Performance Evaluation of Computer Systems

Prof. Krishna Moorthy Sivalingam
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

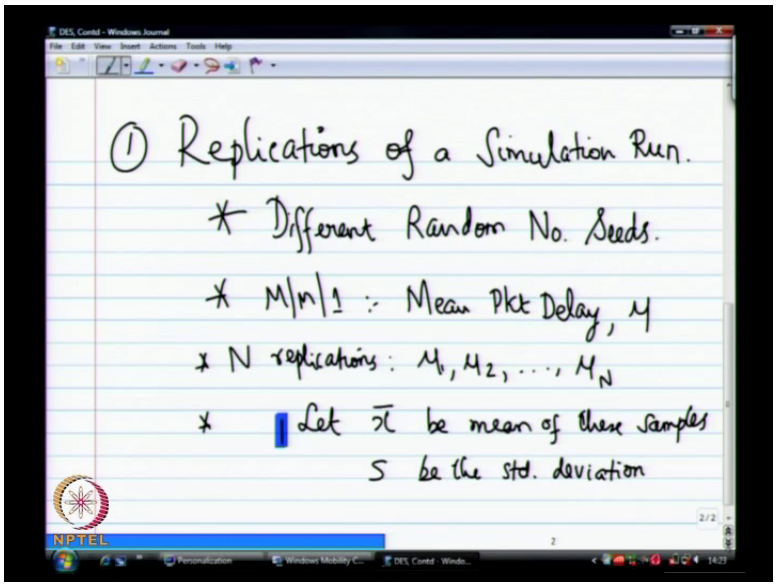
Lecture No. # 22
Simulations - II

So, the first is if you look at any simulation program like what we seen with mm 1.There are of three things, if the run the experiment just once; let say if I am asking you for the mean delay with some lambda mu values. So if you run it exactly once, then you are only getting the value of mean for a particular seed value for a given random number see. And the depending on the value of a random number seed, you know that random streams can be different. Just reporting value based on one exactly random seed is not accurate in the case of a system such as this, where the randomness can have impact on your system performance.

Simple thing is let say are trying to do topology generation, in a trying to do some set of you know protocol analysis, manatee protocol analysis. So, huge randomly generate a first topology first with some. So, there the nodes are placed in random positions, and then you give them random speed and so on, then if you are going to tell me that run the simulation of compared a o d v, and wireless r; and wireless r is having better delay than a o d v by margin of 10 percent. And therefore wireless area indeed greater than dv. So, the first question that will be asking sub in this is at that is for one particular random deployment of topology and for one to particular rate generation of packets.

There so many thing that are random in the system, the way the nodes of placed the mobility, therefore there are the way the nodes move, and also the rate in which the packet generation, different packets come it different point in times. So, there is the possibility that is the start lucky with this particular simulation, you run it again you find that you a o d v actually better than o l s on then and so on. So, you have to essentially run multiple replication of the same system.

(Refer Slide Time: 01:47)



So, if you simply run mm 1 with just one value that is one seed value, which is not adequate. You have to the able to so you have to have mechanism to change the seed and in the case of this modify version of reactimately, we have to sixteen random streams we can chosen. But some time it is important to use a same seed. So, you should know what the seed values. If you so how to be initializing the random number as seed in UNIX. For example, you can use time of the day that is what we normally do, but they say that really not action thatgood see but we normally use that.

So, if I use time of the day and then you run the simulation, once simulation trashes then you when the simulation again know got simulation trash, let may run it again. Second time for some reason simulation does not trash, why because a random seed was different. Therefore, the arrival pattern of packet was different your queue. So, may be something that happened in your queue, when it the queue sort of overflowing. For example, in the first case it is possible that you queue are overflowing in did not really capture that condition some vary set someupper limit is queue running around memory; and therefore, the system trashes.

Second case, just magically way the packets arrived there was no real overflow of the queue therefore, a queue was running perfectly of code was working perfectly. So, in that case, then you are running it trouble, because you did not know what the first seed was, you know that some time it crashes, sometime is does not then all known that in the t a as runs as it does not run and otherwise it runs perfectly in a room. So, that is sometimes whether t aas probably

