SYSTEM SIMULATION AND MODELLING

LECTURE 1

Section B

TOPIC COVERED: Monte Carlo Method, Types of System Simulations.

MONTE CARLO METHOD

- Monte Carlo simulation, also known as the Monte Carlo method, originated in the 1940s at Los Alamos National Laboratory. Physicists Stanislaw Ulman, Enrico Fermi, John von Neumann, and Nicholas Metropolis had to perform repeated simulations of their atomic physics models to understand how these models would behave given the large number of uncertain input variable values.

Monte Carlo simulation **is based on the concept of** *repeated random samples of model input variables over many simulation runs.*

When setting up a Monte Carlo simulation or employing the Monte Carlo Method, one follows a four-step process. These four steps are:

Step 1: Define a distribution of possible inputs for each input random variable.

Step 2: Generate inputs randomly from those distributions.

Step 3: Perform a deterministic computation using that set of inputs.

Step 4: Aggregate the results of the individual computations into the final result.

While these steps may seem overly simplistic, they are necessary to capture the essence of how Monte Carlo simulations are set up and run. This four-step method requires having the necessary components in place to achieve the final result. These components may include:

1. probability distribution functions (pdfs) for each random variable
2. a random number generator
3. a sampling rule—a prescription for sampling from the pdfs
4. scoring—a method for combining the results of each run into the final result

5. error estimation—an estimate of the statistical error of the simulation output as a function of the number of simulation runs and other parameters.

Step 1: Requires the modeler to match a statistical distribution to each input random variable. If this distribution is known or sufficient data exist to derive it, then this step is straightforward. However, if the behavior of an input variable is not well understood, then the modeler might have to estimate this distribution based on empirical observation or subject matter expertise. The modeler may also use a uniform distribution if he or she is lacking any specific knowledge of the variable's characteristics. When additional information is gathered to define the variable, then the uniform distribution can be replaced.

Step 2: Requires randomly sampling each input variable's distribution many times to develop a vector of inputs for each variable. Suppose we have two input random variables X and Z. After sampling n times, we have $X = (x_1, x_2, \ldots, x_n)$ and $Z = (z_1, z_2, \ldots, z_n)$. Elements from these vectors are then sequentially chosen as inputs to the function defining the model. The question of how large n should be is an important one because the number of samples determines the power of the output test statistic. As the number of samples increases, the standard deviation of the test statistic decreases. In other words, there is less variance in the output with larger sample sizes. However, the increase in power is not linear with the number of samples. The incremental improvement of power decreases by a factor of about $1n$, so there is a point when more sampling provides little improvement.

Step 3: Is straightforward. It involves sequentially choosing elements from the randomly generated input vectors and computing the value of the output variable or variables until all n outputs are generated for each output variable.

Step 4: Involves aggregating all these outputs. Suppose we have one output variable Y. Then we would have as a result of step 4 an output vector $Y = (y_1, y_2, \ldots, y_n)$. We can then perform a variety of statistical tests on Y to analyze this output.