# MONTE CARLO RISK AND DECISION ANALYSIS


3/17/2013

## INTRODUCTION

The risk generated and the ensuing economic impact of incorrect values used in decision-making models is incalculable. The models could be correct, but are usually fed average values as input data assuming linearity of the models. However, most models upon which business, government, the military and individuals base their decisions are nonlinear, leading to erroneous decisions. Reaching management decisions based on assumed linear models, when reality mandates the use the nonlinear models is called the "Flaw of Averages," or the Jensen's inequality. Monte Carlo simulations are inherently capable of dealing with nonlinear problems, and are considered a valued tool in non-linear decision-making. Since in management, spreadsheets are normally used, a spreadsheet Monte Carlo simulation demonstrates the Flaw of Averages concept, suggesting remediation.

## MONTE CARLO DECISION SUPPORT AND FORECASTING

Monte Carlo offers a future solution of what Sam Savage from Princeton calls the "Flaw of Averages" problem in Decision Support systems and Forecasting. Mean values that are used in today's decision support systems and forecasting can be very misleading. What engineers, scientists and economists have considered as exact until fairly recently are now recognized as mere mean values. In the real world and nature, the laws of probability and possibility theories prevail.

Consider the case of placing one foot in a bucket of hot water at 100 degrees, and the other foot in cold water at 0 degrees. The mean value of 50 degrees should keep a person very comfortable. But the real situation is doubly uncomfortable from cold and heat.

Taking the case of an engineer or a product manager who is asked by their plant manager to forecast the yearly demand for a newly designed microprocessor. He may come up with a forecast for the yearly demand in the number of processors in a range of:

$$[50{,}000\ ,\ 150{,}000].$$

The plant manager uses the average of the forecast:

$$\mu = \frac{50{,}000 + 150{,}000}{2} = 100{,}000\,,$$

to build a facility producing 100,000 processors per year.

He reports, based on this average, and an expected profit of 100 dollars per processor to his board of directors an average expected yearly profit of:

$$p = 100\mu = 10x10^6 \ \ [\frac{dollars}{year}].$$

Assuming that demand is the only uncertainty, this estimate of the expected yearly profit is not the best guess for average profit.

The Flaw of Averages ensures here that the average profit has to be *less* than the profit associated with the predicted average demand. If demand is less than 100,000 units, the profits will be *lower* than 10 million dollars, but the profits can *never* be higher than 10 million dollars, because the maximum capacity of the plant is based on a flawed average. As a result, the plant manager's forecast of average demand leads to an inflated forecast of average profit. What he came out with is more like an estimate of maximum possible profit rather than an average expected profit.

Even when given good data, people make bad decisions. They misunderstand, misinterpret and mismanage important problems. The cause is that our thought processes have bugs in them. To fix those bugs, decision support software latest incarnation generates thousands of scenarios covering real world contingencies. Software products such as XL Sim and @risk allow people to test their assumptions through random sampling and Monte Carlo simulations, turning spreadsheets into wind tunnels for testing hypotheses and designing good decisions.

To test their intuitive assumptions about average returns and average losses, executives managing plant capacity, investors with stock portfolios, employees with retirements funds, would be running spreadsheet Monte Carlo simulations.

## THE FLAW OF AVERAGES: JENSEN'S INEQUALITY

The Flaw of Averages concept is commonly expressed in terms of inputs and outputs to a system as:

*"Average inputs do not necessarily yield average outputs."*

,
In terms of demand and cost in inventory analysis it is expressed as:

*"Average demand does not necessarily yield average cost."*

If we consider a model of a system as:

$$y = f(x) \tag{1}$$

the average value, first moment, mean value, mathematical expectation, or just expected value is:

$$\bar{x} = E[x] = \frac{\int xf(x)dx}{\int f(x)dx} \tag{2}$$

where the denominator normalizes the function f(x) and turns it into a probability density function.

For a model f(x), we can write that the mean value or expected value of f(x) is:

$$E[f(x)] = \frac{\int f(x) f(x)dx}{\int f(x)dx} = \frac{\int f^2(x)dx}{\int f(x)dx} \tag{3}$$

We can also write according to Eqn. 2:

$$f([E(x)] = f(\frac{\int xf(x)dx}{\int f(x)dx}), \tag{4}$$

which may or may not be equal to the RHS of Eqn. 3, depending whether the model f(x) is linear or nonlinear.

To study the effect of model linearity or nonlinearity on Eqns. 3, and 4, we consider the two following cases.

