Project Risk Management

Teaching Notes - Topic 11

QUANTITATIVE RISK ANALYSIS

MONTE CARLO SIMULATION

Contents

Topic 11: Quantitative Risk Analysis - Monte Carlo Simulation

Introduction to topic

Simulation is a technique for emulating a process, usually conducted a considerable number of times to understand the process better and measure its outcomes under different assumptions. Monte Carlo simulation is a technique that allows us to examine what would happen if we undertook many trials of a project. The repetition of calculations allows one to artificially live through a project a number of times. Monte Carlo simulation is essentially a computer-based technique and has been available since the 1960's. This Topics use knowledge from previous Topics to describe the process of Monte Carlo simulation

Learning outcomes from this topic

On completion of this topic, you should:

- Apply the process of Monte Carlo simulation
- Appreciate the significance of correlation in Monte Carlo simulation
- Describe the advantages and disadvantages of Monte Carlo simulation

Copyright Notice

1.0. MONTE CARLO SIMULATION (MCS) - INTRODUCTION

1.1. Monte Carlo Simulation – Definition

MCS is a probabilistic method for analysing risk. MCS runs a repetition of calculations that allows one to artificially live through a project a number of times. John von Neumann has been credited with naming and popularising MCS while participating in the design of the atomic bomb in the 1940s. It is essentially a computer-based technique and has been available since the 1960's. MCS is so called because it involves the randomly selection of values, similar to how a ball on a roulette wheel randomly stops at a number. The basics of MCS are easy to understand and its use for quantitative risk analysis in projects is growing and probably represents best practice.

1.2. Monte Carlo Simulation – Modelling Risk

MCS uses probability distributions (see Topic 10) to model uncertainty. There are two *types* of uncertainty to be modelled for MCS, which were described in detail in Topic 1:

i. *Accomplishment Uncertainty*

This is where we are estimating the quantity of a variable, such as the cost of a project activity, which has some degree of uncertainty. There is *no* uncertainty of the *existence* of the variable i.e. a cost variable will have a final actual cost (100% certainty) we just are uncertain as to what that cost will be. For example, there may be uncertainty about labour productivity for a project activity. We know there will be some level of labour productivity but we do not know what that level will be.

A common approach is for this uncertainty to be expressed as a range of values (e.g. minimum, most likely, maximum) and represented by a probability distribution. So for a cost element we may set a minimum, most likely and maximum value of $900, $1000, and $1200 respectively. For example, uncertainty could be in relation to productivity so that the minimum value of $900 may reflect good productivity, and most likely value is the expected level of productivity while the maximum possible cost of $1200 reflects poor productivity.

Accomplishment uncertainty may conceptually also encompass the effects of many minor uncertainty *events* and opportunities that may not have been identified as separate event uncertainties but would specifically exclude events identified separately as event uncertainty (see next).

ii. *Event Uncertainty,*

Risk events have a less than 100% likelihood of occurring. i.e. may or may not occur e.g. a cyclone There are two components of uncertainty for a risk event:

- *Possibility* – There is uncertainty of occurrence. There is a *degree* of probability of a risk event occurring or not occurring. This is commonly represented by a binomial discrete probability distribution e.g. for a cyclone, we may determine this risk event has a 25% probability of occurring and a 75% of not occurring.
- *Consequences* – There is uncertainty of impact, which is accomplishment uncertainty. There is a range of possible consequences if an event eventuates, and each consequence has its own degree of probability or likelihood. A probability distribution can represent the consequences if the risk event occurs e.g. if the cyclone occurs, the cost may range between a minimum of $15,000, most likely of $50,000 and maximum of $100,000.

These risk events can be the high-risks qualitatively analysed in the risk register. The register is re-visited and the qualitative ratings from likelihood and consequences for each risk event are converted into quantitative probability distributions.

1.2.1. Sources of uncertainty

The *types* of uncertainty we are interested in – accomplishment and event – have two *causes/sources* of uncertainty – aleatoric (natural variability) and epistemic (knowledge) (explained in Topic 1). A key source of uncertainty is the degree of scope development (epistemic uncertainty) so the range of values could reflect assumptions of the nature of the final scope of a variable So a probability distribution for accomplishment or event will represent aleatoric (natural viability) uncertainty and/or epidemic (knowledge) uncertainty. Accomplishment or event uncertainty can have a probability distribution can represent a combination of these two *causes/sources* of uncertainty. As stated in Topic 1, it is sometimes difficult to distinguish between aleatoric uncertainty and epistemic uncertainty, particularly as both can be expressed in the same probability distribution. The separation of aleatoric and epistemic uncertainty is often subjective and is more an analytical convenience than a fact of the world and dividing sources of uncertainty between aleatoric and epistemic components is an overt choice by the risk analyst.

2.0. MONTE CARLO SIMULATION (MCS) – THE PROCESS

2.1. Monte Carlo Simulation – The Process

STEP 1- CREATE A PROJECT MODEL

The scope of the project model must be agreed. A risk model for projects is often based on project plans (e.g. cost estimate, network schedule, investment analysis). This is the most challenging aspect of MCS and requires collaboration with stakeholders and experts. The usefulness of MCS is significantly dependent upon the quality of the model. Broadly, the model contains two categories of uncertainty – accomplishment and events. Raydugin (2013) highlights the need to have the consistent level of detail throughout the model. For example, in a cost model there should not be *hundreds* of cost variables (accomplishment uncertainty) and *just only a few* risk events (event uncertainty) of high probabilities and consequences. This may result in the output probability distribution having two peaks, which is evidence of this mismatch of level of detail. One solution is the have both accomplishment and event uncertainties at a high-level, rolled-up level e.g. 15-50 cost variables and 15-30 high-rated risk events.