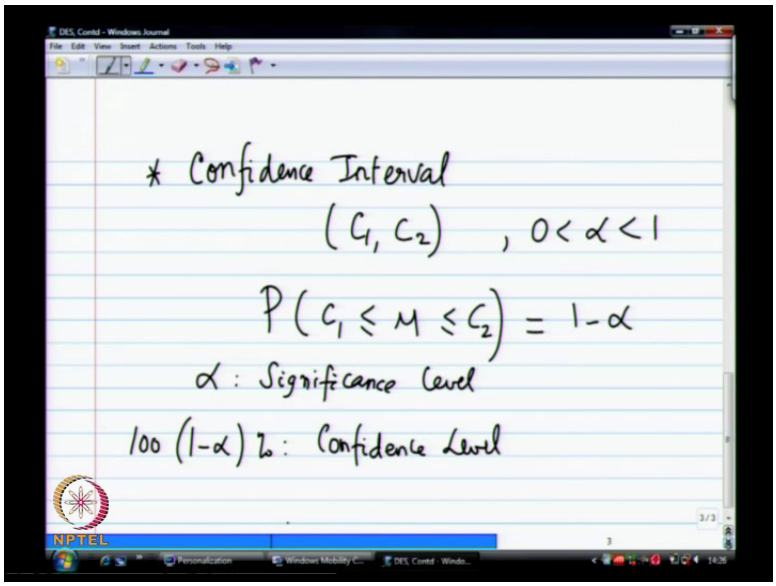
will, but not you getyou know has knows that these thing will definitely casual program. Therefore, the give the verses in programs, so that is boundary condition normally fail, but in this case are not even a boundary condition, it is just the two different random values.

So, you know thatonesends you want to be able to replicate the same set of event,whenin run it againyou shouldbe able to specify a given random seed. So, normally simply said seed out seed to zero and then you know that same sequence of random number will be there, same arrival of packets will be there that you can check with fix your random seed. That is for debugging proposes, but debugging proposes fix your random seed knowing what the random seeds, whether it is same stream of packet is getting generated.Once your figure out this work then; of course, you have to change random seed, then it crashes again you know that for this particular random seeds crash.

So, you have to go and in to the system in find outin this case for soonce all that is done then you want to do this replication with different random number of seeds. There are also other things that look at the interrupt of your system we can look at yours slashing. What is the thing you can this you take your look at what other things can be used in you are the UNIX system itself. Look at some of current system state, essentially try to take that capture that is the seedit is absolutelythere is the absolute randomness, when your selecting a particular streams gets totwo different stream haveno similarity of top. So, let us say that I have computed the run the simulation for different random seeds then a measuring some entities.

So, let us look at once particular entity in the case of will mm 1.We are measuring the mean delay, mean packet delay which is say mu. So, this is for every application your getting a set of sample your let say you run it forn replications. So, your sample space is mu 1, mu 2, and mu 3 and so on. So, you have a set ofn values that represent what could have in the essential that actual sample actual mean of the system itself. So, will say that let us let x bar, be the mean of these and s be the simplest thing is to compute the mean of these means have done x number of replications, and then compute a mean of these mean, then I say the this is the mean of in terms of packet delay.

(Refer Slide Time: 06:33)



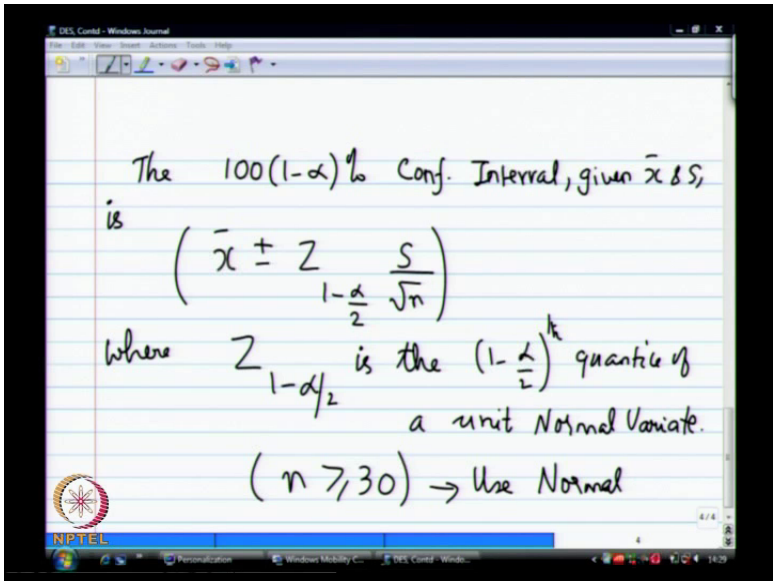
But you can goone step further actually compute the confidence interval. So, the term confidence interval is now the new some of you headed some of you not some take information takes. So, how do compute the confident interval, that what you if you how do you computed. What is the definition for the confidence interval? So, the interval it waiting probability which clients spend them, that particular confidence interval you needs to pdf, you needs to pdf in one minus to dependent. So, you are given the confidence interval basically is definitelytwo values to ensure the intervals C 1 C 2 and the probability, and usually there is also another parameter which will curve as you know alpha. Alpha will be usually less than alpha features between 0 and 1.

