Data and Decision Making 36109

Assignment 1: Monte Carlo simulation of a project

Declan Stockdale

11214549

Introduction

This report details the process behind the Monte Carlo simulation of a system implementation. It consists of 4 tasks where task 2 runs in parallel, and the other tasks run one after the other. The values of the minimum, most likely and maximum time for each task to be completed can be seen in figure 1.

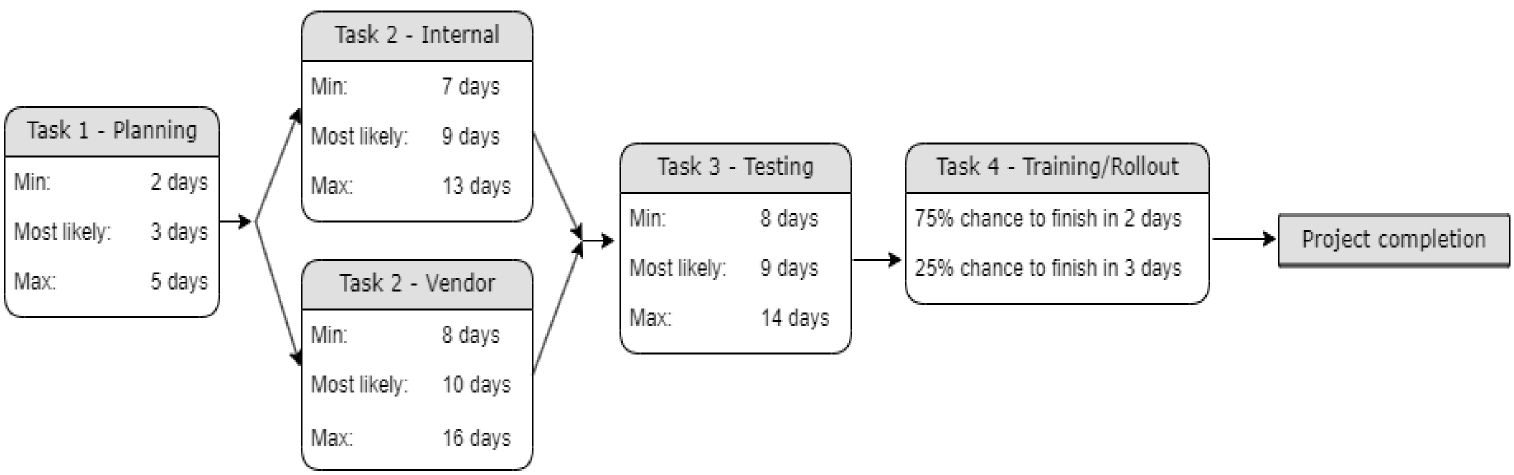

Figure 1. Project sequence diagram

Project timeline simulation

A Monte Carlo simulation using 1000 iterations was used to sample various potential project completion lengths. The minimum, most likely and maximum day was used to construct a triangular distribution for each task. The simulation was coded in R and the script will delivered in addition to this report. Monte Carlo is a stochastic process; the seed has been set to 1 to ensure reproducibility.

Results

After running the simulation 1000 times a probability distribution histogram and curve (PDH and PDC) (figure 2) was created as well as the cumulative probability distribution curve (CPD) (figure 3).

