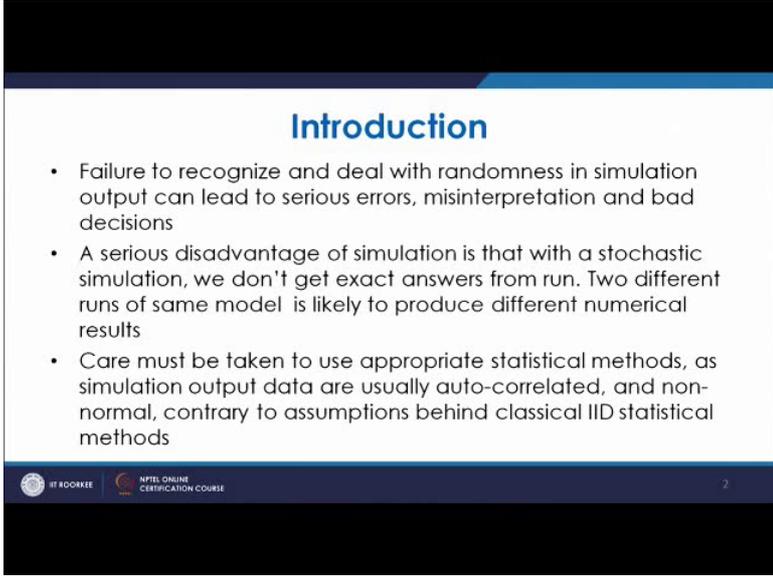**Modeling & Simulation of Discrete Event Systems**
Dr. Pradeep K Jha
Department of Mechanical and Industrial Engineering
Indian Institute of Technology, Roorkee

Lecture - 26
Output Analysis of a Single System: Introduction

Welcome to the lecture on Output Analysis of a Single System. So, in this lecture you will have the introduction about the output analysis for a single system, when for any system if it runs and we get the performance measures, then how to estimate, how to find the estimate about the performance measures. And how can we say with some confidence that if we get this value particular value, we can be sure that it will be close to it is mean or so in that will estimate the mean of that particular system. Let us have the introduction to what it is.

So the first thing what is there, that you know in simulation you get random outputs. So, failure to recognize and deal with randomness in summation output can lead to serious errors.
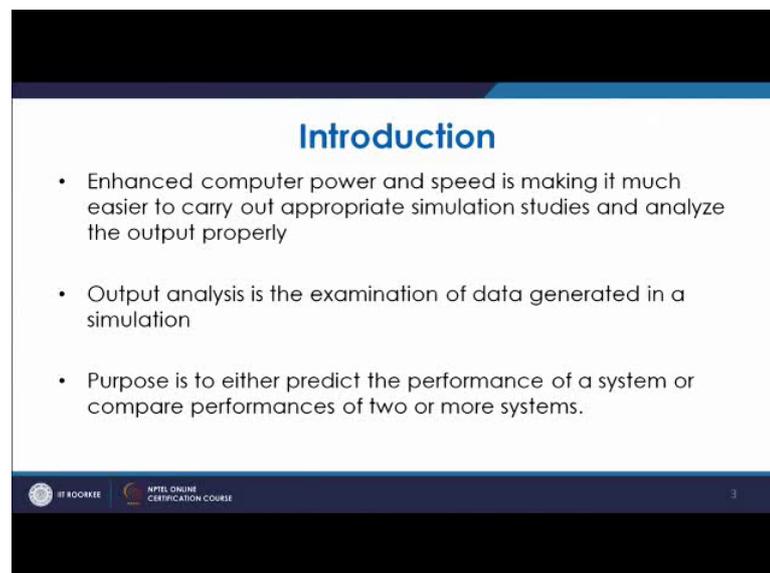
(Refer Slide Time: 01:21)

Like, we simulation means we run the computer programs we get the results. Now, if we do not get good results, if we do not get meaning full results and if those results are interpreted then in that sense if you see the company which likes to further you know copy that further take steps for improving the system based on the simulation outputs. If

the output itself is wrong and it is not in a for the goodness of the system in that case that will be catastrophe.

