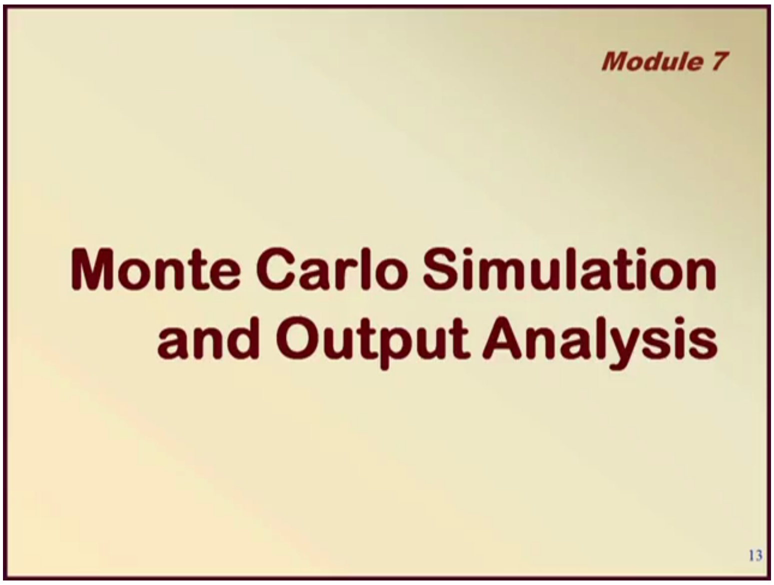
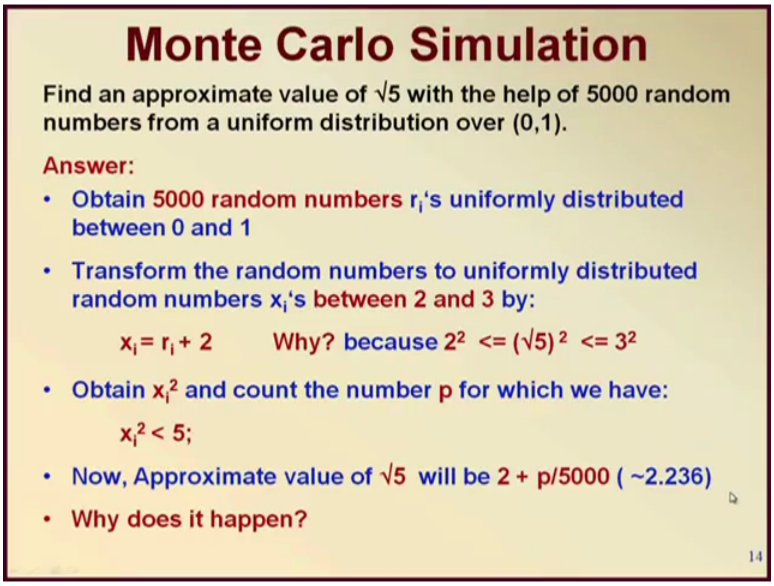
Course on Decision Modeling
By Professor Biswajit Mahanty
Department of Industrial and System Engineering
Indian Institute of Technology Kharagpur
Lecture No 27
Monte Carlo Simulation and Output Analysis

(Refer slide time 0:27)

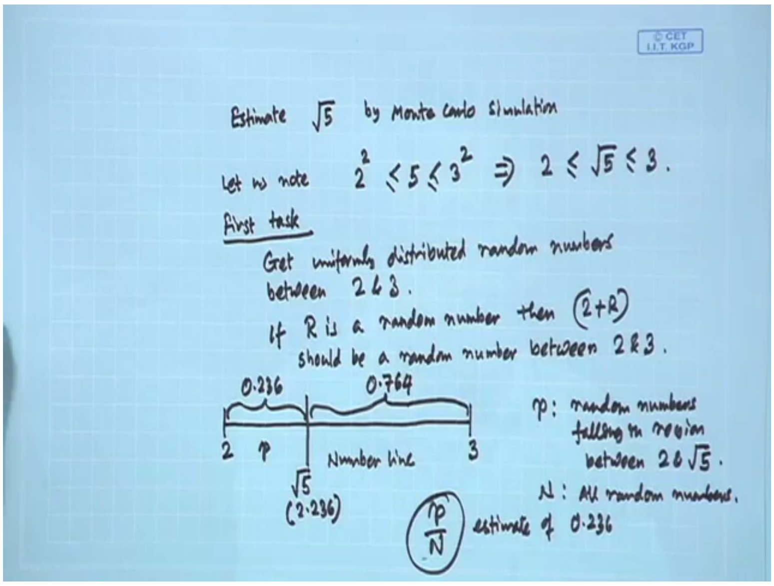


(Refer slide time 0:35)



So in this particular section we are going to have first a few Monte Carlo simulation examples which we could not cover in our previous class and then we start a very important topic on output analysis. First let us see some Monte Carlo simulation examples; earlier we have seen an example where we have computed the value of pi. So in this particular thing you know let us see find an approximate value of root 5 with the help of 5000 random numbers from uniform distribution from 0 to 1.