STEP 2 SPECIFY INPUT & OUTPUT VARIABLES

2A. Assign input probability distributions for uncertainty
Probability distributions are the basic building blocks for MCS. Appropriate probability distributions are allocated for each accomplishment variable and each risk event variable. [Paper 10 described some common probability distributions]. In selecting a probability distribution for a variable, the following should be considered:

- list everything known about the variable - Continuous or discrete? Bounded or unbounded?
- understand the basic types of probability distribution e.g. triangular, uniform, normal, betaPERT
- select the distribution that characterises the variable under consideration. i.e., shape, spread, skewness.

Fundamentally, the selected probability distribution should be easy to understand with the aim of keeping the risk analysis process as simple as practically possible. While there are several different types of probability distributions – e.g. discrete, triangular, uniform, betaPERT - the selection of the appropriate *shape* for a probability distribution has less importance of the outcome results than their *parameters* that describe the distribution e.g. minimum, maximum and most likely values. Most commonly, input probability distributions are selected as skewed to the right towards the higher values because experience shows us that there is invariably more prospect of a variable exceeding most likely values then achieving minimum values. The choice of a specific type of distribution for each element of uncertainty has only a small effect on results whilst the correct assessment of *correlation* among variables is far more significant in terms of the accuracy of the result (see section 2.2. Correlations).

Minimum and maximum values for a variable could be derived from:

- reviewing past data for minimum and maximum values for similar variables i.e. objective
- expert opinion for minimum and maximum values i.e. subjective
- expert opinion for degree of uncertainty i.e. subjective - linked to guidelines (Oracle, 2009):

Degree of uncertainty	Minimum	Most Likely	Maximum
Very High	50%	100%	200%
High	75%	100%	150%
Medium	85%	100%	125%
Low	90%	100%	115%
Very Low	95%	100%	110%

Developing parameters for probability distributions is often obtained through a facilitated group process, including scope clarification, discussion on reasons for maximum, most likely and minimum values, and correlations between variables.

2B. Select Output Variable(s) that are of interest for decision making. For projects this is invariably the bottom-line value of total cost, or total duration, or profit.

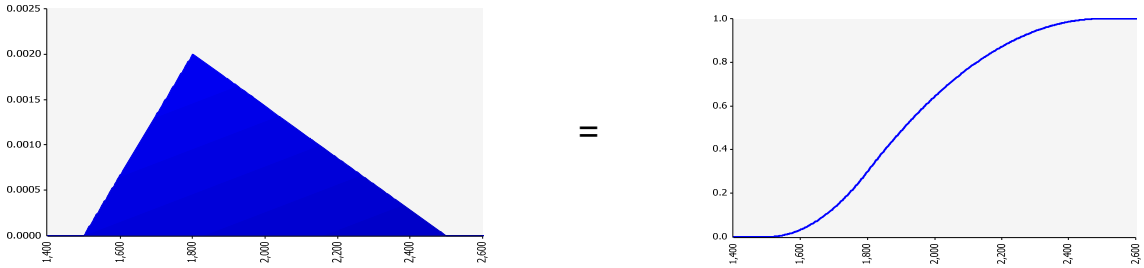
STEP 3 - CORRELATION

Correlation refers to the relationship between input variables. Variables in a model can be either "independent" or "dependent". An independent variable is totally unaffected by any other variable within your model. A dependent variable, in contrast, is determined in full or in part by one or more other variables in your model. You have to decide which variables, if any, are dependent and these must be correlated. For more information see section 2.2 in this paper.

Step 4 - SAMPLING

Sampling refers to selecting values from distributions. MCS is based on random sampling from the input probability distributions in the model. Sampling selects one value for each variable in the model and presents the result, which for projects is typically the final cost, final duration or profit. This one combination of one set of sampled values is called an *iteration* and could be viewed as a deterministic estimate as it is based on one set of values. This one iteration is saying "this could be the final cost/duration/profit of the project". Then numerous iterations are run and the results are analysed for decision making.

Before sampling, the input probability distributions (e.g. triangular, uniform) are converted into cumulative distributions:



To sample values from these input distributions a *random number generator* is needed, which is an algorithm to generate a long list of random numbers. There are various types of random number generators, the most common producing each random number based on the previous one. So if the first random number is known (called the *seed*), the random number series can be known. The seed is often chosen at random from the computer's internal clock. The most common random number generator is the Mersenne Twister, developed in 1997, which is the default choice in @risk software. The sampling of a value from an input probability distribution occurs as follows: for each variable, a computer-based random number generator randomly generates a number between 0 and 1.0 value, which is then placed on the vertical axis of the variable's probability cumulative distribution and translated into its equivalent value (e.g. dollar value for a cost variable). For example, see Figure 1: if the random number is 0.3215, then the dollar value ($1804) at the 32.15 percentile of the input distribution is selected/sampled:

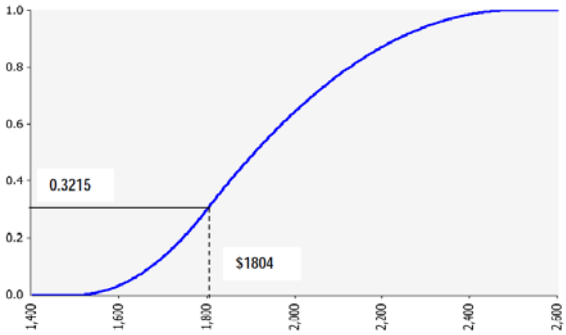


Figure 1: Randomly selecting a value from an input probability distribution

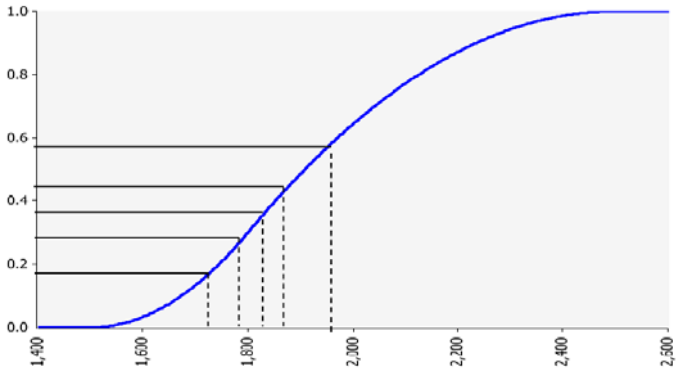
It is important to note that the random selection of values for each variable closely follow its probability distribution. The sampling process is guided and confined by the chosen probability distribution for the variable, so that the frequency with which values are selected should generally conform to the distributions e.g. values closer to the most likely should be selected more often than values towards the minimum and maximum values, so that the set of selected values will resemble the distribution that produced them. So when values are randomly selected, these values have the same probability of being selected as is indicated in their probability distribution. As a simple example, image a dartboard that is divided into three coloured zones; red (50% of area), blue (25%) and green (25%). If a blindfolded person now throws 100 darts, we could reasonably expect that approximately 50 darts would land in red, 25 in blue and 25 in green. So the results from the dart throwing (i.e. sampling) mimics the predetermined colour distribution on the dartboard.

[Note: Selecting a fixed seed: MSC software, such as @risk, allows you to nominate a fixed seed. This means that if you run one simulation of a model for say 5000 iterations, then you run a second simulation of the same model for 5000 iterations, the results will be identical if a same fixed seed is used for both simulations. This is because the fixed seed determines the same sequence of random numbers generated in each simulation so that the same values are selected. So why set a fixed seed? Once one simulation of say 5000 iterations a model has been run, it is quite common to change parts of the model and then run another simulation with these changes in the model to see their effects on the output results. By using the same seed for each simulation, we know that the difference in the output results between the two simulations is solely due to changes between the two models. If we did not have a fixed seed, then the difference in the output results between the two simulations could be due to each simulation starting with a different seed and consequently a different set of random generated numbers.

There are two alternative methods for randomly sampling a value from a probability distribution

i. *Monte Carlo Sampling* (MC) (Figure 2):

This randomly samples any value within the input probability distribution. Samples are more likely to be drawn in areas of the distribution which have higher probabilities (i.e. close to most likely value). With enough iterations (this is equivalent to the number of darts thrown), there will be random selections throughout the complete range of the probability distribution, so that MC sampling recreates the full input distribution. But a problem of clustering can arise when only a small number of iterations are performed. Where input probability distribution have some low probability values (at tails of distribution), a small number of iterations may result in clustering of sampled values around the centre of the distribution and so not sample sufficient quantities of these outcomes to accurately represent the probability distribution of the input variable. This led to the development of Latin Hypercube sampling.

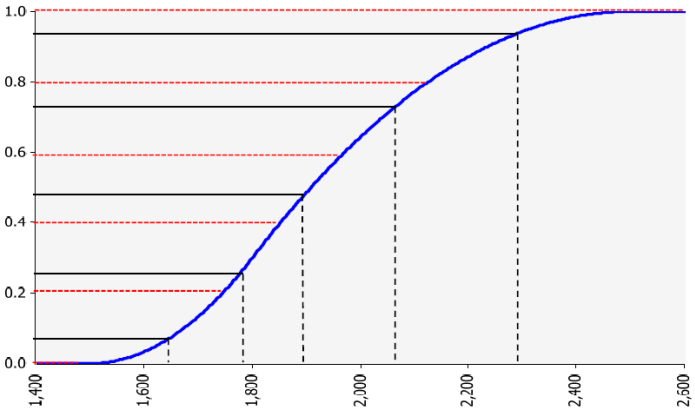


Solid line = random number selection ; dotted line = value selected

Figure 2 – Monte Carlo Sampling – 5 iterations - clustering

ii. *Latin Hypercube Sampling* (LH) (Figure 3)

LH was invented in the 1970s when computer processing was relatively limited and slow as LH requires less iterations than MC to produce adequate results. It uses stratified sampling to sample over the full range of the input distribution in fewer iterations when compared with MC Sampling. LH sampling divides the PD into equal probability intervals based on the number of iterations e.g. there are 100 iterations there are 100 intervals each of 5% probability. Then a sample is randomly selected *once* from each 5% stratification over the 100 iterations. LH sampling is forced to represent values in each interval and thus forced to sample for the full range of the input distribution. Consequently, the sampling more accurately reflects the distribution of values in the input probability distribution. LH sampling provides increased sampling efficiency and faster runtimes. It also aids the analysis of situations where low probability outcomes are represented in input probability distributions. So by forcing the sampling to include the low probability values, it ensures these values are accurately represented in the simulation outputs. With today's computer power there is less justification for LH over MC, and when the number of iterations is large (say 2000+) there is no practical difference between the results of the two sampling methods, and MC can be preferred as it is a more natural random process.