So, basically you must be able to see that whatever output you are getting how much it is close to reality how much it is authentic. Now the serious disadvantage in simulation is that I mean in stochastic simulation, we do not get exact answers from every run. So, if you do the next run for the same thing we cannot get the same results. So, we do not get the similar result every time, because if we are getting the random numbers because random number which you are using they will be every time different. So, the result which you are getting they will be somewhat different.

So, two different runs of the same model they will be producing different numerical results. So, you have to take the care so that you use the appropriate statistical methods. Because many a times the simulation output data they are normally auto correlated and also they are not normally distributed. So, normally you have the assumption that the data is not having any autocorrelation they are independent and as well as they are from identical distribution. So, many a times you do not get the data like that and you must use you must be careful that your data must be independent and from identical distribution.

(Refer Slide Time: 03:39)

Now, since in the present day we have the enhanced computer power and is speed. So, we can do you know large number of simulation studies and the time is not the constraint in earlier days we did not have that very high end of computer or simulation facilities.

So, we can analyze the output properly. So, output analysis what is that output analysis. So, basically when we do the simulation we get certain output like, if we run a queue program we will get certain outputs, like average queue length average delay of the customer in the queue all that. So, once we get these results now this result is required to be analyzed what kind of result you have got. How whether this result can be expect can be expect this result to be you know trust worthy how much can we trust over these results this is basically this examination is the output analysis.

So, the purpose is to either predict the performance of a system or many a times we have two or more systems. So, we have also to compare between them we have to compare that how they are giving the different results in how much is the difference. So, basically this simulation is used either to predict the performance of a single system or it can also be used to compare the performance of two or more than two systems. So, what is the nature of simulation output, what we see that if we you know get the results in any particular you know simulation?

(Refer Slide Time: 05:14)

Nature of simulation output

- If Y_1, Y_2, ... Are output parameter from a single simulation run, random variables Y_i's are generally neither independent nor identically distributed (nor normally distributed). Hence classical IID normal-theory statistical methods don't apply directly to the Y_i's

- y_{11}, y_{12}, ..., y_{1m} be a realization of the random variables Y_1, Y_2, ..., Y_m from a single simulation run of length m observations, and from another simulation run of this same length (using different set of random variates, rang being same), we get y_{21}, y_{22}, ..., y_{2m} (independent of, but identically distributed to, y_{11}, y_{12}, ..., y_{1m})

Suppose we do the simulation of queueing. And if we are getting certain results for certain output. Now we cannot say that they are IID, they are basically independent because these all these results have been calculate based on certain parameters. So, if the y 1 y 2 are output parameters from a single simulation run then the y is which we get. So, basically we get you know across the row we get for a particular simulation, we get

certain results and these results are not necessarily always IID they are neither independent nor identically distributed they are also not from the same type of distribution curve. So, they are neither independent because they depend upon certain values and based on that you are getting those results.

So, basically the theory or all those rules which we apply based on this assumption. They are not very much valid for that. Now suppose you have the y 1 1 y 1 2 y 1 m. These are basically the realization of a random variable y 1. So, for y 1 y 2 or y m now in that case what happens suppose you do the first replication you get y 1 1 y 1 2 y 1 3 y 1 4, like that then for the second replication it will 2 1 2 2 2 3 like that.

