MA 323 - Monte Carlo Simulation Assignment - 5

VISHISHT PRIYADARSHI
180123053

1 QUESTION - 1:

(a) **For Box-Muller Method:**

Sample Size	Theoretical Mean	Sample Mean	Theoretical Variance	Sample Variance
100	0	0.081953294901575	1	1.038751640282734
10000	0	-0.00916512448576	1	1.012785820982868

For Marsaglia and Bray Method:

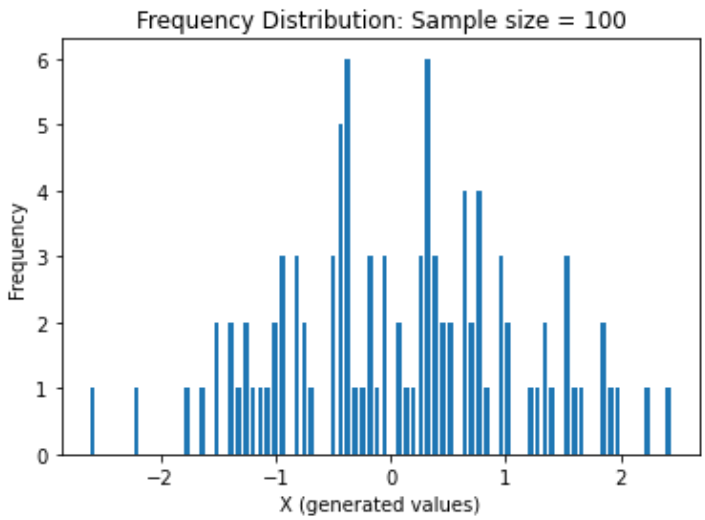
Sample Size	Theoretical Mean	Sample Mean	Theoretical Variance	Sample Variance
100	0	0.04178055012509841	1	0.9989293514443901
10000	0	0.008204305075926377	1	1.0019874304236942

Observations -

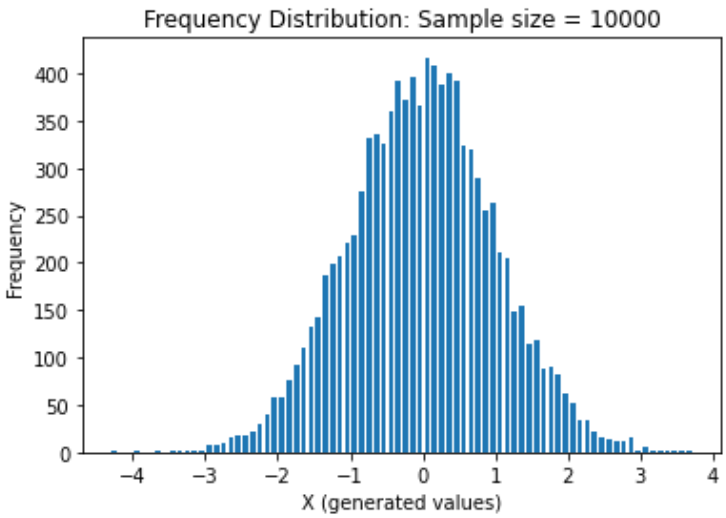
As we can see that the sample mean and sample variance are very close to their theoretical values. The error decreases substantially when the sample size is increased, i.e., more random numbers are generated.

(b) **For Box-Muller Method:**

i) **Case 1: Sample size = 100 –**

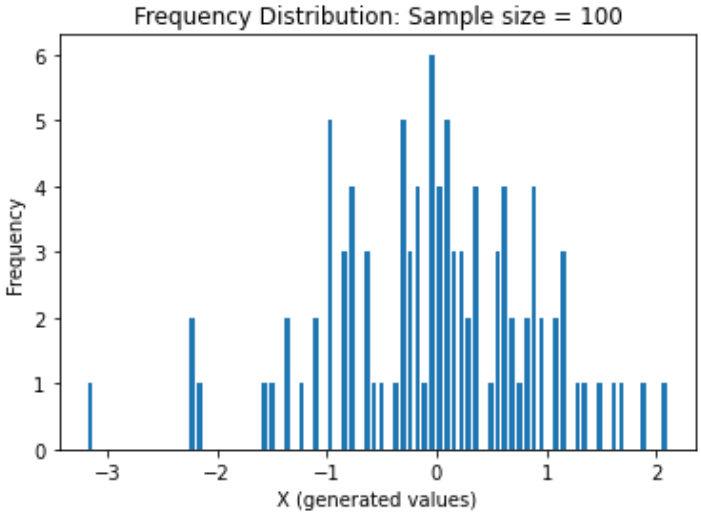


ii) Case 2: Sample size = 10000 –

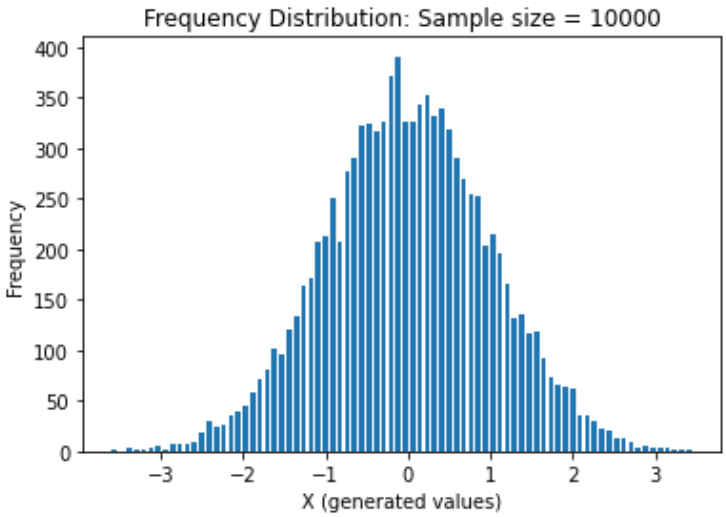


For Marsaglia and Bray Method:

i) Case 1: Sample size = 100 –



ii) Case 2: Sample size = 10000 –



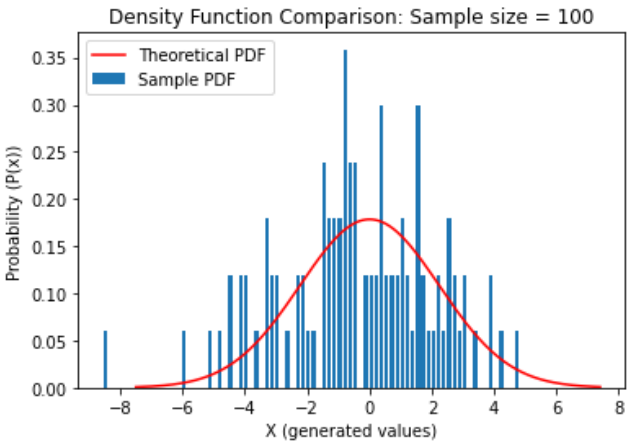
(c) The given Normal distributions can be easily generated from N(0, 1) using the following formula:

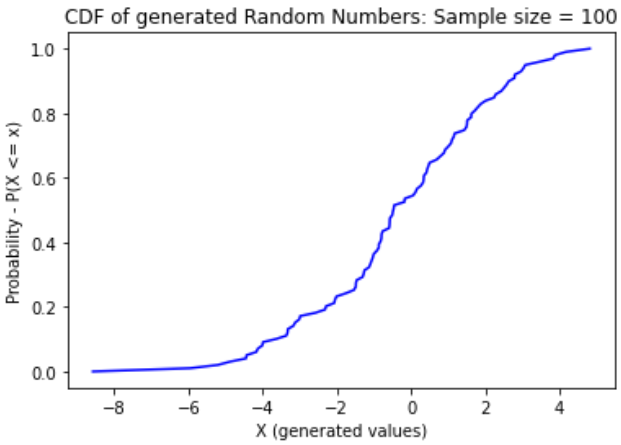
$$N(\mu, \sigma^2) = \mu + \sigma . N(0, 1)$$

Case 1: Samples from N(0, 5)

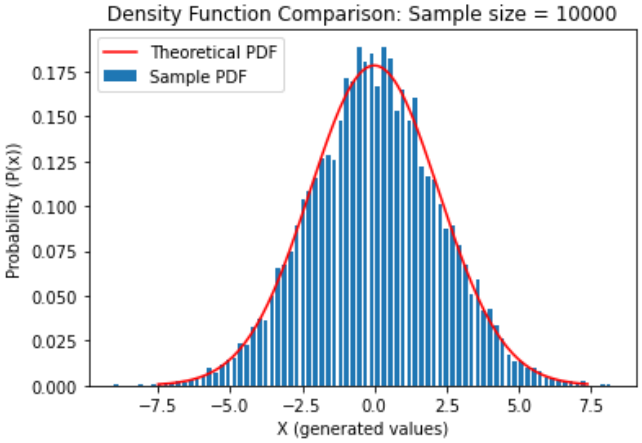
(a) For Box-Muller Method:

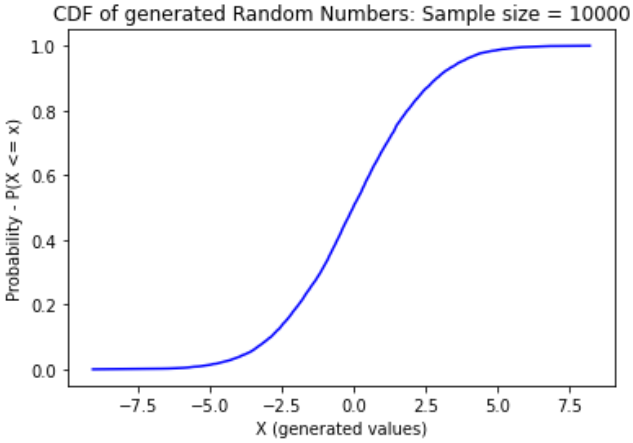
(1) Case 1: Sample size = 100 –





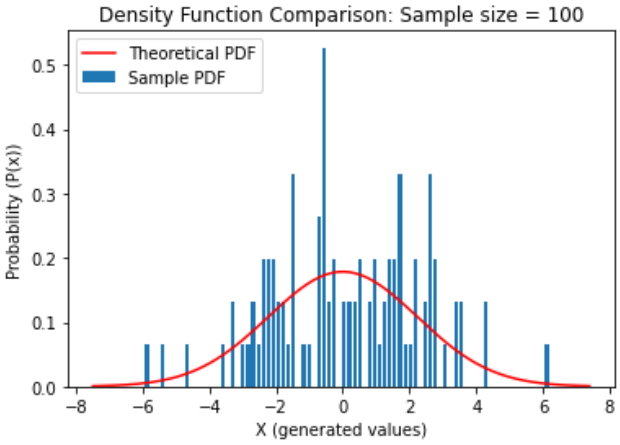
(2) Case 2: Sample size = 10000 –

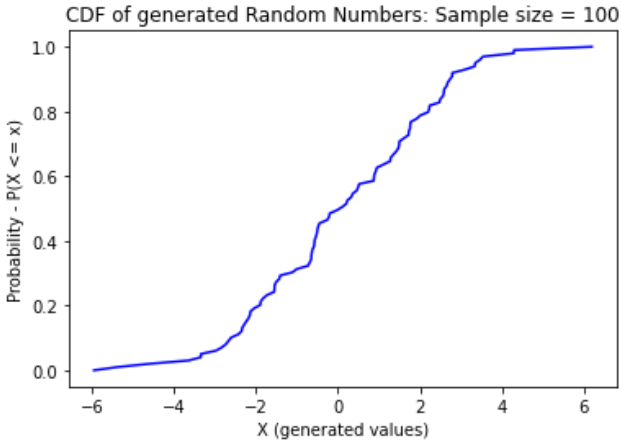




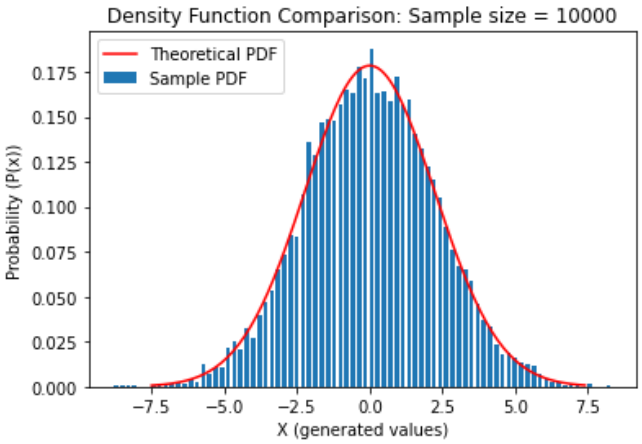
(b) For Marsaglia and Bray Method:

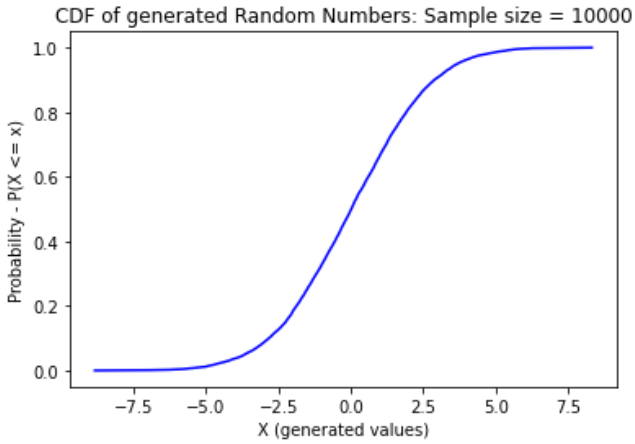
(1) Case 1: Sample size = 100 –





(2) Case 2: Sample size = 10000 –

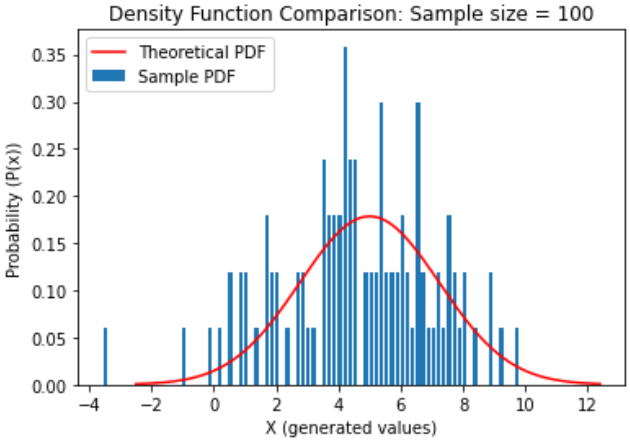


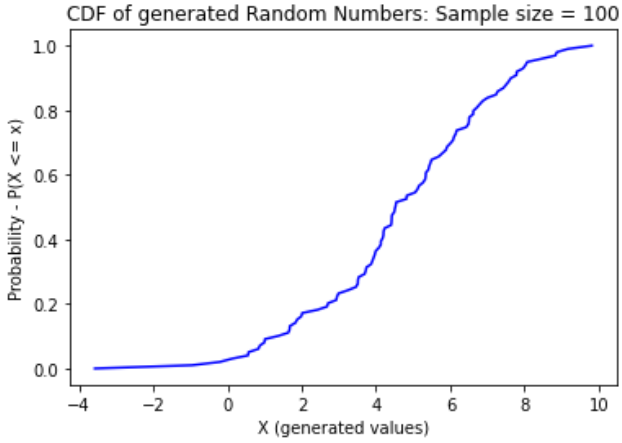


Case 2: Samples from N(5, 5)

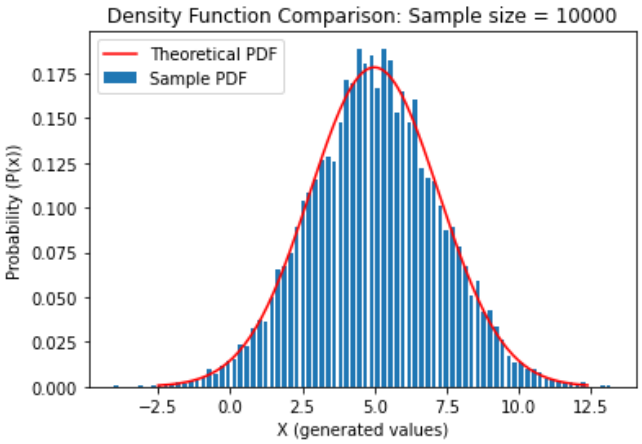
(a) For Box-Muller Method:

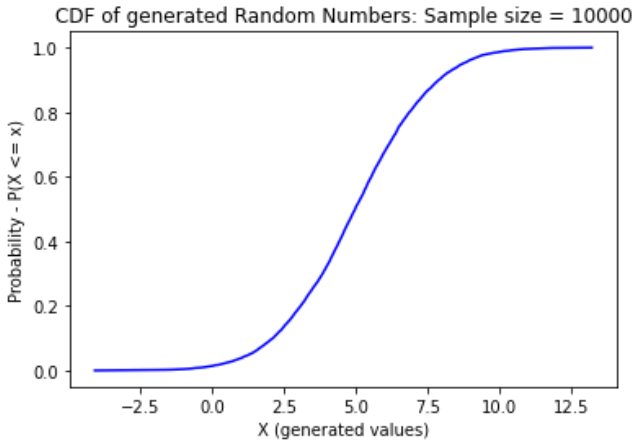
(1) Case 1: Sample size = 100 –





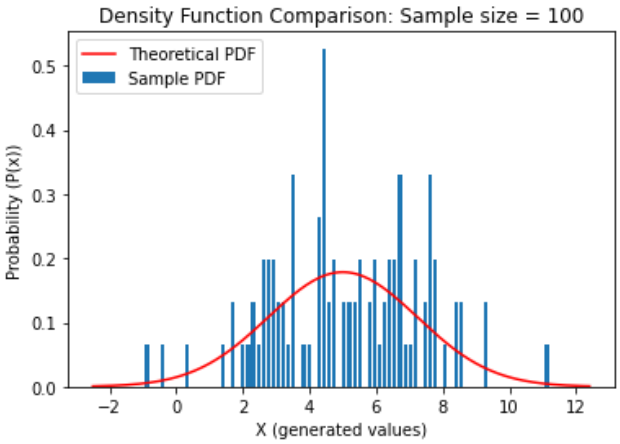
(2) Case 2: Sample size = 10000 –

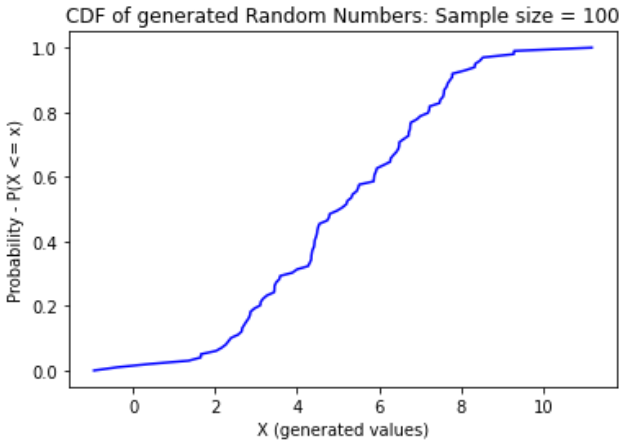




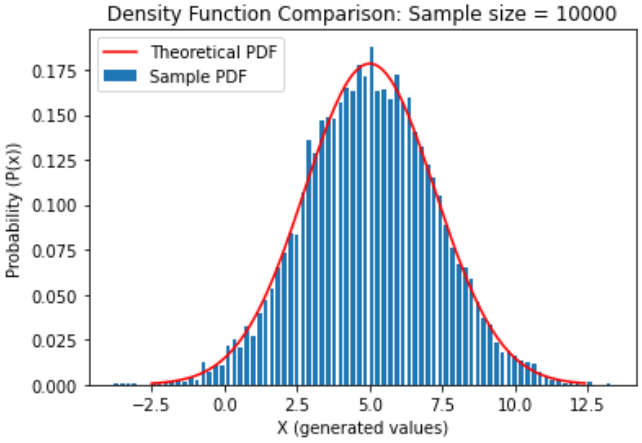
(b) For Marsaglia and Bray Method:

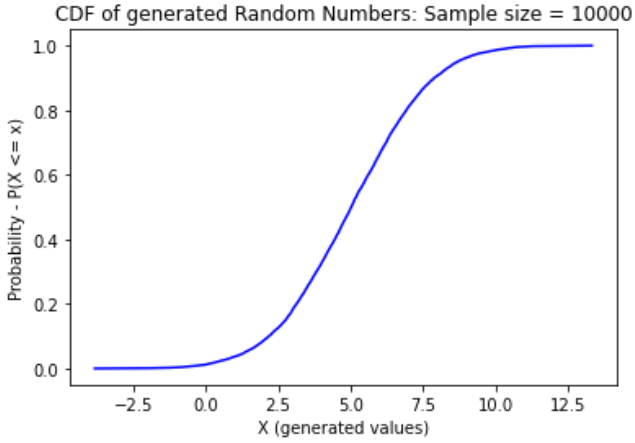
(1) Case 1: Sample size = 100 –





(2) Case 2: Sample size = 10000 –





Observations -

1. In both the cases, as the sample size is increases the PDF generated from the sample of random variable approaches the theoretical PDF. Same is true for the CDF of the sample random numbers. Initially when the sample size if less, there is a large deviation from the expected curve.
2. It is clearly seen that the mean and variance varies with the graph as Normal distribution is changed. The peak is attained at the mean, and the spread of the values depend on the variance. In both the cases of N(0, 5) and N(5, 5), the plots convey this information clearly.
   The plots for N(5, 5) is shifted to 5 units to the right of the N(0, 5) and its variance is same as that of N(0, 5).
3. When the sample size is less (size = 100), the irregularities in the plots are more dominant and it deviates slightly from the normal distribution.

2 QUESTION - 2:

The computation time is as follows (for sample generated from N (0, 1)) :

Size of Sample	Box-Muller Method	Marsaglia and Bray Method
100	0.000141639709472656 sec	0.00010860919952392 57sec
10000	0.006637425422668457 sec	0.011284046173095703 sec

Observations -

After running the simulations for a lot of times and taking average of the resultant values, we can say that the Marsaglia and Bray method is faster than the Box-Muller method, when the sample size = 100, and opposite behaviour is observed for sample size = 10000.

Theoretically, Marsaglia and Bray method is faster than the Box-Muller method as we avoid the evaluation of the “cos” and “sin” functions. But in practical cases, we may not always observe this trend. This is because in the Acceptance Rejection technique (which is used in Marsaglia and Bray Method), we are looping through to accept only the suitable values and rejecting the undesired ones. As the sample size increases, this leads to significant overhead, as a result of which Marsaglia method becomes slightly slower than the Box-Muller method.

So, it depends on the way of implementation. If sin and cos functions are implemented in such a way that they run faster than the Acceptance rejection technique, then Box-Muller method will be faster. But this is not the case always as at lower sample size (here, size = 100), we can clearly see that Marsgalia method beats Box method. But when sample size increases to a very large number, similar trend is difficult to observe.

3 QUESTION – 3 :

The comparison table for sample size = 100 for various iterations is :

Iteration No	Proportion of Values rejected	Theoretical value (1 – π/4)	Error Percentage
1	0.15254237288135597	0.21460183660255172	28.9 %
2	0.21875	0.21460183660255172	1.93 %
3	0.2753623188405797	0.21460183660255172	28.31 %
4	0.2063492063492064	0.21460183660255172	3.85 %
5	0.21875	0.21460183660255172	1.93 %
6	0.24242424242424243	0.21460183660255172	12.96 %
7	0.16666666666666663	0.21460183660255172	22.37 %
8	0.25373134328358204	0.21460183660255172	18.23 %
9	0.25373134328358204	0.21460183660255172	18.23 %
10	0.180327868852459	0.21460183660255172	15.97 %

The comparison table for sample size = 10000 for various iterations is :

Iteration No	Proportion of Values rejected	Theoretical value (1 – π/4)	Error Percentage
1	0.2088607594936709	0.21460183660255172	2.68 %
2	0.21531701192718145	0.21460183660255172	0.33 %
3	0.21654653713569416	0.21460183660255172	0.91 %
4	0.21098311503866185	0.21460183660255172	1.69 %
5	0.21123205552926327	0.21460183660255172	1.57 %
6	0.21494740147589886	0.21460183660255172	0.16 %
7	0.21691464369616287	0.21460183660255172	1.08 %
8	0.2064751626725917	0.21460183660255172	3.79 %
9	0.21408362150267213	0.21460183660255172	0.24 %
10	0.20810896420652514	0.21460183660255172	3.04 %

Observations –

Theoretically, the proportion of values rejected should be almost equal to the quantity 1 – π/4. This is because this is the area of the discarded region from a box of unit area. We are choosing only those random numbers s.t, they lie inside a circle which is inscribed in a square of unit area. As a result, the area of the remaining square turns out to be 1 - π/4, which measures the proportion of the values rejected.

When the sample size is large (size = 10000), the deviation from the theoretical value is very negligible, but when the sample size is less (size = 100), there is a slight deviation from the observed values which varies on various iterations of the simulation.