Solid line = random number selection; dotted line = value selected; Red= equal probability intervals

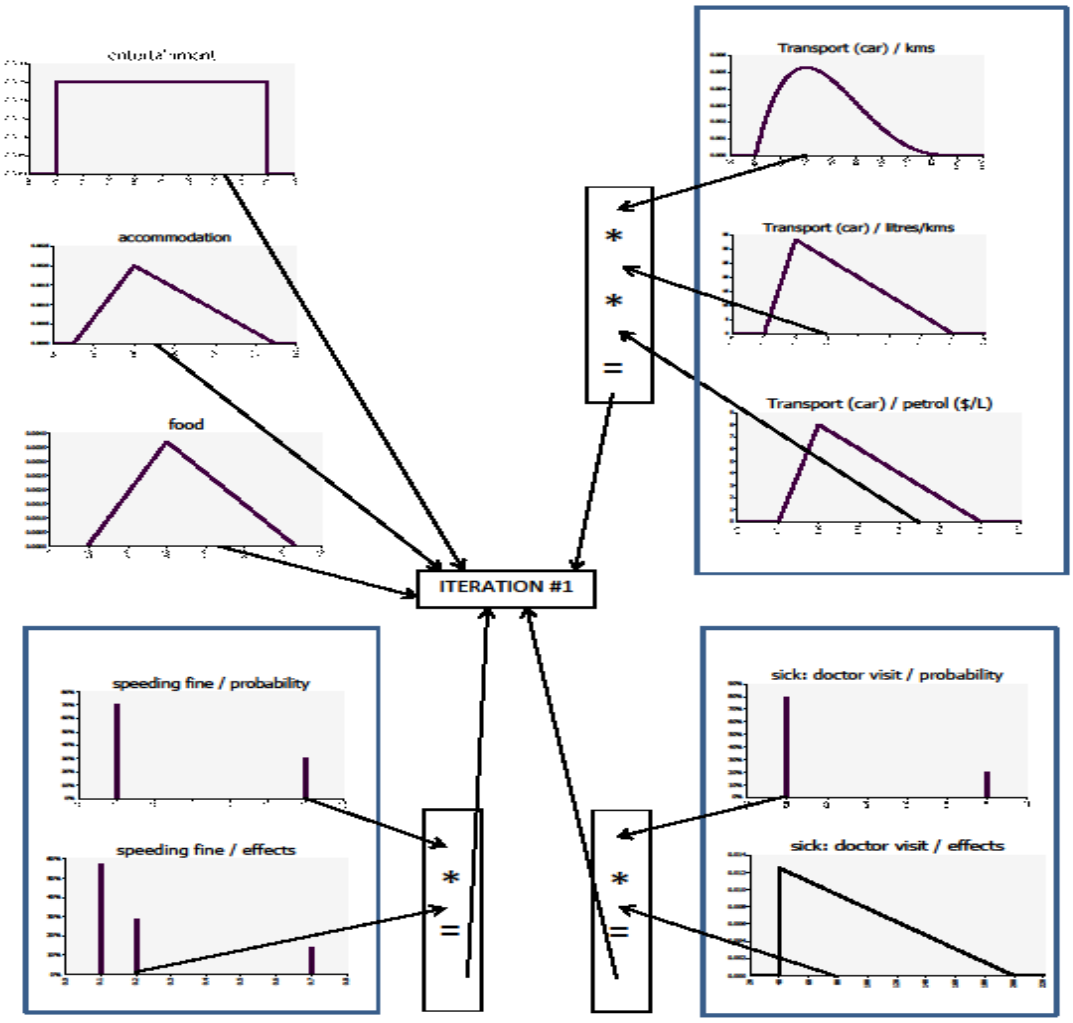
Figure 3 – Latin Hypercube Sampling – 5 iterations

Iterations

For one iteration, one value is randomly sampled from each input probability distribution for all uncertain variables – accomplishment and event. From this, the project's output value (e.g., profit, time, cost) is calculated (see diagram next page). This one iteration is just one plausible scenario of the 'bottom line' outcome. It is preferred that a large number of iterations be undertaken so that a more representative final probability distribution of the output variable is produced. The number of iterations required depends on the size of the model (number of variables) or where the probability distributions have some small probabilities that need to be sampled. A typical number of iterations is between 1000-5000. For each iteration through the model the computer may randomly select values towards the pessimistic end of some variable's range and the optimistic end for others. In this way MCS mimics reality because it unusual that a project will experience all high values or all low values.

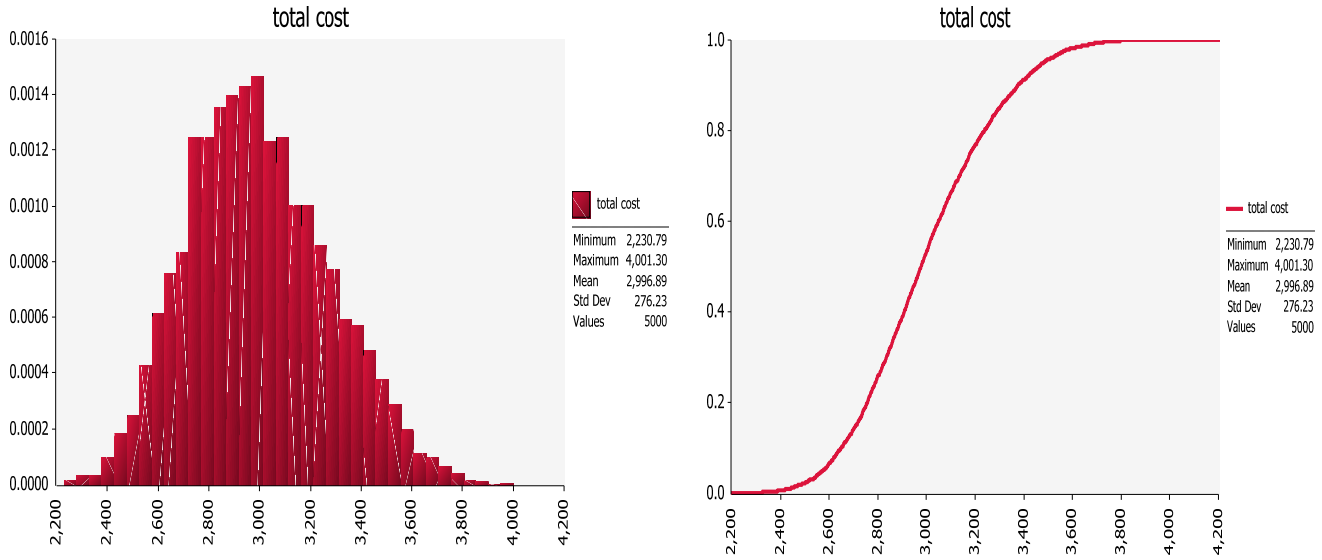
Example – Iterations
This diagram below shows one iteration. Values are extract from all input variables and totalled.
The table shows the results of the first 5 iterations
The cost model is set out in the appendix