So, you are getting for a particular parameter you are getting along the column, you are getting the values. Now in that case along the row the values are not in independent whereas, along the column they can be thought of as an independent. So, that is what, it is written that if we do the single simulation run of m observations and if we do. And from another simulation run of the same length, if you are getting the other I mean we use the same random variates and we are getting like I am y 2 1 y 2 two or so, this way we are getting.
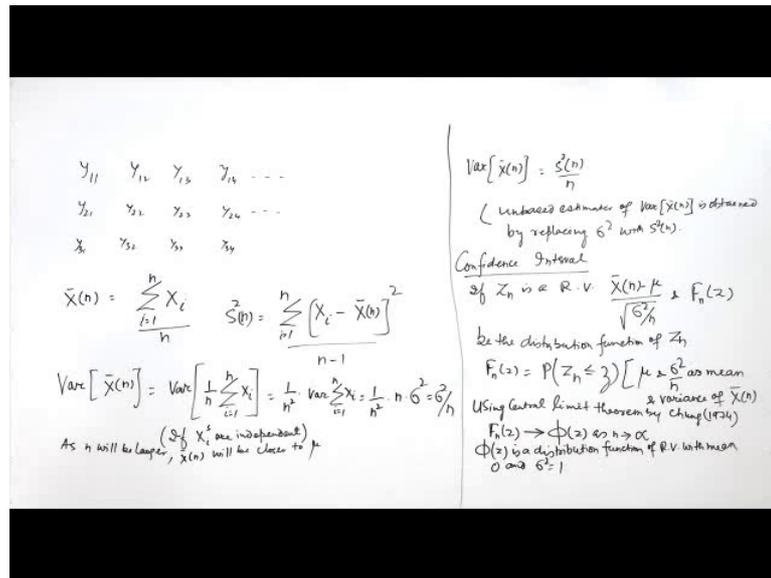
(Refer Slide Time: 07:49)

In that case. So, what we get is. So, in a particular replication in a particular replication the values are not basically considered to be IID, whereas along the row they are independent. So, just for example, we can say that if you have the first replication.
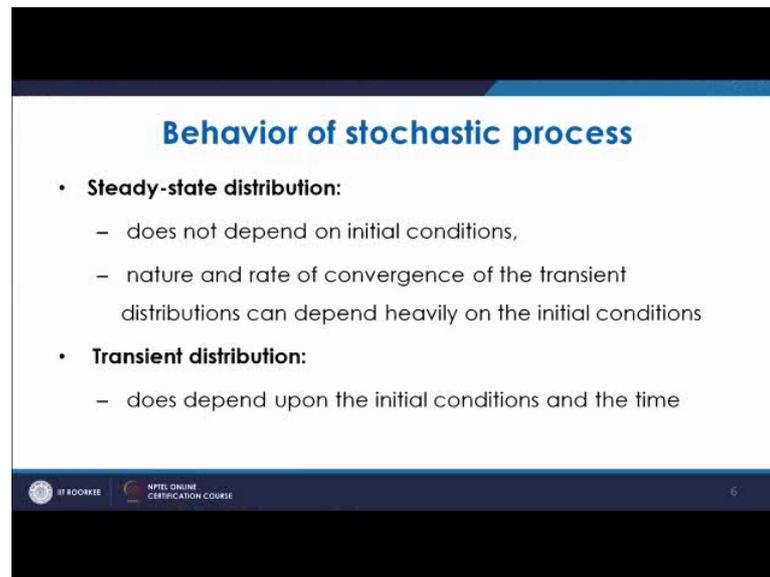
We get suppose the value 1 1, then 1 2 1 3 1 4 or so similarly it will be y 2 1 y 2 2 y 2 3 y 2 4; so similarly y 3 one y 3 2 y 3 3 y 3 4.

So, what we say is that the first simulation because of certain input parameters, we get all these results this may be different parameters like q length. This may be the delay this may be the average number of customers all that. So, basically across the row they may not necessarily be independent, whereas if you take across the column basically they can be considered to be an independent type of random variables now. So, that is what we see that among across the row they may not be IID, but across the column they are basically IIDs.

Now, what we see is in case of or put analysis, what we try to see that if we have the data. Now the goal is to use the observation and to draw inferences about the random variables whatever the random variable, we are getting now what can we come out of that. So, how can we relate them how can we say that the result which we have got what is the true you know what can be said true about it. How much it is it is seems to be accurate or so.

(Refer Slide Time: 09:48)

Now, the behavior of stochastic process, so this is something about what kind of behavior you have two types of process. You have you may have steady state process you have a transient process. So, as we know that in the steady state case, we I mean do not care about the initial phases or we do not care because at the with time then parameters do not change with time after certain time. So, that case is basically the steady state case. And transient distribution is it will basically be depending upon the initial condition and the time. So, basically in the steady state distribution also we have to reach that stage when there is steady state distribution.

