Monte Carlo simulation in MS Excel

The Monte Carlo method is based on the generation of multiple trials to determine the expected value of a random variable. The basis of the method is provided by the following relationship:

$$\Pr\left\{\left|\frac{1}{N}\sum_{N}\xi - \mu\right| < \frac{3\sigma}{\sqrt{N}}\right\} \approx 99.8\%$$

There are a number of commercial packages that run Monte Carlo simulation, however a basic spreadsheet program can be used to run a simulation. In this case the generation of multiple trials is implemented by propagating a basic formula as many times as the number of iterations required by the model.

Lets assume that a project has six activities. Each activity has a total cost in a specified range. In some cases the value is fixed (Activity B), but generally it can assume any value within an interval. Each of these variables can have a unique distribution, but as we will see in a moment, we can safely assume a uniform distribution without compromising the result.

One important assumption that we make is that each of these variables is independent of the others. This means that the cost of any activity is not influenced by the cost of any other activity.

Activity	Minimum	Maximum
A	10,000	20,000
B	15,000	15,000
C	7,500	12,000
D	4,800	6,200
E	20,000	25,000
F	5,000	7,000
Total	62,300	85,200

The total cost of the project is a random variable with a value between the minimum and the maximum. This variable will be normally distributed since it is the sum of a number of random variables. This is also the reason why the individual distribution of each variable is not important.

The general scheme of the Monte Carlo method is as follows:

- Generate random values for each of the activity costs
- Add each series of random values to arrive at a total project cost.
- The expected project cost is the average of these values.

There are a number of parameters that can be calculated to assess the goodness of the solution. We will discuss some of these parameters later on.

The first step is to generate random values for each of the activity costs. Assuming a uniform distribution, we can use the RAND() function to generate random numbers in the interval (0,1) and multiply these by the range of each variable. The range is the difference between the maximum value and the minimum value.

The random cost for Activity A will look like this:

    =RAND()*(20,000-10,000)+10,000

This formula generates a random value between 10,000 and 20,000

If we create one of these formulas for each of the activities, the total project cost will be the sum of all these functions. We can propagate the row of formulas for each iteration of the model.

The following table is a sample of the model showing the first seven iterations. Row 4 is the basic model, which is propagated as required.

	A	B	C	D	E	F	G	H
1	**Activity**	**A**	**B**	**C**	**D**	**E**	**F**	**Total**
2	**Minimum**	10,000	15,000	7,500	4,800	20,000	5,000	62,300
3	**Maximum**	20,000	15,000	12,000	6,200	25,000	7,000	85,200
4		14,167	15,000	7,773	5,682	20,788	5,837	69,247
5		15,556	15,000	10,456	4,943	20,957	6,239	73,148
6		19,970	=RAND()*(A$3-A$2)+A$2		5,136	24,853	6,474	81,149
7		16,811	15,000	11,405	5,252	=SUM(B4:G4)		76,467
8		14,783	15,000	10,030	4,870		45	71,898
9		18,002	15,000	8,435	4,808	20,890	5,211	72,346
10		17,628	15,000	9,808	5,112	22,612	5,138	75,297

Determining the number of iterations

The Monte Carlo method provides an estimate of the expected value of a random variable and also predicts the estimation error, which is proportional to the number of iterations.

The total error is given by: $\varepsilon = \dfrac{3\sigma}{\sqrt{N}}$ where b is the standard deviation of the random variable, and N is the number of iterations. We can estimate an upper bound of b by calculating the standard deviation between the maximum, the minimum and average values of the random variable:

σ = STDEVP(H2:H3,AVERAGE(H2:H3)) = 9,349

Note that we used the function STDEVP, which calculates the standard deviation of the entire population, in this case only 2 values.

Lets determine the number of iterations required for an error of less than 2%. A gross estimation of the random variable is the average of the maximum value and the minimum value. An absolute error of 2% is this average divided by 50:

ε = AVERAGE(H2:H3)/50 = 1,475

Therefore the number of iterations to obtain a result with an error of less than 2% is:

$$N = \left(\frac{3 \times 9,349}{1,475} \right)^2 = 362$$

The Model is ran by propagating line 4 a total of 362 times. The expected value of the random variable is the average of the Total column:

Expected project cost = AVERAGE(H4:H547) = 73,428

Given that the variable is normally distributed, the median should be very close to the mean

MEDIAN(H4:H547) = 73,304. A difference of only 0.2%

The standard deviation is calculated over the entire population:

STDEVP(H4:H547)=3,604

We can now review the true error of the estimate

$$\varepsilon = \frac{3\sigma}{\sqrt{N}} = \frac{3 \times 3,604}{\sqrt{362}} = 568.3$$

This is an error of only 0.8%.

Other useful information is the Kurtosis and the Skewness of the distribution. The Kurtosis is a relative measure of the shape compared with the shape of a normal distribution. The normal distribution has a Kurtosis of 0.

=KURT(H4:H547) = -0.296

This indicates that the distribution is somewhat flatter than a normal distribution.

Skewness is a measure of asymmetry. The normal distribution has a skewness of 0.

=SKEW(H4:H547) = 0.061

This indicates that the tail of the distribution extends towards the right.

The results can be easily plotted to produce the following chart:

Frequency/Cumulative Chart
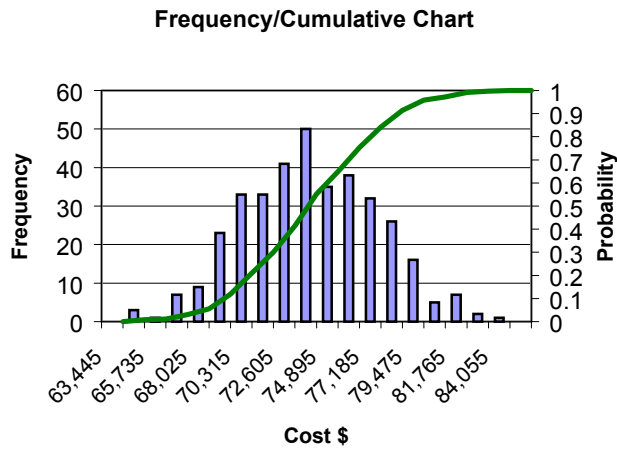
© Rob Jeges
www.projectware.com.au

Microsoft Project (MS Project) is a trademark of the Microsoft Corporation.