Advanced Business Decision Support Systems

Professor Deepu Philip
Department of Industrial Engineering and Management Engineering
Indian Institute of Technology, Kanpur
Professor Amandeep Singh
Imagineering Laboratory
Dr. Prabal Pratap Singh
Indian Institute of Technology, Kanpur

Lecture 08

Monte Carlo Sampling and Usages

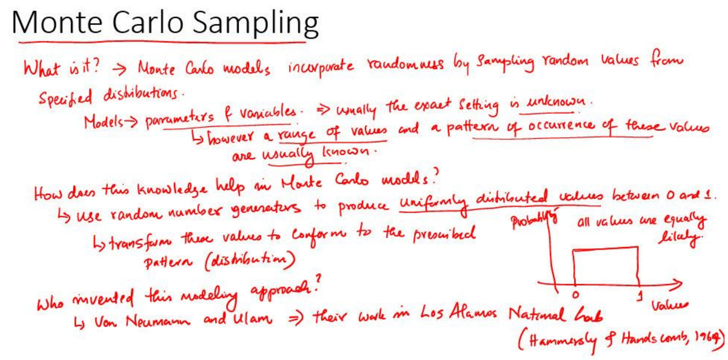
Good evening, everyone. Welcome to yet another lecture of the week 2 of the course Business Decision Support System and as you widely know, by now, this is an advanced course of the basic course which is called the Web Based Decision Support System which was also floated in NPTEL MOOC's in the last semester. So, this is an advanced level course of it.

So, unless you are already taken that course or you are reasonably good with the Decision Support System especially, Web-Based Decision Support System. Some of the stuff that we are going to discuss in later in this course will be sounds like Greek and Latin to you. But under all circumstances we would like to make it as easy as possible and convenient for practitioners and academicians alike.

And, we have already seen various aspects in this course especially, about we overviewed the Decision Support System, we looked at how the problem-solving approach, the type of decisions and how the decision-making process happens and then, we now talked about what is a system, what is a model, what is a modelling process and, in that process, we heard across a name called Monte Carlo sampling or Monte Carlo simulation model. And, briefly I just mentioned it in the previous class.

So, in today's class, we will learn Monte Carlo sampling as the first model that is used in Business Decision Support Systems. So, without further delay, I am Dr. Deepu Philip from IIT Kanpur and the other two constructors of this course are Dr. Prabal Pratap Singh from IIT Kanpur and Dr. Amandeep Singh Oberoi, he is also from IIT Kanpur and they will also be co-teaching other modules of the course.

So, this is again week 2 lecture and let us look into the slides now. So, Monte Carlo sampling and its usages are the major contents today and after we go through the presentation, I will be using a excel sheet to show how to do Monte Carlo simulation or Monte Carlo sampling.



So, what is Monte Carlo sampling right? So, the question is what is it? The answer is Monte Carlo models is a model. Some people call it as Monte Carlo sampling, Monte Carlo simulation, these are all names same. They incorporate randomness by sampling random values from specified distributions.

So, we had mentioned already in the previous things that when we make models, they will have parameters and variables. And, usually the exact setting is unknown.

However, range of values and a pattern of occurrence of these values are usually known. We have parameters and variables in the models, but the exact setting is usually unknown. It is very difficult to understand the beginning itself what those value will be. But we may be able to say that the range of the values they can this way, both parameters and the variables, the value cannot occur within this range.

And, within this range, this is the pattern in which these values will occur that much might be known to us. So, then, how does this knowledge help in Monte Carlo models? That is the second question. If we know this pattern and the range how is it going to help. So, the answer to that, use random number generators.

