

STA 141A

Fall 2016

Homework 1

Due: October 12 (Wednesday)

Need to submit the assignment both electronically (through smartsite) and by submitting the printed copy.

1. **Markov chain:** Write a function to generate a sequence of random variables $\{X_i\}_{i=1}^n$ taking values 0 and 1, following the probability model

$$\mathbb{P}(X_{i+1} = l | X_i = k) = \pi_{kl} \quad k, l = 0, 1;$$

where $0 < \pi_{kl} < 1$, and $\pi_{00} + \pi_{01} = 1$ and $\pi_{10} + \pi_{11} = 1$.

2. Write a function that takes a sequence (vector) of 0 or 1 as input and returns the starting locations of runs of 0's and runs of 1's, where the length of a run is set to an integer $K \geq 1$.
3. Write a function that takes a sequence (vector) of 0 or 1 as input and returns the starting location of all subsequences that start and end with prespecified motifs (vectors of 0's and 1's).
4. Generate sequences of length 10000 from the Markov chain described in Problem 1, by using two different *transition probability matrices*:

(a)

$$\begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

(b)

$$\begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}$$

For each of the cases (a) and (b), compute the following:

- (i) Starting location of runs of 0's and 1's of lengths up to 10. What is a good representation of this data? Do you see any difference in patterns between runs of 0's and runs of 1's?
- (ii) Starting locations of all subsequences of length 200 that start and end with the motifs 0000 and 1111, respectively.
- (iii) Perform a descriptive statistical analysis to compare your results in cases (a) and (b). Write a brief report summarizing your findings.

5. Give an informative graphical statistical summary of the following datasets (available with base R). In each case, write very brief (maximum of 80 words) description highlighting the findings.
- (i) `AirPassengers` : Monthly airline passenger numbers during 1949–1960.
 - (ii) `EuStockMarkets` : Daily closing prices of major European stock indices during 1991–1998.
 - (iii) `PlantGrowth` : Results from a controlled experiment on plant growth.
 - (iv) `trees` : Girth, weight and volume for Black Cherry trees.
 - (v) `airquality` : Data on New York air quality measurements.