Simulation & Risk Analysis

Objectives

- Understand how Monte Carlo simulation can be used for Risk Analysis
- Be able to simulate with Excel and MCSim
- Understand how to test data to see if it fits known probability distributions
- Understand statistical issues that need to be addressed when using simulation
- Awareness of dynamic simulation

Types of Simulations

- Monte Carlo
  - based on repeated sampling from prob. dist. of model inputs to determine dist. of model outputs
  - mainly used with static models
- Systems (Event Based Monte Carlo)
  - models the dynamics & behaviour of a system **over time**

Simulation Uses

- **Explanatory devices** - understand a system or problem; gain insights
- **Analysis vehicles** - determine critical elements, assess uncertainty, find good solutions; what-if, comparative & sensitivity analysis

Finance Example

- Choice of a single investment. If future cash flows are not known with certainty, the choice is unclear. Simulation can help you answer:
  - Which project is riskiest?
  - What is the probability than an investment will yield at least a 20% return?
  - What is the probability that the investment will have a NPV < $10 million?

Corporate Applications

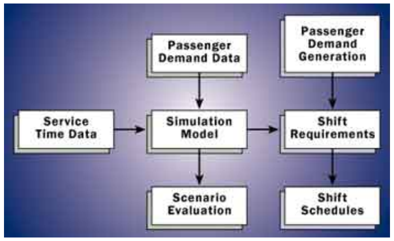
- Merck (Fortune, May 30, 1994)
  - Simulation model to determine whether Merck should pay $6.6 billion to acquire Medco. Inputs included:
    - » possible futures of U.S. healthcare system
    - » possible future changes in mix of generic and brand-name drugs
    - » probability distributions of profit margins for each product
    - » assumptions of competitors' behavior
- Chrysler Corporation
  - By using simulation, CC reduced time from concept to showroom from 7 yrs to 1 ½ yrs
  - CC's design capabilities was one of the aspects that attracted Daimler-Benz

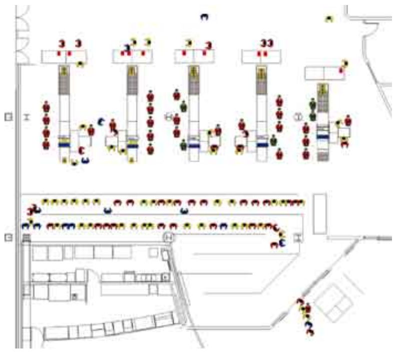
Vancouver Intl Airport

- Preboard screening problem
- Simulation model of preboard screening security checkpoints
  - Goal: use model to determine configuration for acceptable service levels
  - UBC's Centre for Operations Excellence
    - » First observed current situation, collected data; then developed alternatives
- 90% passengers wait < 10 minutes

Atkins et al, "Right on Queue", OR/MS Today (April 2003)

(cont.)





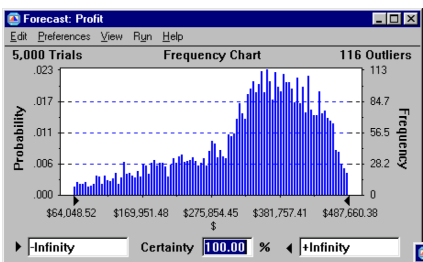
Benefits of Simulation

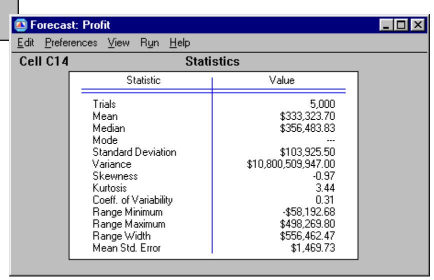
- Understand systems
  - without building them, if they are proposed
  - without disturbing them, if it is costly or unsafe to do so
  - without destroying them, if the objective is to determine their limits of stress
- Simulation can model any assumption and is usually easier to explain than other approaches
- VIM/GUI is latest improvement in s/w

Disadvantages of Simulation

- Can be costly for custom simulations
- Potential for programming errors
  - Verify model
- “Mushy” answers; no guarantee of optimality
- Important (even critical) inputs may be missed in formulation
  - Validate model
- Longer computer running times for complex models
- Sampling error
  - Can deal with this

Actual vs Average (Sim in Excel)





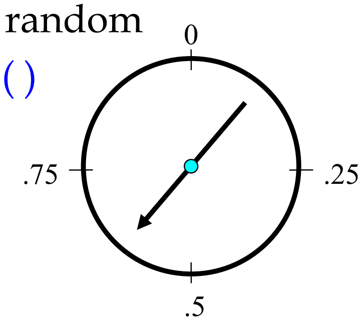
The ‘Flaw of Averages’

- Previous example: average **output** was lower than that associated with average **inputs**; can also go other way.
  - Jensen's Inequality (of interest to statisticians)
- Only time average inputs are guaranteed to result in average outputs is when spreadsheet model is **linear in all uncertain variables**
  - Hard to determine, even for experts
  - Best advice: be careful with averages!