(Refer slide time 1:12)



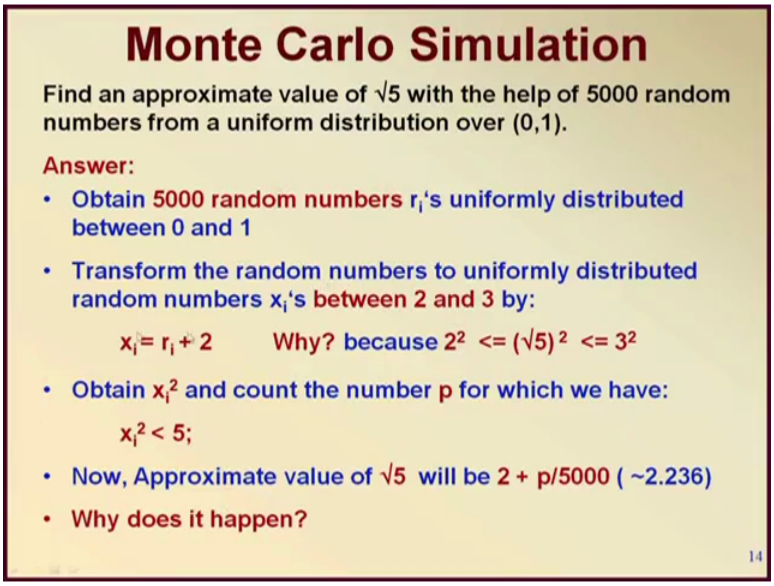
So let us see how we can do such computations? Let us say that we have to estimate root 5 by Monte Carlo simulation. So if we have to estimate root 5 by Monte Carlo similar nation, first let us note let us note that 2 square that is 4 is less than equal to 5 is less than equal to 3 square, alright. so we can therefore from here we can write 2, here we can say that because 5 is between 4 and 9 so we can say that root 5 will be between 2 and 3.

So as our first task get uniformly distributed random numbers between 2 and 3, how to do that? So if R is a random number then 2 plus R should be a random number between 2 and 3. So this is the first thing, right? Get uniformly distributed random numbers between 2 and 3 and if R is a random number then 2 plus R should be a random number between 2 and 3 because R is between 0 and 1, if you add 2 to that it will become 2 to 3.

Then another interesting thing that you note, suppose this is number line and say this is 2 and this is 3, alright. So if numbers are, you see if a random uniformly distributed random's are taken between 2 and 3, each number between 2 and 3 are equally likely, is it not? So supposing if this point is root 5 then say we know root 5 value, so let us write it 2.236, we know it and it is theoretical value of root 5, so I just wrote it.

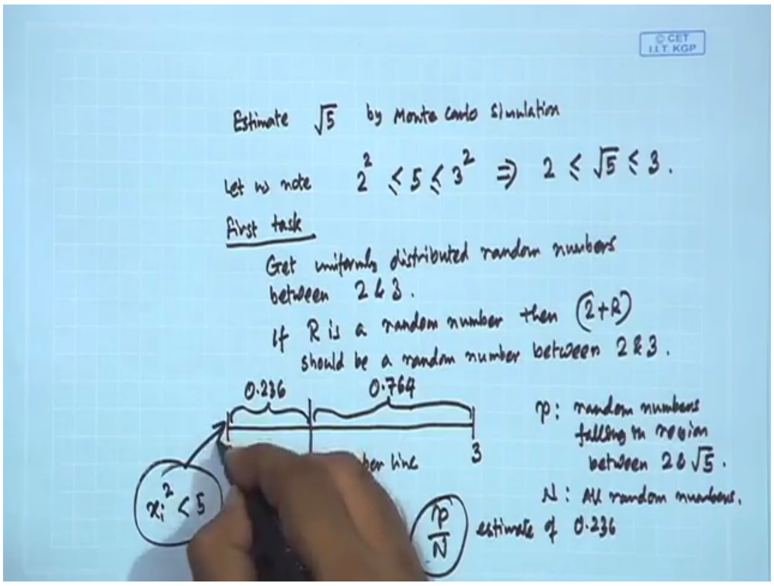
So you see how much is this portion? This portion as 0.236 and how much is this portion that is the rest? That is 0.764, is it not? So you see the random number that will be falling in this region, right? Random number that will be falling in this region compared to the entire number. So suppose P is the, so what is P by N? P by N estimate of you know, say roughly 0.236. So if you add 2 to this number then you will get a root 5, is it all right? So that is the essential idea you know that is the essential idea, look at the slide now.

(Refer slide time 6:41)



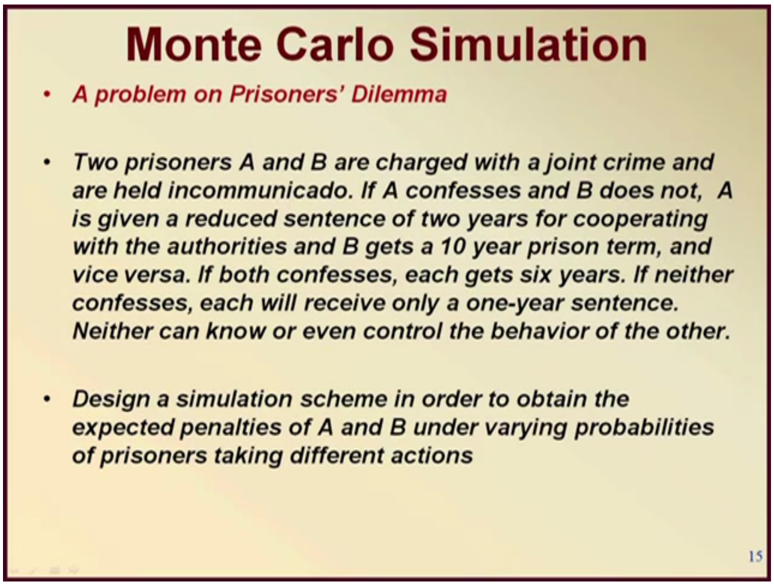
So basically what we have to do? Obtain you know transform the random numbers to uniformly distributed random numbers xi's between 2 and 3 xi equal to ri plus 2 and then obtain xi square and count the numbers p for which we have xi square is less than 5, so why this criteria? Because look at this particular you know diagram would have drawn that this is root 5, this is 2.

