Counting K-mers

This program counts the K-Mers of the randomly created nucleotides and plots the **time vs K** curves. The input of the nucleotides are generated randomly

Program Inputs

1. Length of nucleotides
2. Value of K

Description: Long nucleotides chains generated by random function and return the long chain. Another function takes the input **K** which generates the permuted sequence and passes to main function a L length sequence and permuted sequence and value of K

K-mer: ['TTT', 'TTA', 'TTC', 'TTG', 'TAT', 'TAA', 'TAC', 'TAG'] for K = 3

Sequence: AATTGATATAGCAATGCGCGGCAGGGCTTTCAAGGTTCT

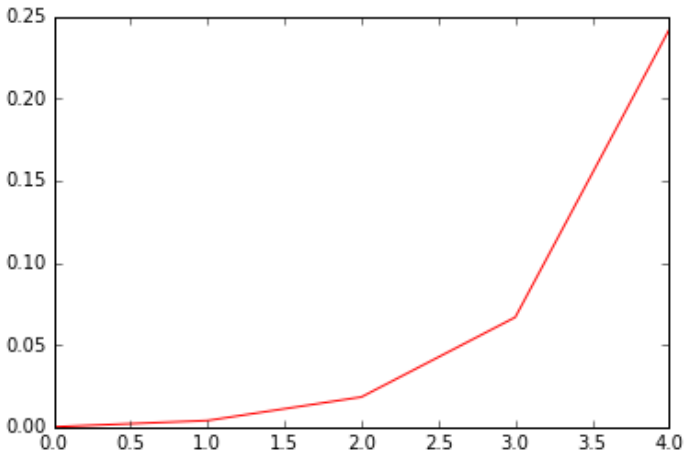
Complexity of the Program

Complexity calculated for the program is:

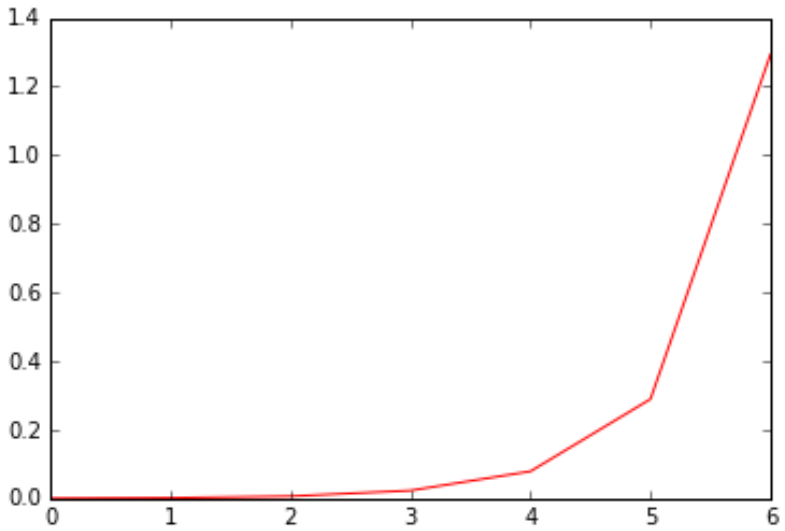
K: Length of polymers, L= total sequence length

From program outer loop iterate **L-K+1** times and the 2nd loop iterate 4 k times. The count function have complexity of **O(L)** in this case. To combine the overall complexity it becomes $\boldsymbol{O(4^k \, . \, L^2)}$**.** Here if we consider K as constant then our complexity becomes $\boldsymbol{O(L^2)}$

Graphs



Time plot for K=1:5 and L =100000 1



Time plot for K=1:7 and L=32000 1