Random Numbers & Random Variates drive Simulations

Random numbers

- **Uniformly distributed** (usually between 0.000 and 0.999…)
- **No discernible pattern**; appear random
- **Pseudo random** vs random
- In EXCEL: =RAND( )

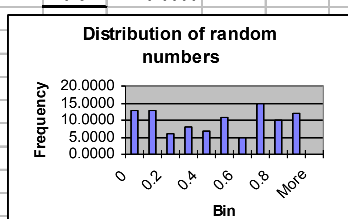


Random number simulation

100 Random numbers

0.1370	0.4774	0.7266	0.7221	0.0677
0.4002	0.7335	0.5866	0.9650	0.5399
0.1429	0.8011	0.0615	0.1680	0.9701
0.0155	0.8830	0.2129	0.1674	0.3714
0.0405	0.5635	0.3845	0.9672	0.0959
0.4421	0.8014	0.9773	0.3664	0.7012
0.5839	0.7654	0.2076	0.3704	0.6827
0.2139	0.1286	0.4663	0.8880	0.8277
0.1666	0.7972	0.0093	0.1490	0.6807
0.2791	0.9690	0.0533	0.2799	0.1768
0.3371	0.7894	0.4200	0.5542	0.9804
0.9715	0.0759	0.3771	0.6517	0.6087
0.9177	0.0594	0.0241	0.5160	0.5266
0.6223	0.4849	0.3272	0.1294	0.1589
0.7665	0.5832	0.8870	0.2588	0.8788
0.7740	0.1716	0.0588	0.4581	0.5311
0.0923	0.7762	0.7986	0.9441	0.5426
0.7270	0.8146	0.5106	0.1057	0.9911
0.9814	0.3179	0.8111	0.1776	0.0367
0.7572	0.7786	0.7140	0.8429	0.9528

Bin	*Frequency*
0	13.0000
0.1	13.0000
0.2	6.0000
0.3	8.0000
0.4	7.0000
0.5	11.0000
0.6	5.0000
0.7	15.0000
0.8	10.0000
0.9	12.0000
More	0.0000



Random Variates

- **Definition:** Outcomes generated from a probability distribution
- Random variates are assigned **proportional to relative frequency** we want to generate in simulation
  - key is to build a **cumulative distribution**
- Correspondence between random variates and actual outcomes is key to valid, credible simulation models!

Example: Sales Distribution

Sales	Prob.	Range	Proportion
80,000	.2	0.0 - 0.2	20%
100,000	.5	0.2 - 0.7	50%
120,000	.3	0.7 - 1.0	30%

Probability Distribution

Cumulative Distribution

What sales levels would these random numbers generate? 0.00, 0.11, 0.27, 0.38, 0.53, 0.74, 0.99, 1.00

Excel LOOKUP Tables

Random Number Range		Sales
0	0.2	80,000
0.2	0.7	100,000
0.7	1.0	120,000
Trial	Random no.	Sales
1	0.616676	100000
2	0.389207	100000
3	0.438059	100000
4	0.274028	100000
5	0.800230	120000
6	0.740820	120000
7	0.165618	80000
8	0.697073	100000
9	0.927857	120000
10	0.008219	80000

=VLOOKUP(B7,$A$2:$C$4,3)

NB: range goes up to, but does not include, the upper limit

Simulating with Excel

- Formulate problem & build spreadsheet model
- Specify probabilistic assumptions; assign probability distribution(s) to input variable(s)
- Implement model
  - sample input values from probability dist.
  - Compute output variables & record results
  - repeat above until sufficient # of trials (replications) generate useful dist. of outputs

Freezing Random Numbers

- Automatic recalculation of random numbers is a blessing / curse
  - may want to keep data generated; may want to manually control regeneration
  - can do first by “freezing” data
- Select Range; Copy; Paste Special (from Edit menu), select Values option
  - changes formulas to numbers
- Tools > Options > Calculation
  - select automatic or manual

Replication Techniques

- Manual (F9 key); turn off automatic recalculation
  - Excel: Tools > Options > Calculation . …
- **Replicate** formulas within the spreadsheet
  - limited only by spreadsheet size
- Special software (e.g., @Risk, Crystal Ball, MCSim, etc)
  - can quickly replicate 1,000 or 100,000 times!