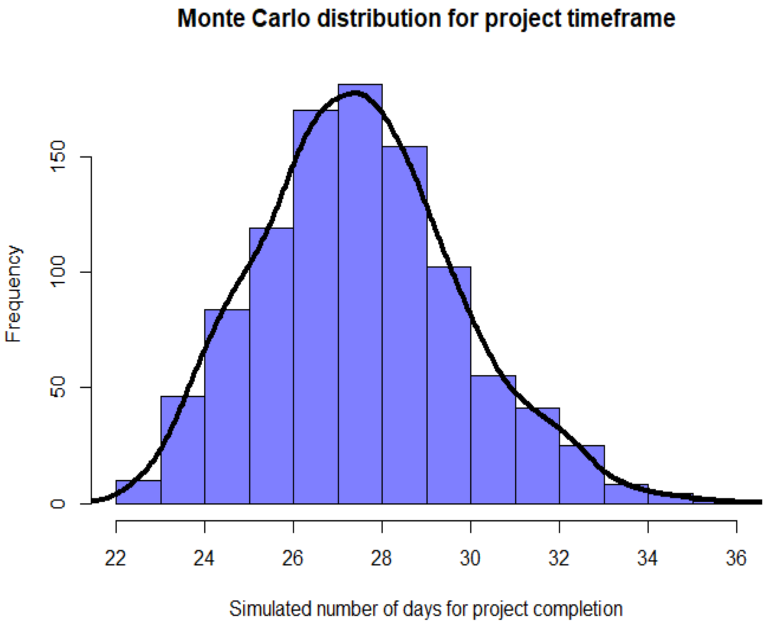

Figure 2. Histogram of number of days for project completion

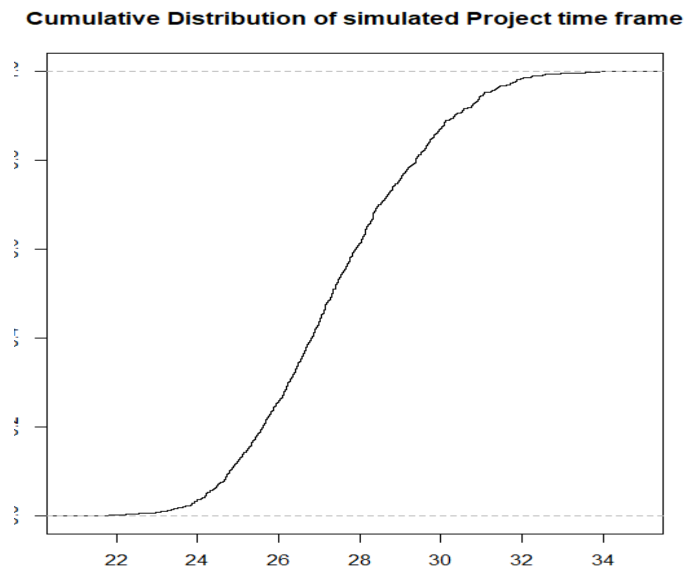

Figure 3. Cumulative distribution of project estimations

Project estimation dates should be made using the CPD as it's the probability of completion by a certain day rather than on a certain day. The diction is that while inspecting figure 2, 27 is the most likely day for project completion but using that gives no indication of the project was completed before the 27th day whereas the CPD provides the probability that it will be completed before or on a specified date.

Cumulative project completion probabilities

Day	22	23	24	25	26	27	28	29
Cum Prob %	0.0	1.0	5.6	14.0	25.9	42.9	61.0	76.4

Day	30	31	32	33	34	35	36
Cum Prob %	86.6	92.1	96.2	98.7	99.5	99.9	100

Table 1 Cumulative probability of project completion

Using the cumulative probability in table 1, the likelyhood of project completion by day 27 is 42.9%

Probability of completion %	10	20	30	40	50	60	70	80	90	100
Days	24.6	25.6	26.2	26.8	27.4	27.9	28.6	29.3	30.5	35.6
Days (rounded up)	25	26	27	27	28	28	29	30	31	36

Table 2. Probability of project completion

Table 2 shows the days for given project completion probabilities. The days have been rounded up in the last row as the project must be completed by a given day, not partially through a day.

Model Rationale and Assumptions

The rationale is that the process of knowing the completion date is very complex due to substantial uncertainties. To get around this, simulations will be required. To simulate a single task, a random number is generated which is then converted to a number of days for that task to be completed. By performing the same process for all tasks n times, in our case 1000 times we are able to generate 1000 potential project completion timeframes which we can then make informed estimates using the cumulative distribution curve above (fig 2).

Various assumption have been made in the model. The first is that a triangular distribution is a realistic representation of the probability of the project completion timeframe. Other distribution such as PERT are available which may be more suitable to this task. Another assumption is that as soon as one task is completed, work on the next task will begin. This may not be realistic as workers may “wind down” if a task is completed near to close of business, waiting until the next day to start the next task. A final assumption is that the estimations given by each team are accurate. This may be wrong as they could be poor estimators, or it’s based on historical data which may not reflect current practices.

Accounting for late projects

Delays will likely have knock on effects to subsequent stages. A brief explanation of the process implemented to model this is as follows.

1. Generate matrix of size tasks (5) x simulations (1000)
2. Each task in each row s assigned a random number between 0 and 1.
3. Calculate the Prob of the most likely day occurring using triangular distribution starting at second column
   a) If the random number is less, then move to next column
   b) If not resample random number between prob of the most likely day and 1
      a. Move to next column and repeat process

This method ensures that if any task occurs above the most likely time, all subsequent tasks do so as well.