So, the probability that actual mean this is not the sample mean the actual population mean lies between these two interval these two values C1C 2 is equal to 1 minus alpha. So, if I tell you what alpha usually should be other wait, tell you what the alpha is then based on the means of the means.We can actually tell that is this is the probability, that the mean for the system will live between these two values with some probe with this one minus mu. It is specification that you are actually provided again internals some time. So, this alpha is all the significant level then your 1 minus alpha. So, we normally could that in a percentage is called a confidence level.

So, when I tell that mean of the system of the average delay of the system is given this confidence will say that you know lies between 10 and 20 with 95 percent confident interval

is basically these 2 values 10 and 25. So, there is some way to is essentially computing this modeling, this is either ast distribution or a normal distribution.

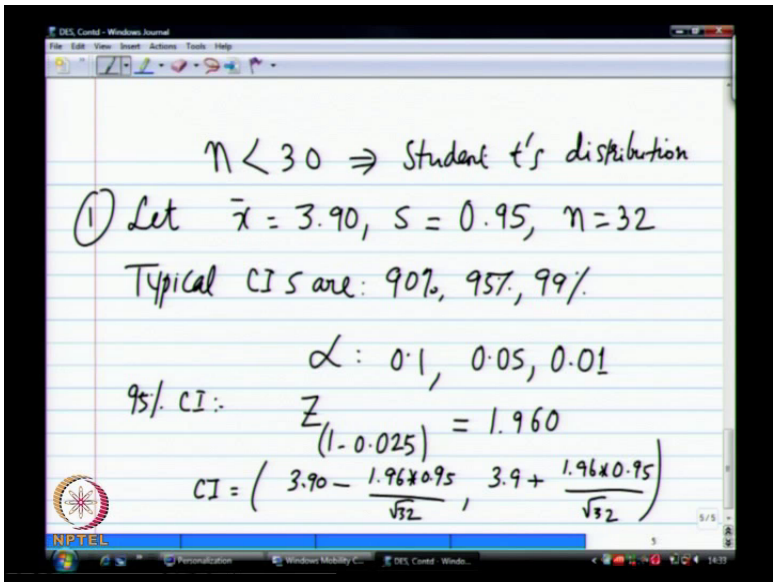
(Refer Slide Time: 09:25)



So, the confidence it given your values ofgiven that x bar and is given byjust make it expression. So, repute times here which I want attempt to explain now. So, need to go through that once morefor now I will just take the formula as such in a comeback with where actual figure of how this maps to unit normal variate and so on. So, in normally if you have more than 30 samples, thisis if your n is greater thanbe go to the unit normal variate and then there is the table usually that tells, if you remember of mole fraction class table andsoon.

That tells you for the given value of alpha, what is the corresponding value for this z. Let, if one example end of this book you will find when on the wept will find this values for z, there unit normally variate I just yet those valuation plus 10 wants to time which is little available, and if it is less than 30 then we have to go to a different distribution called the t distribution. So, if it isthisare we reduce the normal table.

(Refer Slide Time: 11:51)



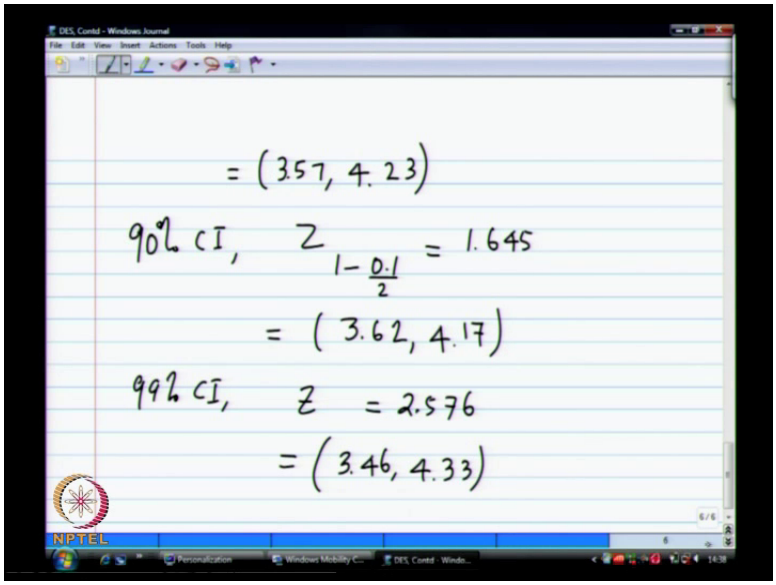
And if it is n is less than 30, then we normally is the student t is the distribution. So, why this is important will justtaken moment. So, in this a previous example that we hadwe run say any examples, so for which look atone numerical example. So, letyou know your x bar equals 3.9 this standard deviation be set 0.95 and n equals to 32 replications and computed the mean value of across all the mean packet delays, what the system will could be any per many metric, you looking at one metric that your measuring and then this is mean across all those metrics.

