# Probability and Statistics: To $p$, or not to $p$?

## Module Leader: Dr James Abdey

## 6.5 Monte Carlo simulation

We are forever bound by **uncertainty**. To study, or not to study? To invest, or not to invest? To marry, or not to marry? Other examples are:

- what will be the profits of our new product?
- what are the best logistics for our new supply chain?

Often, several different system configurations are available. For example, from our R&D portfolio, there are several products we could release, but we have to select just one due to our marketing budget constraint.

Trying different solutions is **expensive** and/or **risky**. For example:

- testing different products in the market
- designing a new schedule of arrivals/departures for an airport.

**Improving the system** is difficult. For example:

- which is the most profitable product to be released in the market?
- what is the optimal flight schedule for an airport?

Finding a **robust system** design is a challenge. For example:

- a product which sells well under tough market conditions
- an airport schedule which can deal with delays and disruptions.

**Monte Carlo simulation** works by simulating multiple hypothetical future worlds reflecting that the outcome variable of interest is a *random* variable with a probability distribution. We can then determine the **expected outcome**, as well as a **quantification of risk**, using the usual statistics of mean and variance. Decision-makers can then make an informed judgement about the best course of action **based on their risk appetite**.

To perform a Monte Carlo simulation we proceed as follows.

1. Associate a (pseudo-)random number generator for each input variable.
2. Assign a range of random numbers for each input variable (according to some assumed probability distribution).
3. For each input variable:
   - generate a random number
   - from the random number, select the respective variable value.
4. Calculate the outcome, $x$, and record it.
5. Repeat '3.' and '4.' until the desired number of iterations, $N$, is reached.
6. Draw a histogram of the outcomes and determine the mean and variance of the simulated outputs:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

   and:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2.$$

   The mean gives our estimate of the expected outcome, while the variance is our measure of risk.

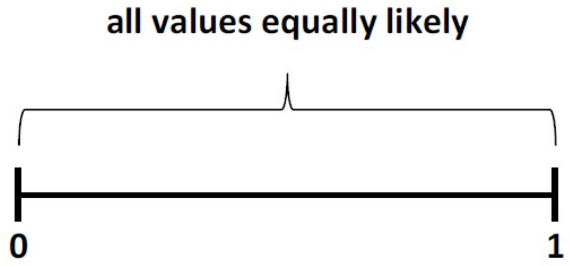
## Example

A company wants to analyse the following investment with the following uncertain revenues and costs given by the probability distributions below. Assume, for simplicity, that the revenues and costs are uncorrelated.

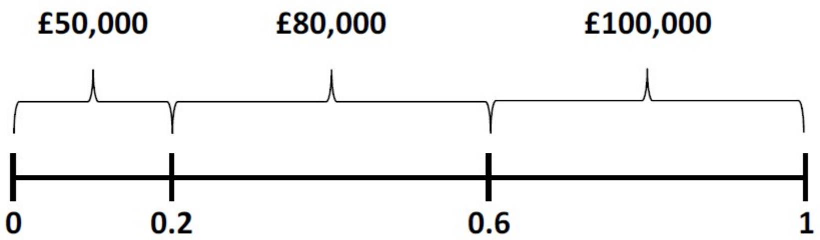
| Revenues | £50,000 | £80,000 | £100,000 |
|---|---|---|---|
| $P$(Revenues) | 0.2 | 0.4 | 0.4 |

| Costs | £30,000 | £60,000 | £80,000 |
|---|---|---|---|
| $P$(Costs) | 0.2 | 0.6 | 0.2 |

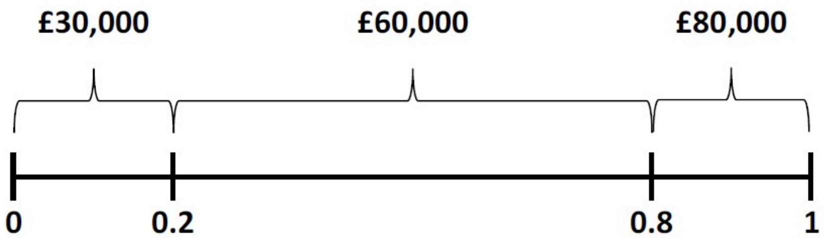
By using Microsoft Excel's '`=RAND()`' function, we can randomly select a number equally likely to be between 0 and 1:




Given the probability distribution for revenues, we could simulate revenues depending on the value returned by '`=RAND()`' according to:




which ensures each revenue value occurs with the required probabilities. Similarly, for costs, we could use:




Suppose for the first simulation we had random numbers for revenues and costs of 0.7467 and 0.1672, respectively. This correspends to:

$$\text{Revenue} = £100{,}000$$

$$\text{Costs} = £30{,}000$$

$$\text{Profit} = £70{,}000.$$

Suppose for the second simulation we had random numbers for revenues and costs of 0.0384 and 0.9056, respectively. This corresponds to:

$$\text{Revenue} = \pounds 50{,}000$$

$$\text{Costs} = \pounds 80{,}000$$

$$\text{Profit} = -\pounds 30{,}000 \text{ (a loss).}$$

This process would then be replicated ideally thousands of times ($N$ = the number of simulations), resulting in the calculation of the mean and variance of profit to estimate the expected outcome and risk.

Of course, the quality of these expected outcome and risk estimates is only as good as the model used. What if the true probability distributions were different? What if revenues and costs were correlated? What if other factors affected profit as well?

Clearly, in practice we would wish to conduct a **sensitivity analysis** to see how sensitive (or robust) the distribution of the outcome variable is to such issues.