Assumptions around this process

The implementation does not distinguish between task 2 internal and vendor however if either is late, the parallel task is late so the distinction between the columns is irrelevant. Also, the last column is left alone as it can only have values 2 and will have minimal impact on the final result.

Late Results

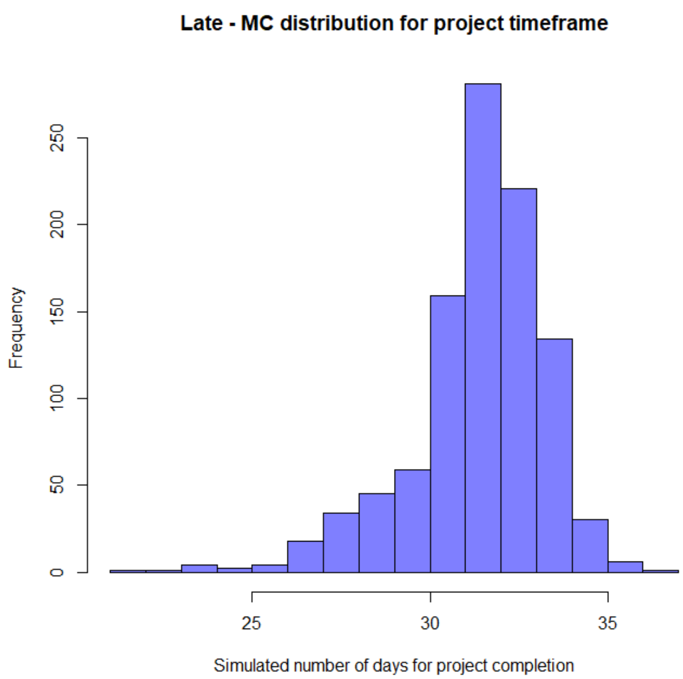

Figure 4. Histogram of simulated projects using modified method

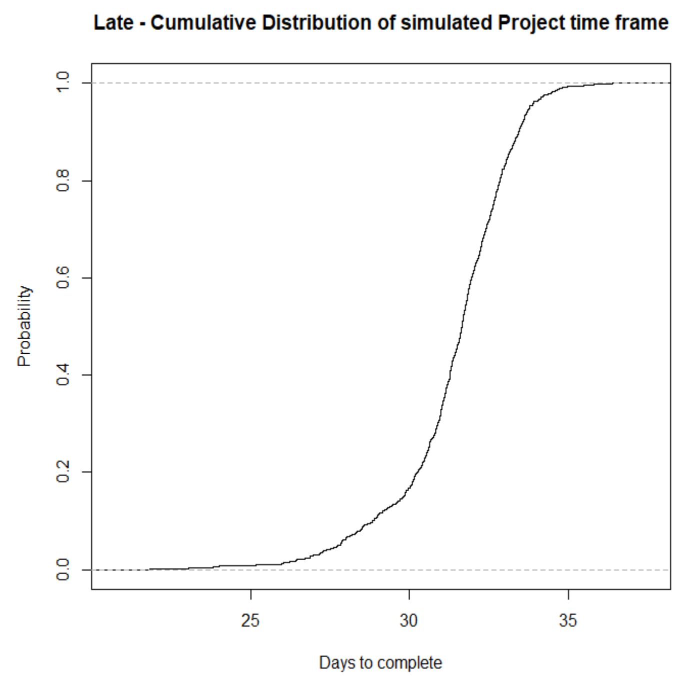

Figure 5. CDF for late projects using modified method

It's clear that any delay especially early on impacts the total number of days until prpoject completion. Figure 4 has a distinct right skew towards higher completion days and is also evident in the sharp rise in the curve in figure 5.

The tables below provideadditional information about the cumulative probability (table 3) and the probability of completion (table 4).

Day	21	22	23	24	25	26	27	28	29
Cum Prob %	0	0.1	0.2	0.6	0.8	1.2	3.0	6.4	10.9

Day	30	31	32	33	34	35	36	37
Cum Prob %	16.8	32.7	60.8	82.9	96.3	99.3	99.9	100

Table 3. Cumulative probabilities of project completion

Probability of completion %	10	20	30	40	50	60	70	80	90	100
Days	28.8	30.3	30.9	31.3	31.7	32.0	32.4	32.9	33.4	36.4
Days (rounded up)	29	31	31	32	32	32	33	33	34	37

Table 4. Probability of project being completed by a given day