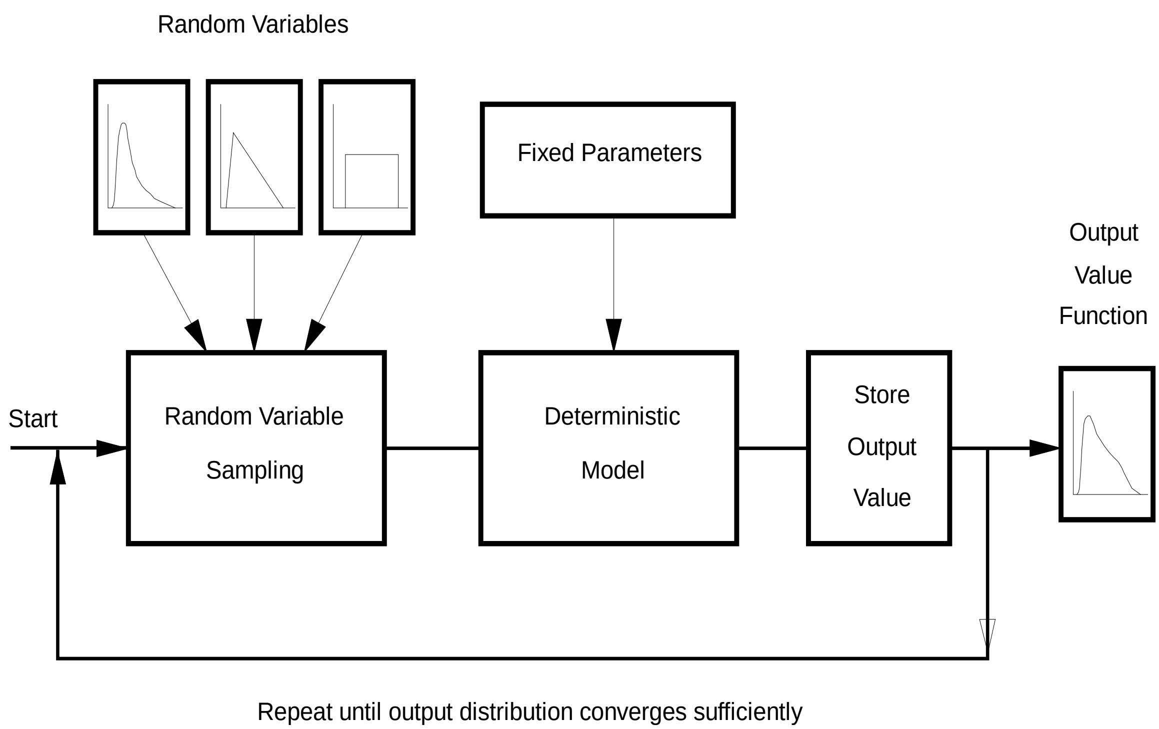
5. SIMULATION

Monte Carlo simulation is an extension of deterministic modeling.[1] Simulation, like decision trees, allows judgments about risks and uncertainties to be incorporated explicitly into the analysis. Probability distributions represent these judgements in a convenient form. Unfortunately, mathematical operations with probability distributions are notoriously difficult. Simulation provides an elegant and simple way to solve forecasting models where some of the parameter inputs are expressed as probability distributions.

The simulation process is conceptually quite simple. Key judgment variables in the conventional, deterministic business model are identified, often with sensitivity analysis. These are the variables that can cause large changes in the output value, either because input values range widely or because of the model's extreme sensitivity. Input variables in the model that can assume a range of values are called stochastic or ***random variables***. The remaining are called fixed or deterministic variables.

The figure shows how a simulation analysis is structured:



[1] Monte Carlo simulation is what I mean by “simulation.” However, not all simulation models are stochastic (e.g., the “black oil simulator” familiar to reservoir engineers).

random sampling

In a simulation analysis, each random variable is represented by a probability distribution. During a simulation run, these probability distributions are “randomly” sampled to determine trial values for the variables. They are called *random samples* because a random number generator drives the sampling process for obtaining trial values. This is the key to the simulation process. Though picked “randomly,” the frequency distributions of sample values closely match the input probability distributions.

