Probability

Probability is central to project management. We need to understand the probability of achieving cost or time outcomes and the probability of a risk event occurring. Probability is also closely aligned with possibility and creativity. Certainty eliminates the need to consider options. Only if you consider there is a possibility the current situation / solution is not optimal can you investigate possible improvements.

The problem with probability is understanding what we know and what we don't – society is built on certainty. Certainly Western culture for the last 2400 years has built on the thinking of Socrates, Plato and Aristotle that uses logic and argument to determine the truth. These ideas were central to the Churches of the Middle Ages (it's very difficult to be a martyr for a probability) and is a core doctrine of our judicial systems and modern management. Argument based on logic defines the truth!

Probability theory

Based on this very successful paradigm, modern risk management practices have developed analytical methodologies to determine the probability of events occurring (or not occurring) that allows contingencies to be calculated based on mathematical certainties. At a simple level is a perfectly sensible process. The probability of a 'six' showing if you role one dice once is 1/6 (there are 6 sides and the answer will be 1, 2, 3, 4, 5, or 6). If you roll a pair of dice, the probability of one 'six' showing is 11/36 (not 2/6 or 1/3) the reason for this is there is a possibility of two sixes showing. The calculation is:

 - The probability of the first dice showing a 'six' is 1/6

 - The probability of the second dice showing a 'six' is 1/6

 - However, the probability of both dice showing 'six' is 1/36

There are 36 possible combinations: (1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)
(2, 1) (2, 2) (2, 3) (2, 4) (2, 5) (2, 6)
(3, 1) (3, 2) (3, 3) (3, 4) (3, 5) (3, 6)
(4, 1) (4, 2) (4, 3) (4, 4) (4, 5) (4, 6)
(5, 1) (5, 2) (5, 3) (5, 4) (5, 5) (5, 6)
(6, 1) (6, 2) (6, 3) (6, 4) (6, 5) **(6, 6)**

Therefore the probability of only 1 six showing is: 6/36 + 6/36 - 1/36 = 11/36[1]   If the question asked for the probability of at least one 'six' (ie, either 1 or 2 sixes), the answer would be: 1/6 + 1/6 = 1/3 the 'two sixes' option would not matter.

The basis of most probability calculations are similar to the above and rely on two factors. The first is the 'classification' of the event; the second is the assumption there is a bounded range of outcomes. However, even in these circumstances, there is a significant difference between the 'chance' of an event occurring and the 'probability' of it occurring.

Every one knows the chance of a coin when it is tossed landing on 'heads' is 50%; the coin will either land on 'heads' or 'tails'. The first vital consideration is that each throw of the coin is independent. Therefore for any single toss of the coin there is always a 50% chance of 'heads' being the outcome even if the coin has landed on 'heads' in 10 or even 100 previous throws. However, whilst the 'odds' in favour of any single toss landing on 'heads' is 1 to 1 (or evens) the probability of 10 consecutive throws landing on 'heads' is extremely low. The probability of 'heads' being tossed 10 times in a row is 1/2 to the power of 10 = 0.00048828125 (or roughly 1/2000). But remember whilst the probability of tossing 10 'heads' in a row is very low, the chance of any single toss of the coin coming up 'heads' remains 50/50.

---

[1] This is part of 'Probability Theory' when two events occur the probabilities can be added if they are independent (the roll of the first dice has no effect on the roll of the second) but should be multiplied if they are interdependent.

Therefore, whilst over many thousands of rolls of the pair of dice, the average number of times a single six will appear will be close to 11/36 and the chance of any one roll of the dice showing a single six is 11/36; the probability of seeing 11 single 'sixes' from just 36 rolls of the dice is extremely low. This is part of the 'Law of Large Numbers'.

Similarly, if a test has a 99% chance of being successful, the probability of 50 successful tests in a row is 61.73% this is calculated by multiplying 0.99 * 0.99 fifty times. If the probability of a successful outcome is 98% the probability of 50 successful tests in a row reduces to 37.92%

Bounded v Unbounded Ranges

The problem with project data is that any probability calculation is based on the assumption of a finite range of variables; there are only 36 possible options for a pair of dice. Conversely, there are no limits to many aspects of project risk. Consider the following:

- You plot the distribution and average the weight of 1000 adult males. Adding another person, even if he is the heaviest person in the world only makes a small difference to the average. No one weighs a ton! The results are normal (Gaussian-Poisson) and theorems in probability such as the Law of Large Numbers and Least Squares (Standard Deviation) apply.