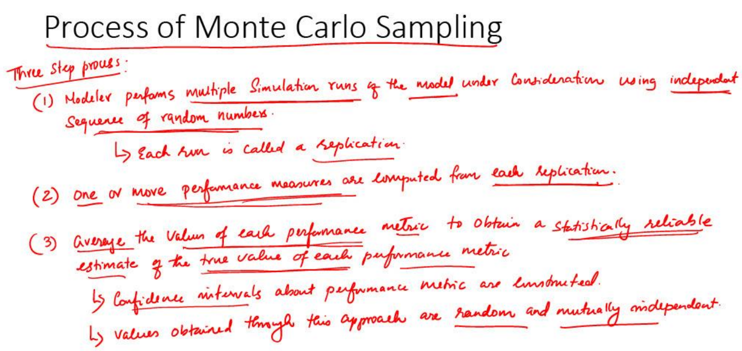
These random number generators are associated with every programming values. A programming software including excel to produce uniformly distributed values between 0 and 1. So, what do you mean by uniformly distributed values? Uniformly distributed values are if this is the values, this is the probability.

Then, the uniform distribution values are something like this between 0, 1, let us say take it this way. So, this is the probability. So, that means, all values are equally likely. Any value can be taken, they all have the same probability of occurrence.

Then, once you generate this uniformly distributed random values between 0 and 1, what do you do is you transform these values to confirm to a prescribed distribution or pattern. So, that is what we will do.

We will take this uniform 0, 1 value and then make it into follow a particular shape. So, then who invented this? This modeling approach that is the another question as part of it. The answer is Von Neumann and Ulm. And, this is attributed to their work in Los Alamos National Laboratory. So, the reference you can look at it is Hammersley and Hanscom in 1964.

This will give you more details about the work and if you want to study how this actually happened, the history behind it, you can do it. So, that is the fundamentals of this Monte Carlo sampling.



So, how do we do this? So, it is a process of Monte Carlo sampling. And, this we call it as a 3-step process. So, 3-step process. So, the first step number 1, Modeller performs multiple simulation runs of the model under consideration using independent sequence of random numbers.

So, multiple simulation runs of the model is conducted and we want to create independent sequence of random numbers. So, each run using the model is an independent replication.

So, what we call it as, we use this word earlier, each run is called a replication. So, multiple independent sequence of random numbers, we create one replication each. Now, one or more performance measures are computed from each replication. You calculate at least one or more than one performance measures from each replication, that is the second step.

The third step is, average the values of each performance metric to obtain a statistically reliable estimate of the true value of each performance metric.

So, what you are saying is that, we average the values of each performance metric. So, ask each, we do one replication, we take the values and we keep on averaging them and what do we do to obtain a statistically reliable estimate of the true value of each performance metric.

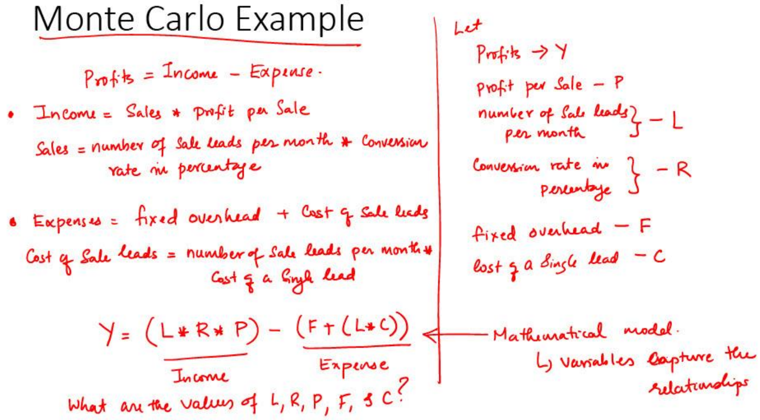
So, you have some guidelines, some performance limits within which, you are trying to calculate each replication, you will try to get an averaged value. Then, by doing this, what can you do? We can construct something called as Confidence Intervals about performance metric are constructed.

You construct a confidence level about the performance metric and also the values obtained through this, approach values obtained through this approach are random and mutually independent.

These values are random and they are mutually independent. So, the three steps the first one is the modular performs the multiple simulation runs. So, most of the time there is a model available to you most of the time in this case, a model is available to you but you really do not know the parameters and the values, you have some range. So, that is how it works.