(Refer slide time 7:16)



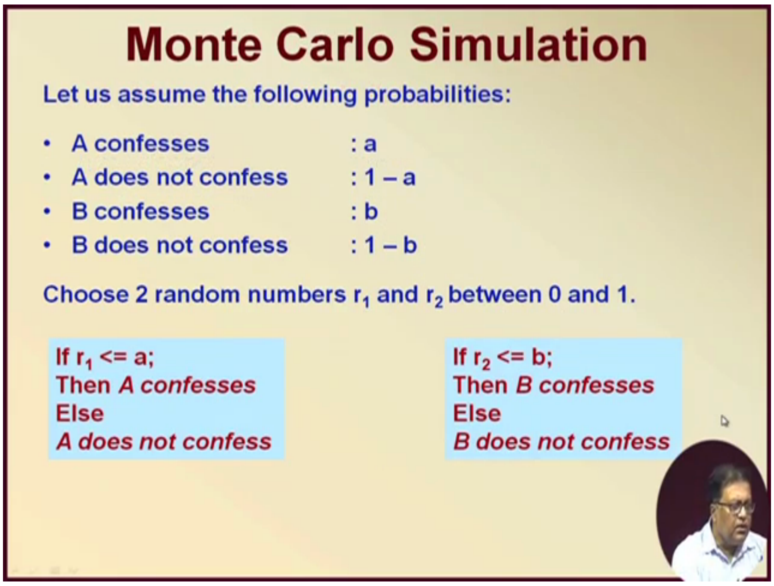
So this, if we put xi square less than 5 this basically falls to this region because that will be this region and number falling here is P, total number taken is N, right? Total number is N, so P by N will be this portion, so that is the idea. So if total number N here is 5000, so approximate value of root 5 will be 2 plus P by 5000, it will come roughly 2.236 but not just in one trial, you have to do several trials and then they are called replications and all these replications we have to then take the average of, is it all right?

(Refer slide time 7:57)

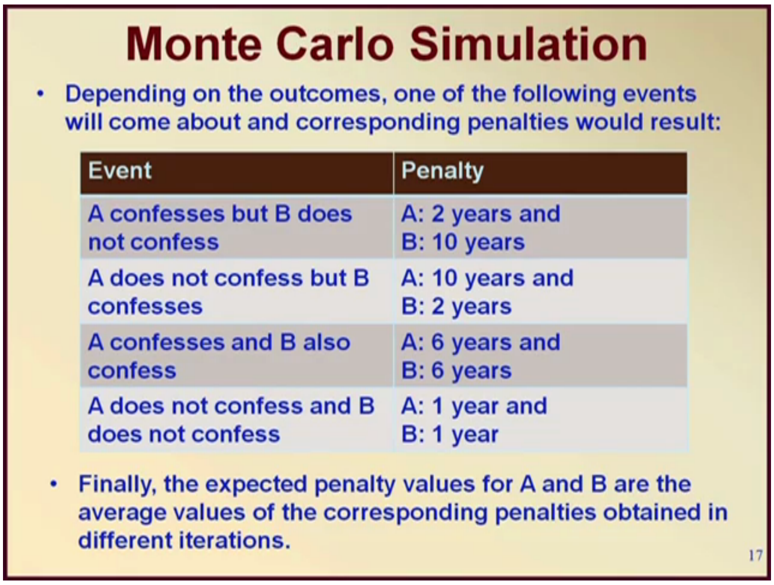


Let's take second example, in this example I have discussed earlier and I gave the problem but did not solve. This problem is called the "Prisoners dilemma", what happens? The 2 prisoners A and B charged with a joint crime and they are held in such a manner that they cannot talk to each other. Now happens they have done a joint crime, so if A confesses and B does not, A is given 2 years and B will be given 10 years and vice versa. If both confess both will get 6 years but if neither confesses then they will receive only one year and neither can know or even control the behaviour of the other. So how do you simulate such a situation?

(Refer slide time 8:47)

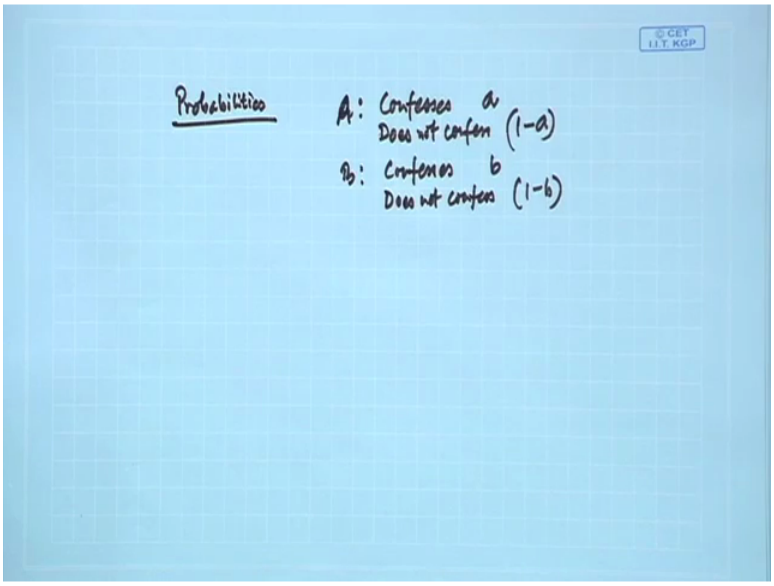


(Refer slide time 9:29)



