

3E3 PROBABILITY AND STATISTICS LABORATORY

Department of Electronic and Electrical Engineering

(MATLAB e-Report submission)

## Assignments:

1. Homogeneous Markov chain modelling of a human gene DNA base sequence
2. Empirical and Exponential modelling of survival data in a large mechanical assembly
3. Empirical and normal modelling of multidimensional biosensor data

Submitted by:

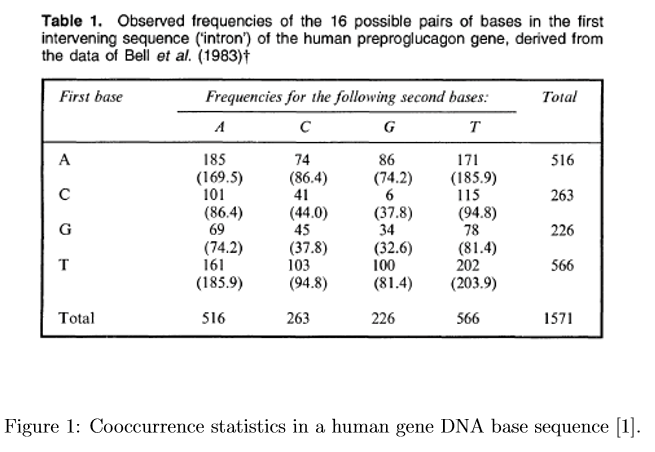
**Harsh Dhingra**

**19323904**

# Assignment 1:

# Homogeneous Markov chain modelling of a human gene DNA base sequence

Assignment stats are provided as follows:



1. In this task, with given total. We get the count N = total + 1, i.e. N = 1571 + 1 = 1572.

The transition probability matrix generated is as follows:

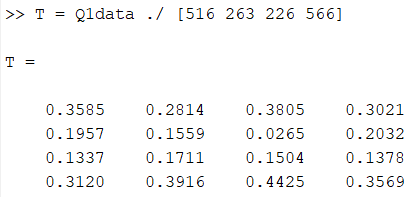


Figure 1.1 – Transition Probability Matrix (T) of the HMC

Here, the total probability vector i.e. **pi = [516 263 226 566]**, used for generation of Transition Probability Matrix.

The initial probability vector i.e. p1X (where X = {A, C, G, T}) can be taken at random as we are not specified with selection. Hence, after selection it’s useful for modelling dynamics of the sequence using HMC.

1. In this task, we modelled the transition probability matrix T as a graphical model. The plotted direct graph is as follows:

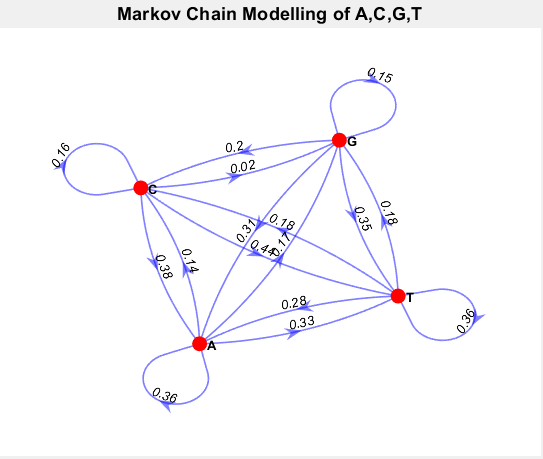


Figure 1.2 – Markov Chain – Directed Graphical Plot

After verification of data, the plotted directed graph is accurate up to 2 decimal places.

1. In this task we generated the sequence using simulate function and then using stairs created the DNA sequence as follows:

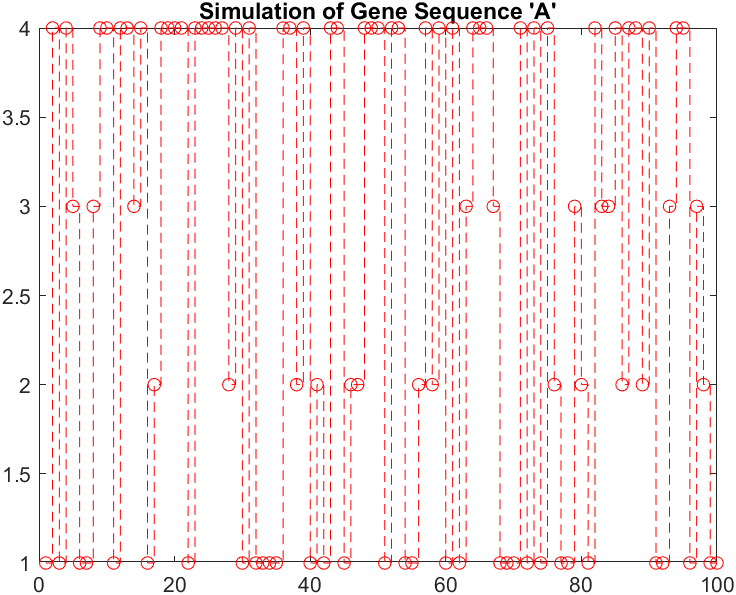


Figure 1.3.1 – Simulation of DNA Sequence ‘A’