So, in this casenow this confident interval we normally specify in terms of 90 percent, 95 percent, and 99 percent and so on to the typical confidence interval specifications. So, has the higher are confidence interval requirement of the confidence interval level, confidents level, then your window will interval bigger will be larger. Because you want to make sure that 0.99 percent of all from how with probability 0.99 your system going to haveyour interval represent the tag of the probability, that 99 percent of the time your values will be mean value will be this event range.

So, the corresponding value for alpha is 0.1 and 1 minus alpha. So, say start with the law confidence interval compute the alpha, then we going to look of this stable; so let say for thewant to compute the 95 percent confidence interval for this particular system. So, I need to compute Z of 0.1 minus alpha by 2. So, that is 1 minus 0.025 which terms have to be, somany

times is simply,they have just out you really want to do it in thesafe roughly proportional to equal to 2.

(Refer Slide Time: 15:44)



Then your confidence interval is basically 3.9 minus, say under the single value computer discrete 3.9 essentially are window for 95 percent interval is 3.57, 4.23. So, this means that 95 percent of the times mean value will be with this particular range, this is the first values. If you want to have 90 percent confidence intervals then you have Z 1 minus 0.1 by 2 that is 1.64; and therefore, the confidence interval actually will work out to be the 90 percent in sent to be narrower is your giving more fullrewindin terms of the expectation of varies out.

What the mean valuation beenthis in one the table yes say it mu, is not just in either it will live inthis range this probabilitymu is the mean delay, because we never need you know what the real delay event willlooking at this is of behave a set of samples. You could run this for a million times the larger you run, then your smaller the ranges are if run it for 10 million times. For example, and compute the mu is strength, the average delay mu 1 mu 2 and so on.

You will find that the actual variation will have if you look at the average of that there will probably more be more representative, and therefore when I have a larger sample size and less standard deviation. If all the mu values they for the mean value computed at end of each experiment are fairly closed each other, you are deviation will be less variants will be less in s twill also be less which case again your inference intervals going to be very small.

We should terminate the 95 present.

So what, now we will get the that is the batch means spot which have state then run, that is that isthere I am not doingreplication. I am just doing means batches within their simulation and I am computing the batch means;

In you m m 1 queue sir you are have the replicate flag. What is the mean that e of n list of staging, then you terminate the programreduce the repetition value.

Normally, I use the replicate flag then it will try to auto terminate then use the order terminate, other will simply run for the duration that is specify.

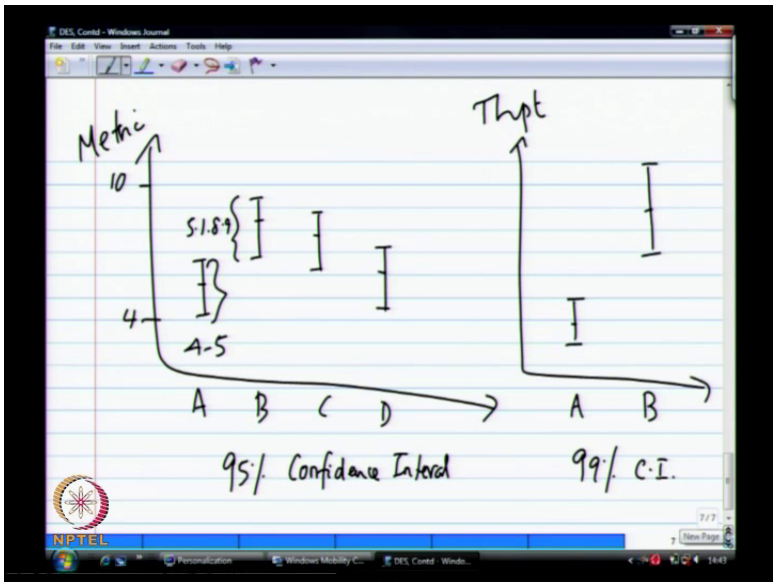
I will just come to that termination part just taken you are here to the set of sample, when here is the interval 90 percent confidence interval based on the set of samples mu and sigma values. So, if you go to the 99 percent then you the Z value for this corresponds to will be larger. So, is the general has a confidence interval or level increases a corresponding Z value will also increase which cases a window corresponding get last.