So, because there will be initial transience. And after sometime we reach at a stage where the initial conditions and time they do not play any role. So, that then we say that we have reached the steady state situation. So, basically in in simulation when we have to do the steady state analysis basically there are many approaches and the thing is that we have to wait till that steady state is achieved. So, basically we have may have to delete certain initial readings.

So, in discrete or continuous you can have some conditions specified because of which we can say that up to this time, or up to this condition we can say that we are going to delete we are not going to consider them. And then after that we feel that this state is the steady state. So, whatever parameters we are trying to you know get or through performance measures, they do not now further change with time the transient

distribution. However, that basically in that the values which are there basically they are effected by the initial conditions and the time. So, that is steady state and transient behavior of the stochastic process.
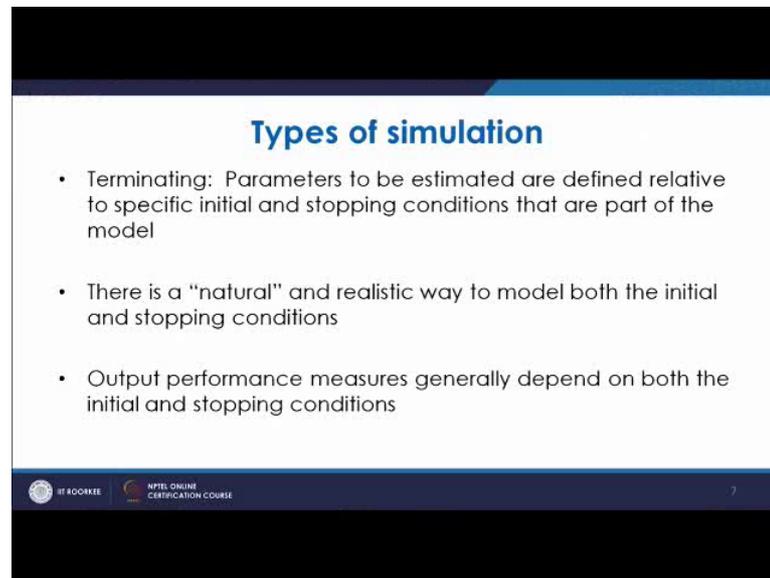
Now, types of simulations. So, when we talk about the simulation we have two types of simulation. One is terminating simulation and another is nonterminating simulation. So, terminating simulation means the parameters to be estimated are defined related to specific initial and stopping conditions that are part of the model. So, in that case basically you have the initial condition as well as the stopping condition. A condition is given when it tells that when the simulation has to terminate. Suppose we do the queuing model in that if we give the condition that after 25 delays we have stop. So, that is a case of terminating simulation. So, there we have started with certain initial condition and the stopping condition is given that the simulation has to stop at the time of suppose you know 25 delays getting observed.

So, there is a natural and realistic way to model both the initial and stopping conditions. And output performance measure will be depending upon the both initial condition as well as the stopping condition. So, what is a initial condition based on that the output performance measure will be you know depending, it will be effected. Similarly, the stopping condition will also effect that output performance measures if the stopping

condition is varied or changed in that case also the output parameters are going to be changed.

(Refer Slide Time: 13:40)
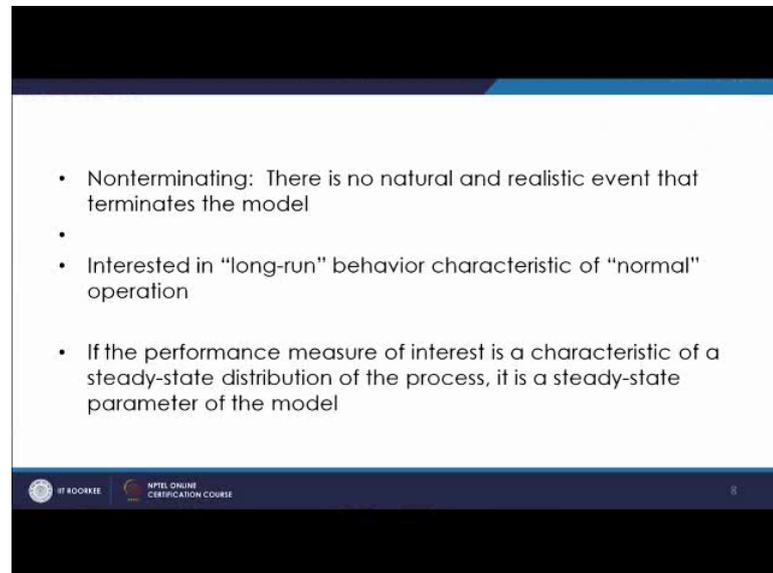
- Nonterminating: There is no natural and realistic event that terminates the model
-
- Interested in "long-run" behavior characteristic of "normal" operation