## LINEAR MODELING

We here consider for the model f(x), the equation of a straight line with slope m defined over the unit interval:

$$f(x) = mx, \quad x \in [0,1] \tag{5}$$

According to Eqn. 2, the expected value of the parameter x as a cumulative distribution function is:

$$E[x] = \frac{\int_0^x x(mx)dx}{\int_0^1 mxdx} = \frac{\frac{x^3}{3}}{\left[\frac{x^2}{2}\right]_0^1} = \frac{2x^3}{3} \tag{6}$$

According to Eqn. 4:

$$f([E(x)] = m\frac{2x^3}{3} = \frac{2}{3}mx^3 \tag{7}$$

The expected value of the model f(x) is according to Eqn. 3:

$$E[f(x)] = \frac{\int_0^x (mx)^2 dx}{\int_0^1 mxdx} = \frac{\frac{mx^3}{3}}{\left[\frac{x^2}{2}\right]_0^1} = \frac{2}{3}mx^3 \qquad (8$$

Comparing Eqns. 7 and 8, we can deduce that for the linear system considered:

$$E[f(x)] = f(E[x]) \qquad (9)$$

Thus, for a linear model f(x), average inputs E[x] yield average outputs E[f(x)]=f(E[x]). In fact that condition is weaker than it appears, because it applies here to a straight line going through the origin, and not for a straight line that intersects the y-axis. Thus we can say that it applies under the condition of proportionality rather than linearity.

As a simple numerical example, we consider the linear model represented by the straight line f(x)=x through the origin with unit slope, defined over the unit interval. We can write:

$$E[x] = \frac{\int_0^1 xf(x)dx}{\int_0^1 f(x)dx} = \frac{\int_0^1 x^2 dx}{\int_0^1 xdx} = \frac{2}{3}$$

$$f(E[x]) = \frac{2}{3}$$

$$E[(f(x)] = \frac{\int_0^1 f(x)f(x)dx}{\int_0^1 f(x)dx} = \frac{\int_0^1 x^2 dx}{\int_0^1 xdx} = \frac{2}{3}$$

Thus: $$E[f(x)] = f(E[x]) = \frac{2}{3}$$

The same result could be obtained from Eqns. 7 and 8, by substituting x=m=1.

## NONLINEAR MODELING

We here consider for the model f(x), the equation of a parabola defined over the unit interval:

$$f(x) = bx^2,\ x \in [0,1] \tag{10}$$

According to Eqn. 2, the expected value of the parameter x as a cumulative distribution function is:

$$E[x] = \frac{\int_0^x xbx^2 dx}{\int_0^1 bx^2 dx} = \frac{\frac{x^4}{4}}{\left[\frac{x^3}{3}\right]_0^1} = \frac{3x^4}{4} \tag{11}$$

According to Eqn. 4:

$$f([E(x)] = b\left(\frac{3x^4}{4}\right)^2 = \frac{9}{16}bx^8 \tag{12}$$

The expected value of the model f(x) is according to Eqn. 3:

$$E[f(x)] = \frac{\int_0^x (bx^2)^2 dx}{\int_0^1 bx^2 dx} = \frac{\frac{bx^5}{5}}{\left[\frac{x^3}{3}\right]_0^1} = \frac{3}{5}bx^5 \tag{13}$$

Comparing Eqns. 12 and 13, we can deduce that for the linear system considered the Jensen's inequality applies:

$$E[f(x)] \neq f(E[x]) \tag{14}$$

Thus, for a nonlinear model f(x), average inputs E[x] do not yield average outputs.

As a simple numerical example, we consider the linear model represented by the parabola $f(x)=x^2$ through the origin with unit slope, defined over the unit interval. We can write:

$$E[x] = \frac{\int_0^1 xf(x)dx}{\int_0^1 f(x)dx} = \frac{\int_0^1 x^3 dx}{\int_0^1 x^2 dx} = \frac{3}{4}$$

$$f(E[x]) = \left(\frac{3}{4}\right)^2 = \frac{9}{16}$$

$$E[(f(x)] = \frac{\int_0^1 f(x)f(x)dx}{\int_0^1 f(x)dx} = \frac{\int_0^1 x^4 dx}{\int_0^1 x^2 dx} = \frac{3}{5}$$

Thus we get the Jensen's inequality:

$$E[f(x)] \neq f(E[x])$$

The same result could be obtained from Eqns. 12 and 13, by substituting x=b=1.

