Homework assignments

Imagine that you are a poor web admin named Joe. You are running very popular web server which for some reason has only four pages. For speeding up the system it is required to know which page a user will probably ask next. To do this you decided to use ordinary Markov chains. As your data set you are using web server logs which is a time ordered sequence of page requests.

You are given two sequences consisting of four symbols. One is used for training and the other is used for testing. Sequences are generated from the same n-order Markov chain, where $1 \leq n \leq 3$. The task is to find out the parameters of Markov chain and the order n.

1.
Construct ML-estimates using the training sequence for first, second and third order Markov chains. Print your parameters as a result.

Hint: Tweaking indices may cause a severe head trauma. The easiest way to calculate the estimates is to write the estimation code for each Markov chain separately. You can add small ϵ to the estimates if the normalisation phase fails.

2.
Calculate the log-likelihoods for the training and testing data. Explain the phenomenon. What is the order n?

You will receive one point for completing question 1 and two points for completing both questions.