- If the performance measure of interest is a characteristic of a steady-state distribution of the process, it is a steady-state parameter of the model

Then you have the non terminating type of simulation, where we do not have any natural or realistic event which terminates the model. So, in that case that is not specified and basically we are interested in the long run behavior. So long run average number of customers in the queue like that. So, these type of terminologies are used. So, we always think of the long run behavior characteristics and if the performance measure of interest is a characteristic of steady distribution of the process. So, that is why the in that case we tell it as a steady state parameter of that model.

(Refer Slide Time: 14:22)
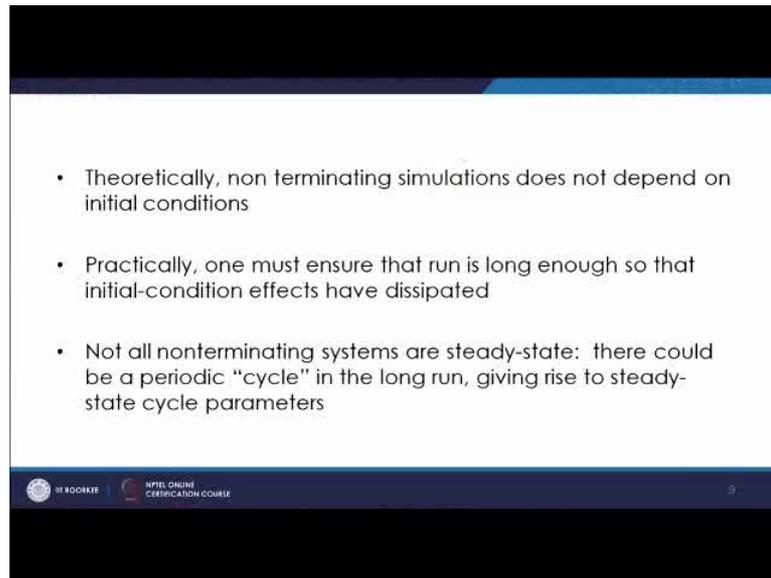
Now, theoretically non terminating simulations, they are not depending upon the initial conditions. So, basically as we discussed when we say that steady state I mean has achieved in that case the simulation do not depend upon the initial conditions. So, basically you must ensure that you are making the run long enough so that initial condition which effect it is giving the effect which is given by initial conditions that should be vanished. So, that should dissipate.

So, that is why in the initial periods you have certain time where you have to ignore, you have to delete those values. So, you must have that that you have to run it for quite long time. All non terminating systems are not basically steady state and in many cases you see that the results will be cycling. So, in that case you have steady state cycle parameters the parameter values which you are basically finding they may be basically you know cycling they may be repeating. So, it is that they are on a all not may be among them you have this kind of behavior also.

(Refer Slide Time: 15:49)

Physical model	Terminating estimand	Steady-state estimand
Single-server queue	Expected average delay in queue of first 25 customers, given empty-and-idle initial conditions	Long-run expected delay in queue of a customer
Manufacturing system	Expected daily production, given some number of workpieces in process initially	Expected long-run daily production

Now, So, what we see that if we do the model like single server queue in that you see that when you have the expected average delay in queue of first 25 customers, when we give this type of conditions like you have empty and ideal initial conditions. So, initial condition is given and then stopping condition is given like you have the average delay of 25 customers. So, when we tell that in a single server queue then that is a case of terminating estimand. So, we are whatever estimand we are getting what we are whatever we are estimating that is a case of terminating simulation.