And if your access very small which means that the relative deviation between the different runs, between the different replications; therefore, you know that right the range which youractual valuable fall is going to within this very small value. So, this is square root of n is there the formula square root of n is needed, as not needed divided by square root of n. Because later on and the our same large invoke in top of square root n s 2 different ratio for the simple form what you know the square root of n.

So, there are still some errors that event after multiple additions there are not fixed this are most introduction. So, what is the correspondinginterval here? In other words when you report values in pieces or papers, whatever it is, we just current we just say thatthis is my one single mean value. Then compare we look at comparing two different systemlater on, this is simply comparing to means is not adequate.

Even though, I am returning the replication you also have to look at the confidence interval for both those two different system a system a system b, and you says system a is mean is better. I have run this 1000 times can run the mean is alone is simply cannot compare, you should also compare based on the confidence interval only then can conclusively we say that a is better that b, so will come to that later on.

(Refer Slide Time: 20:33)



So, sometimesyou may have seen group plots,where actually plot.

(( ))

How to do say that digital printer then it higher the value it is how to be know…

The interval should not overlap there simplest thing is the interval should not over lap, if I have two systems there is comparing in this case, for a let say I am comparingfour systems.

This is the access simply the system type and this is your metric. So, you would have seen you can do is the plot very easily; if just have one more values, so if x y and then the third things your confidenceinterval, it automatically compute the plus or minus for the confidence interval.

So, this is the mean I am not run it perfectly, but the main is in the middle exact middle of these bar. So now if you look at it right visually says speaking, you find that there will overlap it they are apparent of this system overlap this is the system. So, therefore is the non-zero probability that the system that actually be operating, my end of the delay for the system or metric for the system, actually be here and the delay for the system is actually here, in which case their not statistic statistically speaking there not significantly different they have to send. So, this is no static, if I really want to say that system A is better than system B, and then I want something like this. Let say this is throughput;so you want higher throughput should be better if this is A and this is B.

So, I implement system A, implement system B and then there is the clear visual separation. If you two intervals then you can say that one is better than the other. Especially, larger the interrupt confidence level, if it is 90 percent confidence level in say may there is the small chance that you know. If I look up expand the confidence interval when it is possible there will be some over lock 90 percent there is no over lock, but 99 percent that might be over lock in the system.

So, you to be really say the one is better than the other. We should have this there are other smaller conditions too, but is the simplest one that one can visualize;if you can get a graph like this set compares two system that to after large number of replication. So, more the replication better the confidence in your system, you running say some starting algorithm A and B and then you are running it, but several different data structure million different data sets. Then there is more confident strangers running with different dada sets, because as you increase a data set say number of data sets your run value goes up in; therefore, you have confidence intervalalso start shrinking.

So, that therefore, you are range will smaller. So, that is the first in that one has torealize that when you report it, it has to be reported with replications and the corresponding confidence interval for a given metrics. So, that is first point that I want to take today questions. In your next assignment there is no comparison of system, third assignment you will do comparison of system.

Here, the confidence interval issame of for all the system.

This is the different A, B, C, D are different confidence that is say, this is delay. So, this is set timing and this is side over on something like that. So, this is this guides interval is between 4 and say in order on this is 5 and this is say between you know 5.12 over you know 8.9 and thing say that. So, these are the actual confidence interval. So, what I am plotting here is the actual intervals right the corresponding to the x is always.

(( ))

This is let a say this is the 95 percent type let us say. I am plotting the 95 percent confidence interval, I am running this four system and I am let say this is sorting because A, B, C, D, and computing the time taken to sort and I am just I am taking in to sort is the metric do for each in replication run this over several million data sets. Then I compute the mean and compute

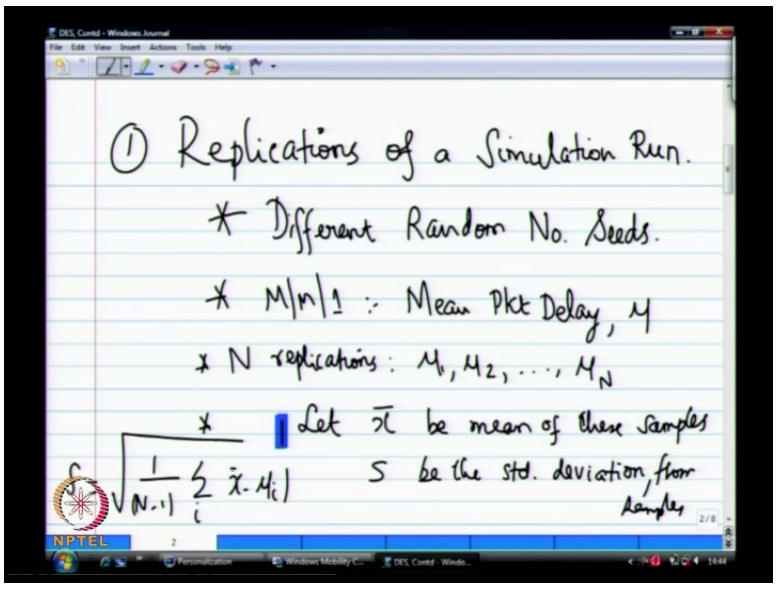
the C I based on the standard deviation for that particular data set for particular algorithm a, and that is what I am plotting here in this manner.

The system A should be all the A, B, C, D all those based on 95 percent on only 80 percent and then 90 percent.

That is correct, that was simply long short they should be all on the same time. Let say this is 99 percent, so the take then is of the requirementis your pieces the report find that 99 percent; there is over lock then go for 90. Then they where there is no over lock. So, therefore, you can say your system is better than there. So, depends on when you want to finish your pieces, and that is the firstaspect of what I want to talk about it is the mean, it is the sample standard deviation say based on the samples.

So, it is mean of the samples and of this samples this not the sample this is not the population standard deviation, because this solve achieve no. You compute the standard deviation and this we can complete. What estimated exactly? Could you just write down s in terms of sigma, what we what we estimate as population standard deviation from the sample standard deviation, ask the 10 minus 1 by n factor, this is this is your standard deviation as computed from the sample.

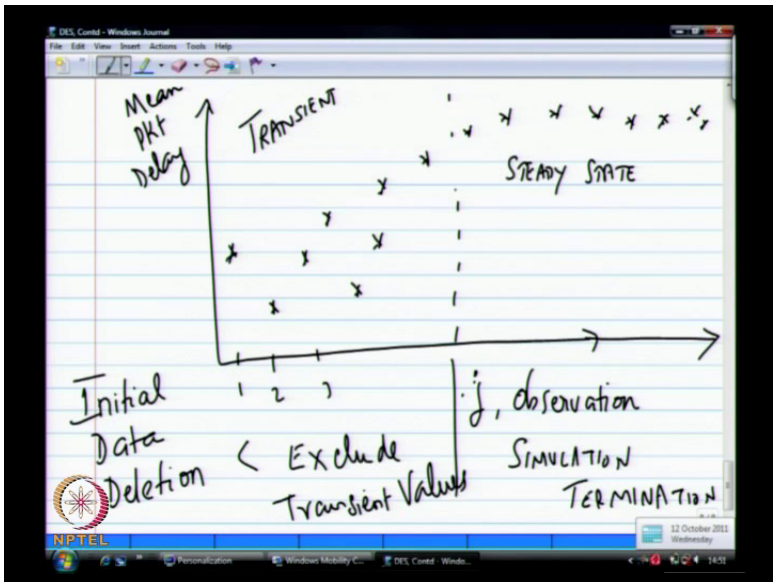
(Refer Slide Time: 26:11)



So, if you guess we are variants, sos is square root of 1 over N minus 1. This is what I am looking its say this is just based on the sample. I do notthink, going the estimate in the you

are packing about estimating the sample the population standard deviation from the sample. Next, so there are two aspects of running simulations. So, if you look atsay measuring that is say we are measuring they act the averagedelay for customer for packet on the system.

(Refer Slide Time: 27:06)



Let say this is j your observe observation count or basically j customer whose departing this system, and you are now plotting this with respect to increasing j for the first customer the mean delay is whatever value that is solve. So, this is simply one say on this is the mean packet delay. So, probably now this alreadyfirst will be simply whatever value, that you get from verse, and then this value will be just going up and down the mean value will be unstable for avail are not unstable will be set that transient state of the system. There the values can of jump widely up and down, and has your j value increases then you are, and if you system is getting stable then slowly yoursystem will finally, convergence of some point.

There is will be some variation, but over time your steady state will be reached and this is essentially what you will call as the steady statemean delay mean packet delay for this particular system. So, your systems will always especially if you are looking at the network queues. The queues are all empty when you start with and then as a packets of arrive in to this system.Then the queue starts getting populated and thethen slowly the system queue start getting built up. Once queues are building up to certain point then addition of new customers does not really change the behavior of the system.Your system is essentially entering steady

state. So, you arebehavior of the system is essentially this is your transient state or initial state and then this is finally, where you enter steady state.