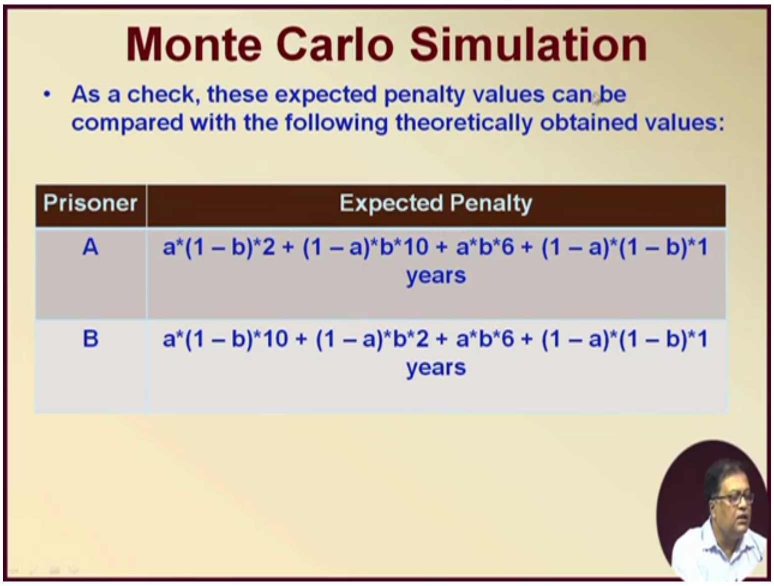
So now let's look at the answer, suppose let us have this 4 probabilities a there is A confess, 1 minus a if A does not confess, b if B confess, 1 minus b if B doesn't not confess, so what we have to do? Let us choose to random numbers r1 and r2, so if r1 is less than equal to a then A confesses else A does not confess, same thing about b with the random number r2 then after that, you know let us look at this chart that what happens if A confesses but B does not confess, so what is that probability?

(Refer slide time 10:13)



Let us let us look at that let us look at and put this in a chart this can be put this way, so first of all probabilities A, 1 minus A and B if b, so these are the probabilities. So if A confesses and B does not, so look at the first situation A confesses but B does not then A gets 2 years and B gets 10 years. So A confess and B does not the probability is A confess a, B does not so 1 minus b, right? So that is the probability of these 2.

(Refer slide time 11:06)



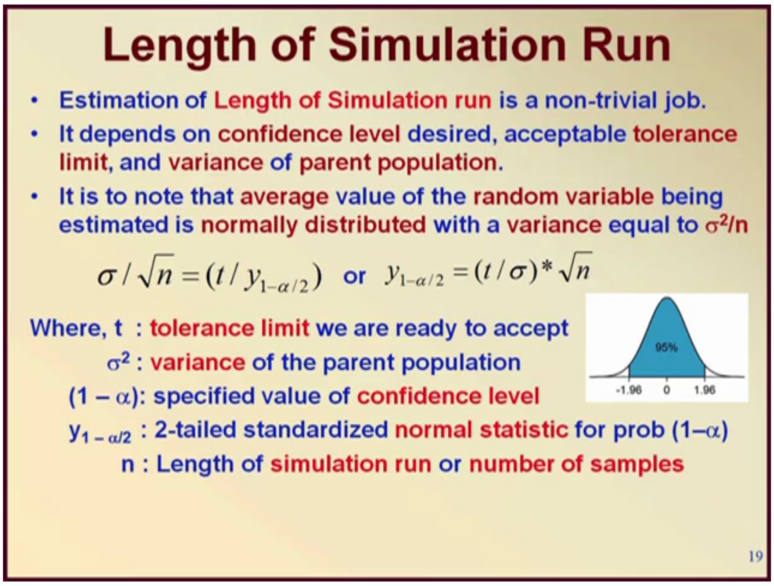
So like this finally we can do all these calculations and we can actually put you know this sort of scenarios. So prisoner versus expected penalty, so I said A confesses and B does not that probability is a into1 minus b, what was the penalty? A 2 years and B 10 years. So a into 1 minus b star 2 and a into 1 minus b star 10 that is the case for you know A and B, is it all right? Similarly 1 minus a star b into 10 and 1 minus a star b that is the reverse, so 10 and 2.

And if both confess each gets 6 years and if both do not confess each get one year. So as a rational decision maker we really compute what is known as expected values, so these are the expected penalty values that we can obtain and these figures will very much depend on the values of A and B is it all right? So if you know depends on what are the different A and B values that we obtain.