Similarly, in the case of steady state system you will have to deal with the parameters like long run expected delay in the queue of a customer. So, that will be basically steady state you know simulation just in a example of steady state type of process. So, similarly in the manufacturing system like expected daily production given some number of work pieces in process initially. So, like this is the example of terminating type of process simulation and then expected long run daily productions. So, this is example of non terminating type of simulation nonterminating cases like. Now this is like a steady state estimand. So, this is is how it is specified.

(Refer Slide Time: 17:17)

Now, how to do the analysis for a terminating simulation; so what is there that after making an independent replications. If x j is the performance measure from the jth replication. And approximate 100 into 1 minus alpha percent confidence interval for the unbiased point estimator of mu will be given by x bar n plus minus t n minus 1 1 minus alpha by 2. And then s n upon under root n this is the basically confidence interval. And the second part that is this part this is known as the half length. So, t n minus 1 1 minus alpha by 2 and s n upon under root n this is basically known as the half length.

So, you have the sample mean and from there for certain precision for certain confidence level what should be the half length. So, that if any value comes you can say that this is the percentage chance that the expected value or the next value will be in between these two limits. So, that is why there is plus and minus. And this x bar and this is basically known as the unbiased estimator of mu unbiased estimator we have already discussed that if we the n is done. So, many times this x bar n will be close to the mean mu and this is defined as the summation of all the values j equal to 1 to n x j divided by n. And then this is known as unbiased estimator of variance x j and this is computed like that.

So, we have to see how we can get these results. Now let us see that what we say we get x bar n as summation of x i, i will be varying from one to n divided by n. So, that is how this sample mean is basically we are computing. And we are also computing the s square n. So, s square n is given and this is basically computed as summation of i equal to 1 to

n. And then xi minus x bar n and then it is a square divided by n minus 1. So, as we know that one is sample unbiased estimator of mu and one is an unbiased estimator of variance x j.

Now, if we have to see we have to find basically the confidence interval for the mean x bar n. So, for that we have to find the variance of x bar n. So, variance of x bar n this will be variance of 1 by n summation i equal to 1 2 n x i. So, we can further if we take this 1 by n outside it will be 1 by n square. So, it will be 1 by n square and then variance summation i equal to 1 to n x i. So, we can further write it as 1 by n square into n or n sigma square.

Basically, if the x is are independent if x is are independent in that case we can take variance xi will go here and in that case it will be n times sigma square. So, it will be equal to sigma square by n. So, what we get is variance x bar n will be sigma square by m. So, it means what we see that larger will be n, as n gets larger this variance will go on reducing. So, we will be more and more close to the mean estimated mean or so as n will be larger x bar n will be closer to mu.

Now, further what we get now, we have to find the un biased estimator of variance x bar n. So, that n biased estimator of x bar n variance x bar n for that we are replacing this sigma square with s square n. So sigma square can be replaced with s square n. Because this is s square n will be the unbiased estimator because for large value of n this s square n be taken as sigma square. So, what we can write is variance x bar n. It will be basically s square n by n. Because unbiased estimator of variance x bar n is obtained by replacing sigma square with s square n. So, that is why we can write variance x bar n as s square n upon n.

Now, we have to find the confidence interval. Now in that basically we are defining one random variable that is, if z n is defined as a random variable, z n is a random variable with that is x bar n minus mu upon under root sigma square by n. And f n z be the distribution function of z n. So, what happens it is nothing, but it is giving you a distribution like f n z will be probability that z n is less then equal to z. So, in that case it is a distribution with mu and sigma square by n as mean and variance of x bar n. So, basically what we see in the case of normal distribution we define x minus mu by sigma. So, basically for x bar n you have the mean as mu and this is this as the variance.

So, this is the variance and this is the standard deviation. Now in that case using the central limit theorem by Chung, what we get is that this 5 z, if we can write that this f n z will be told as phi z as n tends to infinity. And the phi z is a distribution function of random variable with mean 0 and sigma square as 1. So, basically that is known as the standard normal variable what we get in normals, you know distribution we get the normal variate. So, you get this known as standard normal distribution and in that case this phi z will be. So, f n z can be taken as phi z when n is tending towards infinity.

