Simulation of Business Systems

Prof. Deepu Philip
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture - 25
Probability Distributions – 2

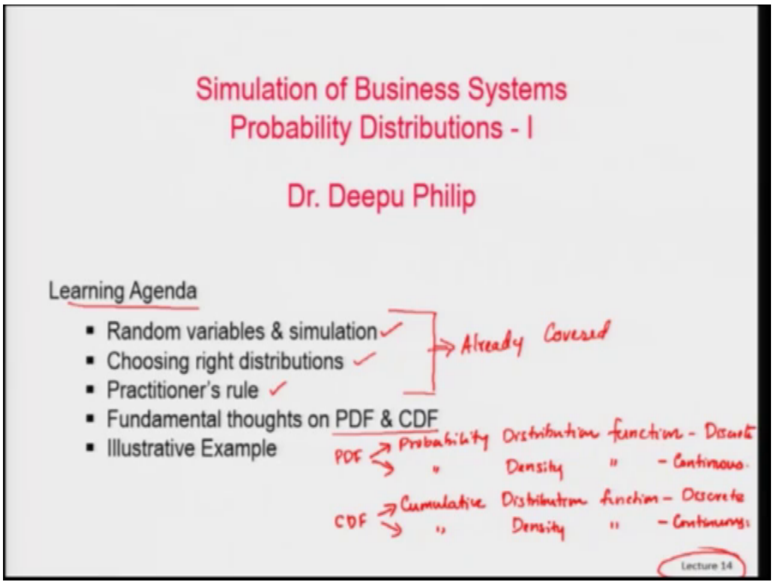
Good evening students. Welcome to towards the end of the concluding lectures. So, we are now that end of this course of Simulation of Business System. And we are now trying to end of the course and we are trying to look into some of the basis some more of the fundamental things that you require to build basic simulation models.

While this course is not the complete course yet, but this is a starting point for you guys to get into the field of simulation, which is going to be a pretty important tool for the business decisions in the coming years. The next wave after analytics will be prescriptive analytics which is which will be dependent on simulation and that is why we decided to offer this course.

So, today we are going to get in to the concluding side of the probability distribution concepts that are required for the simulation simulating different business systems. Please understand that this is not a probability statistics course.

But the problem that we understand from many of the students were coming on to the forum is there require some more additional information in probability and statistics that is why this lecture are included in this course. So, that the remaining topics that we are discussing the high level topics we are discussing kind of make sense to you.

(Refer Slide Time: 01:37)

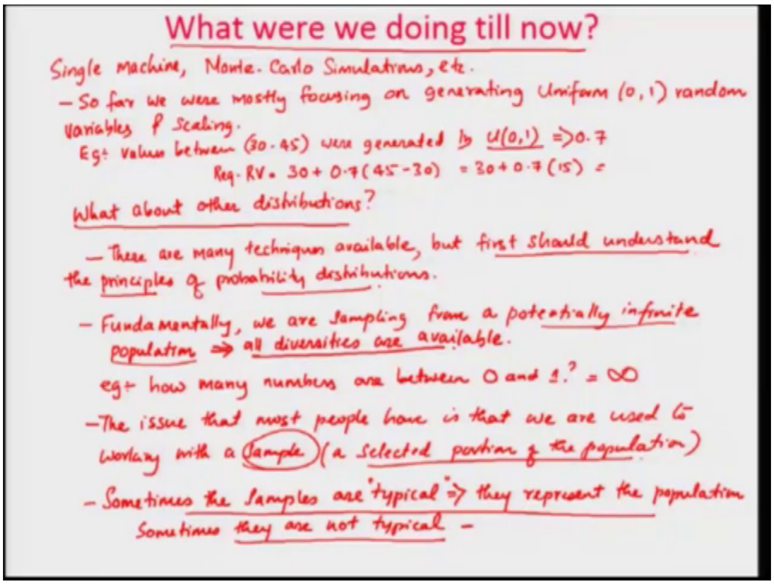


So, if you look into the agenda learning agenda for today. You can see that we have already covered the learning agenda of this, we already covered the random variables in simulation. We have covered how to choose the right distributions and you also seen what are the practitioners rules of the simulation. So, this was how do you choose a probability distribution this was already covered. So, this is already covered that was in lecture 13.

So, in lecture 14 today we are going to look into the some fundamental thoughts on PDF and CDF. And I said earlier PDF stands for two things probability distribution function this is in the case of discrete. This is the density function in case of continuous.

Similarly CDF has also two things cumulative distribution function. In the case of discrete and cumulative density function in the case of continuous ok. And we also will look into a simple illustrative example today.

(Refer Slide Time: 03:11)



So, what were we doing till now? Up to this point what there we saw different type of simulations? We saw single machine models, we saw what we call as the Monte Carlo simulation etcetera ok. So, what were we doing there ok? So, the main questions the question the answer to our question is this. So far we were mostly focusing on generating; we are mostly focusing on generating uniform 0, 1 random variables; the one random variables and scaling.

So, if you remember we have to come up with the random numbers between example values between let us say 30 and 45 were generated by; we were generated by uniform 0, 1. So, let us say you got a value of 0.7. So, then what we would end up doing is this will give you the required value required random variable equal to 30 plus 0.7 times 45 minus 30 is what we ended up doing or 30 plus 0.7 times 15 which will give you what they random variable that you want to do ok.

So, the question we have is then what about other distributions? This is the most important question that we need to understand. What about the other distribution? Why only uniform 0, 1? Why not the other distributions ok? So, the answer is there are many techniques available.

There are many techniques available for generating random variables, but first should understand the principles of; understand the principles of, what principles you need to understand? Principles of probability distributions.

So, some fundamental basic principles should be understood in the case of probability distributions to study about the other type of distributions. Fundamentally or basically what we are doing is we are sampling from a potentially infinite population. So, what does this mean? This means all diversities are available; that means, you are going to sample from an infinite population which means the whole population itself is quite diverse.

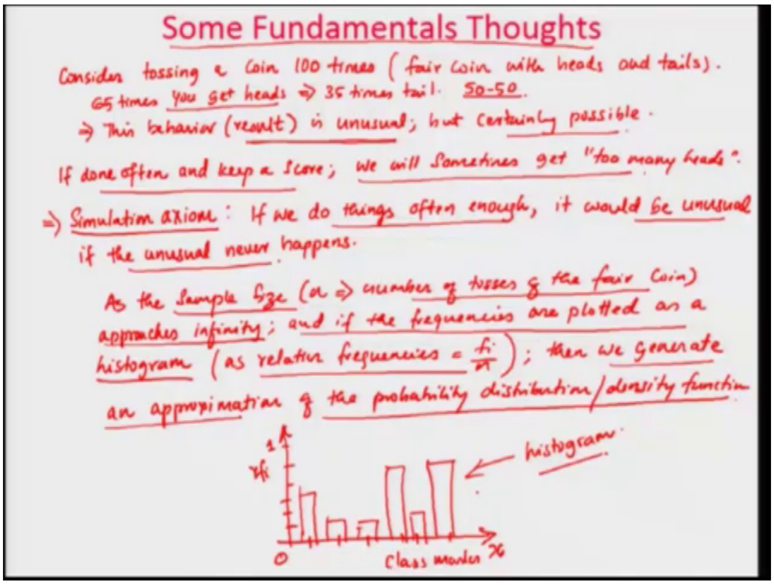
So, example how many numbers are between 0 and 1 ok? How many numbers are in between? If you look into the real number the answer to that it is there infinite number of numbers in between. There is quite a lot of large numbers in between right ok.

The issue that most people have is that we are used to used to working with a sample, you are used to working with the sample and sample means a selected portion of the population. So, you are working on a selected portion of the population and but fundamentally ideally we are supposed to work with a potentially infinite population.

But we end up choosing with a sample usually work with the sample ok. Sometimes the samples are typical which means they represent the population. And sometimes they are not, sometimes the samples are not typical. So, which means they do not represent the population ok.

So, sometimes samples are typical means you work with the sample sometimes you are sample will be representative of the population, sometimes it is not representative of the population. And when it is not representative of the population then you will have all sort of issues in getting understanding that, understanding the behaviour of the system.

(Refer Slide Time: 09:28)



So, let me give you some basic examples ok. Consider tossing a coin; consider tossing a coin 100 times ok. So, the coins as a fair coin with heads and tails ok. And let us you toss this coin 100 times 65 times you get heads, which implies what? 35 times tail ok.

What we are trying to see here is ideally speaking in a coin you are supposed to see 50 50 ok. But you see a different number here ok. This behaviour or result is unusual; it is unusual, but certainly possible ok. If you are tossing the coin 100 times and you see 65 times you get the heads ok. This result is unusual means yes it is possible you can easily get this is doable, but it is rare ok.

So, what we are trying to say is if done often if we keep on tossing the coin many times and keep a score. We will sometimes get too many heads it is possible. If you keep on repeating this experiment tossing the coin 100 times and you keep a score sometimes you will get too many heads ok.

The reason that you will get too many heads is because it is an unusual event this is not very common, but it is possible certainly possible. So, this gives you the simulation axiom, this is where one of the fundamental observations or simulations comes into picture.

What it says is if we do things often things often enough, if we do things often enough, it would be unusual if the unusual never happens ok. So, the simulation axiom says if you do things often enough then it would be unusual that is the unusual never happens. This is the very important aspect because let us assume that you are riding a motorcycle every day daily basis without wearing helmet ok.

Then one of the unusual thing that should happen is an accident happening at the following of the motor cycle and gets severely injured. And if you keep on doing this for a long time period the law of large numbers will catch up with you at some point of time you will actually see that accident happening. So, hence this is why we people say that if you keep on doing things too often enough then the chances that if the unusual never happened then that will be quite unusual. So, the unusual will happen ok.

So, in another way to say it is as the sample size in this case n which is the number of tosses of the coin, of the fair coin. As the sample size increases approaches infinity as a sample size reaches infinity and if the frequencies are plotted, frequencies are floated as a histogram as a histogram. We already seen what is a histogram earlier in the previous lectures, I will have one more example in the next lecture about histogram because there are so many questions about it.

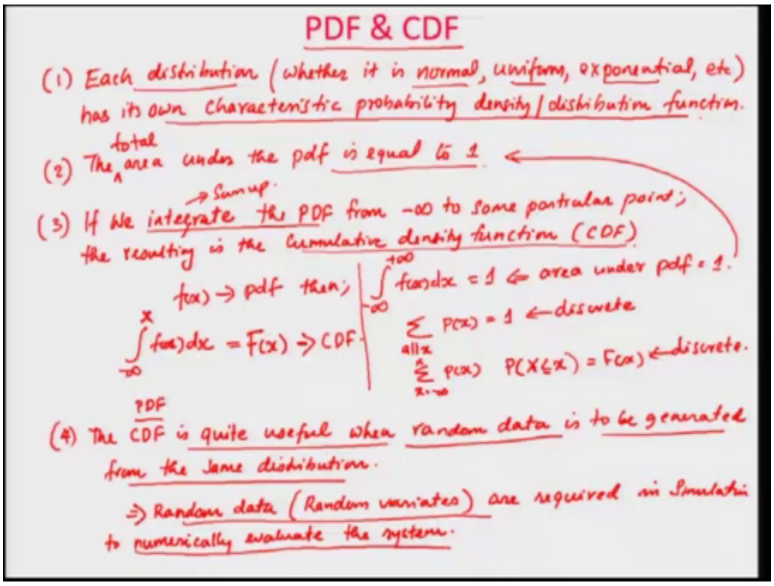
But if you plot this frequency as if this is a histogram as relative frequencies say ok. Relative frequencies relative frequency is f i divided by n, the frequency divided by the number total number is a relative frequency ok. Then we generate an approximation; then we generate an approximation of the probability distribution, or density function ok.

So, what we are saying here is that as the sample size the n. For example, let us talk about this the number of tosses in the fair coin approaches infinity. And if the frequencies are plotted we plot the frequencies of receiving the head and obtain how many heads are seen in the tosses and we are take the frequencies and then we calculate the relative frequency which is the number of tosses divided by n.