Iteration	Total cost	food	enter	accom	kms	l/kms	$/L	fine / prob	fine / effects	sick: prob	sick: effects	Comment
1	2894	260	675	1852	824	9.4	1.4	0	100	0	56.72	No risk occurred
2	2787	212	694	1738	767	0.1	1.4	0	200	0	144.14	No risk occurred
3	2324	272	409	1539	793	9.8	1.3	1	100	0	50.76	Speeding Fine
4	3128	348	482	2077	618	0.1	1.3	1	100	1	119.54	Both risks occurred
5	3309	577	487	2056	709	0.1	1.3	0	400	1	97.12	Sick



STEP 5 - RESULTS

MCS culminates in the production of a probability distribution for the output variable of interest (eg total cost, total duration) and related statistics (e.g. mean, standard deviation). As each input variable in the model is drawn from a probability distribution then the characteristics of the output distribution is determined by the characteristics of the distributions of each input variable. This output probability distribution is a description and pictorial representation of the project's intrinsic uncertainty. (Note: if the all the variables are *independent* then the Central Limit Theorem implies so that the resultant distribution will approximate a Normal distribution). The resultant probability distribution is typically presented in two forms: relative (histogram) and cumulative (graph):

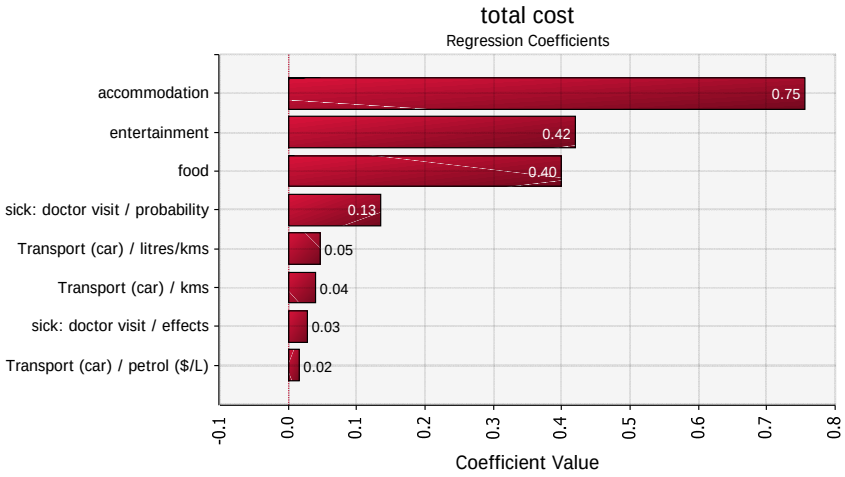


The resultant probability distribution provides information:

- *mode, median, and mean(expected value).* These show the centre of the distribution and act as a basis for evaluating extreme outcomes
- *standard deviation.* This shows how widely dispersed the distribution is about the mean. It is an indication of the uncertainty of the distribution.
- *skewness.* This is the degree of deviation from being symmetrical. If greater than +/-1, it indicates a highly skewed distribution.
- *kurtosis*. This describes the shape of the distribution. The higher its value, the more peaked the distribution is (i.e. narrower the range between maximum and minimum values)
- *range*. This is the different between the minimum and maximum values and establishes the 100% interval of all possible values of the simulation

STEP 6 - SENSITIVITY ANALYSIS

It is possible to identify the input variables that are significant in determining the value of the output variable. One method for this is rank order correlation, whereby correlation coefficients are calculated between the output value and the sampled input values. The higher the correlation coefficient between the input and output, the more significant the input is in determining the output. The result can be displayed as a tornado chart. – see below. The longer the bar, the greater the effect that variable is having on the model's output. So that, the bar length represents the degree of correlation between the input variable and the output. The higher the degree of correlation between the input and output variables (calculated using rank order correlation), the more the input variable is affecting the output. The identified critical input variables should be focused on when making decisions based on the model. Typically, the variables are displayed from the top down in decreasing size of correlation. Where there are positive and negative correlations, the result appears somewhat like a tornado, hence the name.



STEP 7 – MAKE DECISIONS

The ultimate purpose of MCS is to support decision making and the output probability distribution produced by MCS is a core source for this. The decisions can be made by the same team who create and run the mole, or by management.

This information can help answer such project management questions as:

- What is the probability of making a loss?
- What is the most likely value for a project's cost or duration?
- What is the probability of the project exceeding its baseline budget or deadline? ie confidence level
- How risky is the project (i.e. standard deviation)?
- What contingency must be added to the budget or schedule for a desired confidence level?
- What are the sensitive variables and what shall we do about them?

