Assignment 3: due date October 30 11:59pm

Write a Python program and call it *assignment3.py* that can calculate the time to the most recent common ancestor (TMRCA) and present a distribution of the TMRCA.

Input:

3 parameters: population size, number of samples, number of trials.

(use the mutation scaled population size Θ , good values are are around 0.01 to 0.05; sample size n can be any number between 2 and, say, 1000; number of trials t could be any number between 1 and 1000000)

Required output:

- Mean, Median, and Mode
- $\bullet~2.5$ and 97.5 % percentiles
- number of trials, number of samples, and population size.

The central part of the program will be a loop that starts with the number of sample given on the commandline, every iteration is reducing the sample by 1, and stops when the sample size is 1. In each iteration it will calculate the time until coalescent and will sum up these times for all intervals (at the end of the loop this sum is the TMRCA). To get the time of an interval you will need to solve

$$r = \exp\left(-t\frac{k(k-1)}{\Theta}\right)$$

where r is a uniform random number, k the size of the sample in the interval, and t is the interval time (you will need to solve the equation so that you get t =)

There will be an outer loop over all trials. You will need to record all TMRCAs for the histogram, and to calculate the mean, mode, and median and percentiles. The mean is simple: average over the TMRCA list, the median is the 50% value of the sorted TMRCA list, the mode can be evaluated from the histogram intermediate output (it is at the highest value on the histogram). The percentiles, can be taken from the sorted list (pick the 2.5 amd 97.5 % percentiles). The report includes the following:

- Source file,
- A short description (README) on how to run the program including the input/output requirements, and also containing an example of the values presented on the screen (do copy/paste from the terminal output)
- The resulting histogram as a PDF