## SPREADSHEET MONTE CARLO SIMULATIONS

In financial management and forecasting the use of spreadsheets is a favored method of problem solving. Monte Carlo can be used within the context of spreadsheet calculations as we attempt to use it here.

We use a stochastic cost model suggested by Sam Savage as follows. A supply firm has an average random demand for a piece of equipment such as a lawn mower. To satisfy this demand, the firm must either:

1. Stock a sufficient number of this piece of equipment and incur the associated modest storage cost, in this case its available supply satisfies the demand, or:
2. Incur a large shipping cost by airfreight every time the demand exceeds the available supply.

Balancing the stored supply with its associated small storage cost against the demand is key to achieving profits, since the shipping of these units would cost more than the storage cost reducing profits.

To achieve this balancing of supply and demand, the firm builds a spreadsheet inventory model as shown in Fig. 1. The model is governed by two simple relationships:

1. It the demand is larger than the number available on stock, the firm has to incur an airfreight cost and promptly ship the unit to its customer.
2. If the demand is less than the number available on stock, the firm will incur a storage cost, that is nevertheless smaller than the shipping cost on each unit in its supply that exceeds the demand.

Matching the stored supply to the expected demand becomes a key to achieving profit for this firm.

In the spreadsheet model, the storage cost is taken as $50 per unit per month, and the air freight shipping cost is taken as $100 per unit per month. Expecting an *average* demand of 5 units per month, the firm stocks 5 units as a supply. Entering these two numbers in the spreadsheet of Fig. 1, the average cost appears as zero, which would maximize profits. So everything looks fine.

**Inventory Model**

| | | |
|---|---|---|
| **Demand** | | **Supply** |
| 5 | | 5 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 0 |
| **Shipping Cost** | $100.00 | 0 |
| **Total Cost** | | 0 |

**Governing Relations**

IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=5

C5=5

Figure 1. Spreadsheet for linear inventory model, with input value = 5. Supply equals Demand, A5 = C5.

**Inventory Model**

| | | |
|---|---|---|
| **Demand** | | **Supply** |
| 10 | | 5 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 0 |
| **Shipping Cost** | $100.00 | 500 |
| **Total Cost** | | 500 |

**Governing Relations**

IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=10

C5=5

Figure 2. Spreadsheet for nonlinear inventory model, with linear input with mean = 10. Demand exceeds Supply, A5 > C5.

**Inventory Model**

| | | |
|---|---|---|
| **Demand** | | **Supply** |
| 5 | | 10 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 50 |
| **Shipping Cost** | $100.00 | 0 |
| | | |
| **Total Cost** | | 50 |

**Governing Relations**

IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=5

C5=10

Figure 3. Spreadsheet for nonlinear inventory model, with linear input with mean=10. Supply exceeds Demand = 5, C5 > A5.

**Monte Carlo Inventory Model**

| | | |
|---|---|---|
| **Demand** | | **Supply** |
| 1 | | 5 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 40 |
| **Shipping Cost** | $100.00 | 0 |
| **Total Cost** | | 40 |

**Governing Relations**

```
IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=NORMDIST(200,5,2,TRUE)

C5=5
```

Figure 4. Spreadsheet for nonlinear inventory model, with Normal Distribution input with mean = 5.

**Monte Carlo Inventory Model**

| | | |
|---|---|---|
| **Demand** | | **Supply** |
| 0.559285809 | | 5 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 44.40714191 |
| **Shipping Cost** | $100.00 | 0 |
| **Total Cost** | | 44.40714191 |

**Governing Relations**

```
IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=LOGNORMDIST(200,5,2)

C5=5
```

Figure 5. Spreadsheet for nonlinear inventory model, with LogNormal Distribution input with median = 5.

**Monte Carlo Inventory Model**

| | | |
|---|---|---|
| **Demand** | | **Supply** |
| 1 | | 5 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 40 |
| **Shipping Cost** | $100.00 | 0 |
| **Total Cost** | | 40 |

**Governing Relations**

```
IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=POISSON(100,5,TRUE)

C5=5
```

Figure 6. Spreadsheet for nonlinear inventory model, with Poisson distribution input with mean = variance = 5.

**Monte Carlo Inventory Model**

| | | |
|---|---|---|
| **Demand** | | **Supply** |
| 1 | | 5 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 40 |
| **Shipping Cost** | $100.00 | 0 |
| **Total Cost** | | 40 |

**Governing Relations**

IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=POISSON(100,10,TRUE)

C5=5

Figure 7. Spreadsheet for nonlinear inventory model, with Poisson distribution input with mean = variance = 10. Supply = 5.

