Kerstin Fontus

ISYE 6644

Group 101

A Study of the Tausworthe Pseudo-Random Number Generator

Abstract

The Tausworthe pseudo-random number generator simulates a Uniform(0,1) distribution by randomly generating bits and then converting them into base-2 numbers, which are then divided by their maximum value to get a number between 0 and 1. Implementing the algorithm and studying its output reveals that using certain values of the arguments needed to generate a sequence of numbers (*q*, *r;* and *l*) results in an identical and independent distribution.

Background

Robert C. Tausworthe published the first version of this generator in 1965 to help address situations where a mathematical model needs a random sequence of numbers [3]. At this time, he was working as a Research Engineer in the Jet Propulsion Laboratory in addition to his duties as a professor at the University of Southern California, so there is no doubt that he was encountering these problems often [2].

The Tausworthe Generator is a recursive generator, where each generated bit is a linear function of previous bits. In other words, we have a sequence a made up of k binary bits, where each element in the sequence a_i can be defined as $a_i = (c_1 * a_{i-1} + c_2 * a_{i-2} + \cdots + c_n * a_{i-n}) mod 2$ where the set of integers c_j, $j = 1,2, \ldots n$, are also binary and $c_n = 1$. The resulting sequence has period $p = 2^n - 1$ if the polynomial $f(x) = 1 + c_1 * x + c_2 * x^2 + \cdots + x^n$ is primitive, meaning that it is the lowest degree polynomial having coefficients 0 or 1 within the Galois field with order 2. This field contains two elements (0 and 1 in our case). Its additive properties are $0 + 0 = 1 + 1 = 0$ and$1 + 0 = 1$, which is analogous to the logical XOR operation. We are usually left with two non-zero coefficients c_i leaving us with $a_i = (a_{i-r} + a_{i-q}) mod 2$ where $0 < r < q$ [1]. Therefore, given an initialized sequence of q bits, we can define $a_i = a_{i-r}\ XOR\ a_{i-q}$.

We can then string together l consecutive a_i's, convert it into a base-2 integer, and divide it by 2^l to get a number between 0 and 1. Thus we have an algorithm for generating pseudo-random numbers between 0 and 1.

Implementation

I implemented the Tausworthe Generator with a series of functions in an R script, culminating in a single method to get pseudo-random numbers (PRN) and plots of adjacent and three times adjacent PRN's.

The function initial_binary_rand takes in a value *q* and returns a sequence of length *q* of random binary bits using the sample() function. This forms the basis for our recursive, "random" assignation of bits.

The function binary_sequence_xor takes in the initial binary sequence and values for *q*, *r*, and *n*, and uses the equation $a_i = a_{i-r}\ XOR\ a_{i-q}$ to return a sequence of bits, *a*, with length *n.*

The function bin_to_num takes in any sequence of binary bits and the length of the sequence *l,* and converts it to a base-2 number that is then divided by 2^l to obtain a number between 0 and 1.

The function bin_rand_gen takes in the sequence *a* and the value *l* that denotes how many bits that we want to use to create our PRN's, and returns a sequence of *n* PRN's.

The function taus_period calculates the period of the resulting sequence of PRN's. It is accurate assuming that certain requirements for *q*, *r,* and *l* are met.

The function taus_gen takes four arguments: *N* is the number of PRN's we want to generate, *L* is the number of bits we want to use to calculate each PRN, *Q* is the length of the initial binary vector we want to generate, and *R* is a number strictly between 0 and *Q* that is used to generate bits. It returns a vector of length *N* of Uniform(0,1) PRN's. It also outputs a PNG file of a histogram showing the distribution of the PRN's.

The function plot taus_2d takes in a vector of PRN's, plots PRN *i* against PRN *i+1* on a unit square, and saves it as a PNG file. Similarly, plot_taus_3d takes in a vector of PRN's and plots PRN's *i, i+1* and *i+2* on a unit cube, which is displayed by R. In both types of plots, the color gradient of the points is based on the frequency of its occurrence, i.e., when the period of the PRN's is less than the number generated. When there is no repetition, the points appear to be one shade of blue. Otherwise, a darker blue indicates a more frequently occurring point. The function plot_taus runs both of the above functions at once.

