**SIMULATION**

We are all aware of simulation models in our daily lives. An aircraft manufacturing firm commonly builds simulation models of their proposed jet aircraft and then tests the aerodynamic properties using these abstract models rather than experimenting with the real aircraft or even with an expensive physical model. A flight simulator may be used to train pilots rather than risking their lives in training or the cost of the actual airplane. This might be a physical simulator as used by commercial airlines and the US Air Force or it could be a computer simulation such as a computer game. The US army might simulate an enemy attack in order to study the best defense strategy.

Simulation as a management tool tries to create an abstract, idealized representation of the real system under study, generally using mathematical variables, equations and logical relationships which can imitate the behavior of the real system rather closely under various input conditions. A simulation model has two kinds of inputs—controllable inputs that the analyst selects and inputs representing uncontrollable factors which may often also be only statistically known *i.e.* by specifying probability distribution. In conducting a simulation experiment the analyst selects the controllable inputs (decisions) while uncontrollable probabilistic inputs are generated randomly. The simulation model uses these inputs, the equations and the logical relationships of the model to compute the values of the outputs (performance criteria such as cost, revenue, etc.). Conducting a series of experiments using a variety of controllable variable values (strategies) the analyst learns how the values of the controllable variables affect the values of the outputs—various measures of performance. After studying the simulation results, the analyst will be able to recommend decisions for the controllable variables that will provide the desired outputs for the real system.

In the context of quantitative analysis, simulation means experimentation with an abstract (often mathematical) representation of a real system-- a model. Although direct optimization (such as linear or non-linear optimization) or direct analysis methods (such as queuing theory) may be available, in many cases the problem may be of such complexity that such direct approaches may not be applicable. Consider a linear programming optimization where the objective function coefficients of the variables are only known statistically—as probability distributions. Linear programming using the mean values of these distributions will give a solution which is optimal “on average,” but will not be able to assess the risks associated with the “optimal” solution stemming from the uncertainty.

**Simulation versus “direct” optimization**:

* In direct optimization the values of the decision variables are *outputs*. The model computes a set of values for the decision variables that maximize of minimize (optimize) the value of the objective.
* In a simulation model the decision variables are *inputs* to the experiment. The model evaluates the objective function for a particular set of decision values.

To see what this means, consider the following example from the so-called queuing theory. Suppose that a supermarket wants to decide how to assign checkout personnel during the weekend. The goal is to minimize labor cost, subject to the restrictions imposed by the labor contract and the policy constraint that customers should not have to wait too long. If we had a direct optimization model, we would need to supply the model with the appropriate parameters. Perhaps these would be quantities such as the description of the random arrival patterns of the customers, the probability distribution of the time it takes to check out a customer with and without the use of a bagger, and so on. When the model, if one existed, is solved, the answer would include the best way to assign personnel, (perhaps the number of check-out lines to operate and whether or not to use a bagger) and the cor­responding value of the objective function (the total cost) as well as other performance criteria such as average time customers wait in line, average line length or the probability that someone waits more than say 15 minutes.

In a simulation model, the inputs would include the parameters we described above (customer arrival rates and the like), an expression for the objective function (total costs), and *a possible assignment of personnel* (the number of check-out lines to operate and whether there is a bagger or not). The model would produce a specific set of results showing how the system performed by various measures, such as total cost, customer waiting time, staff utilization, and so on. In general, unlike the direct models, a simulation experiment assesses the quality of the suggested solution as well as how much variability there might be in the various performance measures due to randomness in some of the inputs. Simulation allows for a lot of experimentation and interac­tion with the modeler, but it will not necessarily optimize the goal of interest. The analyst will have to use the simulation model judiciously to search by trial and error for a satisfactory (not necessarily a proven optimal) solution. The simulator is usually a much cheaper and faster way to experiment with many factors of interest.

When should simulation be used then? From this brief description, it seems as if no one would ever want to use a simulation model. Why not use a model that always yields the best answer, that is, an optimization model? Indeed, in the past simulation was often seen as the technique of last resort, to be used only when analytical (optimization) methods failed. It is true that if a realistic enough, analytical direct model is available, exact optimal results can be obtained quickly. Nowadays simulation modeling is one of the most frequently used tools of quantitative analysis. Why are simulation models so popular?

