MAT013 Coursework

Deadline: 8/5/2013 at 0900

Instructions

The outputs of this coursework will be:

- A written report describing your code (SAS and R) to be handed in to Joanna Emery.
- An appendix containing a commented version of your code (SAS and R) to be handed in to Joanna Emery.
- A file containing the required SAS code. Name this file SAS-lastname-STUDENTNUMBER (eg. Knight-123456) and email it to Joanna Emery with MAT013 as the subject. Note that all operations needed to complete the coursework should be included in the SAS code.
- A file containing the required R code. Name this file R-lastname-STUDENTNUMBER (eg. Knight-123456) and email it to Joanna Emery with MAT013 as the subject. Note that all operations needed to complete the coursework should be included in the R code.

Coursework

1. Using both SAS and R (in other words attempt this question using SAS and then using R):

Write code (in SAS: a macro, in R: a function) that will reproduce a mathematical procedure covered in MAT001 or MAT002. Clearly document this procedure in your report.

[20]

2. Using SAS:

The data files Greedy.csv, Random.csv, Longest.csv and Shortest.csv contain data relevant to the experimental play of 4 strategies for the game Shut the Box (you do not necessarily need to know of this game to complete this coursework).

The four strategies will be referred to as:

- `Greedy`

- `Random`
- `Longest`
- `Shortest`

The data file contains two variables for each strategy: `Score` and `Length`. The aim of the game is to have the lowest score (a minimum of 0).

i. Obtain plots of the distribution of the `Score` variable for each strategy (represent these distributions on the same graph);
ii. Obtain plots of the distribution of the `Length` variable for each strategy (represent these distributions on the same graph);
iii. For each method are `Score` and `Length` related?
iv. Do the strategies give different outcomes and if so which strategy seems to be the best?

[25]

3. Using R:

Write a function that will return the nth Fibonacci number, $F(n)$.

Modify the function so that it returns the nth number of the sequence defined by:

$$\begin{aligned} K(0) &= a \\ K(1) &= b \\ K(n) &= \alpha K(n-1) + \beta K(n-2) \end{aligned}$$

Where a, b, α and β are input parameters.

Adapt your function so that it will write all numbers less than k to a csv file. The name of the csv file must not be an input parameter to the function but include the parameters a, b, α and β as well as the date on which the code was run. For example: `general_fib_for-a=2-b=3-alpha=10-beta=-2_1984-14-02.csv`.

[25]

4. Using SAS or R.

The file Solution_Space_Exploration.csv contains experimental results pertaining to two approaches to solving an optimisation problem (aiming to minimize a cost function). These two approaches will be referred to as approach `A` and approach `B`. Approach `B` involves searching a space that contains the solution space that approach `A` searches. Thus approach `B` can at least match approach `A`.

Every row of the data file contains 6 variables which are (in order):

- A boolean variable indicating `True` if approach `B` finds a better solution than approach `A`: `B_optimal`;
- The first dimension of the problem: `m`;
- The second dimension of the problem: `n`;
- A further problem parameter: `tau`;
- The optimal cost function obtained using method `A`: `A_Cost`;
- The optimal cost function obtained using method `B`: `B_Cost`;

i. Give summary statistics for all the variables. [5]

ii. Obtain a 3 dimensional representation (eg surface or contour) showing the proportion of times that method `B` finds a better solution based on the dimensions of the problem. [5]

iii. Obtain a distribution of the gains made by method `B` over method `A`. [10]

iv. Explore and attempt to indicate parameters that influence the performance of either method (and when method `B` is better). [10]