And then we plot them as a histogram remember histogram look something like this you have your x axis on which the different classes are class marks are seen like this plus marks ok. And here you can have your relative frequency which is the frequency divided by n and these values will be vary from 0 to 1. If you are using relative frequency and you might see graph like this something like this ok. Some stuff like that, so this is what we talked about relative frequency integers ok. So, this kind of a distribution is of diagram is called as a histogram. I will give you one more example because people are lot of questions about the relative, because since that many people are not worked on the frequency tables or frequency distribution here.

So, we will get into one more example of this, but understand that if we do this, if this graph is plotted. Then what we are getting is we generate an approximation somewhat similar to what the probability distribution function of the process that we are looking into ok.

(Refer Slide Time: 16:51)



So, let us talk about the PDF and CDF ok. As I said earlier PDF, and CDF are probability density function, or distribution function, and cumulative density function and distribution function ok.

So, the first thing that you need to understand the most important thing that you need to understand is each distribution each distribution ok whether it is normal, uniform, exponential etcetera. Whatever be the distribution each distribution whether it is normal, uniform, exponential etcetera has its own characteristic probability density slash distribution function ok.

So, whatever be the distribution whether it is normal, uniform, a exponential it does not matters. It has it is own characteristic probability density function. The probability density distribution function is unique for each one of those probability distribution. Second thing you should understand is the area under the PDF is equal to 1 ok. The total area total area under the PDF is equal to 1 ok.

So, third one if we integrate the PDF from minus infinity to some point some particular point. The resulting is the cumulative density function which is what we call as CDF. So, if we integrate or integrate in a way in other way to think about it is if we sum up ok. If it is a continuous function if you have f of x this is the PDF ok. Then integral minus infinity to plus infinity f of x d x equal to 1 which means area; area under PDF is equal to 1 ok. This is what the second point I was talking about ok.

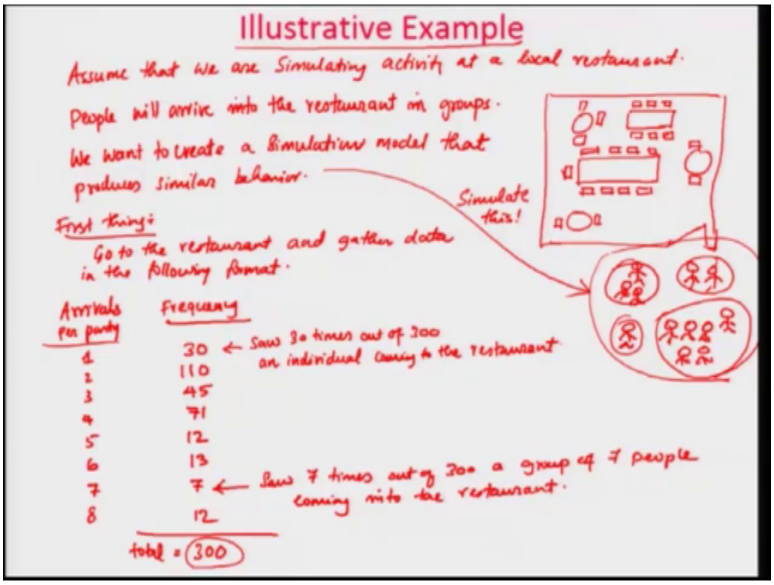
The other thing is it is if you take integral minus infinity to some value x f of x d x. If you do this to some point x if you are integrating this then this is denoted by capital f x which is the CDF. So, integrating from minus infinity or in the case this is the case of continuous distributions. In the case of discrete cases you can write the same thing as sigma all x the value of probability of x is equal to 1. This is the case of your discrete ok.

Another example is we can say summation of x equal to minus infinity to sum x P x gives you probability of x less than or equal to sum x this is your F x in the case of discrete ok. So, the fourth point that you need to understand is the CDF is useful ok, is quite useful, not just CDF the PDF also. The PDF, CDF is quite useful when random data is to be generated from the same distribution ok.

So, as I said earlier that what is the PDF is? The it is own characteristics how the probability of the random variable varies and if you integrate the PDF from minus infinity some particular point you get the cumulative density function. And I showed you what this means actually this is the continuous case and this is the discrete case.