2.2 Monte Carlo Simulation - Correlation

2.2.1 Correlation - Introduction

Variables in a model can be either "independent" or "dependent". An independent variable means it is totally unaffected by any other variable within your model. For example, if you had a financial model evaluating the profitability of an agricultural crop, you might include an uncertain variable called Amount of Rainfall. It is reasonable to assume that other variables in your model such as Crop Price and Fertilizer Cost would have no effect on the amount of rain — Amount of Rainfall is an independent variable. A dependent variable, in contrast, is determined in full or in part by one or more other variables in your model. For example, a variable called Crop Yield should be expected to *dependent* with the variable Amount of Rainfall. If there's too little or too much rain, then the crop yield is low. If there's an amount of rain that is about normal, then the crop yield would be anywhere from below average to well above average.

When setting up a model (e.g. cost estimate, schedule) you have to decide which variables are dependent and therefore need to correlated. It is extremely important to recognize correlations between variables or the model might generate nonsensical results. For example, if you ignored the relationship between Amount of Rainfall and Crop Yield, the sampling process might choose a low value for the rainfall at the same time it choose a high value for the crop yield — clearly something nature wouldn't allow. Similarly, suppose we may encounter high costs of steel in the world market place. If this occurred, we would experience high than expected costs for all items in a project that contained significant amounts of steel e.g. structural steel (columns, beams), tanks, piping, etc. So, we could have inconsistent sampling by selecting a high cost for structural steel but a low cost for steel piping or tanks. Setting correlations prevents this inconsistency from occurring.

Reasons for correlation between variables include:

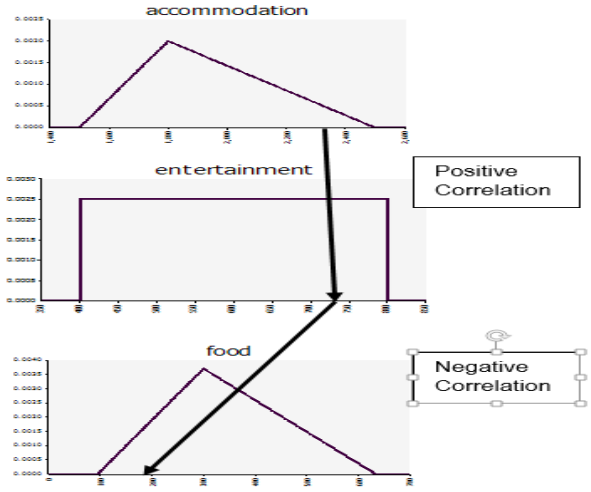
- *Common cause* – Variables may have a common cause or driver that affects each is a similar way e.g. poor productivity of the same resource or technology being applied to more than one variable; or variables have a common component (e.g. steel beams and steel columns)
- *Cause*-and-effect –eg interest rates affect sale prices, rainfall affects crop yield

Correlation refers to the relationship between dependent variables. Correlated occurs in one of two ways:

- *Positive* – Positive correlation means an increase (or decrease) in one variable might be generally associated with an increase (or decrease) in another variable. So, when the value of a variable is say high then there is a higher chance of a high value also occurring for dependent variables. For example, if item A costs more than expected because of market pressure, and B is associated with the same market, then the cost of B will be positively correlated with that of A. Positive correlation is common in project cost models.
- *Negative* – Negative correlation means an increase in one variable might be generally associated with a decrease in another variable i.e. if one variable is say low then dependent variables will be high. For example, in a property development model, if interest rates are high, then sale prices for houses are likely to be low.

Furthermore, the strength of correlation may be:

- *Total* (100%).
- *Partial* (i.e. less than 100%) – As the value of one variable varies the dependent variables tend to follow but they are not completely locked together. So, for example with partial positive correlation, a high value for one variable will occasionally be accompanied by a low value of a correlated variable. The stronger the correlation the less likely this will happen, but it is always possible



The degree of correlation between variables is represented by a *correlation coefficient.* This ranges from +1 if variables have a perfect positive correlation to –1 when they have a perfectly negative correlation. A zero correlation means there is no correlation.

Correlation is common in cost and schedule models in projects but often ignored in risk modelling by naively assuming *independence* amongst variables due to the difficulty of modelling dependence. In fact correlation between variables in almost inevitable - "*An assumption of high dependence generally has a much closer correspondence to reality than an assumption of independence. To minimise bias it is usually better to begin with an assumption of significant dependence … in practice assuming perfect positive correlation of costs items is usually closer to the truth than assuming independence*" (Chapman & Ward, 2011). In fact, the assessment of correlation among dependent variables in more critical in terms of the accuracy of the result that the choice of shape for input probability distributions.

It is important to not confuse correlation with mathematical relationships in a model. For example, a cost model has two random variables to calculate the cost of painting: area of painting (m2) * cost/m2. The cost of painting is derived by multiplying these two variables. These two variables have a mathematical relationship, not a correlated relationship, i.e. the value of one variable (area) is *independent* of the other value (cost/m2). The mathematical link between these two variables is represented in the model by multiplying the two variables, not by correlation coefficients.

2.2.2. Correlation – Effects on Monte Carlo Simulation

Correlation between variables must be taken into account in MCS so that incompatible values are not selected for correlated variables. In effect the random selection process is restricted. So if we determine in advance that two or more variables are correlated then we use a correlation coefficient to represent this in MCS. Consequently variables that are correlated are simulated together so that the value selected from the distribution of one variable will determine which of several conditional distributions should be used to give the value for dependent variables.

