
STAT40780 Data Programming with C (online)

Assignment 2 on Rcpp

Due date: Friday 15th July at 10pm

This assessment is worth 20% of the total grade and is marked out of 100. There are 2 sections to this assignment. For full marks, ensure your code is clear and well-commented. Only the following files should be submitted:

- One R script - **YourStudentNumber_assignment2.R** - containing all of the R code for the assignment
- A single PDF document containing any other required information (e.g. descriptions of code implementations), named in the format **YourStudentNumber_assignment2.pdf**.

Any additional files you submit will not be graded. This assignment should be completed individually. Any student who plagiarises will receive a 0 mark.

1 Random Walk

Please read the full question carefully before attempting the solution. Consider a two-dimensional grid map of a city, with an x-axis (horizontal) of streets and y-axis (vertical) of avenues. Coordinates on the map may be represented as a two-dimensional vector (x,y) , representing the intersection points of the streets and avenues. For this problem, the grid is made of discrete integer values with reference to an origin $(0,0)$. Here $(-1, 1)$ is one street to the left, and one avenue above, the origin $(0, 0)$.

A tourist is standing at coordinates $(0, 0)$, and is looking for the tourist information centre, located at $(-1, 3)$. However, the tourist is completely lost and instead of asking for directions, begins a random walk to search for the tourist information centre.

The tourist moves one road at a time, either horizontally or vertically (but can not move diagonally). Movement can be forwards or backwards (positive or negative, respectively). Therefore, at every point, there are 4 possible moves the tourist can make. For instance, when standing at the origin $(0,0)$, the tourist can move either to $(0, 1)$, $(0, -1)$, $(1, 0)$, or $(-1, 0)$, and has an equal probability of 0.25 of moving in each direction. However, the tourist can easily miss the tourist information centre completely.

Use Rcpp to complete the following exercises. You will be graded on the efficiency and readability of your code.

1. Write a code that simulates a random walk over N roads (streets or avenues) with an output stating if the tourist information centre was reached during the walk. Provide an example of calling this code for $N=100$. For repeatability use the set seed function.

Note: the code should identify if the tourist reaches the tourist centre (yes/no or 1/0 or T/F). After the tourist visits the tourist centre, the tourist should finish travelling the N roads (i.e., the code does not stop at the tourist centre).

[40 marks]

2. What is the probability that this tourist locates the tourist information centre **within** walking along 20 roads (i.e., 20 moves/updates of the algorithm)? Approximate this probability through simulation of 1,000 random walks using your code in (a), each starting at $(0, 0)$.

[5 marks]

3. In the PDF document, describe your implementation for your random walk and any specific reasons that it should be particularly efficient.

[10 marks]

4. Write a second version of your code from part (1) above to estimate the average distance the tourist reaches from (0,0) after walking along 50 roads (i.e., 50 moves/updates of the algorithm). Calculate the average distance over 1,000 simulated random walks. Define 'distance' as the minimum number of moves required to travel from the current/final position back to the origin. This question is based on the Manhattan metric.

[15 marks]

2 BubbleSort

Modify the Bubblesort algorithm (see Lab 6 solution code or accompanying code to Section 4.2) so that it can handle missing values - take an integer vector containing NAs and return an ordered vector with NAs at the end of the ordered vector.

[30 marks]