The PDF or the CDF is quite useful they are very important. When you have to generate random data need to be generated from the same distribution. Random data or what we call as random variates are required in simulation to numerically evaluate the system. So, these random variables or this random data that we generate random data that we generate, it is required for the numerical analysis of the system or numerical evaluation of the system ok.

(Refer Slide Time: 23:09)



So, let me demonstrate this to you using a simple example ok. A simple illustrative example this might make sense to you easily, assume that we are simulating activity at a local restaurant activity at a local restaurant. So, you are studying a local restaurant and if you think about it here is a restaurant ok.

You have table with multiple chairs there could be a another table ok. And there could be a another table here only two people sitting could have a much big table here kind of a thing ok. And let say you have a system like this and here is our restaurant our restaurant has a door ok. So, this is our door this regard ok.

And there is another table also here let just put it this way ok. That means, this implies that people will arrive into the restaurant in groups ok. So, they will come into this, so, the people will be coming in groups. Some people will be in groups of 2, some will be in groups of 3. Some will be wall by themselves, some will be in (Refer Time: 25:08) groups ok. And there will be a person who will be coming all by himself ok. So, this is an individual this is like this you have people coming in different groups ok.

And we want to create a simulation model; create a simulation model that produces similar behaviour. So, you want to create a system which produces the same similar this is the behaviour that you would like to simulate this ok.

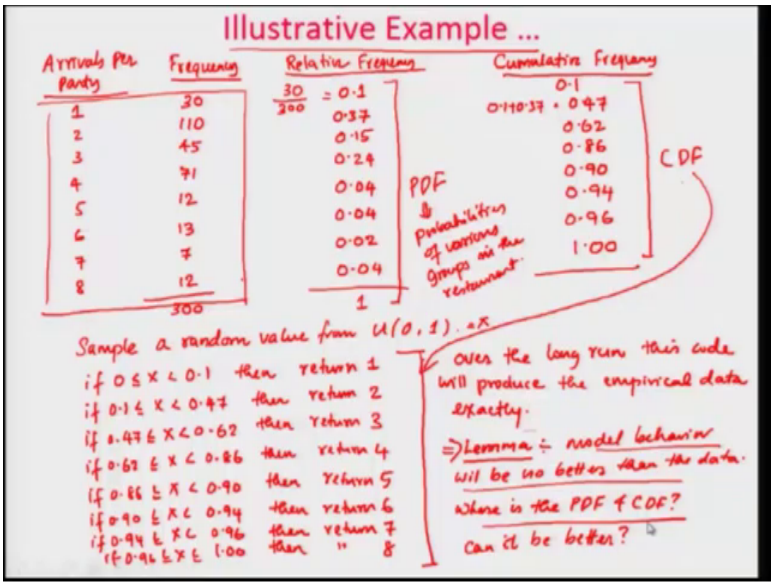
So, first thing you do is; first thing ok. Go to the restaurant and gather data in the following format. So, first thing you do is you go to the restaurant and collect the data in a following format. So, what you will always one is arrivals per party ok. There will be one column then you will have is the frequency ok.

So, arrivals per party you are call it as 1, 2, 3, 4, 5, 6, 7, and 8 ok. So, this is the number of people who arrive this is the number of people that your per party that you have seen. And you seen that individuals arrive in 30. Two people per party you have seen them in 110, 3 is in 45, 4 is in 71, 5 is in 12, 6 is in 13, 7 is in 7, 8 is in 12. So that means, you saw 30 times this implies what? So 30 times ok. So, this total will be equal to 300. So, 30 times out of 300 and individual coming to the, an individual coming to the restaurant that is what you saw alright.

What about this case? So 7 times out of 300 a group of 7 people coming into the restaurant ok. So, that is what this data actually implies to you ok. So, each one of these row implies that in 8 means your seen it all times group of 8 people, the party contains the group 8 people coming into the restaurant.

So, you sat there and you observed 300 times, 300 different cases, or 300 different people coming into them. And these 300 different people were distributed in to different group sizes accordingly alright. So this kind of gives you an idea.