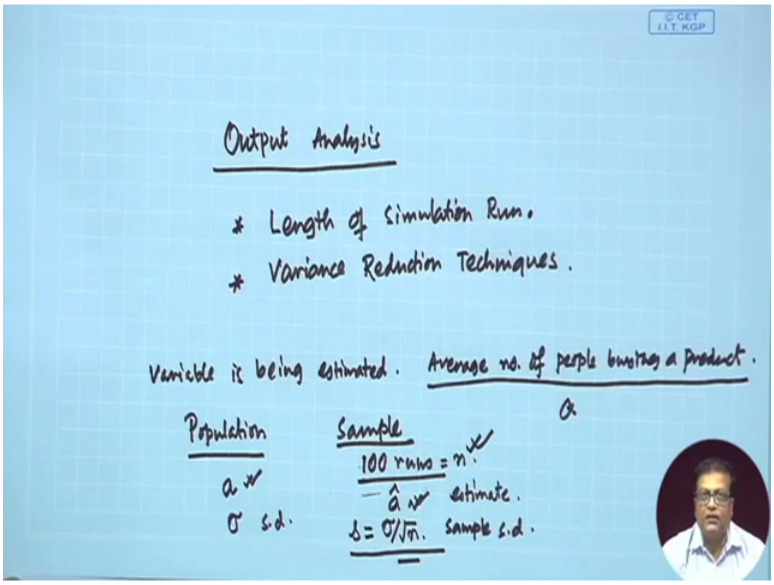
We can do various random experiments simulation and come out with these expected penalty values, right? So you can see also very interestingly if you really do this problem in in a full vigour you can really think supposing I change these probabilities, suppose a person who is A, he would think what should be my value of a? What should I do? What should be my strategy? So what kind of A should I put?

So it's a kind of sensitivity analysis and very interesting results can actually be obtained and you know you can really see one can work out the strategy and all these things really fall into the domains of game theory, is it all right? But we will not go into games theory right here, the essential idea here is really to understand, how Monte Carlo assimilation can be done to obtain some meaning into such situations.

(Refer slide time 13:28)



(Refer slide time 13:37)



Anyhow we will move to what is known as output analyses. So in output analyses we shall do basically 2 types of things, output analyses, 2 topics 2 very important topics with regard to simulation, we shall take one is known as the length of simulation run and second one we shall take is what is known as variance reduction techniques. So these 2 important topics we shall take up now, right?

So length of simulation run and what is known as the variance reduction technique. First of all the length of simulation run is a interesting thing to really look into because we are simulating let us say we did some simulation for 10 customers, 16 customers and sometimes

5000 runs, what is the actual number? How many times do we simulate? This is all right, that's very important.

Obviously I told you that simulating only once is not sufficient, we have to simulate number of times with different set of random numbers and these are called replications. So we have to dig replications also but before that what should, is there an estimate of the length of simulation run? So we have you know a formula that we have and we can actually work it out in this manner, what is that manner?

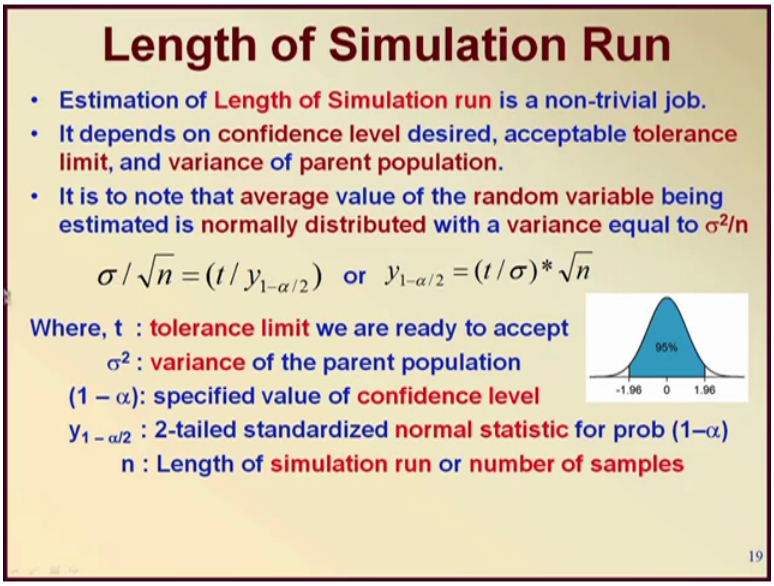
See basically you say, supposing we are actually estimating what is known as an average value? So when we are estimating an average value we should know you see usually what happens when this particular let us say some variable is being estimated, or variable is being estimated, what is that variable? it could be that average buying a product. Average, suppose we are estimating this, average number of people buying a product.

So there are 2 things one is called population another thing is called Sample. A Simulation is like a sampling process, so the original population these are the entire set of people you know who are buying a given product and that Sample is you know maybe we have simulated for hundred runs, that's probably is a Sample, right? So say we are estimating this average number of people buying a product through assimilation and we have taken hundred runs, is it all right?

So this value, suppose we call this value as a then this is the value and here we really get what is known as a cap an estimate of a. A is the actual thing coming out of the population and a cap is an estimate coming out of the simulation. So there is a theorem that says that if the population standard deviation is Sigma than the sample standard deviation is, Sigma by root n, is it all right?

So you know what does it say? So if you take a small sample you may not be able to get a true estimate of the population standard deviation, right? There is a factor root n that comes into the picture and what is n? This n is nothing but the number of runs that is number of samples. So if you have hundred runs n is hundred. So you see why this idea the sample standard deviation is really Sigma by root n you know that will help us in identifying n, so this is the idea this is the essential idea.

(Refer slide time 19:02)

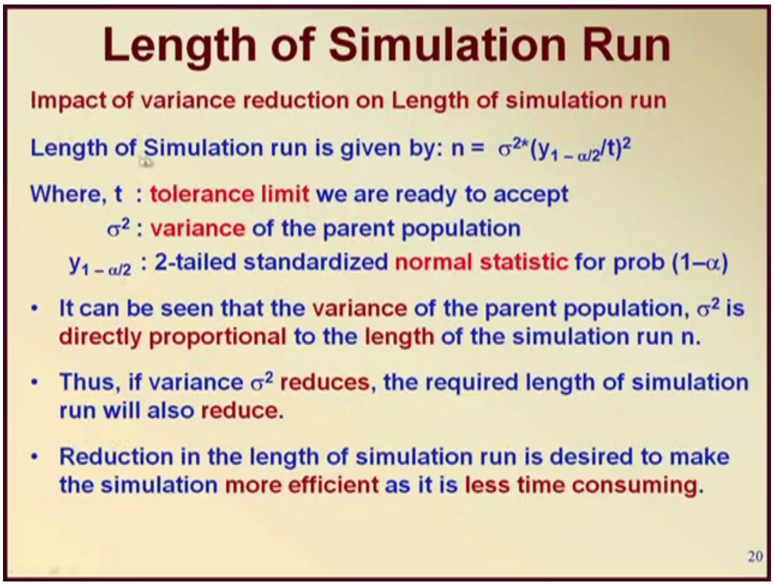


So what we have done? You know what is that estimation? That estimation really can be obtained from the acceptable tolerance limit t, supposing we know that acceptable tolerance limit that is t and we know the 2 tailed standardized normal statistics for a given probability 1 minus alpha, so what is Alpha? Alpha is really you know the specified level it determines the specified level of confidence specified level of confidence.

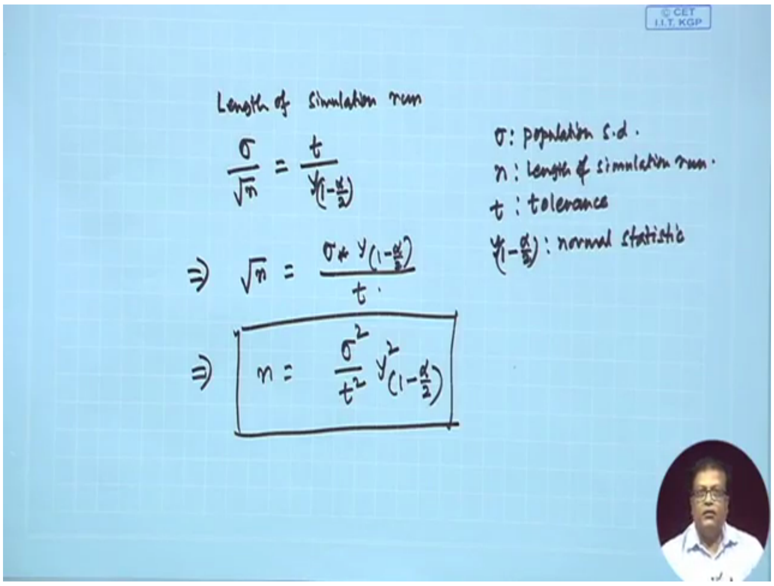
So suppose we really have 95 percent, see basically these small 2 portions are alpha together. So if this is the total normal distribution then with a 95 percent confidence level means the value of alpha it is 5 and what is alpha by 2 it is 2.5. So really speaking if this point is you know the side is 2.5 this side is 2.5 if you add them it will become 50 percent 5 percent area and if you know 95 percent confidence level means within a normal distribution values will fall within this region, probability is 95 percent.

So what is the probability? That we cover 95 percent of the cases, right? That is what we're trying to estimate and if you really and we know those values are minus 1.96 in the standard normal distribution. So this is known, so if this is known at 95 percent so then you know that y that is the value in this case 1.96 that is the Sigma. Say this 1.96 is a kind of Sigma Value standard normal Sigma Value and t by this value y 1 minus alpha that is a 2 tailed standard normal statistic that would be an estimate of the sample standard deviation, is it all right?

(Refer slide time 21:13)



(Refer slide time 21:33)



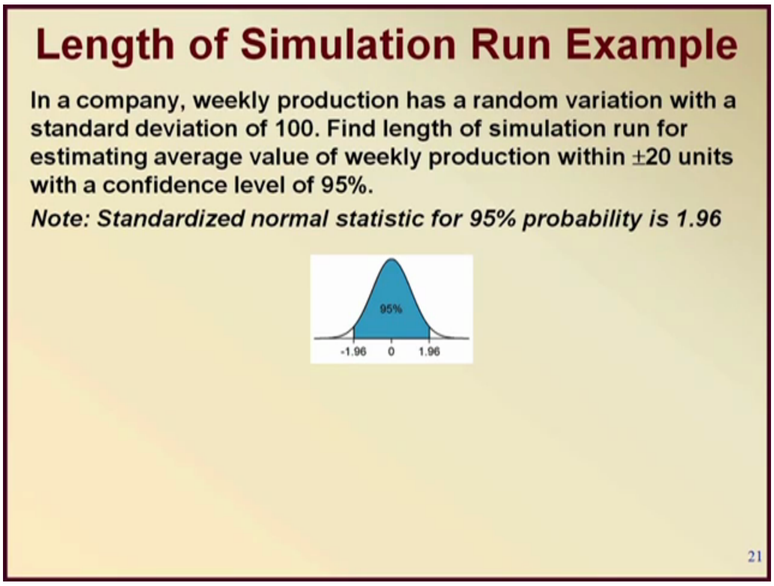
So this is the formula which actually relates n and really one can estimate what should be the length of simulation run? So let us look at the next, what happens? So really if you reorganize that formula, so let us do that. How to reorganize that formula? So we have seen already that length the simulation run Sigma by root n equal to t by y 1 minus alpha by 2, right? So here Sigma population standard deviation, n: length of simulation run, t: tolerance, y: normal statistics.