The ignoring of correlations will allow high values of one variable to be randomly cancelled out by low values of another, leaving only central values to dominant the output results. However, if dependent variables are positively correlated – which is a common situation in projects - then generally both be high, or both be low or both be in between. So if correlations are ignored, the resultant probability distribution produced by simulation will not represent the greater variability created by correlated variables i.e. the resultant probability distribution will be *narrower* if all variables area assumed to be *independent* because high values of one variable to be cancelled out by low values, and vice versa. If variables are positively correlated then during one iteration of MCS, if a high value is selected for one variable then other dependent variables will also tend to have high value selected thereby generating an extreme outcome. Similarly, if a low value is selected for one variable then other positively dependent variables will also tend to have low value selected thereby generating an extreme outcome. So positive correlations make extreme outcomes more likely. Conversely negative correlations, which are less common, can substantially reduce variability.

Oracle (2009) compare two alternative projects that highlight the effect of high values and low values cancelling each other out and the consequent need for correlation in MCS. Project #1 consists of a single 100-day (800 hour) task, represented by a 80/100/150% distribution (ie minimum 640 hours, most likely 800 hours, maximum 1200 hours) ; Project #2 has ten 10-day (80-hours task), each task represented by a 80/100/150% (i.e. minimum 64 hours, most likely 80 hours, maximum 120 hours) distribution. MSC is applied to both. The output distribution for Project #2 is much narrower, due to the Central Limit Theorem. But when all the tasks in Project #2 are positively correlated with a coefficient of 0.80, the outputs for both projects are very similar. This shows that the greater the number of independent/uncorrelated tasks, the tighter the output distribution from MCS. So, for example, where a project has numerous cost variables, the greater the need to cater for correlation to offset this distorting narrowing effect.

2.2.3. Correlation - Calculation

Any correlations between dependent variables must be made explicit by defining correlation coefficients. However a major problem is that the determination of correct correlations between variables can be difficult and sometimes impossible because there is insufficient historical data to compute correlation coefficients. Furthermore, where dependence between variables is suspected, it may be difficult because it may not be known whether the correlation is genuine or coincidental. Generally, the greater the number of variables the higher the chance that dependence exists. Therefore, one way to avoid dependence is to segment the variables into fewer groups or by grouping dependent variables into a single variable.

Approaches for the determination of correlations include the following:

- There are often no data on correlation between variables, so subjective expert judgement can be applied to arrive at correlation coefficient. Hudak & Maxwell (2007) provide an approach that relates correlation coefficient values to qualitative descriptions:

Correlation – Qualitative Assessment	Estimated Correlation Coefficient
Very high correlation	0.9
Moderately high correlation	0.7
Moderate correlation	0.5
Minor correlation	0.3
Little or no correlation	0.1

- Chau (1995b) suggest the use of expert opinion for the degree of dependence as strong, medium and weak, each of which is assigned a quantitative value (e.g. 0.85 for strong; 0.55 for medium, 0.25 for weak)
- Most people find the estimation of partial correlation more than it is worth, so Grey (1995) advocates that "*when using correlation in a project cost or schedule risk model, it is sufficient to say it is either zero or 100%"*.
- Chapman & Ward (2011) typically assume, unless there is reason to believe otherwise, a 80% positive correlation between cost items and a 50% correlation for related project duration distributions
- Correlations less than 70% may have negligible effect of MCS results (Oracle 2009) and so ignored
- US Government Accountability Office (2009) -"*it is better to insert an overall nomination of 0.25 than to show no correlation at all*".
- Garlick (2007) suggest some people simply select 0.50, which "*while this lacks any kind of justification, it is probably as good as any other number*"
- Hillson & Simon (2007) - "*in the absence of better data, it is common to use values of 0.7 or 0.8*"
- Raydugin (2013): only specify main correlations ie between 0.8-1.0

Inconsistent Correlation

When correlation coefficient are determined subjectively, it is possible to arrive at correlations that are logically impossible. For example, the following correlation coefficients between 3 cost elements – A.B C – have been subjectively agreed by the project team (Hullett, 2011):

Correlated cost variables	correlation coefficients
A and B	0.9
A and C	0.9
B and C	0.2

If A is strongly correlated to B and C, then logically B and C must be strongly correlated, but the table created by subjectively shows this is not the case. MCS software uses something known as the Eigenvalue Test that alerts the user if the correlation coefficients are inconsistent, and suggest correlation coefficients that are logically consistent. For example, it may suggest that A and B, and A and C should be 0.765, and B and C should be 0.17 (Hullett, 2011).This means that whilst the correlation coefficients may not be correct, they are not mathematically impossible. Using the revised coefficient mean we can proceed with the simulation.

2.3. Monte Carlo Simulation – Applications

Application – Testing Accuracy

One application is to compare the MCS outcomes with the expected accuracy ranges stipulated by the organisation. For example, an organisation may stipulate that an estimate produced at the feasibility phase of a project has an expected accuracy range of –15% / +20% with an 80% confidence level i.e. 80% probability that the cost will be with -15%/+20% range and 20% it will outside this range. Assume that the base estimate (including contingencies) using traditional estimating is $100,000. Then, according to organisational policy, it is expected that there is 80% probability that the final cost will actually be between $85000 (-15%) and $120,000 (+20%).

We can then use the results from MCS to judge the degree of compliance with estimate accuracy policy. For example, from the MCS output distribution for total cost, we can discover the probability of the project cost being between $85000-$120000. :

A. Let's say that the MCS results show that the probability of the project cost being between $85000-$120000 is 70%. So the MCS results suggest that this project has greater variability/uncertainty that expected because there is a 30% probability the project may be outside the –15% / +20%, but policy expects a 20% chance. We then make take actions to improve the accuracy, such as collect more data

B. Let's say that according to the MCS results, the probability of the project cost being between $85000-$120000 is 90%.This means that the MCS results suggest that this project has less variability/uncertainty that expected because there is a 10% probability the project may be outside the –15% / +20%, but policy expects a 20% chance. So no further modelling and analysis is needed

Testing the effects of risk management
MCS can be used to test the effectiveness of risk treatments on risk events by running separate simulations pre-treatment and post-treatment of risk events. So Monte Carlo simulations are run with risk untreated and so probability and consequence values are at their *inherent* values and an output distribution produced. The model is revised to include the *residual* ratings for probability and consequences for risk events after their treatments. The MCS is then run again and the output probability distribution can be compared, particularly in terms of comparing the reduction in the risk profile of the output distributions against the costs of the treatments.