So, then you use independent sequence of random numbers to replicate them or run multiple simulations. So, each run is a replication and then using each replication value, you calculate one or more performance measures from each replication. Once these performance measures are computed for over a large number of such replications, you average them this performance metric and that gives you a statistically reliable true estimate of the performance metric.

And, then you can construct confidence intervals of it and you can actually get random and mutually independent values.



So, let us do an example. So, everybody knows by a simple rule let us call

Profits = Income - Expense

So, let profits is represented by 'y’. Then, we know that we can say that,

Income = Sales * * Profit per sale. So, let us say profit per sale, let us denote it as 'P'.

Then, we can also say that,

Sales = Number of sales leads per month *Conversion rate in percentage.

So, let us call it as number of sales leads per month. Let us call this as 'L' and conversion rate in percentage. Let us call it as 'R'. So, y denotes profits, profits per sale is P. So, number of sale leads per month is L and conversion rate in percentage is R. So, that is the income, that is one part.

Then, we have is,

Expenses = Fixed overhead + Cost of sale leads.

So, let us call the fixed overhead is denoted by ‘F’ and

Cost of sale leads = Number of leads per month * Cost of a single lead.

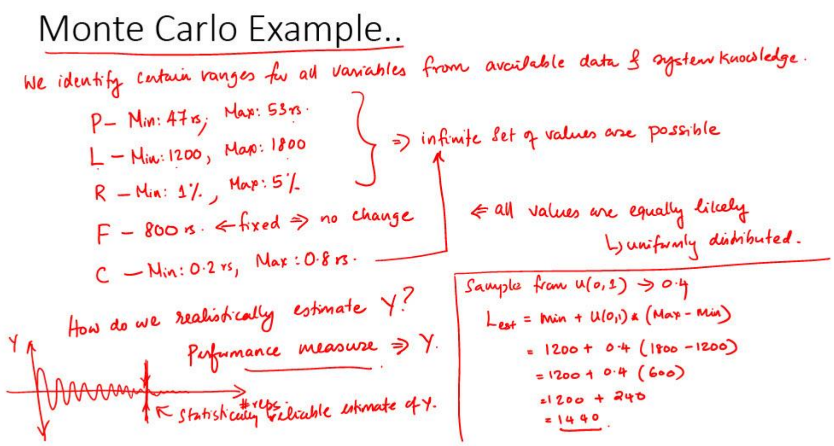
So, we already have the number of leads per month is we defined it as L. So, let us define cost of a single lead, let us call it as 'C'.

So, if you do that then, you can say that

$$y = (L * R * P) - [F + (L*C)]$$

So, (L * R *P) will give your income. [F + (L*C)] is your expense. So, we just converted a knowledge kind of a thing into a mathematical equation at this point. So, this one that you can think is a mathematical model now. It is with where variables express or capture the relationships . So, we have written the example or we have shown how this model is created.

Now, the point is the biggest question is what are the values of L, R, P, F and C? You have no clue what these values are at this point. So, usually in real life, you would not know the exact values of this, but you may know the range of these values. So, let us establish some ranges here.



So, we identify certain ranges for all variables from available data and system knowledge. So, using somebody's knowledge expertise etcetera, we are trying to find out the values of some of these variables.

So, we say that what is P? P is the profit per sale. P can take a minimum value of 47 rupees and maximum value of 53 rupees. So, we are saying that the profit per sale can be anywhere between 42 to 53 rupees.

Now let us take L L is the number of sales leads per month. We will take minimum of 1200 leads per month and maximum of 1800 leads per month. So, in a month, you can get about 1200 to 1800 sale leads. So, we have now fixed the value of P, we have picked the value of L.

Now we need to look at the value of R and we asked around and some people said that R can take a minimum value of 1 percentage and maximum value of 5 percentage. So, we now have R and the C the fixed overhead is fixed. So, if the value would not change and from the database, we are able to find out that the value of F is 800, 800 rupees is fixed implies no change.

And, then the last variable is cost of a single lead and the cost of a single lead C is minimum of 0.2 rupees and maximum of 0.8 rupees that is the cost of a lead. So, people says profit can be anywhere between 47 to 53 that means, here infinite set of values are possible . You only know the except the F this also infinite set of values are possible and you do not know which values and what.

