Statistics and Simulation

By Smita Skrivanek, Principal Statistician, MoreSteam.com LLC

Stochastic Simulation

Stochastic simulation has played an increasingly large role in statistics with rapid increases in computing power. The exercises in the MoreSteam courses are examples of simulation. Simulations are used to model complicated processes, estimate distributions of estimators (using methods such as bootstrap), and have dramatically increased the use of an entire field of statistics, Bayesian statistics, since simulation can be used to estimate the distribution of its statistics, which are usually analytically intractable except in the simplest situations.

So what is 'stochastic simulation'? Stochastic simulation, a.k.a Monte Carlo simulation, is the use of random number generators to produce data from a mathematical model. A random number generator is any process that produces data whose observations are independent and identically distributed (i.i.d) from some distribution. It is a method used to evaluate processes or systems using probabilities based on sampling repeatedly from randomly generated numbers.

Simulation by random sampling was in use during the genesis of probability theory, long before computers. Toss a coin a hundred times and count the number of times it falls Heads. Dividing the total number of times the coin falls Heads by the total number of tosses gives you an estimate of the probability of Heads for that coin, using simulation. The error associated with that estimate is called "Monte Carlo error".

Stochastic simulation was used as far back as in 1777 by Buffon to estimate the probability of a needle falling across a line in a uniform grid of parallel lines, known as Buffon's needle problem. William Gosset (who used the pen-name Student and derived the t-distribution) also used it in the early 1900's for his work on small samples.

The name 'Monte Carlo Simulation' was coined in the 1940's by a group of scientists working on the Manhattan project at Los Alamos National Laboratory who used it to calculate the probability with which a neutron from the fission of one Uranium atom would cause another to fission. The method was named after the casino in Monte Carlo, Monaco and refers to the fact that gambling chances are based on randomness and repetitive sampling, as is Monte Carlo simulation.

Excel can generate random numbers from many distributions, such as the Uniform, Normal, Binomial, etc. The random number generator is available under Data Analysis function in the Tools menu in Excel. The

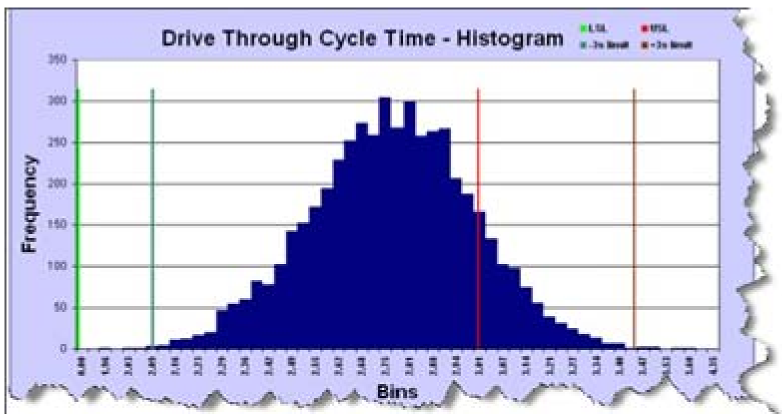
MoreSteam DFSS course includes a toolset on Monte Carlo Analysis and a template for running the simulations, which is also included in our EngineRoom software.

An Example of Monte Carlo Simulation

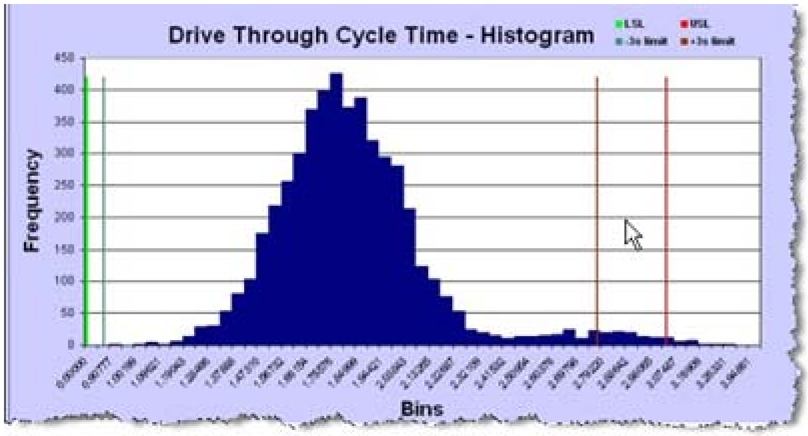
Here's an example of how Monte Carlo simulation helped minimize wait time at a drive-through: a national fast food chain wanted to reduce the amount of time customers spend in the drive-through by limiting the wait to under three minutes at least 90% of the time. The dependent variable drive-through cycle time has four individual components/process steps: Prepare, Cook, Package and Deliver. The drive-through cycle time ($T_{\text{Drive-through}}$) is modeled as the sum of the cycle times of each of the individual steps:

$$T_{\text{drive-through}} = T_{\text{Prepare}} + T_{\text{Cook}} + T_{\text{Package}} + T_{\text{Deliver}}.$$

Using historical data, the underlying distributions for the individual process steps were found to be normal. The normal distribution parameters (means and variances) along with the specification for $T_{\text{drive-through}}$ (upper specification = 3 minutes) were entered into the Monte Carlo template to yield the following histogram:



The red vertical line represents the upper specification of 3 minutes. In the current process only 87% of the wait times are below 3 minutes. Further analysis was done and a breakthrough was achieved when it was found that using burgers that were precooked rather than cooked-to-order drastically reduced the wait time. The improvement was modeled into the simulation and the results confirmed that the wait time was now under 3 minutes 99.4% of the time. The histogram for the new, improved process is shown below:



Summary

This was a simple demonstration of Monte Carlo simulation, to give you an idea of how to use this tool. A primary benefit of this method is that an improvement can be evaluated without disrupting the actual process. Different scenarios can thus be compared and the one resulting in the best results applied to the production process. At the same time, a few things to keep in mind are:

1. Assigning an incorrect distribution or parameters to an input will lead to incorrect predictions.
2. The method assumes that the inputs are independent – this may not be valid.

Still, these are minor drawbacks in a method that is amazingly versatile and powerful for assessing uncertainty in the output of complex stochastic models.