Obviously all of these are true if we have a normal distribution coming up out of these mean value, usually it happens, right? So from here if you really reorganize then you get this vertical value that is root n equal to Sigma into y1 minus alpha by 2 by t or n equal to Sigma square by t square, y square 1 minus alpha by 2. So from here we can estimate the length of simulation run.

so look in the in the slide, so t is the tolerance limit we are ready to accept, Sigma square is the variance of the parent population and this is the 2 tailed standardised normal statistic for probability 1 minus alpha. So it can be seen that the variance of parent population, Sigma square is directly proportional to the length of the simulation run n. Thus if variance Sigma square reduces the required length of simulation run will also reduce.

And reduction in the length of simulation run is desired to make the simulation more efficient as it is less time-consuming, right? You know what is exactly being told here, that first of all the variance of parent population is Sigma square. Suppose you know it is very important that the parent population variance is you know that estimate should be clearly known, is it all right? And if we can have the process by which we can estimate it better with less value, it helps and so how it helps we shall see that later on.

(Refer slide time 24:46)



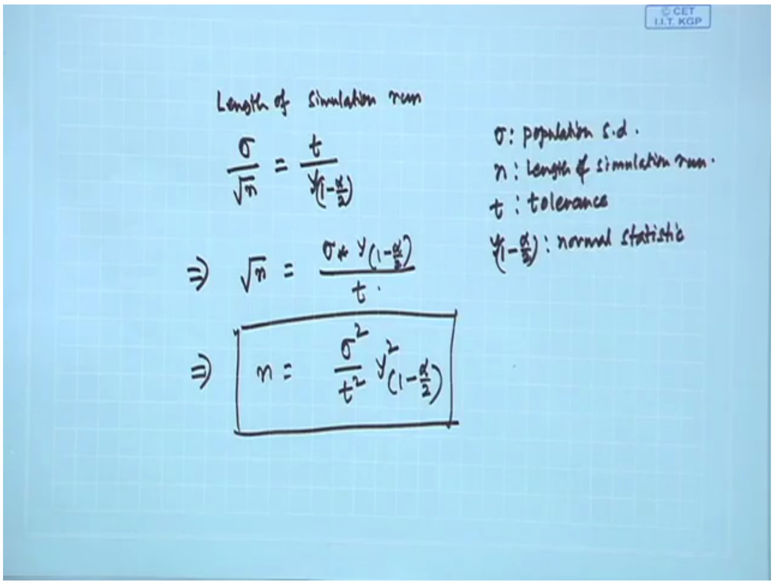
So let us see particular example on the length of simulation run. In a company weekly production has a random variation with standard deviation of hundred. So find the length of simulation run for estimating average value of weekly production within plus minus 20 percent with a confidence level of 95 percent, right?

Note standardized normal statistic for 95 percent probability is 1.96 and here is that plot of this normal distribution. So this is a standard normal distribution, you know what is happens is standard normal distribution that we have mean at 0 and standard deviation is at 1, sigma equal to 1. So if Sigma equals to 1, you know usually between 2,1 Sigma around 66 percent area is covered and in this case it can be seen that at between 1.96 standard deviation 95 percent of the area could be covered.

So how do we go ahead with this particular problem to be solved, right? So first of all the weekly production has a random variation with a standard deviation of hundred. So this is already given and that is about the population. So that means the standard deviation sd which is equal to Sigma that is hundred, right? And what is the tolerance? The tolerance that you are allowed that is within plus minus 20 percent, right?

So these are some of the figures that are given, so let us look at these values, so what is the one minus alpha? One minus alpha is 0.95 and y 1minus alpha by 2 that is a 2 tailed statistical normal statistic that is 1.96 and what is the tolerance limit we are ready to accept that is 20 because it's a 2 tailed and we are taking one tail, is it not? So plus minus 20, so this side 20 and that side 20, so we are taking really 20.