**Monte Carlo Inventory Model**

| Demand | | Supply |
|---|---|---|
| 1 | | 10 |
| | **[$/Item]** | **Storage, Shipping Costs** |
| **Storage Cost** | $10.00 | 90 |
| **Shipping Cost** | $100.00 | 0 |
| **Total Cost** | | 90 |

**Governing Relations**

IF(Supply>Demand,+(Supply-Demand)*Storage_Cost,0)
C8=IF(C5>A5,+(C5-A5)*B8,0)

IF(Supply<Demand,+(Demand-Supply)*Shipping_Cost,0)
C9=IF(C5<A5,+(A5-C5)*B9,0)

Total Cost = Storage_Cost+Shipping_Cost
C11=C8+C9

A5=POISSON(100,5,TRUE)

C5=10

Figure 8. Spreadsheet for nonlinear inventory model, with Poisson distribution input with mean = variance = 5. Supply = 10.

The fact of life is that demand could exceed supply and supply could exceed demand. If the demand at 10 units exceeds the available supply at 5 units, as shown in Fig. 2, the shipping cost becomes $500. If on the other hand the supply is kept at 10 units whilst the demand is only 5 units, as shown in Fig. 3, the storage cost becomes $25. The situation here is that a linear relationship between input and output in the model is assumed. A more realistic simulation would consider a nonlinear model where the demand is represented as a distribution with mean value equal to the average value of 5.

We now consider more realistic input models in Figs. 4-8. In this case Monte Carlo simulations are used to sample probability density functions that represent the actual nature of the demand side as not just an average value. Instead we consider a normal distribution with mean 5 if Fig. 4, a lognormal distribution with median 5 in Fig. 5, and a Poisson distribution with mean and variance of 5 in Fig. 6. In all cases, even though the mean value of 5 was equal to the supply value of 5, the total cost always exceeded the linear model cost, supporting the flaw of average concept.

Unequal value of supply and demand are shown for the Poisson distribution in Figs. 7 and 8, further supporting the need for a Monte Carlo simulation approach for avoiding the flaw of averages where average demand does not necessarily result in average costs. The only time that average inputs are guaranteed to result in average outputs is when the model is linear in all the uncertain variables. The best advice is to avoid average inputs or point estimates.

## MONTE CARLO PERVASIVE COMPUTING

The Flaw of Averages also applies to Pervasive Computing. This is a future where computers are no more residing on a desktop or at a computer center, but are imbedded in the environment including workplaces, transportation means, and even clothing. For instance, a computer could control the temperature of clothing depending on the ambient environment. A computer could adjust the settings of light, communications gear, and climate control in an office, car or plane, depending on the present user of the facility.

At the MIT Sloan School of Management, Dan Ariely is investigating an "electronic wallet." It would advise a user on, for instance, how to spend their money. Such a device would run a Monte Carlo simulation displaying how, based on a person's payment behavior, he would save a sum of $200 over the next few months by using one credit card or the other, check, or cash.

According to Michael Schrage: "… tomorrow's technologies will load the dice in favor of people not repeating the sort of silly statistical mistakes that lead to Nobel Prize winning research. And that will merit a prize of its own."

## LIMITATIONS OF MONTE CARLO SIMULATIONS

Simulations could be misleading if the wrong assumptions are used as inputs to the simulations. If best and worst case returns and rates of inflation are used as inputs, the output results would depend on the chosen inputs.

The numbers of simulations that are carried out are an area of contention. Some think that 500 simulations are adequate, while others advocate the use of 3,000 to 10,000 simulations and still others suggest the use of hundred of thousands of simulations.

Critics on the other hand suggest that running a large number of simulations feeds in too many extreme cases.

Other critics argue that it is not possible to accurately replicate the way markets behave. Markets are suggested to be lumpy and are driven by humans' emotions that drive bull and bear markets in clusters. The incorporation of these extra uncertainty factor would lead to more reliable simulation results.

## PROBLEMS

1. Consider a linear input in the form of a straight line:

$$f(x) = a + mx$$

test whether or not such a linear model would yield:

$$E[f(x)] = f(E[x])$$

2. Instead of a Monte Carlo simulation using a spreadsheet approach, write a high level language procedure to carry the simulation and display the generated probability distributions.