(Refer Slide Time: 29:01)



If that is the case then how can you calculate this or how can you solve this as a computer example, or how can you use a convert this into a computer model. So, let us look into the data once again arrivals per party ok. Then we have is frequency. So, arrivals per party is 1, 2, 3, 4, 5, 6, 7, and 8 ok, frequency was 30, 110, 45, 71, 12, 13, 7, 12 and this number was calculated total to be 300.

So, the first thing that we do is we calculate the relative frequency, relative frequency is calculated first. So, this is given by 30 divided by 300 which is equal to 0.1, this one will be 110 divided by 300 which will be 0.37. Next one will be 45 divided by 300 will be 0.15, 71 divided by 300 will be 0.24; 12 divided by 300 will be 0.04; 13 will also be similar to that 0.04 ok, 7 will be half of that almost 0.02 and 12 will be again be 0.04 ok. If you add all of these things together then you should actually get 1 ok.

Now, we calculate the next one is called as cumulative frequency ok. So, this cumulative frequency this will be 0.1, the first one will be 0.1, the next one will be 0.1 plus 0.37 which will be 0.47. Then it will be 0.62, 0.47 plus 0.15, 0.62 then you will have is 0.8 6 which will be 0.62 plus 0.24 ok. Then will be 0.90 which will be 0.86 plus 0.4, then will be 0.94 which is 0.90 plus 0.04, then will be 0.96 which will be 0.94 plus 0.2, then it will be 1.00 0.96 plus 0.04 ok.

So, if you get a scenario like this, then how would you make a computer program to work on this? So, you can basically say that sample random value from uniform 0, 1 ok, you would sample any value ok. So, let us call this as x. If 0 less than or equal to x less than or equal to 0.1 then return 1. If 0.1 less than or equal to x less than or equal to 0.47 then return 2.

If 0.47 less than or equal to or not less than or equal to it is call it as less than because you do not want values in both cases. If 0.47 less than or equal to x less than 0.62 then return 3, if 0.62 less than or equal to x less than 0.86 then return 4, if 0.86 less than or equal to x less than 0. 90 then return 5, if 0.90 less than or equal x less than 0.94 then return 6, if 0.94 less than or equal to x less than 0.96 then return 7, if 0.9 6 less than or equal to x less than or equal to 1 then return 8.

So, what you are doing is this code over the long run this code will produce will produce the empirical data exactly ok. So, this behaviour of the system whatever data that you collected here ok, this data you can actually reproduce this over the long run using such a system ok. The lemma out of this, the lemma or the conclusion out of this is model behaviour will be no better than the data.

So, what if you follow this approach and this is the approach the most of the time what we do is, if that is what we do then the model behaviour will never be better than the data that you collect. So, obvious question is where is the PDF and CDF ok? Where is the PDF and CDF in this case? Well this is your PDF. How the probabilities are distributed for each case? This is what the PDF is all about, and this is your CDF ok.

I hope you guys understand now what I was talking about. This demonstrate probabilities of various groups in the restaurant. So, the individual probabilities is what we saw in this particular case. Here is how the cumulated, how this probability values are cumulated and you use CDF to obtain this so, that you can use CDF to develop or demonstrate the behaviour of the system ok.

Hope that this kind make sense to you guys and you understand how CDF is quite useful in generating behaviour of the system and we can use a uniform 0, 1 distribution random variable to actually sample the behaviour of the data. So, what we will do now is this kind of shows you the mechanics of how internals of the system works. But lot of the questions student asked the question was that how do; so, they many of them asked sir how do we do the inputs? How do we decide with distribution to be use this input? How do we determine the parameters?

So, our concluding lecture will be focusing on how do we deal with the input data distributions? Or how we decide whether it is exponential to use or viable to use that kind of a things and what are the parameters? So, as of now as I said earlier that this model behaviour will be no better than the data, can you actually create a scenario where you know using PDF and CDF, can it be better ok? Can the system do better and that is what we are going to discuss in the next class. And in the next class with that we will conclude this course and hopefully you guys all had fun learning this one thank you for your patient hearing.

Thank you very much.