So, when you compute the main packet delay. So, final value that compute will also include some of this initialdelays, where the system has been sort of the initial observation there all included in that. So, there are two things achieve you want to do. One is in that fewer go list to only look at state steady state values, you donot wantthe values to be corrupted by the initial transient state of the system. So, you want to be able to figure out, you want exclude these values.See you donot want those observations that all, because you know that it does not matter.

Because the first I could have delay of 0.01 and the system delays actually 10. So, 0.0 just happen that is particular customer had that is 0.01 second of service time. The real service delays this much higher average service less much higher. It is to you wanted exclude the system state, when it is pay head it is in when it is in the transient state. So, that is the first thingcalled initial data deletion.

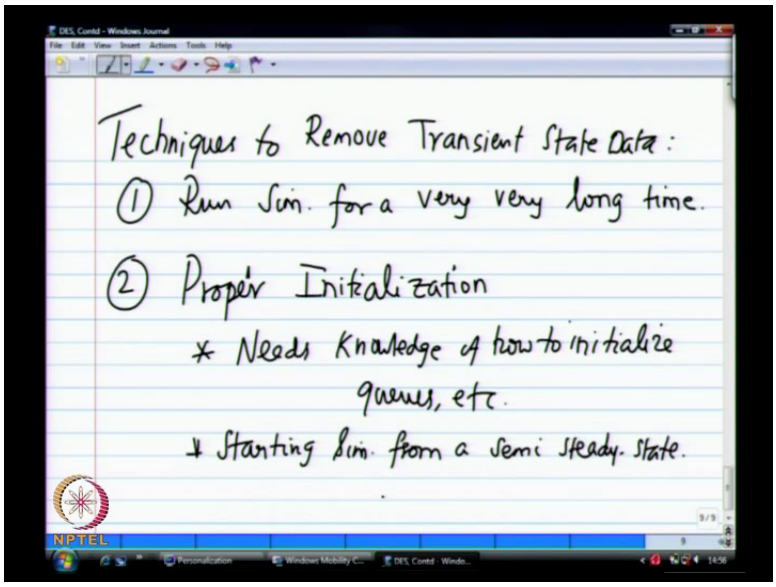
So, that is what you would normally would want to do to get of the transient state, if you are only interest in the steady, sometimes you will only one state the transient state, that is the different steady. But that is then this for steady state analysisand then when you go to steady state also the classic question is when to by stop by simulation. So, simulation termination condition, because if you keep recording the average values is every observation it my varying between some values is not going to be a constant value if it is totally random.

So, it will be varying between a few set of various synchronous up and down, jump up and it is see that and then that went to I say that enough in a run by I will stop my simulation. So, because the simplest thing is for initial relational just ignore first 500 units or 500 customers, because I am learning this for a million customer, so I do not care first 500 customer want to really make it different system performance and likewise termination condition simplest thing is run it long enough.

Usually, it is dependent upon when you want finish your project and leave. So, if I each simulation going to take ten hour, really want proper termination that driver auto terminative will have to draw going to drawn forever. Until, unless gives that other parameter saying, likewise just stop at this pointed compute. So, you want to be every scientifically say the system as the achieved steady state, that is no point in running any more simulation. So,

usually just based on some sort of variants computation is change in a values computation. So, this istwo things that will talk about, how to do this data deletion.Initial data deletion as well as simulation term deletion withoutarbitrarily choosing some values. Easiest thing is to just choose arbitrarily and more head, but if you can possibly do something better than that.

(Refer Slide Time: 31:35)



Then, what are the techniques? A technique to remove transient data, the first is which run the simulation for the very long time, easy for the m m 1 system. Actually, we can do that whether it is simple system, not much of computation as they involved not much of back ground processing is there in terms of state changes and soon.Of course, is question is how long to run. So, this almost this will satisfied both in problem transient state as well as steady state, how long to run just run in it for day and then see whatever but there are some set of much more complex, we have students running simulation will be opponent for analog networks in simulation and soon.

For5 minutes of real time done for a day, if I really want to get higher correction. But 24 hours they will say for run, and then I say run each thing for 10 times or 20 times are replication shake and the said 20 days for one value. And then this one 24 one particular combination of it is, then it is achange the number of users change the combination of traffic that is getting generated.

