MAS6002/6024/468 Assignment 3 2017/18

You should use `R` for this assignment. **Deadline:** 12 noon, Tuesday 23rd January 2018. Submit it using Turnitin on the MAS6002/6024/468 MOLE page.

Part 1

The `R` file `Assignment_3_part_1_data.R` on MOLE contains data relating to an experiment with a single covariate, x, and a single response, y. Consider the 2-parameter (intercept and gradient) linear model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where

- $\boldsymbol{y}$ is the vector of responses;
- $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ is the parameter vector where β_0 is the parameter representing the intercept and β_1 is the parameter representing the gradient;
- X is the corresponding design matrix of dimension $n \times 2$;
- $\boldsymbol{\epsilon}$ is the error vector of length n given by $\boldsymbol{\epsilon} \sim N(0, 3^2 I_n)$ where I_n is the n by n identity matrix.

The log-likelihood of the parameter vector given the data is

$$l(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}) = c - \frac{1}{18}(\boldsymbol{y} - X\boldsymbol{\beta})^T(\boldsymbol{y} - X\boldsymbol{\beta})$$

where c is a constant.

1. Specify the value of c for the data in `Assignment_3_part_1_data.R`.
2. Write a function to calculate the log-likelihood for values of β_0 and β_1 over a $2-$dimensional grid of β_0 and β_1 values. The function should allow the user to specify the relevant features of the grid. If possible avoid using 'for' loops in your code.
3. Using your answer to part (2.) determine $\hat{\beta}_0$ and $\hat{\beta}_1$, the maximum likelihood estimates (MLEs), to 2 decimal places.

Do not use any of the built-in optimization functions to do this, nor the known result for the MLE in a linear model (although it would be sensible to check using these). You are expected to do exactly what is asked of you in the question.

Part 2

Consider an object moving around the four vertices of a square labelled v_1, v_2, v_3, v_4. At time $t = 1$ it starts at a specified vertex and then 'moves' at time $t = 2, 3, ..., n$. At times $t = 2, 3, ..., n$ the object makes either a 'connected move' with probability p or a 'diagonal move' with probability $1 - p$ independently of both the time at which the object makes the move and the current vertex that the object is visiting. These two types of move are defined here:

- A 'connected move' in which the object can move to one of the two vertices it is connected to by an edge (it cannot make diagonal moves across the square). When the object moves, it moves to each connected vertex with probability 0.5. In a 'connected move' the object cannot stay at the same vertex.

- A 'diagonal move' in which the object moves to the vertex diagonally across the square with probability 1.

Let $X_t, t \in \{1, ..., n\}$ represent the vertex that the object is visiting at time t. For example if the object starts at vertex v_3 at time $t = 1$ and moves to vertex v_4 at time $t = 2$ then $X_1 = v_3$ and $X_2 = v_4$.

We are interested in the proportion of the n visits that occur within a certain subset of the vertices (called the 'target set' of vertices H) for a given starting position and value of p. So we are interested in $\frac{\sum_{t=1}^{n} \mathbb{1}(X_t \in H)}{n}$ where $\mathbb{1}(A) = 1$ if A is true and is zero otherwise.

1. Write a function that simulates the movement of the object around the square, records the vertices visited at times $t \in \{1, ..., n\}$ and returns $\frac{\sum_{t=1}^{n} \mathbb{1}(X_t \in H)}{n}$. Your function should allow the user to specify the value of n, p, the vertices in the target set H and the starting vertex.

2. Use your function to assess how $\frac{\sum_{t=1}^{n} \mathbb{1}(X_t \in H)}{n}$ is affected by n and p when H is the same as the starting vertex. Comment on your results.

Hint: There are many valid approaches to writing this function but you may find it easier to consider the vertices of the unit square (i.e. $\{v_1, v_2, v_3, v_4\} = \{(0, 0), (1, 0), (1, 1), (0, 1)\}$. There are functions in `R` that convert binary strings to integers and vice versa. You may find it easier to answer this question if your main function in (1.) calls several other functions that you have written.

A Simple illustrative example: consider an object that starts at vertex $(0, 0)$ so that $X_1 = (0, 0)$ and makes 4 moves. Then we might observe the following sequence of vertices visited:

$$X_1 = (0, 0)$$
$$X_2 = (1, 0)$$
$$X_3 = (1, 1)$$
$$X_4 = (0, 1)$$
$$X_5 = (1, 1)$$

If $H = \{(0, 0), (1, 0)\}$ then $\frac{\sum_{t=1}^{5} \mathbb{1}(X_t \in \{(0,0),(1,0)\})}{n} = 2/5$.

Administrative Information

Marking is anonymous, so **do not** write your name anywhere in the report. Your registration number should be shown on the first page of the report. You should submit a single pdf file for this assignment. The report should be named `studentnumber-modulenumber-A3.pdf`.

Report Structure and Content

There is no page limit for the report but as a guide 7 pages or fewer should be sufficient. Think about how to communicate your answers concisely and clearly whilst providing sufficient detail. Unnecessarily long reports will be penalized. Just answer the questions directly and justify your answers. Include all the `R` code used for each part of the question in the main report at the point it is used. We strongly recommend you use `knitr` to do this. There is a `knitr` assignment template on MOLE. Annotate the R code where you think another MSc student might not easily understand what the R code does.

Assessment and Feedback

The Assignments should be **your own work** and should be undertaken in accordance with the University's rules for non-invigilated assessment. Marks will be given on the scale described in Chapter 1 of the notes. You will receive individual feedback and some general comments will be posted on MOLE.