1. First, analytical models may be difficult or impossible to obtain, depending on com­plicating factors. What a complicating factor is depends on the specific model. Complicating factors for capital budgeting models might include demand uncertainty that direct methods may not be able to handle. Complicating factors for queuing models such as the line you queue up to get a flu shot, random arrivals may not follow certain distributional forms that the direct method may assume. “Balking” (students finding the line to long and not joining) or “reneging” (some getting tired of waiting and leaving the queue) are other realistic queuing system features that most analytical direct queuing models can not handle easily or not all.
2. Analytical models typically predict only average or "steady-state" (long-run) behavior. In real-world models, however, it is often as important to understand the possible variability in the performance measures, or how the performance measures vary in the short run.
3. Simulation can be performed with a great variety of software, from spreadsheets alone to spreadsheet add-ins (Crystal Ball, @Risk) to general computer programming languages (Fortran, Basic, C++) to special-purpose simulation languages: (SIMAN, GPSS, SIMSCRIPT). As simulation models can now be created and run on a PC or a worksta­tion, the level of computing and mathematical skill required to design and run a useful simulation experiment has been substantially reduced. It is now quite reasonable to build and use a simulator even when it is clear that an analytic model could be constructed with more time and effort.

In summary, the ability of simulation models to deal with complexity, to capture various idiosyncrasies of the real world, to display the variability of performance measures, and reproduce short-run system behavior make simulation a powerful tool.

**Simulation of random behaviors (Monte Carlo Method)**

Simulation is often used to analyze decisions under risk where the behavior of one or more of the factors is not known with certainty but only probabilistically. Such factors are referred to as random variables (such as demand for a product next month) and are associated with probability distributions. This type of simulation which tries to explicitly mimic the random behavior of the system is called Monte Carlo Simulation.

 **Generating Random Events**

Suppose one is trying to do a shot-run cash budget to evaluate various financing options. The main source of cash is current sales which is a function of uncertain actual demand. Suppose the monthly demand in the last 100 months has been as in the following frequency distribution.

|  |  |  |  |
| --- | --- | --- | --- |
| Units (x) | Frequency | Probability | Cumulative Probability  |
| 800 | 27 | .27 | .27 |
| 900 | 33 | .33 | .60 |
| 1000 | 15 | .15 | .75 |
| 1100 | 15 | .15 | .90 |
| 1200 | 10 | .10 | 1.00 |
|  | 100 | 1.00 |  |

In other words, in 27 months out of 100, demand has been 800 units, in 33 months it has been 900 units etc. In this case the random behavior of the system pertains to the number of units demanded. To experiment (to try out a given cash management policy) we would like to pretend demand is occurring as per the above distribution. To pretend as if the system is actually operating with a certain cash management policy being tried out (such as taking a chance and trying to get by with rather expensive short term borrowing if and when shortages occur versus borrowing at the beginning of the planning horizon a lump sum at a more favorable rate which may or may not be needed) we will have to generate simulated demand for some number of periods called the simulation length. To present a realistic picture of cash balance scenarios likely to be encountered the demands that we generate should obey the frequencies in the above table.

Suppose we have a spinning wheel divided into 100 equal parts. We could mark the first 27 parts “demand 800” next 33 parts as “demand 900” etc. and spin the wheel, whatever demand number the wheel stops at may be taken as the simulated demand for that month. Spinning the wheel, repeatedly one could generate random pretend demand for a number of months. However, this physical device is awkward to use. Instead, we can rely on “random numbers” that are either available in random number tables or provided by many computer programming environments such as Excel. For instance the Excel function =RAND() returns uniform values between 0 and 1 (that any value in the interval has equal probability of coming up). To use random numbers to generate a demand series we can partition the range between 0 and 1 in a way that the widths of the partitions corresponds to the relative frequency of the demand events, For instance 0 - .2700000 is assigned to demand of 800; .2700001 to .6000000 is assigned to demand of 900; .6000001 to .7500000 to a demand level of 1000; .7500001 - .900000 to 1100 and .9000001 - 1.00000 to a demand of 1200. Notice that the upper limits of these intervals correspond to the cumulative probability distribution of demand given in the last column of the table above. Therefore the standard method of associating ranges of random numbers to represent a simulated random value of a variable is to use the cumulative distribution of the variable.