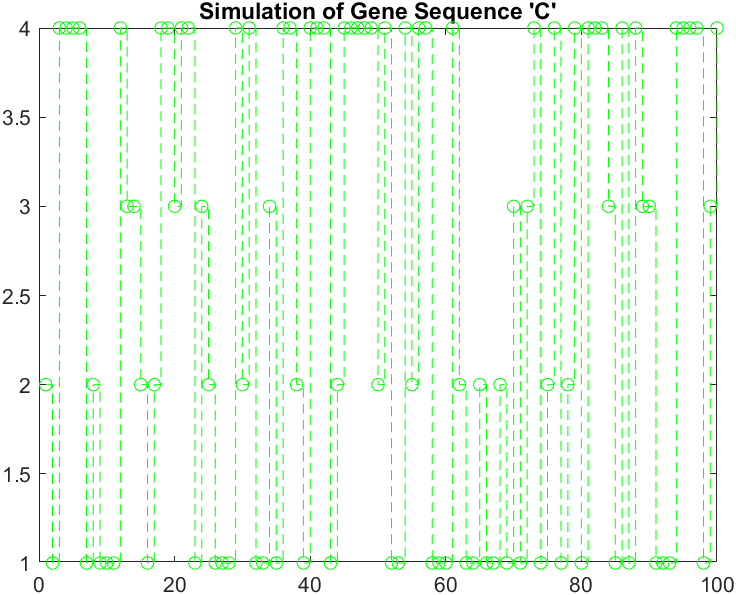
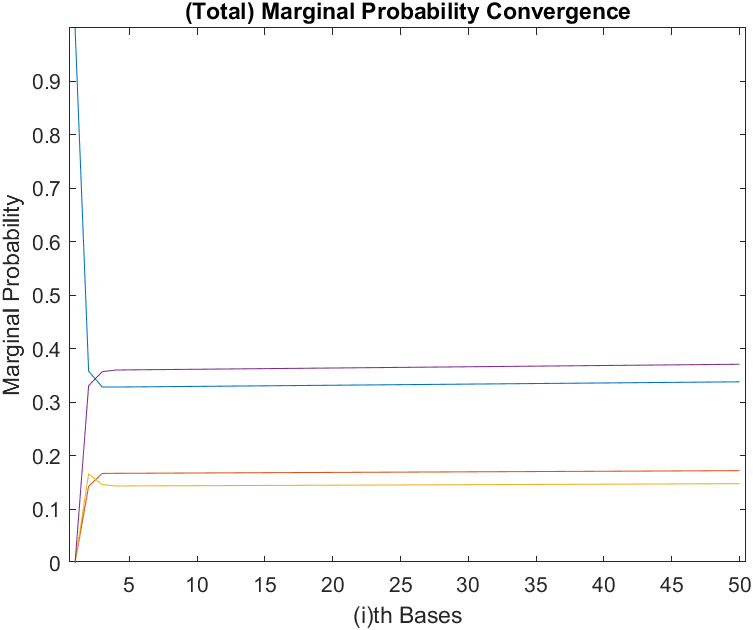


Figure 1.3.2 – Simulation of DNA Sequence ‘C’

1. In this task, Marginal Probability for ith bases was generated as following graphical representation:



The marginal probability converges to stable marginal probability over long run of HMC.

Also, for the task prior probability matrix assumed was **p1A** i.e. = **[1 0 0 0]**.

1. In this task, k-step-ahead T for taken HMC model was evaluated. The results are plotted as follows in graphical representation. (taken upto, k = 6).



Compared to task (iv), the convergence rate of this task is higher than previous one. As seen in the colormap, generated at 3-step ahead the colours are uniform from this stage forward.

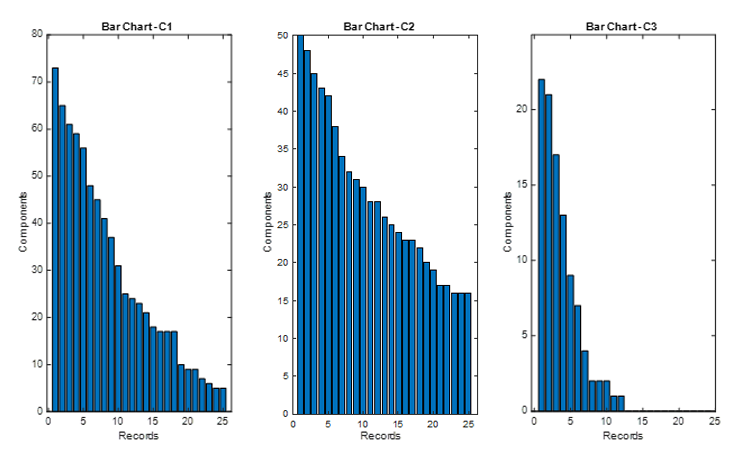
1. In this task, the probability evaluated is as follows.

Pr[Bi+4 = A | Bi = A] = **0.016523**

# Assignment 2:

# Empirical and Exponential modelling of survival data in a large mechanical assembly

1. In this task, we plotted bar chart for the provided data in the Q2stats.mat file. With inclusion of the data in the assignment. The graphical representation is as follows:



From the **Components v/s Record** chart, we can clearly observe that C2 has the most success rate amongst the three components. And the worst success rate is shown by component C3.

1. In this task, failed components are plotted in bar chart and scatter plot. The visualization is as follows:



1. In this task, **T** denotes unknown time to failure. And **I** denote interval of 6 months as natural numbers. So, the relation can be taken as:
2. In this task,



1. In this task,