Then, the question is how do we realistically estimate what are we supposed to estimate profit. So, estimate y show that y is this. So, how do we realistically estimate y . So, our performance measure here measure here is y the profit . So, how do we realistically estimate it? So, that is where we will use the Monte Carlo simulation.

So, just remember this again the value of P can vary between 47 rupees to 53 rupees, the value of L can vary between 1200 to 1800 number of sale leads. The conversion rate is 1 percent and 5 percent minimum 1 percent maximum 5 percent. The fixed sales overhead is 800 rupees it is fixed there is no change in that and the cost of per sale lead is 0.2 rupees to 0.8 rupees and we also assume all values are equally likely.

So, they are uniformly distributed. So, and our aim is to estimate the performance measure y. So, if you try to do this by hand this is very difficult because you have an infinite number of * you have to run this by hand. So, you need a computer program to tackle this. So, how do we do that?

The easiest way to do it is I will keep this one and the easiest way to do it is open an excel sheet. You can see this that the excel sheet is already opened for you and I will zoom it a little bit more so that you can actually see it easily.

So, profit y = Income - Expenses that and income is solely from the sales we are assuming that income is solely from the sales. Income = sales * profit per sale and sales is number of sales leads per month * Conversion rate in percentage and expenses is fixed overhead F + Cost of leads.

Cost of leads is number of sales leads gain, I forgot the sale value, sale leads per month cost of the single lead and we said y = (L * R *P) – [F + (L*C)] this same equation what we wrote. And so, uniform what we do is, there is a function in excel which is called RAND So, the RAND function will return.

So, if I type something here = RAND and do that it will give me a value between 0 and 1 that is what RAND will do to you.

So, what I do is, RAND and I take that value it will be between 0 and 1 and I multiply that between the difference. So, the max - min. So, we know that P, we already discussed 47 rupees to 53 rupees L is 1200 to 1800 R is 1 percent to 5 percent cost of leads is 0.2 to 0.8 and fixed value is 800 right.

So, what we do here is, we take one value. So, the idea here is, that I will go to this one just to explain this to you. So, what do I do is, I sample a value from uniform 0, 1. So, let us take the process here. Sample from u (0,1)? So, let us say we sampled a value of 0.4 for the time being.

So, then if you take an estimate of L,

L_{est} = min + u(0,1) * (max – min)

So, what happens is,

= 1200 + 0.4 * (1800 – 1200)

= 1200 + 0.4 * (600)

= 1200 + 240

= 1440

So, that is how you will estimate the value of one instance of the uniform random number. So, with that we go back to our excel sheet. So, what we do here is, you can see that if you look at this, what I done is B 12 is the 1200, the minimum value for the L. So, you have a set up a table like this and the table is L, R, P, F, C and F is the fixed value of 800. So, nothing changes it is you can say it is just 800 nothing changes there.

L is set up in this min + this order. So, min is 1200 + RAND, it will give you values between 0 and 1 and what I done is C12 - B12 = 1800 - 1200.

So, that if I do, it gives me one value, this is one random instance of L. Similarly, for R if you look into its R is minimum, which is B13 that is 1 % + 0.1 between u(0,1) * (max – min) = 5 % - 1 % = 3.23 %

The P is the profit per sale again, if you look into it is B11 that is 47 + RAND u(0,1) 53 - 47 and I get that value and same way C, if you look into C, it is 0.2 the min value B14 + u(0,1) * 0.8 - 0.2. So, it will give me each one of these values.

So, then I calculate y,

y = y = (L * R *P) that is the first 3 - the fixed value F that is I12 + L * C, J12 that is what this is. So, you get a value here and the average value here is the average of whatever this one thing is.

So, one way to do this is you can see whatever the value changes, the things change here. So, like this. So, I am just going to extend this to let us say some 1000 values. So, the fixed value is 800 already created.

Now, I am going to do is, this going to extend it all the way to 1000. So, these ones are 1000 replications and I am going to calculate Z. So, something like this many values are calculated out. So, here is the profit. So, I am going to now do is calculate the profit also this way right. So, you can see many possible values of profit is already there and you can now see that 715.