To randomly generate a demand, one picks random numbers between 0 and 1; locates it on the vertical axis and draws a horizontal line until the line hits one of the cumulative probability bars. The corresponding demand level to that bar is the simulated demand. In the above suppose the random number we picked is .38 and this corresponds to a simulated demand of 900 units as shown in the graph. The same idea works with continuous distributions as well. Notice how much more efficient this process is compared to the construction and use of a physical spinning wheel. Actually, in a lot of cases we won’t even have to do this. Excel and many other software packages have automated this process in various routines and functions that can be used to generate random values for a variable having various forms of probability distributions. For instance Excel has routines for:

* **Discrete.** Generates discrete random values according to an arbitrary probability distribution such as the demand distribution above. To use this, you must create a range in the worksheet containing the *x* and *P(x)* columns.
* **Uniform.** Generates a desired number of uniformly distributed values in a specified interval
* **Normal**. Generates normally distributed random values with desired mean and standard deviation
* **Bernoulli.** Generates 0’s and 1’s in the desired proportion of 1’s.
* **Binomial.**  Generates a set of number of ‘successes’ in *n* trials with the probability of a single success of *p.*
* **Poisson.** Generates a set of number of occurrences per unit space (time, distance, area) for any desired mean rate.

To use any of these one will choose Data tab 🡪 Data Analysis 🡪 Random Number Generation; choose the distribution; choose how many sets of values you want and how many values you want in each set. Then depending on the distribution enter the parameters; point to where you want the values to be displayed and Excel generates and displays the values.

**A Simple Simulation of Cash Budgeting**

Suppose the uncertain demand for a product follows the above distribution and management wants to study the behavior of cash balances for a six-month planning horizon extending from April to September to study the frequency and magnitude of cash excesses and shortages and to evaluate the performance of various cash management policies such as those mentioned above in terms of certain objectives such as net interest expenses.

The following characterize the situation:

**Demand** is randomly generated by using the Monte Carlo method according to the probability distribution above.

**Collections:** The selling price for the item is $1000. On average 50% of the sales in a given month is collected in that month, 30% next month and the remainder in the second month following the sale. These percentages may also subject to uncertainty and could be simulated using a similar Monte Carlo method.

**Cost-of-goods -sold:** Any sale in any month requires purchases to be made for a cost of $650 per unit.

**Wages and Salaries**: This cost item is semi variable, consisting of $200,000 per month fixed and $120 per unit variable components.

**Debt service**: The Company has an installment of mortgage payments of $150,000 coming due in June.

**Beginning conditions:** The beginning cash balance is $50,000. Demands for February and March were 750 and 800 units. These are relevant since part of the revenue for these past sales will be coming in April and May as per the estimated collection pattern.

**Decision:** To cover possible cash shortages, the Company can borrow any amount up to $60,000 in $10,000 increments at 4% annual interest rate at the beginning of April to be paid off at the end of September. During the period from April to September, any excess cash can be deposited at 2% while any shortages are financed through short term borrowing at 10%. The question is how much should the company borrow in April, if any.

To simulate the cash balances (without using Crystal Ball, for the time being) for a horizon extending from April to September a [table](cash_bud.xls) is created with appropriate Excel formulas conveniently executing the above relationships among the problem variables. The model simulates demand for six months from April to September according to the above probability distribution. To simulate demand we entered the demand distribution in the range A26..B31.

|  |  |
| --- | --- |
| Demand | Probability |
|  800 | 0.27 |
| 900 | 0.33 |
| 1000 | 0.15 |
| 1100 | 0.15 |
| 1200 | 0.1 |

We choose Data tab🡪 Data Analysis🡪 Random Number Generation. In the *Distribution* field we selected “Discrete;” in *Number of Variables* field entered 6 (we want to generate 6 random demands) in *Number of Random Numbers* enter 1 (we are doing one repetition). In the *Value and Probability Input Range* entered or pointed to the address of the above probability distribution that was placed in the range A25..B29 excluding the column labels. In the Output options we pushed *Output Range* radio button and entered or pointed to the address E5..J5 --the range for demand from April to September in the simulation model.

