the expression applies roughly for non-normal population also. So, basically when the population is normal distributions, these are very well acceptable method. But, even though it is non normal also, this method we can apply.

(Refer Slide Time: 50:25)



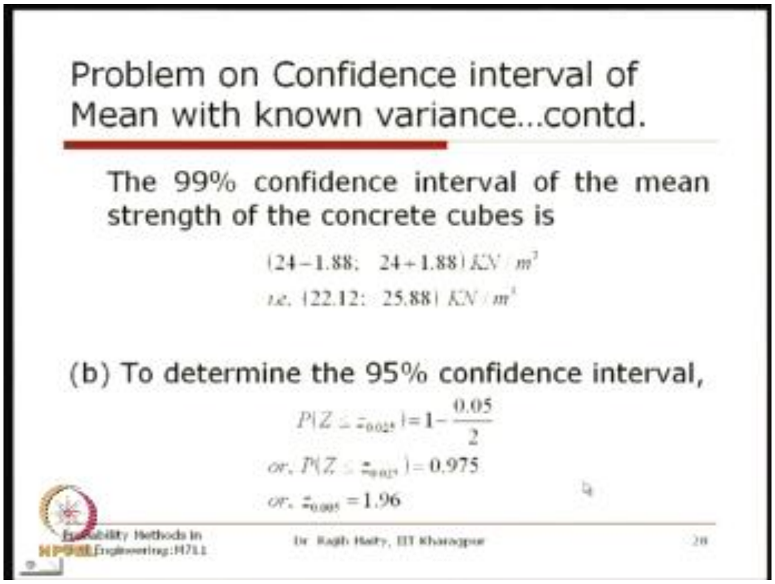
So, we will take one example of this, whatever we have seen, is that 30 concrete cubes prepared under certain condition. The sample mean of these cubes is found to be 24 kilo Newton per meter cube, if the standard deviation is known to be 4 kilo Newton per meter cube, determine the 99 percent and 95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



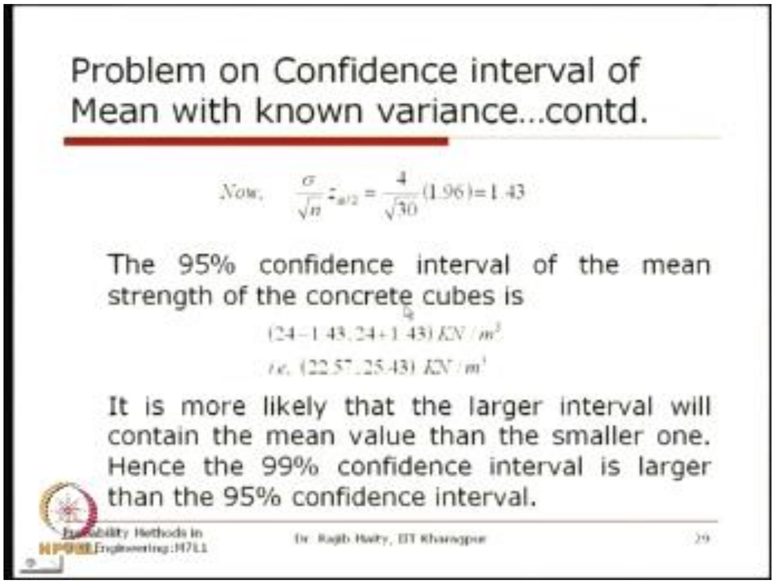
So, if we want to solve this one, then you know, that we first of all, we will get that what should be the quintile value for this z α by 2 and this is for this 99% confidence interval from the standard normal table. You can see that it is 2.575, which is z α by 2 value. So, this sigma by square root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

(Refer Slide Time: 51:35)



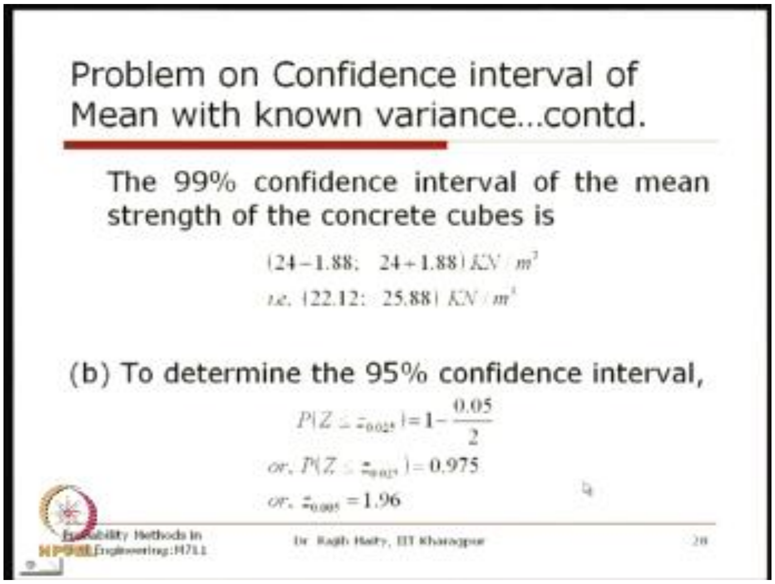
So, the 99 % confidence interval will be the mean - that quantity 1.88 and mean + 1.88. So, the confidence interval is22.12 and 25.88 kilo Newton per meter square so. To determine the 95 % confidence interval, again we have to find out the z α by 2and this is your 1.96, earlier it was 2.575. So, it is now 1.96.