So, if I do it for 100 here, it gives you some value. If I do it for 500 gives me another value, if I do it for 1000, it gives me some other value. So, as I increase the number of replications, this average profit it will start behaving.

So, I will go back to the slide again and just to show it to you the average profit, the Y will behave something like this. There is one value, this is number of replications and this is the Y.

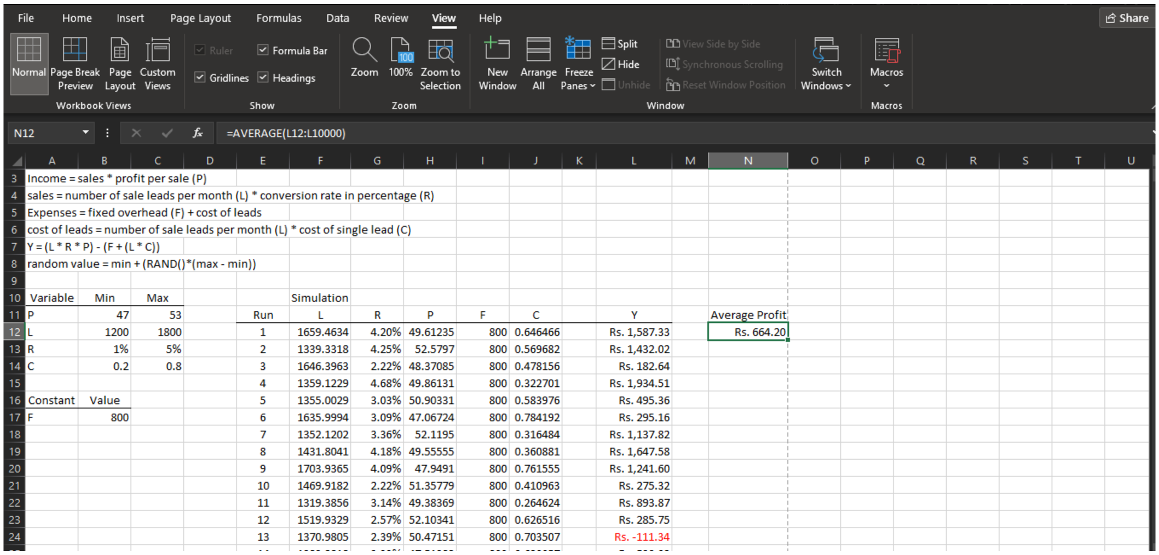
So, it will behave something like this. So, after you need to find out what is that value of replication number where it relatively starts to stabilize out. So, some into like 1000, 10000, 20000 replications, you will get the value stable out and this is that point which where we can say a statistically reliable estimate of Y. So, that is what happens as part of this.

So, if you look into this the excel sheet you can see that as you increase the number of replications. So, let us say I am going to do it as 10000 here and if I increase the number of replications by doing what I just did before by let us take this much and I think if I do this nothing will change yes nothing will not change.

I will do something like this and you can see that now after some point of time this value will start converging around about 686, 650 something like that. So, that is what you can say that the average run, you know this is the profit you would expect out of this organization. So, you have to find out using an excel sheet like this and this approach what we did is.

So, it shows that as the time progresses as you keep on making more and more sales your average profit and the organization will come to 680 whatever number we are getting out of this excel.

So, this is the fundamental of Monte Carlo simulation. So, I will share this excel sheet with you guys and also the slides as part of the presentation that what you just see. So, take a look practice yourself and then based on this Monte Carlo do experiments and then try to create similar problems like this and solve. So, that you can tackle such kind of problems where you do not know the exact values of parameters and variables, but you know the range and you know the distribution and then you can go from there.



-Monte Carlo example in MS Excel

So, thank you very much for your patient hearing and good luck with the rest of the course. I will see you with the next week with more models some AI models and stuff like that and then the rest of the instructors will start introducing you to other aspects of modeling.

Thank you very much.