Using the simulated demand and the assumed collection pattern the model calculates, for each month in the planning horizon, the cash in flows and out flows (COGS, Wages and Salaries and the one time mortgage payment); and finally calculates the running cash balances as beginning balance (end of last period’s balance) plus the in flows minus the out flows. You can create repetitions of this simulation by generating repeated sets of random demands by repeatedly invoking the random number generation routine. It will remember the demand distribution entered. As soon as you hit the OK button a new demand pattern is generated and the entire spread sheet is recalculated accordingly.

 In order to get a reliable picture of the behavior of the system it is important to do this experiment for a fairly large number of repetitions to come up with probability distributions of various results, such as of cash shortages and excesses and net interest expenses etc. With this rather rudimentary model this would be somewhat tedious as we would have to initiate each repetition manually and keep track of the results to summarize in the form of probability distributions later. This is where simulation programs, such as Crystal Ball, come in, to streamline the experimentation and record keeping.

**Introduction to Crystal Ball**

Crystal Ball is a popular and a fairly powerful simulation program which runs as an add-in in the Microsoft Excel environment. It automates the tasks of generating random events and collecting performance data and variety of other functions. Using Crystal Ball you 1) build a spreadsheet model of your problem; (2) simulate its dynamic behavior, and (3) analyze the results. In this note we will review some very basic skills to allow you to run simple simulations using Crystal Ball. There are three special kinds of data in a spreadsheet created for a simulation experiment with Crystal Ball. The first such data is termed in Crystal Ball as **assumption cells** roughly corresponding to uncontrollable variables whose values may be subject to uncertainty and thus must be randomly generated by the Monte Carlo process. In the above example monthly demands are examples of assumption cells. The second is termed the **decision cells.** These are the cells whose values are under the control of the decision maker. Simulation provides information on the performance of the system for various selections of values for these cells. In the above example the initial borrowing at a favorable rate is an example. And the third type of cells is the **forecast cells** which roughly correspond to performance criteria—things whose behavior you wish to analyze under various decision choices and a variety of randomly occurring phenomena. Examples may include the resultant net interest costs, frequency of short term borrowing etc. Crystal Ball not only allows us to estimate the averages of these response variables but also provides their statistical distributions.

 Here is how you might use Crystal Ball to run the above cash management simulation. You will create a spreadsheet model, not unlike the one above, define the appropriate assumption, enter the decision you want to study, create the forecast cells and run the experiment. To try this first save your cash budget model to your desktop. Start Crystal Ball from your computer which you have already installed. Open the file by “file🡪 open” and choose the file you saved on your desktop.

1. Entering the “assumptions”

Move the mouse to E5; select the small green icon (assumption) in the crystal Ball menu bar. Choose from the distribution gallery “custom.” Click OK. In the dialog box enter 800 (first demand is 800) and the corresponding probability (.33). Click Enter. Repeat for all values up to 1200. Click OK after the last entry. Now you have defined in cell E5 a demand distribution. Copy this assumption to cells F5 through J5 as follows (Excel’s copy procedure will not work). While the selection is on E5 choose “CB’s copy data.” Select cells F5 through J5 and click CB’s paste data. Cells E5 through J5 will now be all green.

2. Entering the “decisions”

In Cell C20 we enter the beginning cash balance ($50,000). Move to cell D20; choose decision icon on the Crystal Ball menu bar. Assign a lower limit of $0 and upper limit of $60,000. Choose discrete in the right panel and step size of $10,000. This defines cell D20 which now appears in yellow as a decision which can take discrete values between 0 and 60,000 in increments of 10,000. Enter $0 in the cell representing the no initial borrowing option.

3. Defining performance criteria (“forecasts”)

Our performance criterion—quantity we want to keep track of, is the average interest cost. Net interest cost has there components: cost of initial borrowing (4% of amount borrowed) placed in D23. In row 21 we have calculated the cost of short term borrowing, if need be, at 10%. Row 22 computes the interest earnings due to excess cash. Finally row 23 calculates the short term net interest costs. The net average interest cost is then the sum of D23 and J23 in K23. We therefore define cell K23 as our “forecast cell” by selecting the cell and clicking on “define forecast” icon to enter a name and units for it. The name and the units are used in labeling the results we get from the simulation. This completes the construction of the simulation model. We have assumption cells that will be filled with simulated demand, a decision cell which will be used to denote the initial amount of borrowing and a forecast cell that will keep track of cost of borrowing for us.