(Refer Slide Time: 27:18)

So, this phi z is defined as 1 by root 2 pi and then integral minus infinity to z. And it will be e raise to the power minus y square by 2 dy. And the z value will be from minus infinity to plus infinity. So, this is similar to the normal distribution, what we get in that case we are defining one another normal variable here. And then further the theorem is telling that for very large value of n. So, this z n basically will be distributed as the standard normal variable.

So, whatever be the distribution of x is for this when we define x bar minus mu by under root sigma square by n that will basically be normally distributed. So, in that case for large n what we get for large n. So, when s square n is tending to sigma square, in that case the random variable we define t n as x bar n minus mu upon under root s square n by n. So, because is the estimator of sigma square variance. So, that is why we write like that.

Now, the approximate distributed as normal variable. That is standard normal variable basically. So, this is standard normal variable. So, this is basically this random variable which is defined that is basically approximately distributed as standard normal variable. Now for large n what we get is that we have the probability of minus z, 1 minus alpha by 2. And that will be x bar n minus mu upon under root s square n by n and that will be less than equal to z 1 minus alpha by 2.

Now, what we would say that we know that once we go in the normal distribution to on both the sides, the confidence interval is with confidence level we can say. So, there will be certain probabilities specified. Now if that is said to be like 1 minus alpha in that case what we say in that confidence you see that value should be minus z 1 minus alpha by 2 2 z 1 minus alpha by 2. It should be in between this minus z of 1 minus alpha by 2 to plus z of 1 minus alpha by 2.

Now, it can further be written like this. So, it will be p of x bar n minus z 1 minus alpha by 2 and under root s square n by n. So, it will be less than equal to mu and then that will be less than equal to x bar n plus z 1 minus alpha by 2 under root s square n by n. So, basically this is nothing, but the value of 1 minus alpha. So, it is normally equal to 1 minus alpha. So, we say that because here in this case this z 1 minus alpha by 2. That is basically the lower critical you know point and z 1 minus alpha by 2 that will be the upper critical point for that standard normal variable.

So, for sufficiently large value of n, so we can write that for sufficiently large value of n the this value basically x bar n plus minus z 1 minus alpha by 2 under root s square n by n. This is basically the confidence interval. And this in between that this will be basically the half length of this particular distribution for when we estimate the x bar n then both side you have this value coming up. So, that will be basically the half length. So, what we see that for sufficiently large n, what we get that a at about 100 into 1 minus alpha percent confidence interval for mu is. So, we get x bar n plus minus z, and then this is 1 minus alpha by 2 and under root s square n upon n.

So, in this case you get this you know values. Now the x is are basically the normal random variable of the variable t n. So, now, in this case you have the n minus 1 degree of freedom. So, basically what happens when you have the data lesser in those cases the data may not be so accurate. So, in those cases because for small n the actual coverage of

the confidence interval will be less, when where you have less number of data the actual coverage for the confidence interval will be less. So, basically we do one adjustment and in that case what we do is we are putting another parameter. So, that will be it will be coming like x bar n plus minus t, and that will be n minus 1 1 minus alpha by 2 and then under root s square n by n.

Now, this is because this n is limited, if n goes to infinity that t n minus 1 1 minus alpha by 2 that will be lead leading towards z 1 minus alpha by 2; so to take in account the numbers because number of simulations in normally is small. So, in those cases just to estimate you have to put this parameter this value. And this that is why this t n, n minus 1 is the degree of freedom. So, if we have n values n minus 1 will be degree of freedom and this value you can get from this table.

(Refer Slide Time: 35:21)

So, if you go to this table from here you have this degree of freedom. And this is the confidence interval; so if the 1 minus alpha. So, if 1 minus alpha is suppose you want to have the 90 percent confidence. So, 1 minus alpha is 0.9. So, alpha will be. In fact, point one and minus alpha by 2 will be 0.9 5. So, for any values this t n minus 1 and 1 minus alpha by 2 will be taken from this value like this will be degree of freedom on the first column. And here you will have 1 minus alpha by 2 will be computed. And then you can get the results.

So, in the next lecture we will try to solve the problems based on that and we will try to be more clear about it.

Thank you very much.