After generating a sequence of PRN's, we want to test if they are approximately identically and independently distributed Uniform(0,1) random numbers through a goodness of fit test and some independence tests. The function gof takes in our sequence of PRN's and the number of bins, *k,* we want to use in a Chi-Squared Goodness of Fit Test with α=0.05, and prints out the results. The function up_down_ind takes in our vector of PRN's, runs the Up and Down Test of independence with α=0.05, and then prints out the results. The function mean_ind runs the Above and Below the Mean Test of Independence with α=0.05 on the PRN's and prints out the result. The function ind_and_fit takes our PRN vector and the value *k* and runs all of these tests.

Finally, there are many ways to generate Normal(0,1) variates from Uniform(0,1). The function z_table_norm does a simple "table lookup" by translating the PRN as a percentage to the standard deviation on the Normal(0,1) curve. The function b_m_norm applies the Box-Mueller transformation and outputs a data frame with two columns of Normal(0,1) deviates. The function `polar_norm` applies the polar method of the Box-Mueller transformation and outputs a data frame with two columns of Normal(0,1) deviates. The function `norm_test` outputs the mean and standard deviation of the inputted vector of Normal(0,1) deviates. Finally, the function `norm_dev` generates Normal(0,1) deviates with each method, creates a plot of histograms so we can examine the distribution of each method, and prints out the mean and variance for each generated sequence.

The method `main` runs the full algorithm of generating a sequence of PRN's, plotting them, running independence and goodness of fit tests, and generating normal deviates. It does not take in arguments; the user must define variables within its function definition to be used by all of the function described above. Examples of its output can be found in the Appendix.

Results

I conducted a number of tests using different values of *q, r,* and *l*, using $k = 5$ for all my goodness of fit tests. These tests revealed some criteria for consistently creating an identically and independently distributed sequence of Uniform(0,1) random numbers.

It is best to make sure that the potential period, 2^q-1, is greater than or equal to the amount of PRN's that will be generated. It is possible to get an identically and independently distributed output without meeting this criterion, but if we want to avoid repeating numbers then this is not ideal (Appendix A).

The values *q* and *r* must be co-prime. Otherwise, this results in a much shorter period than what should be outputted by our algorithm (2^q - 1) (Appendix B, C). For a sequence where *q* is a multiple of *r*, we get a sequence with a period equal to *r*, which obviously does not make for an independent, identical sequence (Appendix D).

For small values of *q* and *r*, they should not be close together. This again results in a shorter period than we prefer. In some cases, the resulting PRN's do not consistently exhibit independence and uniformity (Appendix E). For larger *q* and *r,* this is okay sometimes but it can cause us to fail some independence tests (Appendix F, G).

The magnitude of *l* is not too important, but should be scaled in relation to how many PRN's you would like to be generated. There are $2^l + 1$ possible numbers between 0 and 1 inclusive than can be generated. A value as small as 4 works when you have few numbers you need to generate (Appendix H).

Even when following all these rules, we can still be unsuccessful (Appendix I). Large values of *l*, *q*, and *r* are not crucial for generating large sequences of PRN's, but for the purposes of consistently creating unique, non-repeating numbers, it is best to scale those values with the number that you want generated (Appendix J).

Normal(0,1) Deviate Generation

Generating Normal(0,1) deviates from the generated PRNs had varying success, as revealed by the histograms of the sequences and their means and variances. It is possible to have a sequence of PRN's that is neither Uniform(0,1) nor independent, but generates some decent

Normal(0,1) distributions at first glance (Appendix I). However, these cases usually result non-Normal(0,1) looking output (Appendix C). For the most part, the best Normal(0,1) deviates were created from PRN's that are independently and identically distributed to the Uniform(0,1) distribution (Appendix K).

Conclusion

Overall, to get the most out of using the Tausworthe Generator, one should choose l such that $2^l + 1$ is greater than the length of the sequence they want to generate. In addition, they should choose q and r such that they are not close together in value and that they are scaled in relation to the length of the sequence they want. Successfully implementing the algorithm has potential for creating input for simulations or for the generation of other distributions, though this needs to be studied beyond Normal(0,1) deviates.