4. Running the simulation.

We can automatically run this simulation for any number of repetitions. Each time, 6 months of demand is randomly generated according to the defined distribution, appropriate calculations are made and the performance results are collected. Just to see the mechanics of the simulation let’s run it once (one 6 month) Click on the *green triangle with a green bar in front of it*. This is called single stepping and is generally used to make sure the simulation model is working as intended. Examine the results and verify that the model is indeed working as intended.

Although running the system for one trial (one month) gives some useful information, in reality the demand may not occur precisely as generated. What will be more useful is to repeat this run for a large number of times and see what kinds of average results pertaining to the response variable is likely to come up. Now we run the simulation for, say 1000 times (thousand 6-month period). First click on the *double triangles pointing to the left*. This will reset the simulation. Then click on the small *green triangle over a blue square* to set the run preferences. Click on trials tab; enter 1000 for number of trials to run. Click OK. Now click the *single big right pointing green triangle*. When the simulation runs you will see a chart for the performance criteria displaying the frequencies (in thousand runs) of each of the forecast values that occurred.

 After running the simulation for 1000 repetitions, on the “net interest cost” forecast chart click on view, choose “statistics” view to get a variety of statistics (probably more than you care to see). At this point you can terminate the simulation, save your model using a different filename lest you over write the previous model, and go to bed, or change the values in the yellow decision boxes and see what the impact these changes might have on the performance criterion.

5. Analysis

Crystal Ball has quite extensive analysis and report generation capabilities. As you experiment with it you can perhaps discover some of these. Crystal Ball allows you to set up an experiment for up to two decision variables at a time. This automatically generates output for various combinations of the levels of these variables in the form of a decision table—either a one dimensional table *i.e*., a row (if you have one decision) or a two dimensional range (if you have two decision variables). Let’s illustrate decision tables with the cash budgeting problem. Save the model on the blackboard “cash\_bud\_CB-1.xls” on your desk top, then open Crystal Ball and read it into Crystal Ball and reset the simulation. Move the cursor to cell D20 (the only decision cell). Click on the decision icon and observe that we have defined the lower value to be $0 and the upper to be $60. On the right hand side you see that it is declared to be a discrete decision to be changed in steps of $10,000-- the simulation experiments will be done for seven values from $0 to$60,000 in steps of $10,000. Now go to tools and select decision table. In the dialog box you see a list of all the forecasts. You can choose any one of the defined forecast cells to be analyzed—in this case there is only one. Click next to go to the dialog box where you choose up to two decision cells—again in this example we only have the “initial borrowing” quantity. Click on the double arrow in the middle to move this decision cell to the “chosen Decision Variables” panel. Click next to get a preview of the experiments that Crystal Ball is prepared to run. We see that there will be seven tests and each test will be run a thousand replications. Click start. The simulations are run for values of initial borrowing of $0 to $60,000 in steps of $10,000 and a decision table is created on a new work sheet where the expected values of the seven forecasts are displayed in blue. You can get the individual forecast distributions for any, some or all forecast cells by selecting them (pressing ctrl and selecting each) and clicking “Forecast Charts.“ For better comparison purposes, after your selections you can click on “Overlay Chart” to look at these distributions on the same chart distinguished by different colors. Finally you can choose any, some or all of the forecast cells and click on “Trend Chart” to look at several of probability intervals for the forecast cells associated with the selected decision values.

**Additional Examples:**

**Retirement Planning**

*Scenario:*

A thirty year old would like to estimate his accumulation by age 65.

*Known Information*:

* Current annual salary of $30,000
* 7% contribution to the retirement account which is matched by the employer

*Uncertain Information*:

* Average annual salary increase is estimated to be normally distributed with mean 4% and standard deviation of 1%.
* Investment returns are indexed to changes in the S&P 500 over the time period from 1966 to 2005

|  |  |
| --- | --- |
| % change in S&P | Prob |
| -25 | -20 | .025 |
| -20 | -15 | .025 |
| -15 | -10 | .075 |
| -10 | -5 | .075 |
| -5 | 0 | .05 |
| 0 | 5 | .1 |
| 5 | 10 | .1 |
| 10 | 15 | .2 |
| 15 | 20 | .125 |
| 20 | 25 | .075 |
| 25 | 30 | .1 |
| 30 | 35 | .05 |

Note: One can project the *expected* accumulations for the individual at each age till 65 based on the average historical return of about 9.125% per year with 4% average salary increase. However due to highly volatile market returns as seen over the last 40 years, timing of the actual returns and the influence of compounding projected accumulations would have a large variance that this calculation based on the average returns of 9.125% can not reveal.

