`

Important Remarks to the Reader.

In hindsight the simulation could have been improved by a number of things:

Problem 1:

Better Code design:

The initial design was such that the functions generate data by user specification, thereby reducing the need for the user to go and create new data .However ,this being a simulation study , the motive should have been to be able to reproduce the results of simulations . And such, setting seed became an issue and have not been set in the code.

Setting the seed ,will remove all randomness from the norm code created using the initial design phase (due to the dependencies on the calculation of p values for n simulations) This would therefore return the same p value for any no of simulations and a power and size of either 0 or 1 only.

Solution:

Instead of first creating generate data function, and passing it as arguments to smaller functions such as calculate.t.test, calculate.p.value …-> and finally using this to create a wrapper function.

The cycle should have been reversed. The data should have been created with a seed set for a certain no of simulations (say 500), The 500 different sample of data with specific seed numbers should be **passed** as an argument for the **main wrapper/simulation driver** and can be used to test the different scenarios .The lower level functions such as calculate .pvalue functions will therefore, extract columns or vectors accordingly in order to calculate the pvalue for each of the three test.

This would ensure that setting seed at main wrapper function will not affect the no of simulations or the values calculated as the smaller level function it is top down rather than down to top level.

(Please see attached design code drawing for more info)

Problem 2:

Simulation Run time: The Run time generally is reasonable, compared to a few exception when used for the size simulations. Again, any optimization of the code did not provide significant improvement due to the problem raised above.

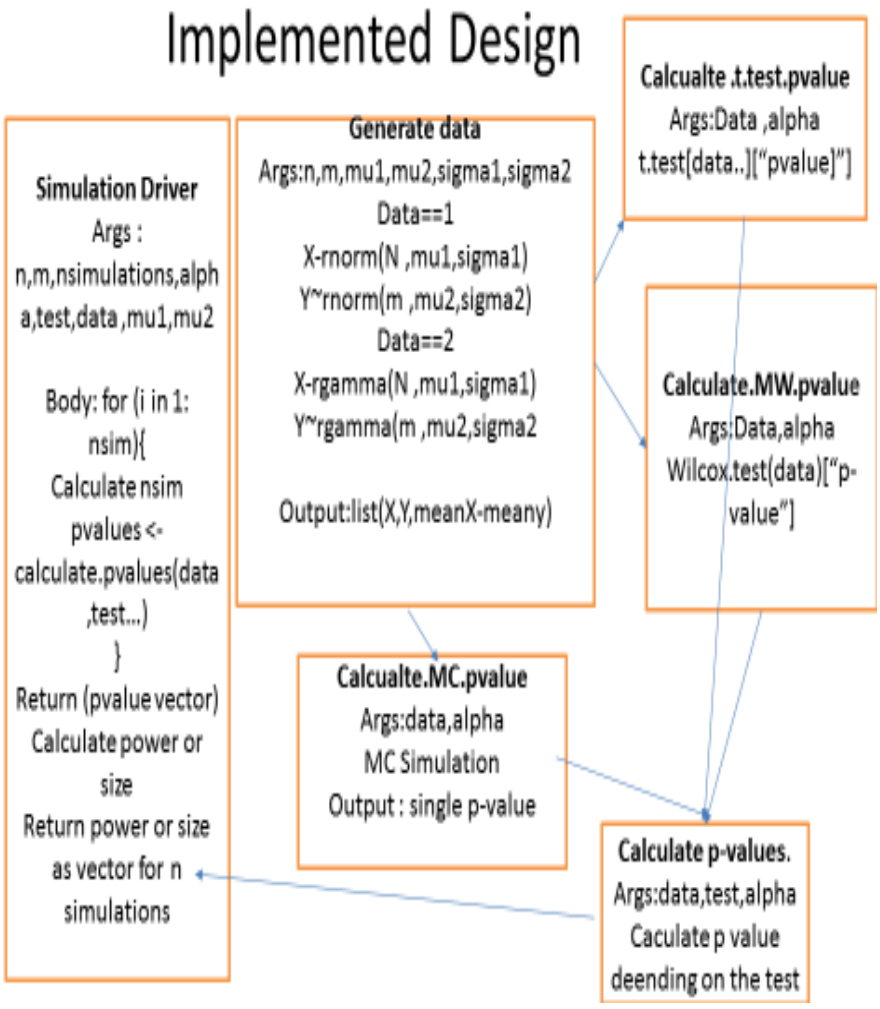
Solution:

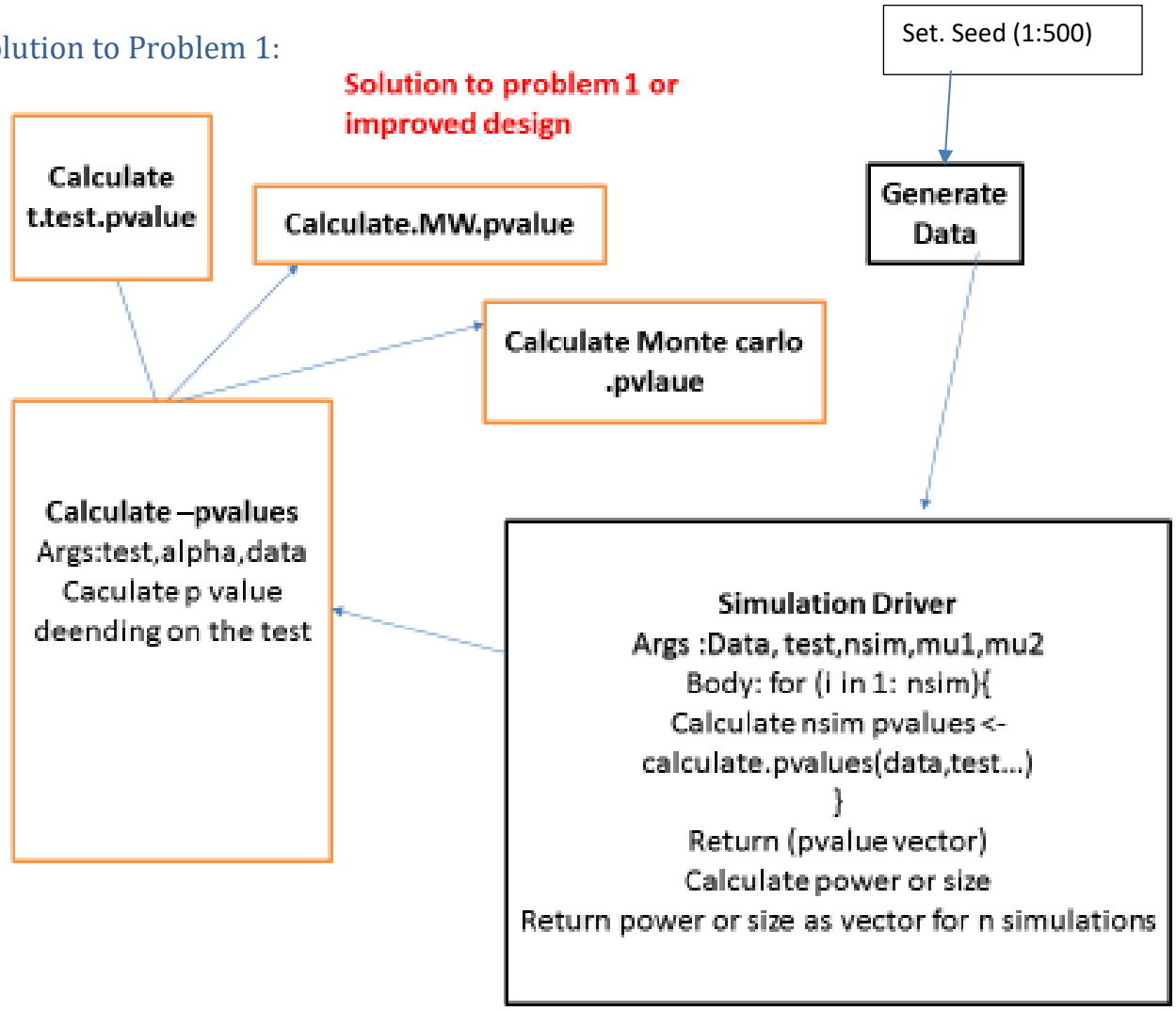
The problem of data being passed to lower functions has had a direct impact on the speed. As seen in the figure below, each simulation call, the function has to pass down all the way through level 0 to generate data for each

To rectify this, the solution to problem 1 has to be rectified. And, one can proceed to save the output for estimates for each simulation in a separate file. And then proceed to look at p values for the desired simulation calling on the file accordingly, thereby avoiding recalculation.

`

Implemented Design



`

Solution to Problem 1: