Monte Carlo Simulation   MA – 323   Lab – 4

Name – Rasesh Srivastava

Roll Number – 210123072

Branch – Mathematics and Computing

Question 1:

1. Use the Box-Muller method and Marsaglia and Bray method to do the following:

   (a) Generate a sample of 100 and 10000 values from $N(0,1)$. Hence, find the sample mean and variance.

   (b) Now plot a two-dimensional graph where the x axis will have the values that have been sampled and y axis will be the frequency or count of those values. Do this for both the cases, namely, 100 and 10000 samples.

   (c) Now use the above generated values to generated samples from $N(0,5)$ and $N(5,5)$. Hence plot the density function from the formula and also plot the sample distribution in the same plot in both the cases. How do these two plots compare in both the cases ?

(a) For Box-Muller Method:

Sample Size	Theoretical Mean	Sample Mean	Theoretical Variance	Sample Variance
100	0	0.05491931458150487	1	0.9717906139809108
10000	0	-0.0027448409261775298	1	0.995404778535134

For Marsaglia and Bray Method:

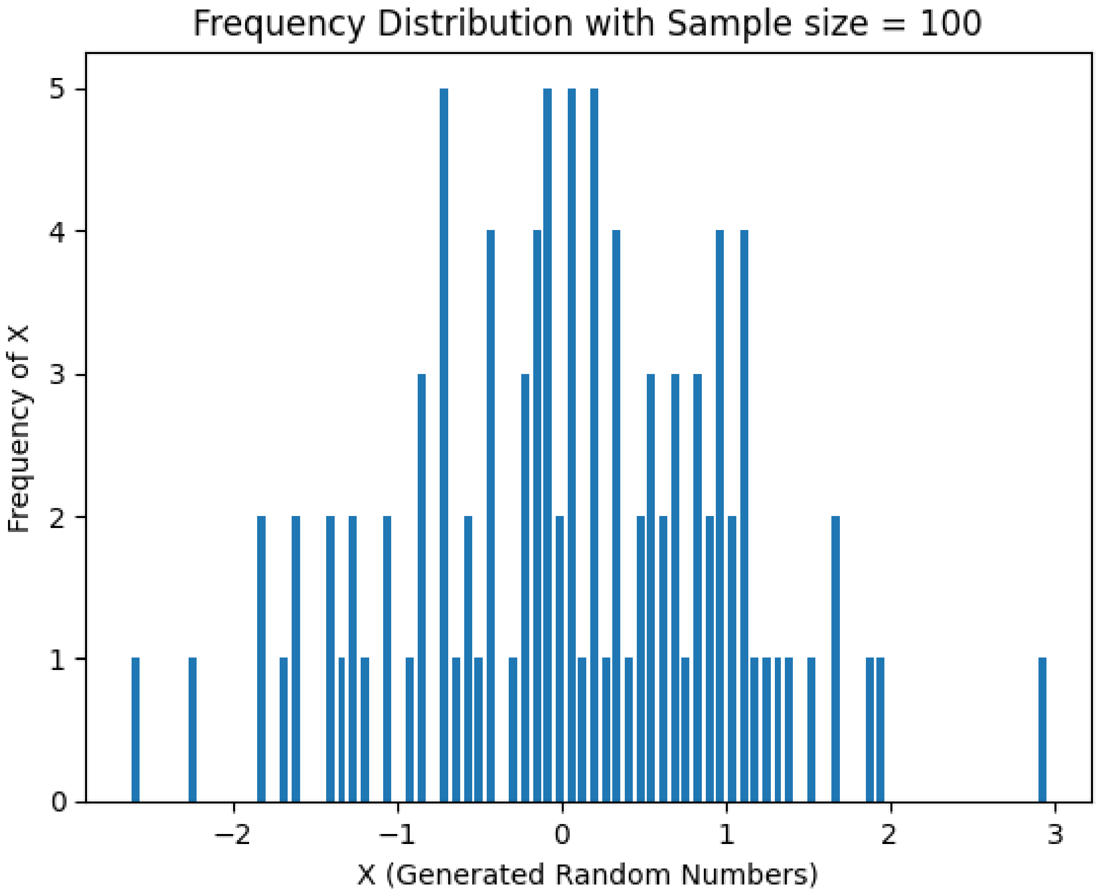
Sample Size	Theoretical Mean	Sample Mean	Theoretical Variance	Sample Variance
100	0	-0.08546233869062776	1	0.9792732319581324
10000	0	0.0035658725833214865	1	0.9851904973614938

Observations:

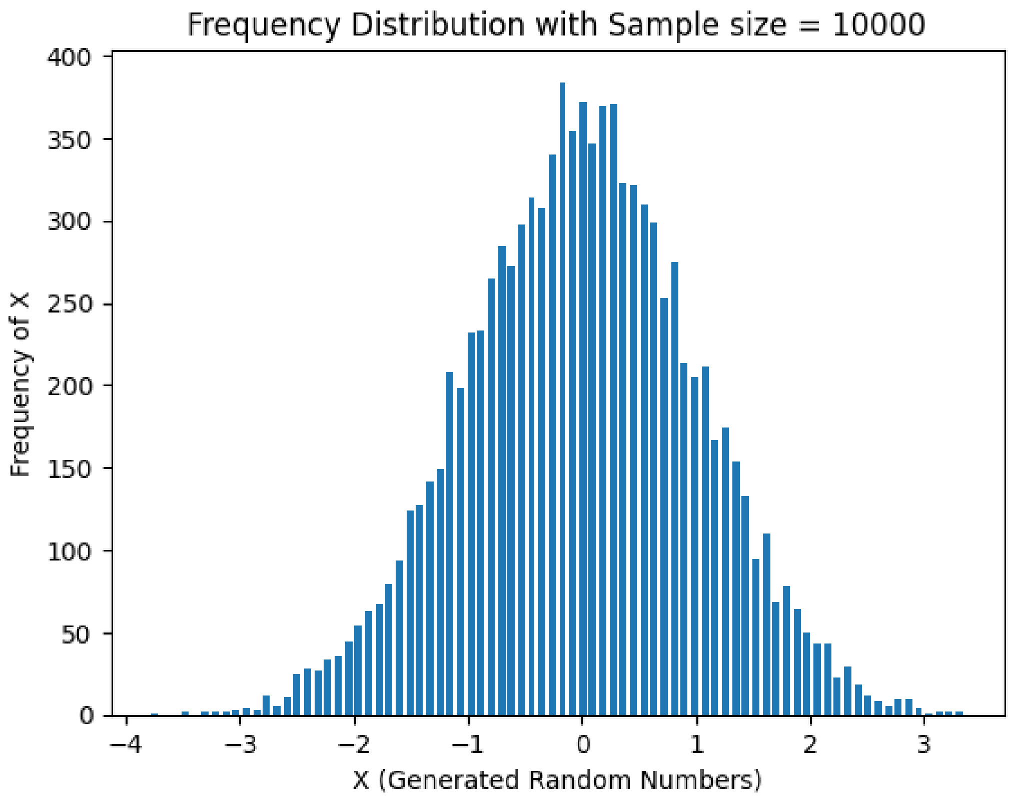
We can observe that the sample mean is very close to the theoretical mean and the sample variance is very close to the theoretical variance. The error in the sample mean and sample variance decreases significantly when more random numbers are generated, that is, when the sample size(N) is increased.

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

(i) When Sample Size = 100,

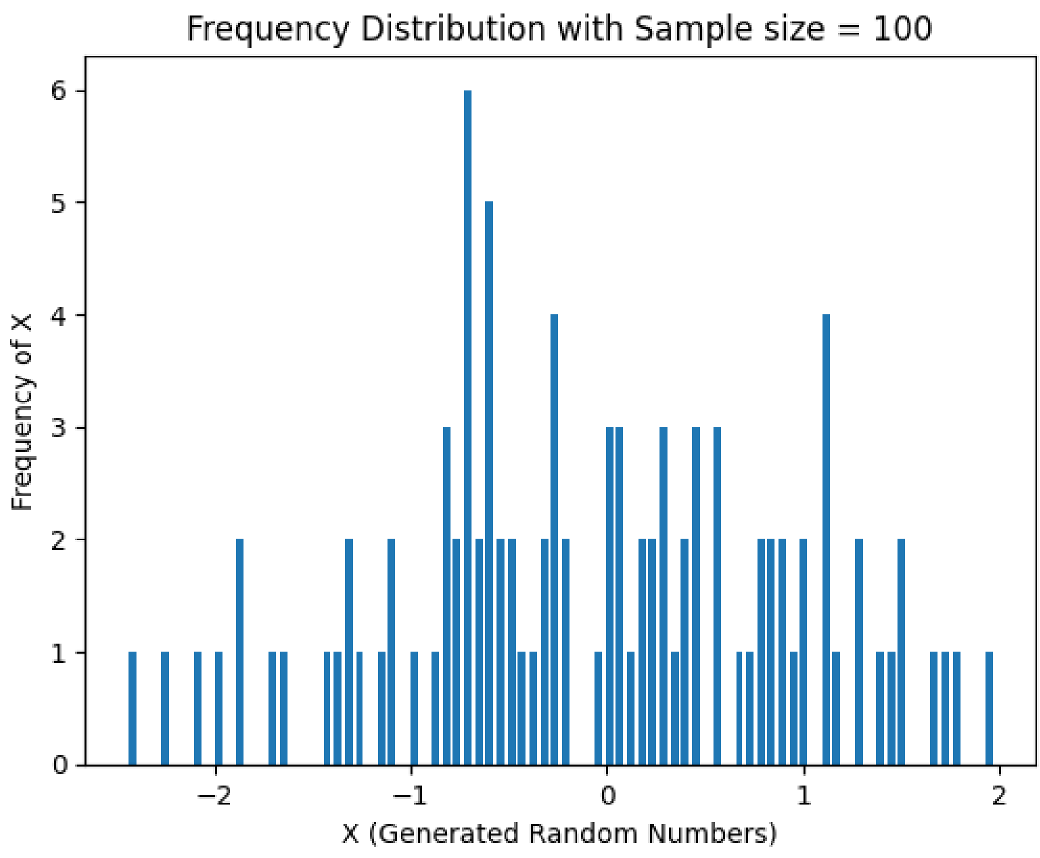


(ii) When Sample Size = 10000,

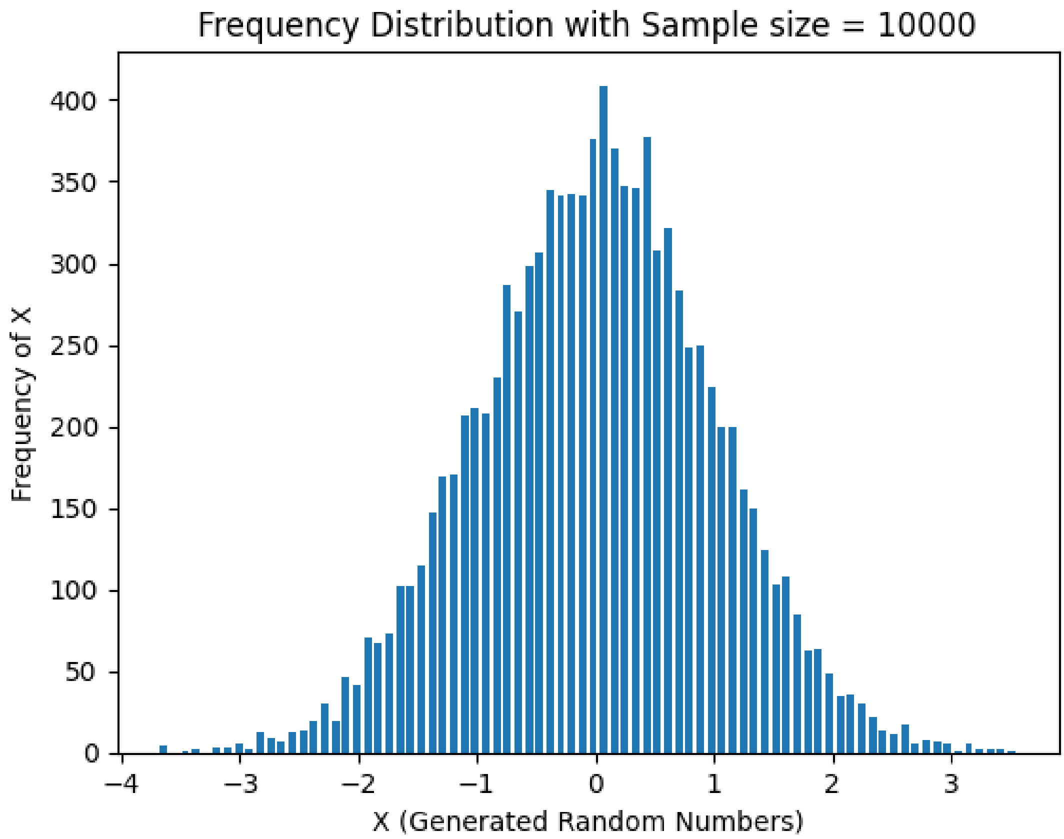


For Marsaglia and Bray Method:

(i) When Sample Size = 100,



(ii) When Sample Size = 10000,



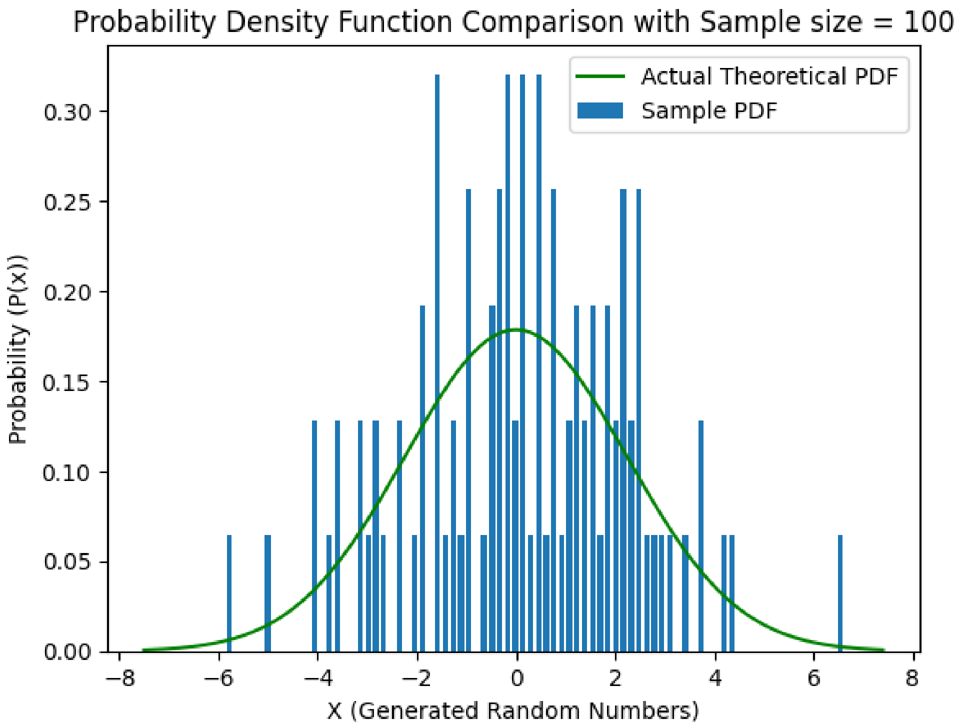
(c) The given Normal distributions N (0, 5) and N (5, 5) can be simply generated from N (0, 1) using the following given formula:

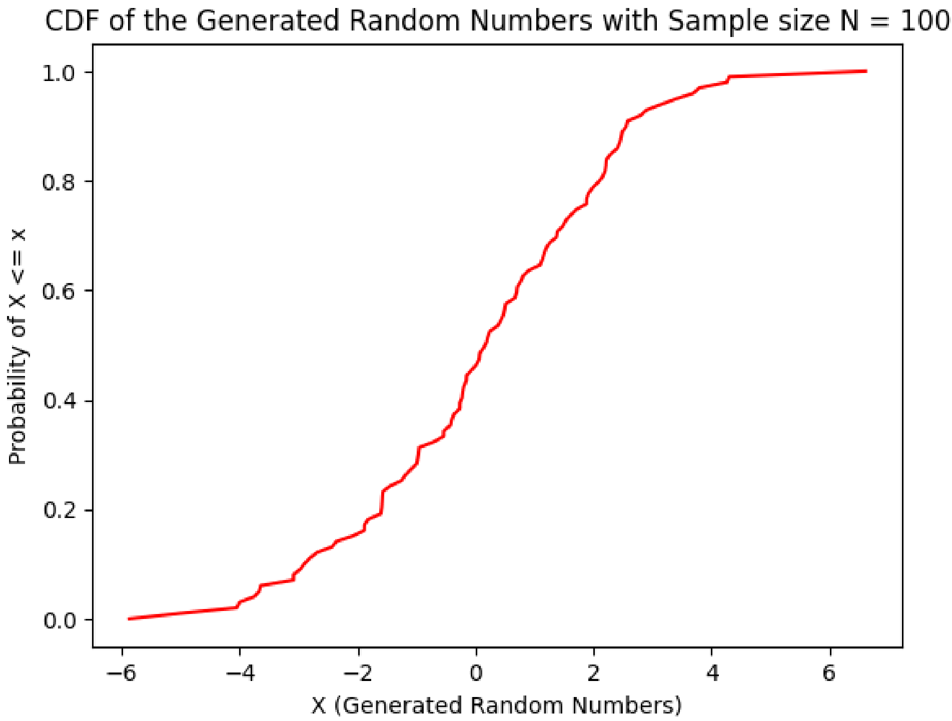
$N(\mu, \sigma^2) = \mu + \sigma . N(0, 1)$

<u>Case – 1st: Samples from N (0,5) distribution:</u>

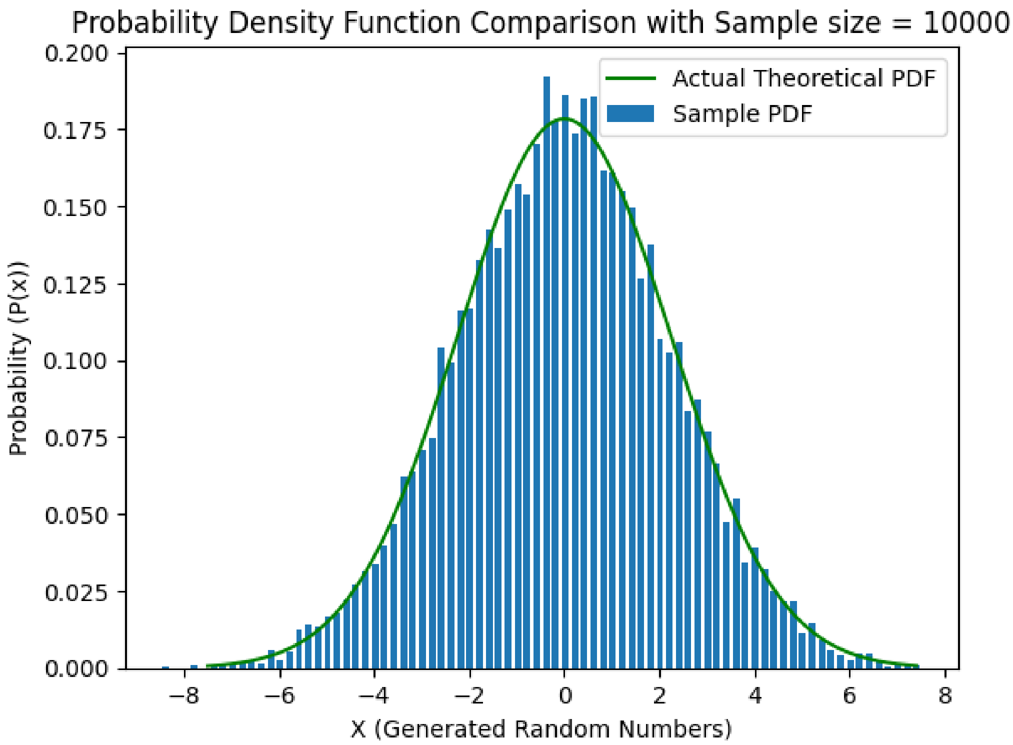
(a) For Box-Muller Method:

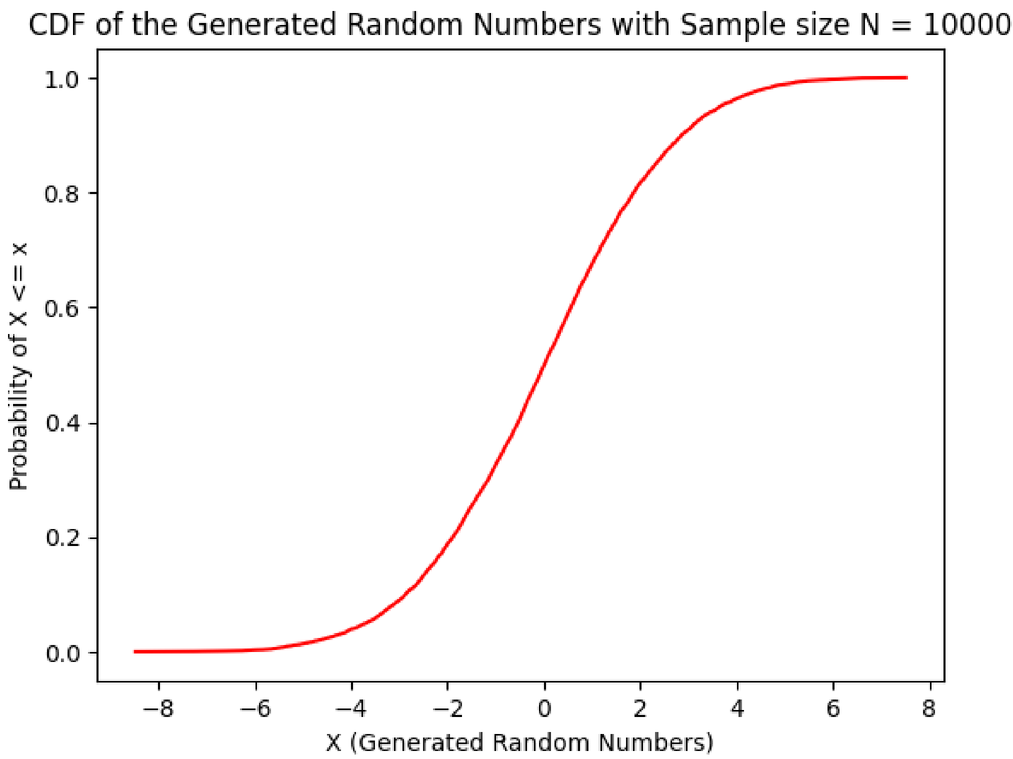
(i) When Sample Size = 100,





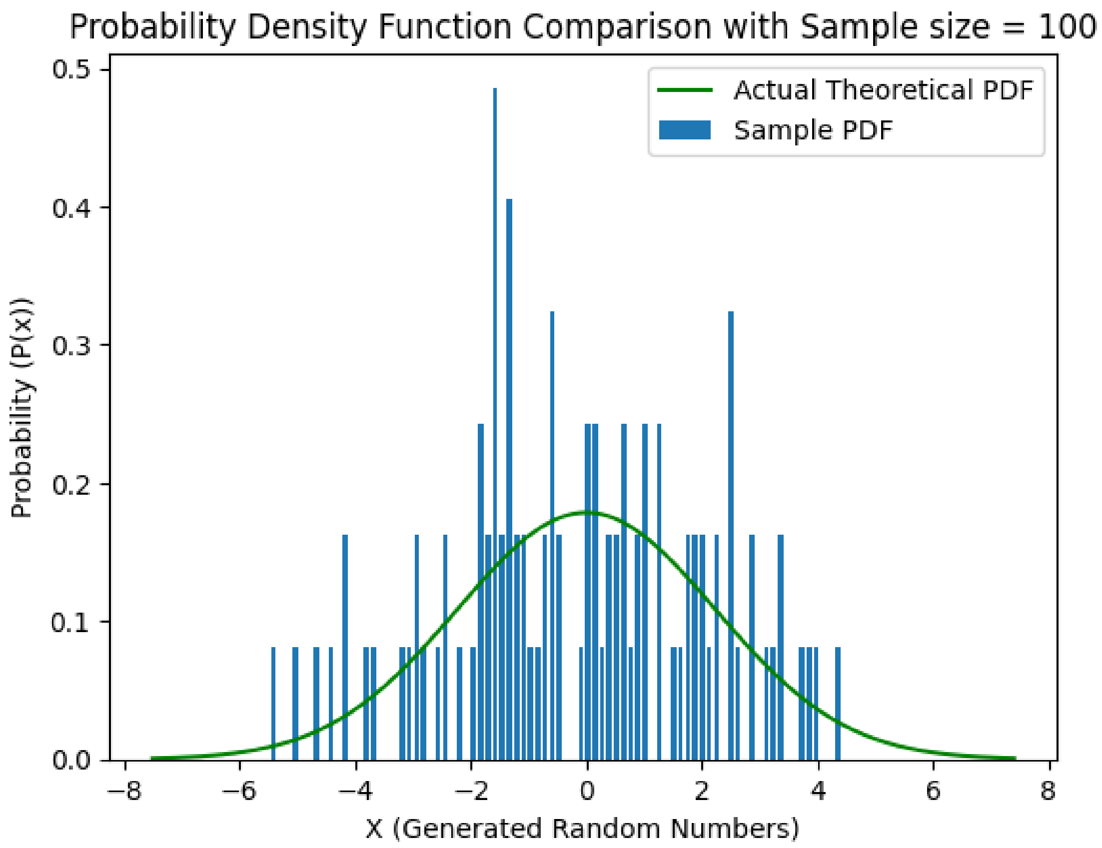
(ii) When Sample Size = 10000,

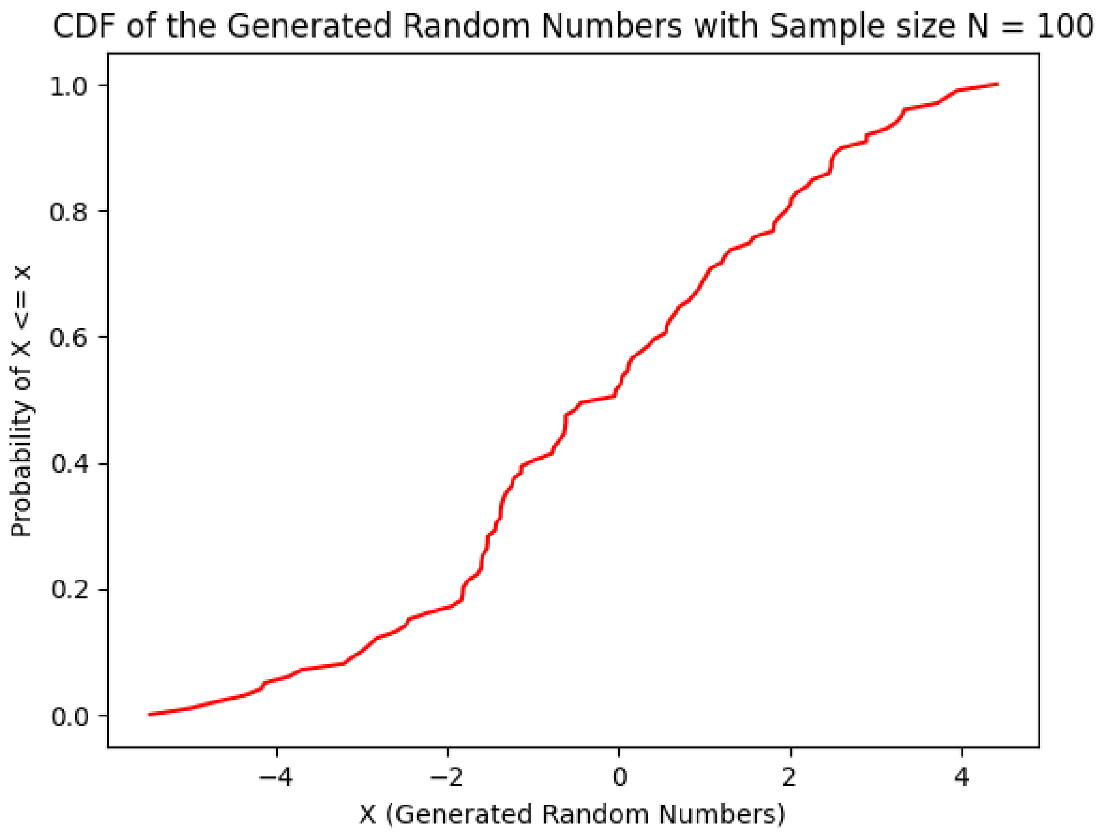




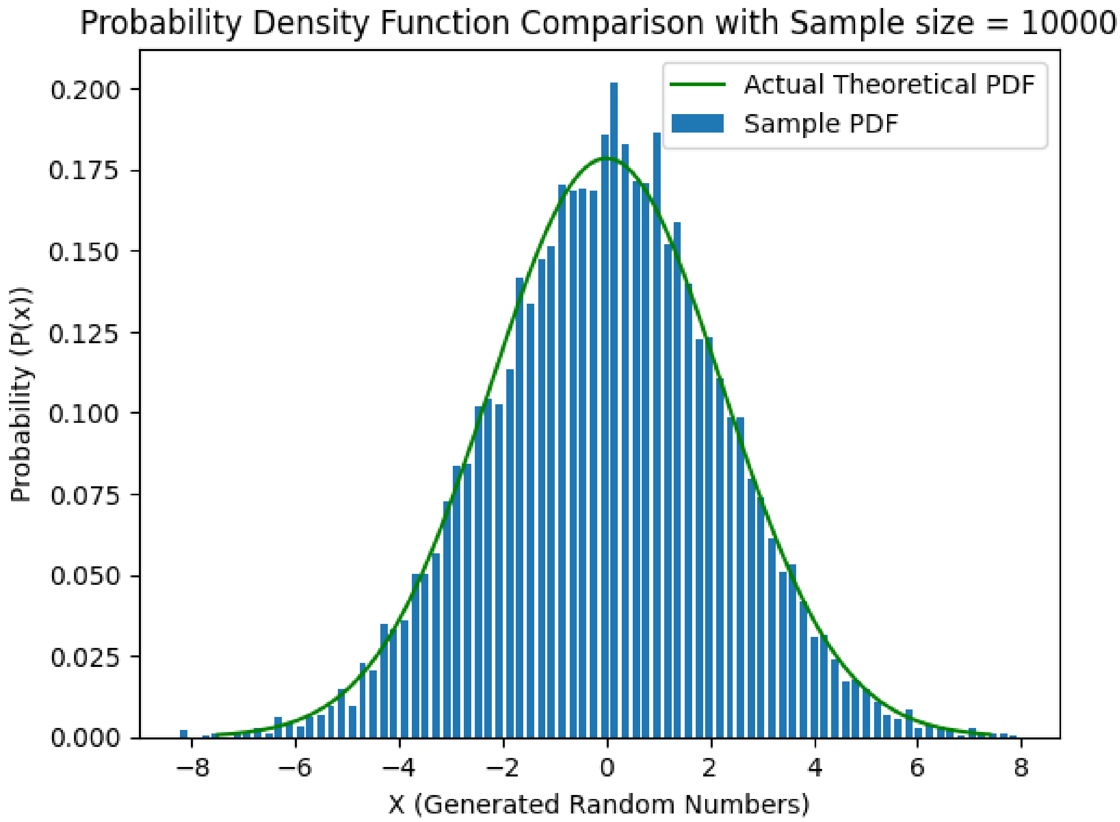
(b) For Marsaglia and Bray Method:

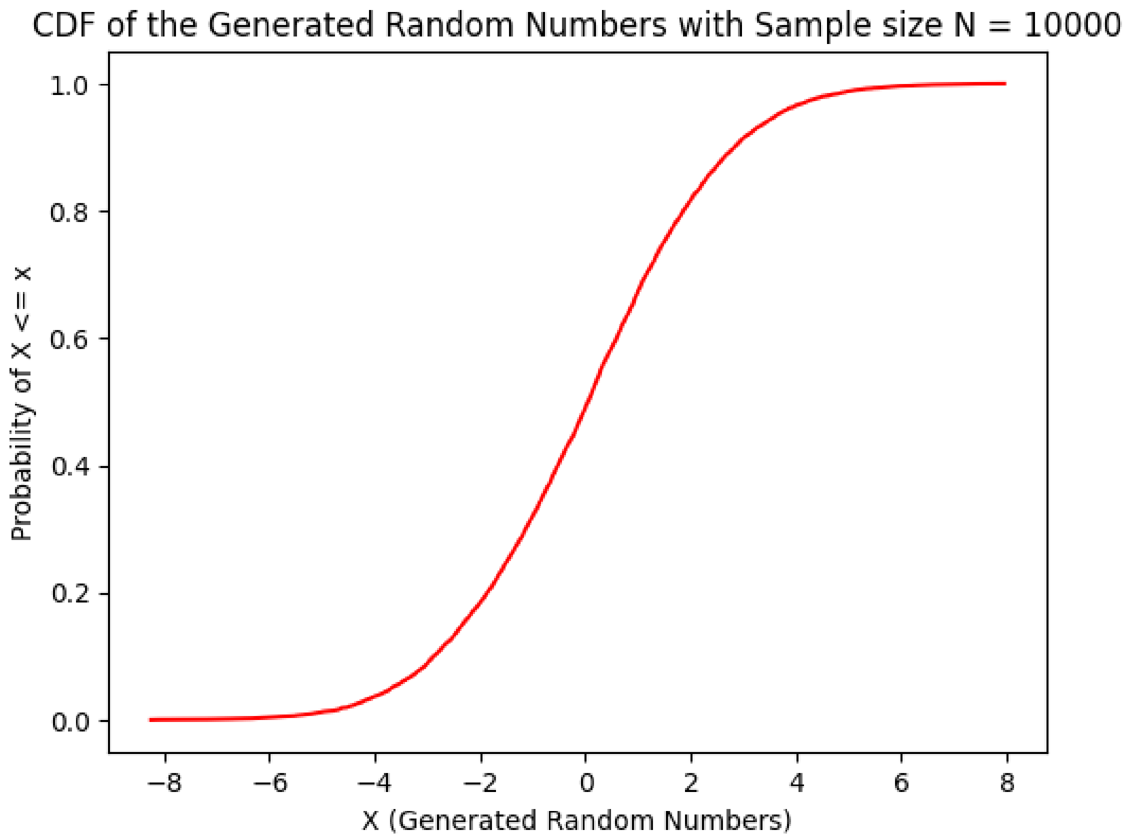
(i) When Sample Size = 100,


Probability Density Function Comparison with Sample size = 100
Actual Theoretical PDF
Sample PDF
Probability (P(x))
X (Generated Random Numbers)
0.0
0.1
0.2
0.3
0.4
0.5
−8
−6
−4
−2
0
2
4
6
8


CDF of the Generated Random Numbers with Sample size N = 100
Probability of X <= x
X (Generated Random Numbers)
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4

(ii) When Sample Size = 10000,

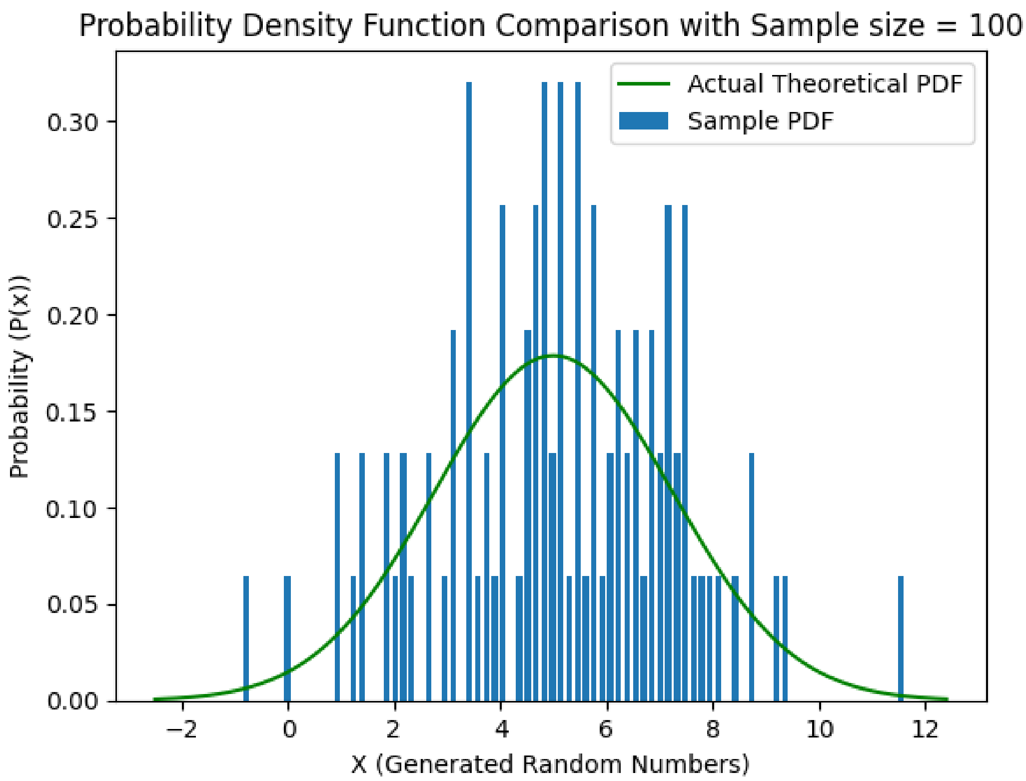


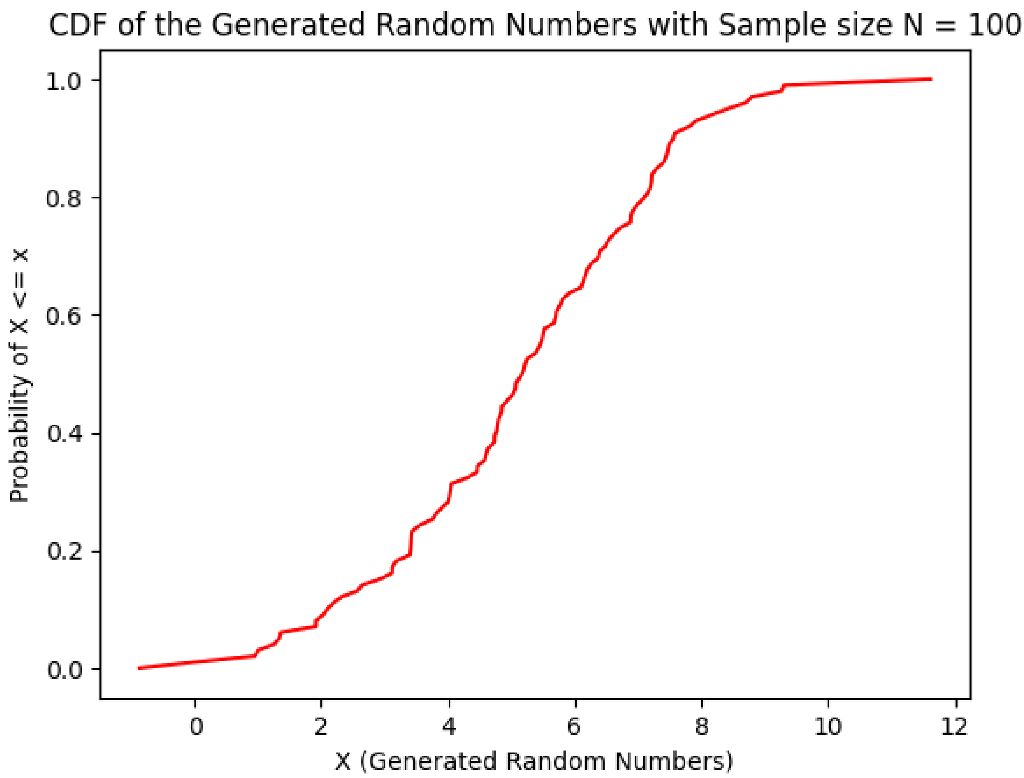


Case – 2nd: Samples from N (5,5) distribution:

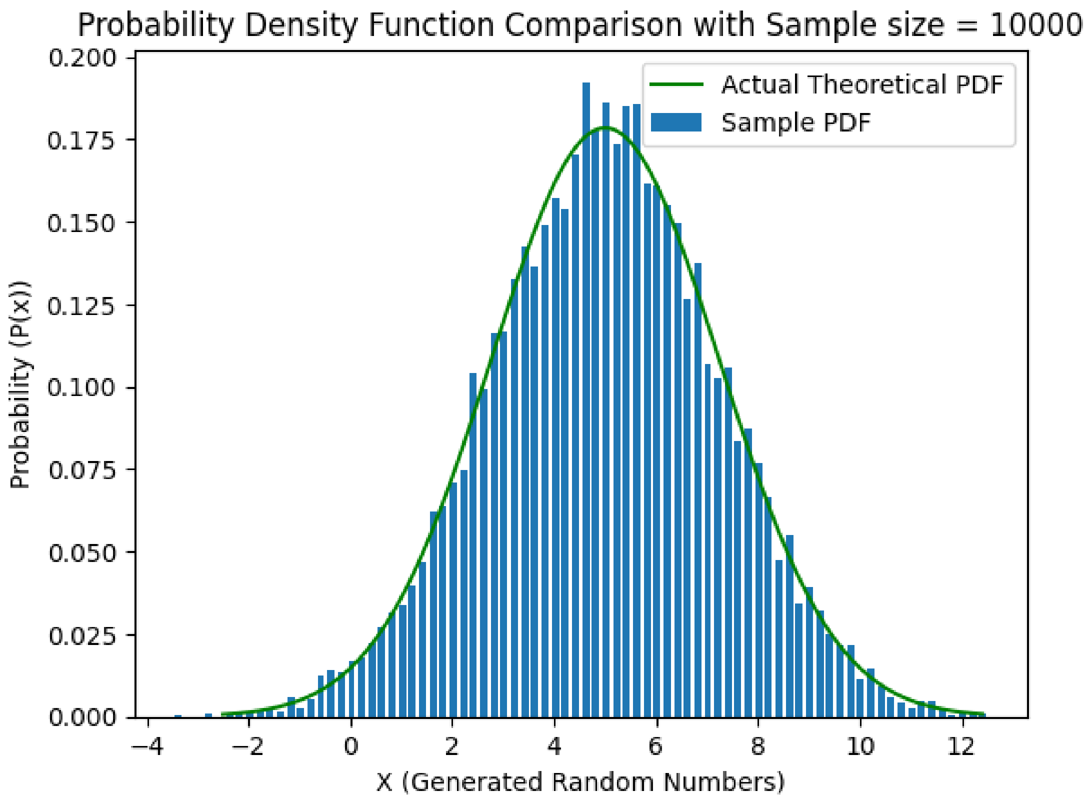
(a) For Box-Muller Method:

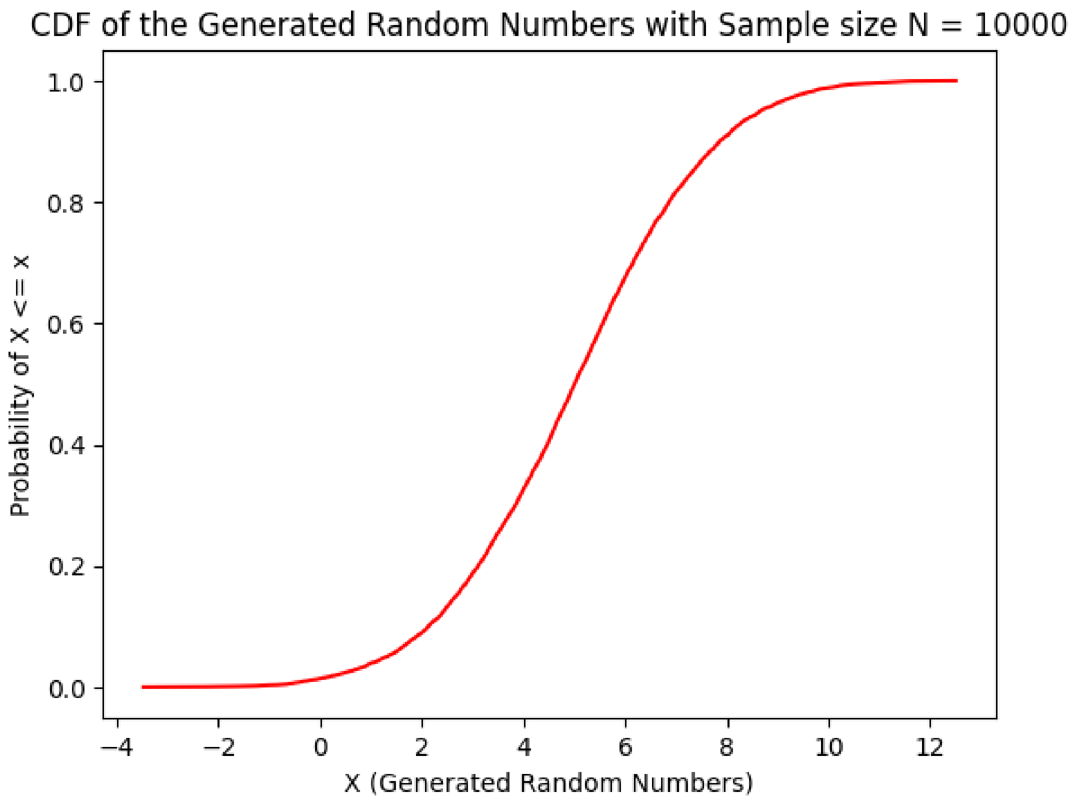
(i) When Sample Size = 100,





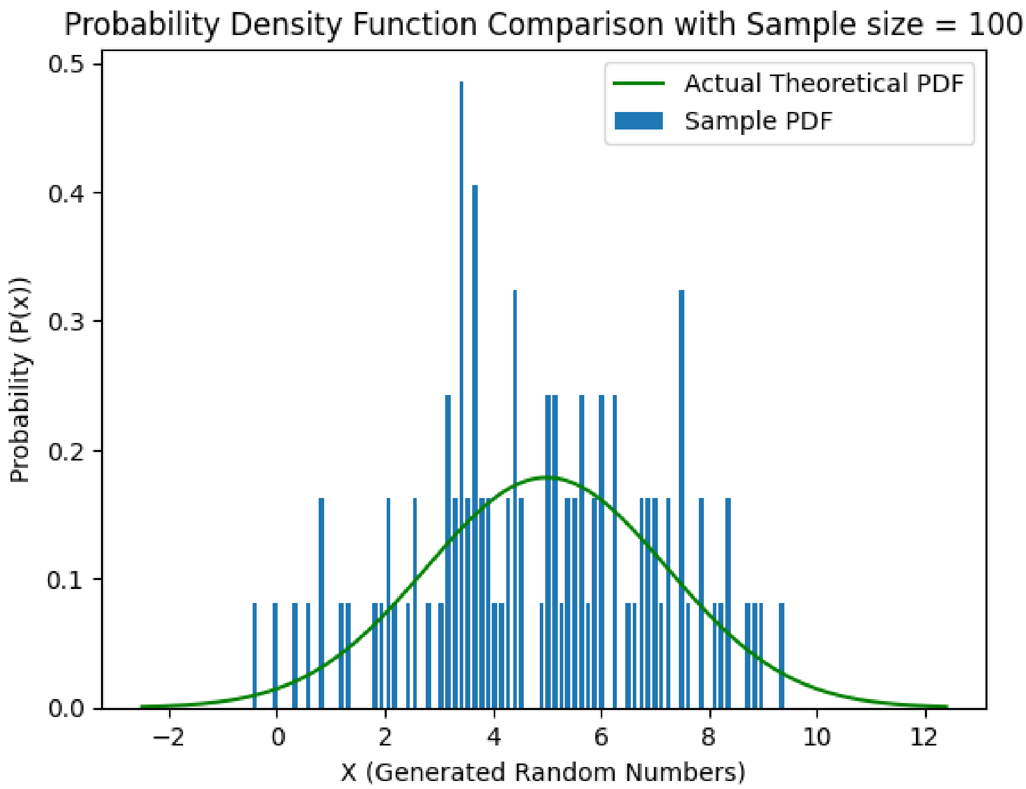
(ii) When Sample Size = 10000,

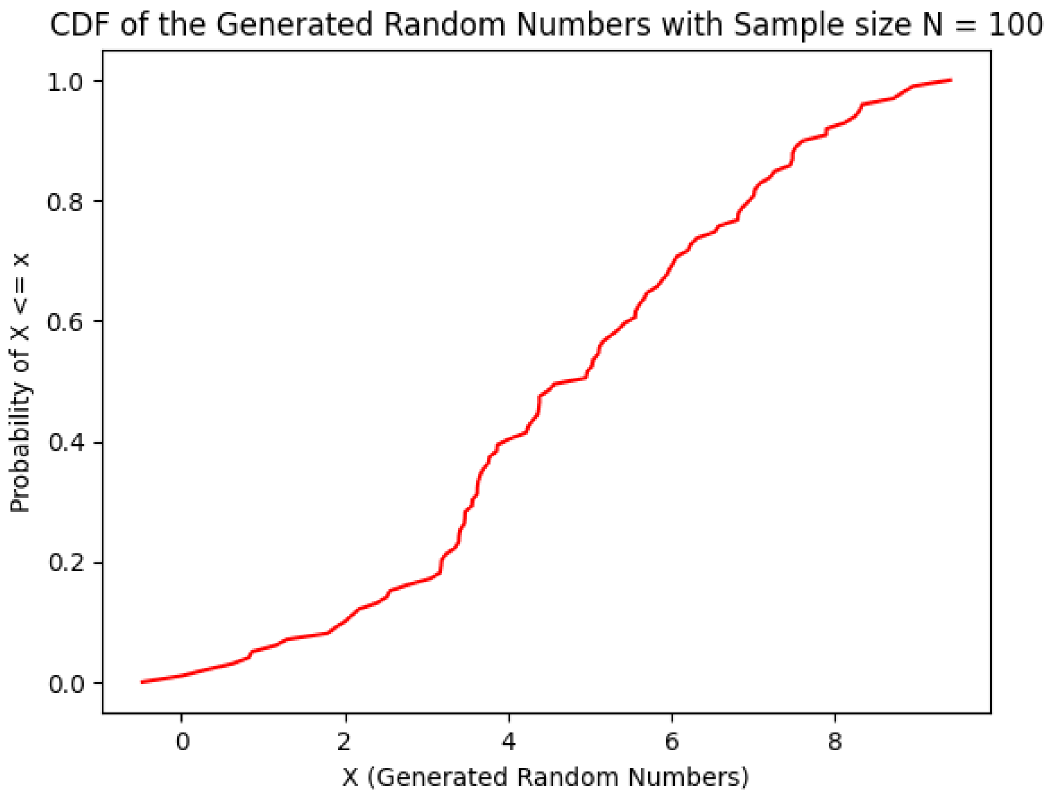




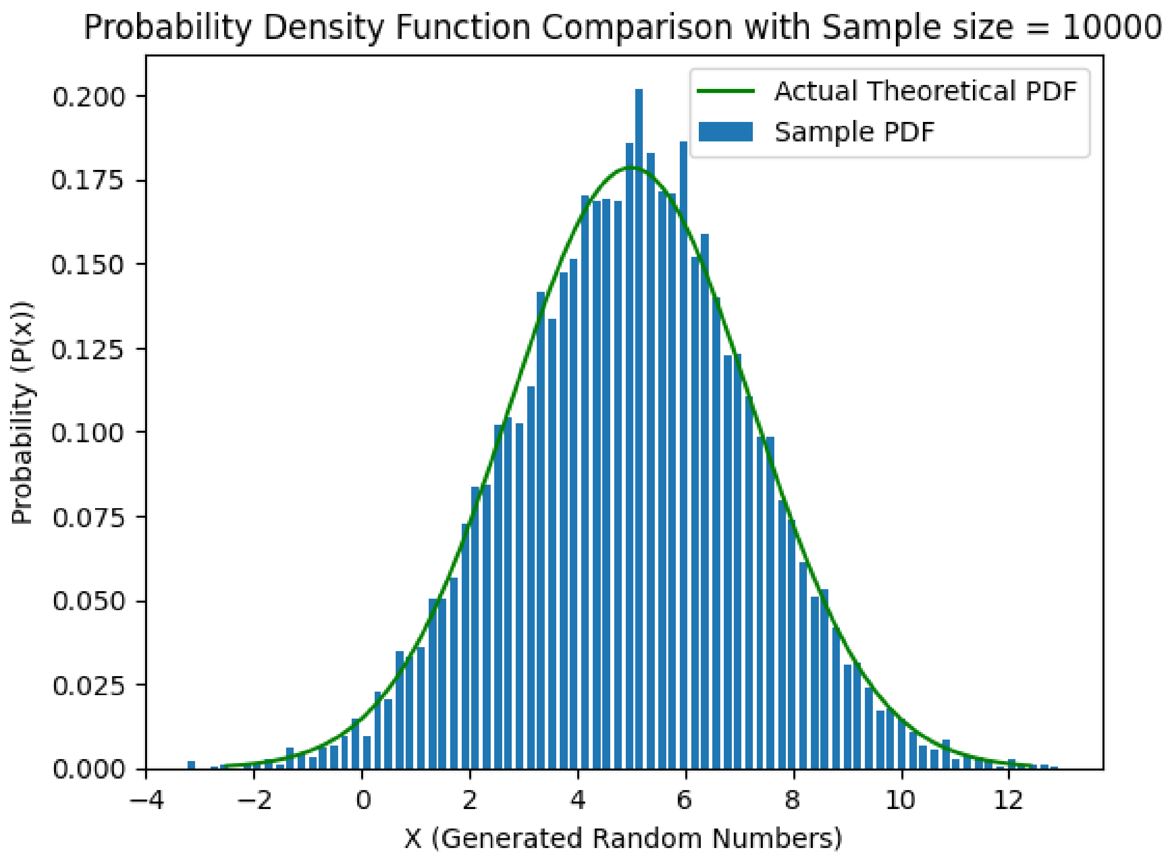
(b) For Marsaglia and Bray Method:

(i) When Sample Size = 100,

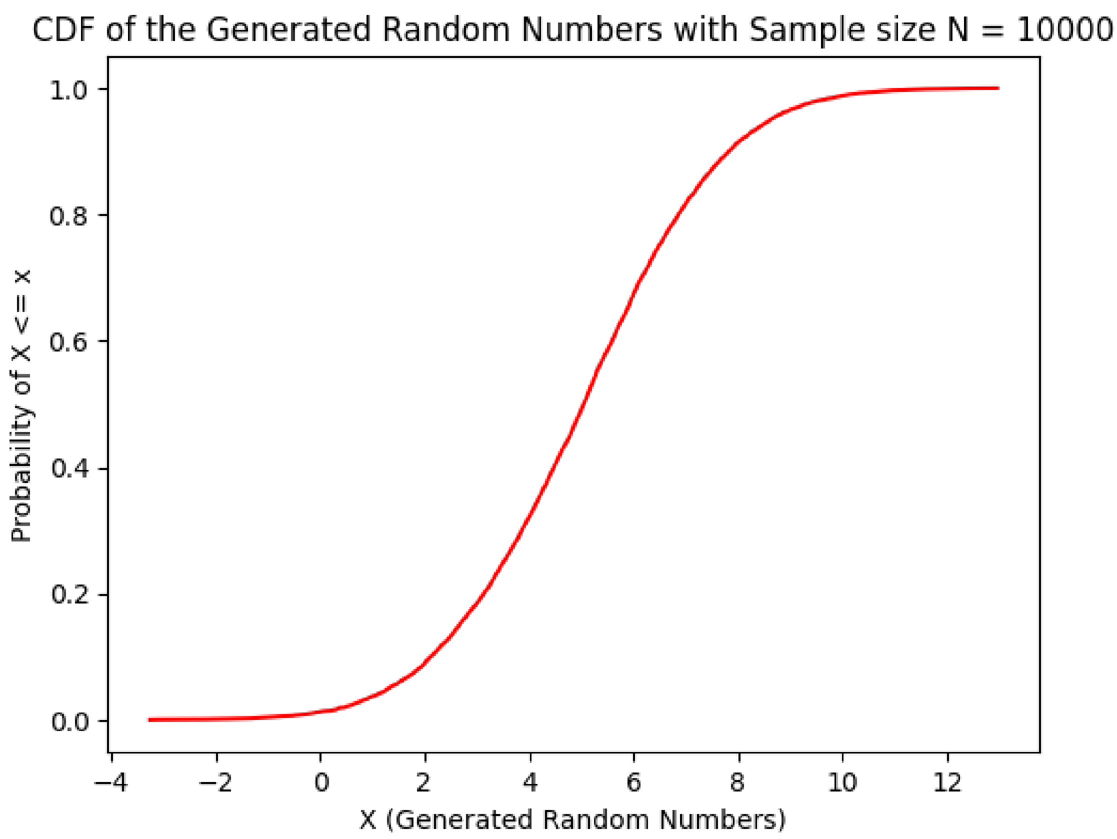




(ii) When Sample Size = 10000,



Probability Density Function Comparison with Sample size = 10000
Actual Theoretical PDF
Sample PDF
Probability (P(x))
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−4
−2
0
2
4
6
8
10
12
X (Generated Random Numbers)



CDF of the Generated Random Numbers with Sample size N = 10000
Probability of X <= x
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
6
8
10
12
X (Generated Random Numbers)

Observations:

1) When the sample size is less (N = 100), the deviations in the graphs from the theoretical graphs are more dominant and it deviates slightly from the normal distribution.
2) We can easily observe that the mean and variance vary with the graph as the Normal distributions are changed. The peak is attained at the mean, and the spread of the values depends on the variance/standard deviation. In both the normal distributions of N (0, 5) and N (5, 5), the graphs demonstrate this information with clarity. The graphs for N (5, 5) are shifted 5 units to the right of the N (0, 5) and its variance is same as that of N (0, 5).
3) In all the Normal Distributions given, as the sample size N increases, the Probability Density Function generated from the sample of random values generated approaches the theoretical Probability Density Function. This same trend is also observed for the Cumulative Distribution Function of the sample of random numbers generated. When the sample size if less (for N = 100), there is a huge deviation from the theoretically expected graphs.

Question 2:

2. Keep a track of the computational time required for both the methods. Which method is faster ?

The computation times of both the Methods are as follows (for samples generated from N (0, 1) distribution):

Sample Size	Box-Muller Method	Marsaglia and Bray Method
100	0.00013020992279052734 sec.	0.00012011051177978515 sec.
10000	0.007882399559020996 sec.	0.016033759117126466 sec.

Observations:

After running the simulations for many times and taking the average of the resultant computation time values, we observe that the Marsaglia and Bray method is faster than the Box-Muller method, when the sample size N = 100, and the Marsaglia and Bray method is slower than the Box-Muller method for sample size N = 10000. Theoretically, Marsaglia and Bray method is faster than the Box-Muller method as we avoid the calculation of the cos and sin functions. But in practical cases, this trend may not be followed always 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 process leads to a substantial overhead time, because of which Marsaglia and Bray method becomes slightly slower than the Box-Muller method. So, the time depends on the method 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. In modern Python libraries like the latest versions of numpy and math, the calculations of sin and cos are done in very efficient and fast way, so the Box-Muller method runs faster than the Acceptance Rejection technique (Marsaglia and Bray Method) for N = 10000. But this is not the case always as at a lower sample size (here, size = 100), we can clearly see that Marsaglia and Bray method beats the Box-Muller method. But when sample size increases to a very large number, such similar trends are difficult to observe.

Question 3:

3. For the Marsaglia and Bray method keep track of the proportion of values rejected. How does it compare with $1 - \frac{\pi}{4}$?

The comparison table for sample size N = 100 for different iterations is:

Iteration Number	Theoretical Value (1 - π/4)	Proportion of Values Rejected	Percentage Error
1	0.21460183660255172	0.180327868852459	15.97 %
2	0.21460183660255172	0.16666666666666663	22.34 %
3	0.21460183660255172	0.23076923076923073	7.53 %
4	0.21460183660255172	0.21875	1.93 %
5	0.21460183660255172	0.2857142857142857	33.14 %
6	0.21460183660255172	0.24242424242424243	12.96 %
7	0.21460183660255172	0.2647058823529411	23.35 %
8	0.21460183660255172	0.1071428571428571	50.07 %
9	0.21460183660255172	0.24242424242424243	12.96 %
10	0.21460183660255172	0.25373134328358204	18.23 %

The comparison table for sample size N = 10000 for different iterations is:

Iteration Number	Theoretical Value (1 - π/4)	Proportion of Values Rejected	Percentage Error
1	0.21460183660255172	0.2084850403672629	2.85 %
2	0.21460183660255172	0.21011058451816744	2.09 %
3	0.21460183660255172	0.2245657568238213	4.64 %
4	0.21460183660255172	0.22275765583708995	3.80 %

5	0.21460183660255172	0.21862791061103293	1.88 %
6	0.21460183660255172	0.21813917122752147	1.65 %
7	0.21460183660255172	0.22081969767804266	2.90 %
8	0.21460183660255172	0.21593225654696568	0.62 %
9	0.21460183660255172	0.20936116382036685	2.44 %
10	0.21460183660255172	0.20546639122834898	4.26 %

Observations:

Theoretically, the proportion of values rejected should be equal to the value $1 - \pi/4$ because this is the area of the discarded region from a square of unit area. We are choosing only those random numbers such that they lie inside a circle which is inscribed in a square having area = 1. So, the area of the remaining portion is equal to $1 - \pi/4$, which measures the proportion of the values rejected. When the sample size is very large (N = 10000), the deviation from the theoretical value is negligible, but when the sample size is less (N = 100), there is a slight deviation from the observed values which varies on different iterations of the Marsaglia and Bray simulation.