IC252:Data Science II

Lab 8

By ADARSH SANTORIA (B21176)

In this assignment, you will verify the phenomena of the "weak law of large numbers (WLLN)" and the "central limit theorem (CLT)" numerically.

1. Generate a random data set $x_1, \ldots, x_m$ of size m for

   (a) exponential r.v. Exp(1),

   (b) uniform r.v. Unif(1, 2),

   (c) Bernoulli r.v. Bern(.2).

   Plot the sample mean $(x_1 + \ldots + x_m)/m$ for $m = 10, 100, 500, 1000, 5000, 10000, 50000$ for (a), (b), (c). This way, verify the WLLN.

2. Generate a random data set of size 1000 for each of independent and identically distributed

   (a) n exponential r.v.s Exp(1),

   (b) n uniform r.v.s Unif(1, 2),

   (c) n Bernoulli r.v.s Bern(.2).

   numerically compute and plot the distribution of the sample mean $(X_1 + \ldots + X_n)/n$ and corresponding "normal" approximation for $n = 1, 2, 4, 8, 16, 32$ for (a), (b), (c). This way, verify the CLT.

Notes:

1. While generating data sets, consider the number (to be with the precision of) decimal place of 1. That is, if the random number generated is 1.4789 then consider it as 1.4.

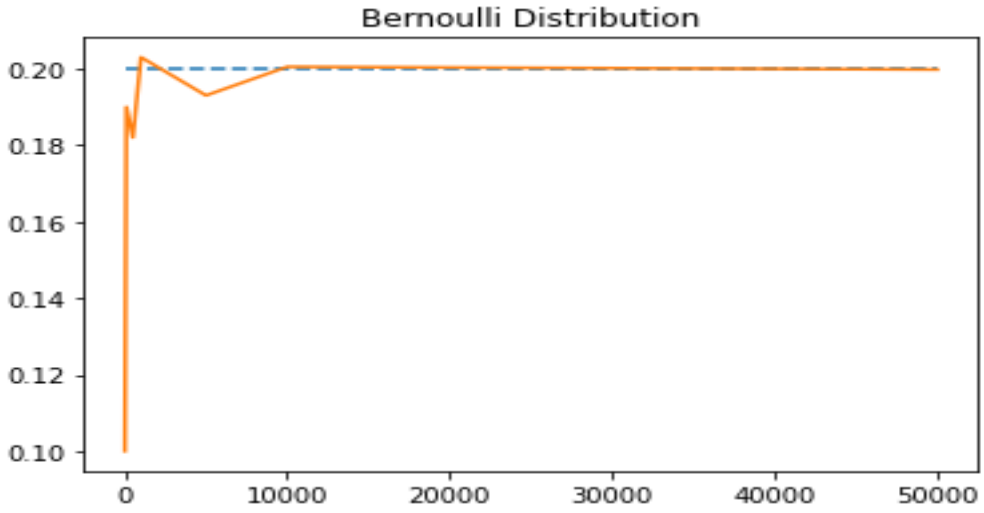
   Optional: To get better, smoother results, (1) in Problems 1 and 2, consider the number (to be with the precision of) decimal place of 2. Also, in Problem 2, you may consider the sample size of 10000.