Example with MCSim

- MCSim.xla is an Excel Add-in
- Makes Monte Carlo simulation easy
- Good Overview of Risk Analysis
  - http://www.crystalball.com/risk-analysis-start.html

Using MCSim

- Develop spreadsheet model
- Define Cells which represent random information
- Define "output" cells to watch
- Set number of replciations
- Run simulation
- Analyze results

Criteria for Decision Making

- How do we select between various options, based on simulation results?
  - Expected Value (largest mean)
  - Optimistic (best scenario, max/min)
  - Pessimistic (worst scenario, max/min)
  - Minimum payoff
    - select option with greatest probability (max $p(x > x_0)$ or $p(x < x_0)$)
  - Minimum variation
  - Management preferences
  - …..

Simulation Languages

- General Purpose
  - Basic, Fortran, Visual Basic, C, C++, Java
- Special Purpose
  - SLAM, GPSS, SIMSCRIPT, GASP, DYNAMO
- Simulators
  - Simfactory, XCELL+, Extend, Micro Saint, Arena
- Excel Add-ins
  - Crystal Ball, @Risk, MCSim, etc

Statistical Issues

- Simulation experiment is a **sample** from an unknown population
- Purpose of simulation is to estimate population parameters (**mean**, **variance**, …) and gain knowledge about distribution of outcome variables in order to assess risk
  - Easy to calculate confidence interval
- Initially need to determine input probability distributions

Inputs: Historical Data

- Assumption: past represents future
- Requires a large amount of data
  - Available through scanners, e-com websites
- Approaches
  - Eyeball test
  - Goodness-of-fit
    - » computer automates this

Eyeball Test

- Good way to start
  - plot & examine histogram of your data (interarrival time, service time, …)
- Visually examine your plot; does it look like a standard distribution?
  - Normal: symmetric, bell-shaped
  - Uniform: flat
  - Exponential: highly skewed
  - Lognormal: skewed, but density falls near zero

Model Verification &Validation

- Challenge: build an **accurate** model & **convince** end users of this
  - Easier said than done, but GIGO
  - Models which do not accurately reflect real world behavior cannot be expected to generate meaningful results
  - Errors in programming can yield nonsensical results
  - Accurate model, which users do not understand, will not be accepted by them
  - Visual Interactive Modeling helps

Verification & Validation ...

- **Verification:** does model perform as intended?
  - Verification is generally done by having an expert review model & computer code
  - Test with dummy data (extremes, averages)
- **Validation:** does conceptual model accurately depict real system?
  - Conduct "walk-thru" with end users
  - If possible, run simulation using actual past data; compare output to historical results

Validation ...

- Validation typically has 3 steps:
  - (1) Develop a model with high **face validity** (looks reasonable to those who understand physical system)
  - (2) Validate model assumptions
  - (3) Validate model output

Experimental Design

- **Experimental design** - an important consideration in simulation process
- Issues to be addressed **prior** to collecting & analyzing output data: input data, length of time of simulation, treatment of initial data outputs from model (**transient** or **start-up** conditions), sample results or analyze all?
- Usually one is interested in results for **steady state** (long run) operation of system

Experimental Design (cont.)

- Initial data inputs to simulation generally represent a **start-up period**; may be important that data outputs for this be neglected if interested only in long run behavior
- For each policy/alternative under consideration by decision maker, simulation is run by considering a long sequence of input data values (given by a pseudo-random number generator)
  - Whenever possible, different policies should be compared by using the same sequence of input data (use same **"seed"**)

of Trials/Replications

- **Different** random numbers yield different results
- **Number** of trials/replications
  - depends on system & objectives
  - larger # provides more accurate estimates of population means
- When comparing alternative models, should use same sets of random numbers, same # of trials

How Many?

- Static Simulation
  - No longer an issue because of Monte Carlo simulation s/w
  - Usually single replication, with large # of trials (100,000+)
- Event-Based MonteCarlo Simulation
  - Usually many replications (dozens, hundreds), where each is a single trial
  - Variance-reduction techniques
    - » Common random numbers most common
  - Computing resources can become an issue

Simulation Types

- **Static Simulation**:
  - independent of time
- **Event-Based MonteCarlo Simulation:**
  - Time Dependent
  - Use a "simulation clock"
  - Time Driven: move clock forward a fixed amount
  - Event Driven: move clock forward to time next event occurs (usual approach)