So, then nobody of not take some product this know known as there is can I something simpler,whether there is no simply event simulation. So, discrete event simulation systems

and if I trying to some those something like law, events law scale system like molecular dynamic something like that. So, there to take a very long time forto do all the computation it is involved. So, thereforerunning simulation for a very long time is easy what it is sometimesnot packet up.

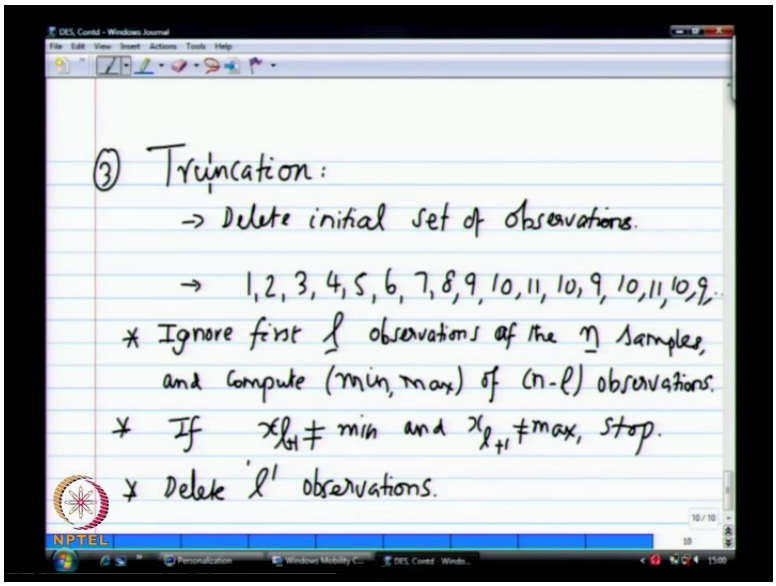
So, that certainlynot going togone and the second base two look at initializing the system with some initial state. Assuming that you know what will be let say that I am looking at network of queues. You normally start a simulation by saying there is no customer on the system, customers will slowly start appearingone byone asking saw in our m m 1 and then I can look at you know the therefore, I canjust start from statewrite downzero.

The other way is to say that I am just fill allthese some of the queue with one or two customers just to split essentially fast holder system to some future point in time, where there are customers sitting in each of those queues with the corresponding service times happen this like that. So, you basicallystart your system with some part of your initializing basically take care of some initial behavior of the system itself. We just moving you aremoving the system forward in time to some relatively immediate futuristic position from which you can then start computing.

So, only thing is how you figure out those initial values that is the hard work. So, if you know sometime, it is possible if you have set trace files. For example, where you can use that probability system and then say this is initial or buffer lock up and see queue occupant see and so on. So, this needs knowledge of how to initialize the queues and soon.

So, basically starting from as semi state; semi steady state is in a have little bit of alcohol, steady state is in a have run. So, I just you that is why call cancelled, if a some lot of many tell you, what is steady state value should slot from then the I will tell you, then there is notion of truncation this will be delete the just the initial set of observations. It is again nothing; just simply run longer other is to general randomly, say first 50 observations, 500 observations. We talked about that we can also delete, but we can little bit more I just mathematical about that.

(Refer Slide Time: 36:44)



So, you have let say set of samplesthat much I can randomly. So, this is the just some values at your measuring from the system. Whatever, this is delay some metricwaiting time. So, the basiclogic is ignore the first one observations, one is some value of the let say you collected n samples total and compute the min and maxof the remaining min and max.

So, l is some index among from 1 to n, for a given l. I am ignoring the first tell and looking at the min and max of the remaining observation, and then if x l this is not equal to the min or x l plus 1and stop and the delete.

It is neither the min are it should be in the range, you should be the range pin. So, met at the say not equal to ten point way it should be within the range they should not be min or it not the another s 2s 3 inthat is all.

So, it wantbecause l plus x plus 1 within your min and max computation, should not be another extremes that is all, because x l plus 1 to x n is what you computing the min, and max over and the x l plus 1 should not be the that the smallest another largest value. I know this not this is not perfect. This is way very crude approximation. It is should be thatthe 10.1 and this values 10.2, and you random up say you know is neither the min nor the maximum. So, this is the very crud way but if you want to look at we basically in this case. So, let see in this particular example what happens when l equals to 2 where l equals 1. I look at 2 2 whatever other remaining samples minimum values 2; therefore, really note that.

So, keep doing this until if finally come to 10 when I look at is l is equal to 9 right, then 10 onwards the set of values is 9, 9, 10 and 11. Therefore, the min is 9 then max is 11, next value happens to be 10 which is neither the min nor max. Therefore, I will stop that is noticeably is not the best method, because your only talking about the two extreme values.