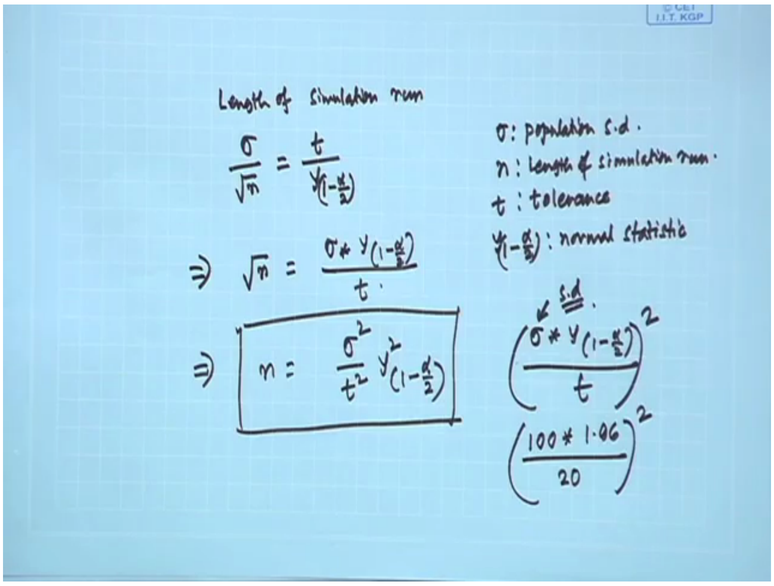
(Refer slide time 27:36)



You see if you divide by a value by the standard normal because 1.96 is a standard normal statistic, 20 units should be you know the value. So if you divide by standard normal that is 20 by 1.96 you know then that will give you the, you know this. Like here we have seen Sigma square by t square. So Sigma square by t square you know t is 20 and Sigma square is you know that is in this case hundred.

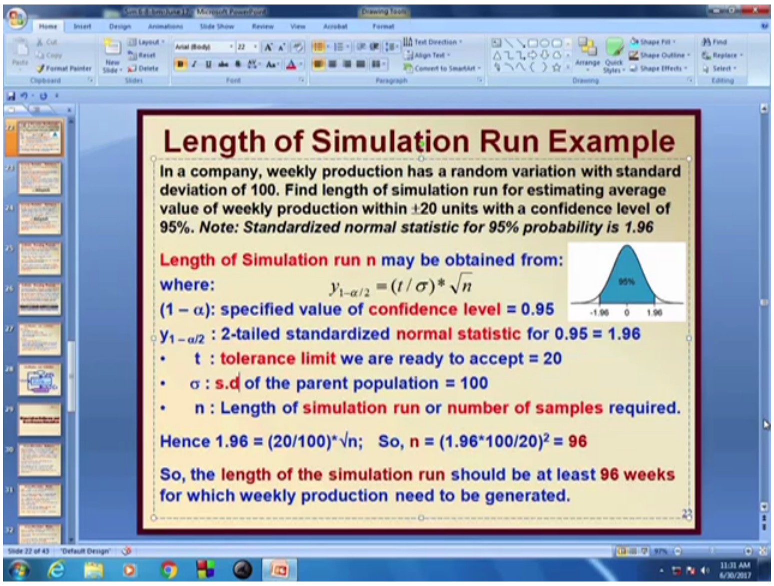
So the variance is hundred, I mean standard deviation it was written that should be variance. So variance is hundred the length of simulation run, no this is something wrong you know, please correct it that standard deviation is hundred. So in this case it is not really variance this is the standard deviation, Sigma is hundred that is why we have you know multiplied.

(Refer slide time 28:26)

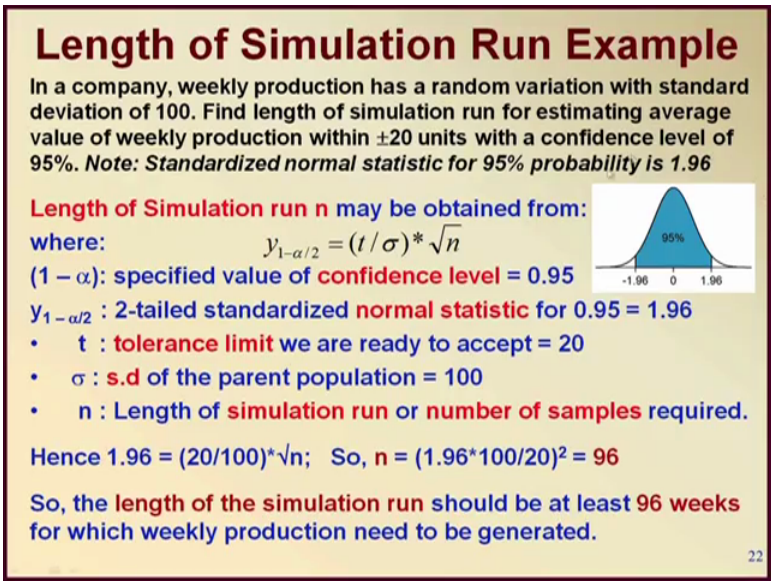


So if you’re really look at this formula the n is n equal to Sigma into y 1 minus alpha by 2 by t whole square, right? So here the Sigma is standard deviation. So in this case we have seen you know this sigma is 100 this is 1.96 and t is 20, so that should be it. So 100 by 1.96 by 2 that should be 96. So length of simulation run in this case will be 96.

(Refer slide time 29:18)



(Refer slide time 29:26)



So you just correct it, so you know you can see that this should not be Sigma square this should be s.d. So this is now corrected, so t is the tolerance limit, Sigma is the standard deviation, n is the length of simulation run or the number of samples. So it should be 1.96 is 20 by 100 that is Sigma into root n, is it all right? So what is 1.96? y1 minus alpha by 2, right? And this is given the standard deviation is hundred so 1.96 is 20 by 100 into root n, so n will be 1.96 into 100 by 20 whole square equal to 96.

So the length of simulation run should be at least 96 weeks for which weekly production need to be generated we stop here and we shall continue our discussions on the output analysis particularly the variance reduction technique the next session, thank you very much.