Contingency
A key use of MCS is to derive the mount of contingency. This is dealt with in detail in Topic 12

2.4. Monte Carlo Simulation - Advantages & Disadvantages

Advantages

- The output probability distribution of the project's 'bottom line' (e.g., cost, duration, profit) provides numerical details and pictorial representation of the project's risk profile.
- It takes explicit consideration that the values of random variables in a project model can vary simultaneously.
- MCS can use almost any type of distribution
- Software is readily available and relatively inexpensive
- It can deal with interdependencies between variables
- It is can be used of most sizes and types of projects
- It can be applied to a wide variety of project models, such as budgets, schedules and investment analysis.
- It is a logical extension to the single-valued deterministic approach

Disadvantages

- Simulation is limited to those situations where probability distributions can be established for a number of variables.
- MCS may be based on poor quality input data and assumptions (ie garbage in, garbage out) particularly as MCS tends to rely to some degree on subjective judgment.

2.5. Monte Carlo Simulation – Challenges in Collecting Data

Hullett (2001; 2011) highlights the problems of collecting data for producing input probability distributions

Individual factors

- Expressing a values of a variable in a range (pessimistic, most likely, optimistic), rather than one value may be seen as admitting one does not know the value of variables .
- The concept of a range of values for a variable, or the probability of a risk event, is new to many people, which make them uncomfortable
- People may find it difficult to propose values from their limited experience
- People do not have training in specifying optimistic and pessimistic values, being more use to making single-point deterministic estimates
- Errors caused by cognitive bias i.e. heuristics

Organisational Culture

- An organisation may have provided information to stakeholders based on specific numbers e.g. obtained $100m loan based on project cost estimate. As risk analysis deals with ranges rather than deterministic values, the organisation may be reluctant to even consider risk analysis.
- The sponsor may have decided the project will cost a specific amount. Contemplating a risk analysis is not comfortable for the project team because it challenges the sponsor's predetermined targets.
- Risk analysis essentially challenges the basis of a cost estimate, which may make some team members uncomfortable
- Risk analysis may be seen as unconventional or too difficult for the organisation. So there is resistant or hostility to efforts to raise the thought that risks may occur.
- The collection of information for risk analysis inevitably interferes with the project as it requires the time of important project people, which may cause animosity or fear. The project manager "*is key in making information collection a priority and a safe thing to do*"
- Biased information may be given either because the participants do not want to participate in the risk analysis process; or, they want to bias the results for their own purposes; or, they deny risk exists for the purpose of gaining competitive advantage

3.0. REFERENCES

CHAPMAN, C & WARD, S. (2011). *How to manage project opportunity and risk.* Wiley, Chichester.
CHAU, K.W. (1995b). *"Monte Carlo Simulation of Construction Costs Using Subjective Data"* Construction Management and Economics. Vol. 13. pp. 369-383.
FLANAGAN, R. & NORMAN, G (1993). Risk Management and Construction. Blackwell Scientific, Oxford.
GOVERNMENT ACCOUNTABILITY OFFICE (2009). *GAO Cost Estimating and Assessment Guide,* GAO-09-3SP
GARLICK, A. (2007). *Estimating Risk: a Management Approach,* Gower, Aldershot, England.
GREY, S (1995). *Practical Risk Assessment for Project Management.* John Wiley. Chichester.
HILLSON, D & SIMON, P. (2007). *Practical Project Risk Management: The ATOM Methodology.* Management Concepts Vienna, VA.
HUDAK, D & MAXWELL, M (2007). A macro approach to estimating correlated random variables in engineering production projects, Construction Management & Economics, 25, 883-892
ORACLE (2009). *Managing Risks in Primavera Risk Analysis Rel. 8.6. – Student Guide*

APPENDIX - 2-week holiday in Margaret River

COST MODEL – Spreadsheet
PD = Probability Distribution;
Red = accomplishment variables;
Green = risk events;
Blue = Excel calculated cell

	A	B	C	D	E	
1	7-day holiday Margaret River				$	excel
2	food				PD	
3	entertainment				PD	
4	accommodation				PD	
5		kms	L/kms	$/L		
6	petrol	PD	PD	PD		E6=B6*C6*E6
7			Prob.	Effect		
8	speeding fine		PD	PD		E8=C8*D8
9	sick: doctor		PD	PD		E9=C9*D9
10				total		E10=sum(E2:E9)

COST MODEL – Probability Distributions

	Prob.	min	Likely	max	shape
food		200	300	500	Trigen
entertainment		400		800	uniform
accommodation		1500	1800	2500	Triangular
Petrol (kms)		600	700	1000	kms-BetaPert
Petrol (L/kms)		0.09	0.10	0.15	L/km-Triangular
Petrol ($/L)		1.25	1.30	1.50	petrol - Triangular
speeding fine	0.30	Speeding: <9km, 9-19km, 20-29km	fine: $100, $200, $400	probability: 0.7, 0.2, 0.1	Fine Probability - binomial Effects - discrete
sick: doctor	0.20	40	40	200	Sick Probability -binomial Effects - Triangular