**Portfolio Selection**

*Scenario:*

An investor wants to invest in four financial instruments with average return and variance (a measure of risk) objectives.

*Known Information:*

* There is $100,000 available for investment
* Four options are Intel, Micro Soft, Procter and Gamble, and Money Market
* The investor does not want invest more than 50% in any instrument.
* There is up to 9 years of return information on the four investment options

|  |  |  |  |
| --- | --- | --- | --- |
| Intel | Micro-soft | P&G | Money Mkt |
|  | 73.0% | 23.3% | 8.27% |
|  | 121.8% | 8.4% | 6.14% |
|  | 15.1% | 14.2% | 3.74% |
| 43.4% | -5.6% | 6.3% | 3.01% |
| -48.5% | 51.6% | 8.8% | 4.08% |
| 77.8% | 43.6% | 33.9% | 5.82% |
| 130.7% | 88.3% | 29.7% | 5.29% |
| 7.3% | 56.4% | 48.3% | 5.44% |
| 68.7% | 114.6% | 14.4% | 5.38% |

*Uncertain Information:*

* The returns are assumed to be normal with means estimated from the historical returns.
* The variance of the portfolio is based on the matrix of co-variances of the returns.

*Decisions:*

* Proportion of investments into the four instruments.

*Performance criteria (forecasts)*:

* Average (weighted) return from the portfolio
* Variance of the returns of the portfolio.

**Airline overbooking**

*Scenario:*

An airline wishes to determine the ideal number of bookings to take for a flight that has 112 seats to maximize expected revenue.

*Known Information:*

* Average fare/seat = $400
* Average cost to bump a customer, if more than 112 show up is $600

*Uncertain information*

* Number of no shows is binomial with a *p* = .15

*Decision:*

* Number to book

*Performance criterion (forecast)*:

* Net profit

**Capital Budgeting:**

*Scenario*:

Airbus Industries is considering adding another jet liner, A3XX to their product line. They want to investigate the economic feasibility of the proposed investment.

*Known information:*

* Start-up cost for similar ventures have resulted in about $150,000 (thousands)
* Selling price $35,000 (thousands) per unit
* Fixed cost of production $15,000 (thousands)
* Tax depreciation is $10,000 (thousands) per year for 4 years—expected product life of A3XX
* Tax rate for the four year horizon is 34%

*Uncertain Information:*

* Best estimates indicate that the demand for these airplanes will follow a Poisson distribution with the mean of the distribution starting at 10 planes and increasing by one per year. So, the demand has Poisson distribution with mean 10 in the first year, Poison with mean 11 in the second year, etc.
* Cost of capital is assumed to be normal with mean 10% and standard deviation of 1%.
* Variable cost is uniform between 75 and 85% of the revenue.

*Decision:*

* Whether or not to undertake this venture

*Performance criteria:*

* Profit after taxes for the 4 year horizon
* NPV of the investment

**Price Change Problem:**

*Scenario:*

A company is considering a price increase for one of their popular models in two years (at the beginning of the third year) that currently sells for $10. They want to investigate the impact of various price alternatives on the net present value of their profit after taxes.

*Known Information:*

* Current price is $10
* Year one demand is 80,000 units
* A dollar increase in price depresses demand by 5,000 unit compared to what it would have been without a price increase
* Corporate tax rate is 34%
* Current fixed costs are $20,000
* Current variable cost is $3 per unit
* Discount factor is 10%

*Uncertain information:*

* Demand growth is normal with mean of 7% and standard deviation of 4% per year
* Variable costs for years 2 through 5 which increases randomly based on a normal distribution with mean 5% and standard deviation of 2%.
* Fixed costs grow at an annual rate which is normal with mean 10% and standard deviation of 3%.

*Decision:*

* New price in period three (to be kept at the same level till the end of year 5.)

*Performance criteria (objective)*

* NPV of after tax profits