- You plot the distribution and average the net wealth of 1000 people. Adding Bill Gates to the group causes a quantum change in the values. Unlike weight, wealth can be unlimited. Gaussian-Poisson theories do not apply!

Most texts and discussion on risk assume reasonable/predictable limits. Managing variables with no known range of results is rarely discussed and many project variables are in this category[2].

Classification of Events

The next problem with project data is finding 'similar data'. Risk management tools and methodologies such as Monte Carlo assume probabilities (or more correctly, probability distributions) for durations, costs, etc to be either known, or they can be accurately determined from relevant historical data. The word relevant is critical: it emphasises that the data used to calculate probabilities (or distributions) should be from situations that are similar to the one at hand. *This apparently simple requirement papers over a fundamental problem in the foundations of probability: the reference class problem*[3].

What does 'similar' mean? Do we look at projects with similar scope, or do we use size (in terms of budget, resources or other measure), or technology or some other measure? There could be a range of criteria that could be used, but one never knows with certainty which are relevant in the particular circumstances. This is an issue because the probability changes depending on the classification criteria used…… this is the reference class problem.

Creating a probability distribution based on data for completed projects that are similar to the one of interest requires:

- Collecting data for a number of similar past projects – these projects form the reference class. The reference class must encompass a sufficient number of projects to produce a meaningful statistical distribution, but individual projects must be similar to the project of interest.

---

[2] For more on probability and risk see:
    The Meaning of Risk in an Uncertain World: http://www.mosaicprojects.com.au/Resources_Papers_040.html
    Scheduling in the Age of Complexity: http://www.mosaicprojects.com.au/Resources_Papers_089.html

[3] Quoting from Eight to Late, *The reference class problem and its implications for project management*:
    http://eight2late.wordpress.com/2010/05/13/the-reference-class-problem-and-its-implications-for-project-management/

- Establishing a probability distribution based on reliable data for the reference class. The challenge here is to get good data for a sufficient number of reference class projects.
- Predicting most likely outcomes for the project of interest based on comparisons with the reference class distribution.

Simple until you realise by definition we never do the same project twice; each project is a *'temporary endeavour, having a defined beginning and end, undertaken to meet unique goals and objectives'*. At best you can gather data from 'similar' projects selected on the basis of common sense or expert judgement or some other subjective approach.

The reference class problem affects most probabilistic methods in project management[4]. It is a problem because it is often impossible to know, which attributes of the projects are the key variables that relate to the event of interest. Consequently it is impossible to determine with certainty whether or not a particular event or project belongs to a defined reference class.

Conclusion

There are no simple solutions to this problem. Basing project decisions on the assumption the data being used is accurate is doomed to failure – it is impossible to predict the future with certainty (if we could casinos and bookmakers would go bankrupt).

However, the alternative of using probability based calculations to define risk and uncertainty has definite limitations. Risk assessments, Monte Carlo analysis and other techniques will provide valuable insights if they are used wisely. What they cannot do is provide certainty.

Unlike motor vehicle and life insurance companies (where the Law of Large Numbers apply), in project management, we simply don't have enough similar data for probabilistic analysis to provide accurate assessments of the probable range of outcomes.

Skilful project managers recognise this lack of certainty, do sufficient analysis to develop a reasonable level of insight (avoiding analysis paralysis) and then manage the ongoing work knowing their estimates are not correct. This allows the key question of 'how wrong are the estimates' to be asked.

Knowing your estimates are wrong allows system to be put in place so that as the expected variances start to emerge, routine adjustments to the project plans are made to lock in gains and mitigate losses. Adapting to an uncertain future requires different skills to those developed by the 'scientific management school' of the 19th Century but are essential in the 'age of complexity'[5].

_____

---

[4] Bayesian Networks offer one solution to this problem but are very complex to establish for short term one-off situations such as a project.

[5] See **Scheduling in the Age of Complexity**: http://www.mosaicprojects.com.au/Resources_Papers_089.html