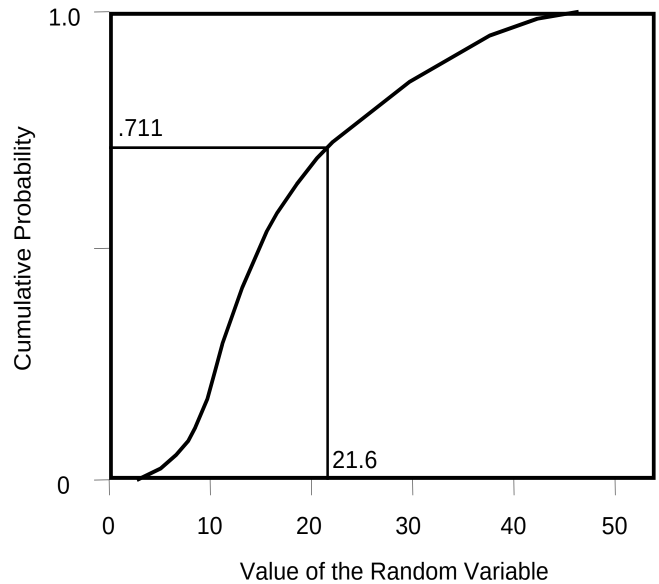
How Sampling Works:

The sample mean $\bar{x}$ approximates the true population mean (μ or EV).

Here’s how it works. We want a sample value for each random variable for each trial in the simulation run. On a particular trial, the input variable trial values are substituted into the cashflow model. PV and other desired outcome results are saved. After performing many such trials, we analyze the distributions of the outcome variables. ***Frequency histograms*** (bar graphs) approximate the probability density distributions. The mean (arithmetic average) of many trials is approximates the expected value; the more trials, the more accurate the approximation. This elegantly simple technique allows us to solve what would otherwise be very difficult or impossible integration problems for expected values.

In conventional Monte Carlo simulation, independent random variables are sampled with replacement. That is, each sampling is independent of the others. The figure on the next page illustrates how a sample is obtained from a probability distribution.

A random number function generates a value between 0 and 1. Conveniently, 0–1 is the probability range on a cumulative probability curve. Enter the graph at a cumulative probability (on the y-axis) equal to the sampling random number. The intersection of a line drawn across to the cumulative probability curve provides the sample value (x-axis) for this trial. On the example shown on the graph, the random number was 0.711. This corresponds to a value of 21.6 for the random variable. The sampling process is repeated for the other random variables in the model. These sample values are used during one trial's calculation pass through the deterministic model.



The sampling-and-calculation process is repeated for many trials. The number of trials is sufficient when the resulting frequency distribution adequately defines the model's true solution probability distribution. Often, we are interested primarily in the expected value, and a few hundred trials are usually sufficient to assure reasonable convergence.

LHS and other Techniques

stopping rule

seed

The standard error of the mean ($\sigma_{\bar{x}}$) statistic can be used to determined the confidence of the expected value estimate.[2] This is a common basis for a ***stopping rule***. When comparing alternatives, using the same ***seed*** value to start the random number generator will reduce the number of trials needed for adequate convergence.[3]

Latin hypercube sampling

A variant of the sampling process described above is Latin Hypercube Sampling (LHS). This fancy-sounding technique is simply a hybrid between *uniform sampling* and conventional *Monte Carlo sampling*. LHS provides faster convergence and, thus, is a very useful time-saver. LHS works by sampling in two steps. The probability range (y-axis) is divided into equally probable sections, like layers of a cake. The first step is to select a section at random, and this is done sampling without replacement. In the second step, the model randomly determines a probability point within the segment. The sections are recycled after all have been used. LHS ensures that all segments along the probability curve are equally sampled.

---

[2] The standard error $\sigma_{\bar{x}} \cong$. With careful use of Latin hypercube sampling, the actual error is typically 1/3 the value indicated from this formula.

[3] Technically it is a *pseudo-random number*, since the calculation is a formula.

Simulation can be thought of as a way to perform many sensitivity analyses automatically. The technique offers additional powerful features:

- It provides an easy way to solve real world problems that are too complex to solve mathematically.
- The outputs are probability distributions showing the nature of the risks and uncertainties.
- More variables can be modeled, and more completely, than is practical with decision tree analysis.

Simulation and decision tree analysis provide a more accurate answer than obtained with a deterministic model. There is often a great error with a deterministic model's single-point result due to correlations and to unequal effects of favorable and unfavorable deviations.