(Refer Slide Time: 52:03)



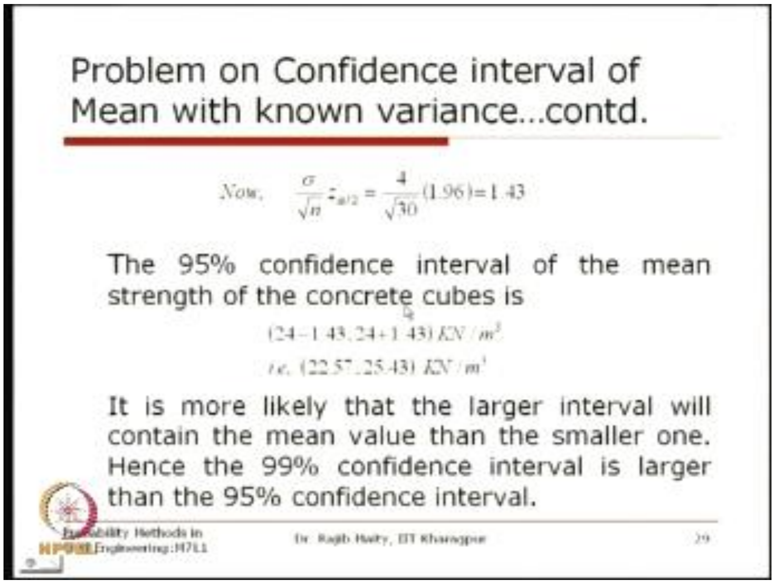
So If we calculate this one, it will become 1.43and the 95 % confidence interval of the mean strength of this concrete will be 24 - 1.43 and 24 + 1.43. So 22.57 and 25.43 kilo Newton per meter square. Sorry. This is or this is the density sorry this is the density. So, this it is kilo meter per meter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

(Refer Slide Time: 52:53)



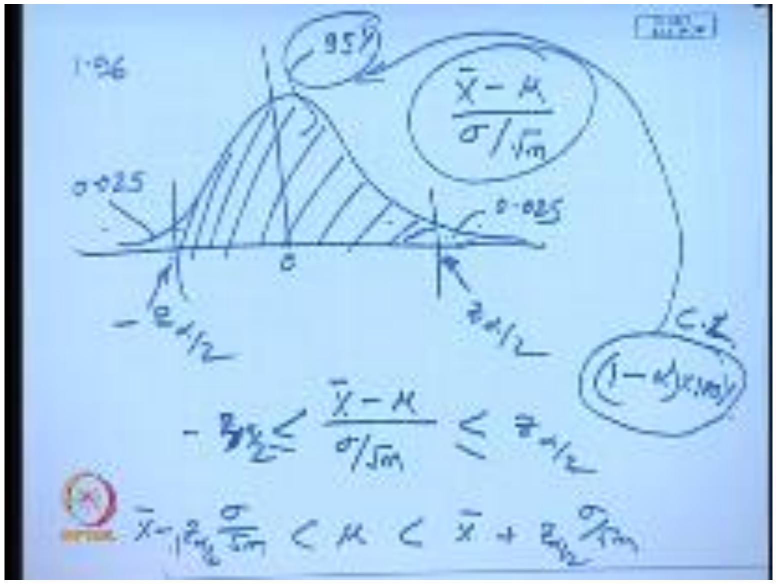
Whereas, the 99 % confidence interval is 22.12 and 25.88.

(Refer Slide Time: 52:59)



So, you can see that this 99 % 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 % confidence interval is larger, when the 95percent confidence interval whereas compare to this 95 % confidence interval. So, this example, that we have used, it is we have taken that a variance is known.

(Refer Slide Time: 53:26)



So, once 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.

Probability Methods in Civil Engineering

End of lecture 35

Next: "Sampling Distribution and Parameter Estimation (Contd.)" in lecture 36

NPTEL Video Recording Team
NPTEL Web Editing Team
Technical Superintendents
Computer Technicians

A IIT Kharagpur Production
www.nptel.iitm.ac.in