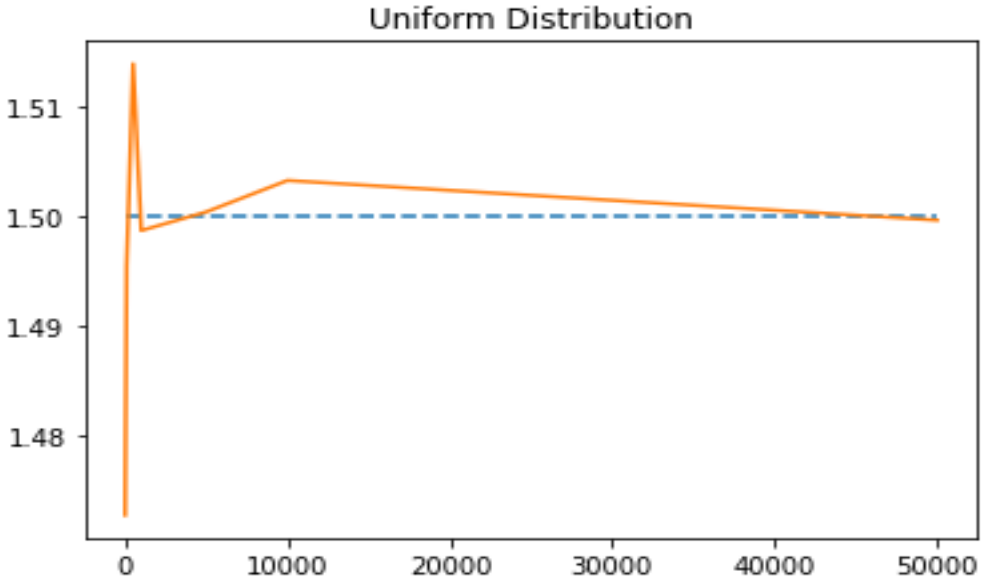
2. You may use in-built functions/library to generate samples from desired underlying distributions.

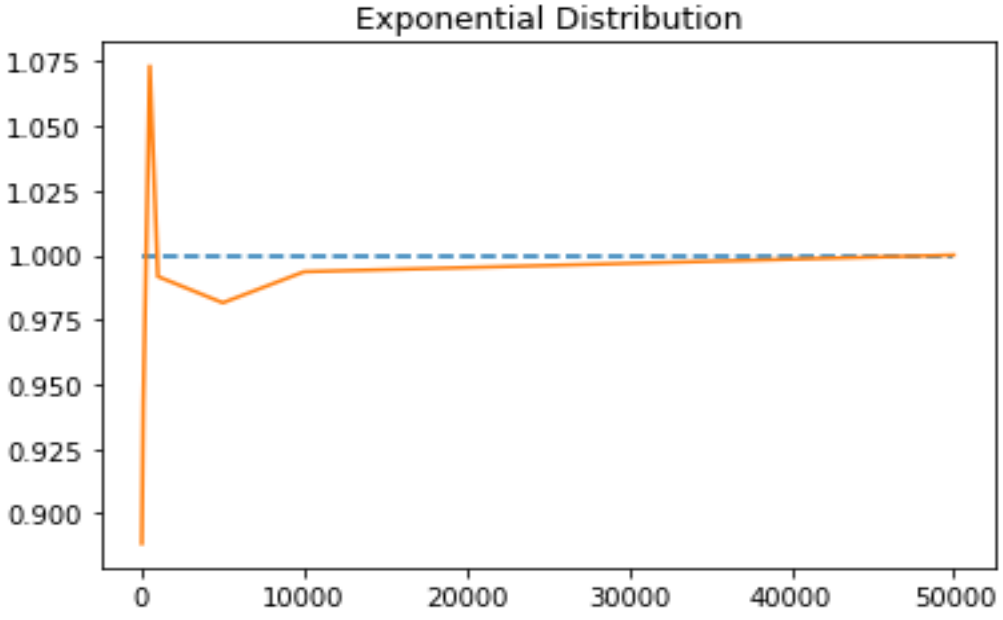
3. In Problem 1, the x-axis of the plot should be the sample size m and y-axis should be the sample mean.

4. In problem 2, numerically compute the distribution of X_i's from a data set of size 1000 and then plot. (In reality, the distribution of X_i is exactly exponential, uniform or Bernoulli depending upon the subproblem (a), (b) or (c). But in this problem, you will be working with numerical data and finding distribution from this numerical data.)

1.


Bernoulli Distribution
0.20
0.18
0.16
0.14
0.12
0.10
0
10000
20000
30000
40000
50000


Uniform Distribution
1.51
1.50
1.49
1.48
0
10000
20000
30000
40000
50000


Exponential Distribution
1.075
1.050
1.025
1.000
0.975
0.950
0.925
0.900
0
10000
20000
30000
40000
50000

2.