References

1. Law, Averill M., W. David Kelton, and W. David Kelton. *Simulation Modeling and Analysis*. Vol. 3. New York: McGraw-Hill, 2007.
2. Robert Tausworthe (n.d.). [LinkedIn page]. LinkedIn. Retrieved November 28, 2023, from https://www.linkedin.com/in/robert-tausworthe-2694a855/.
3. Tausworthe, R.C. "Random Numbers Generated by Linear Recurrence Modulo Two." In *Mathematics and Computation*, volume 19, pages 201-209, 1965.

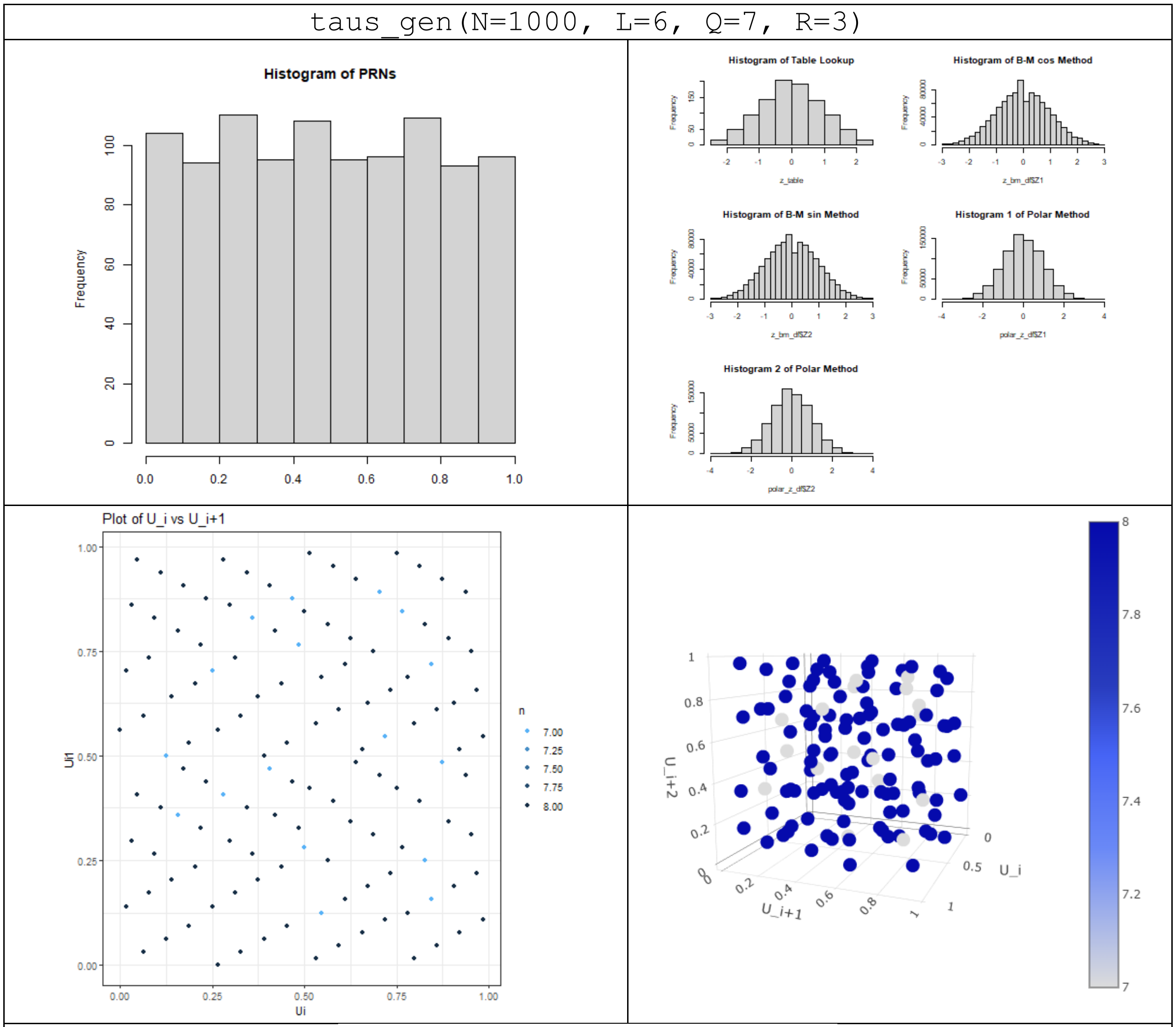
Appendix

All of the below tables have the following layout:

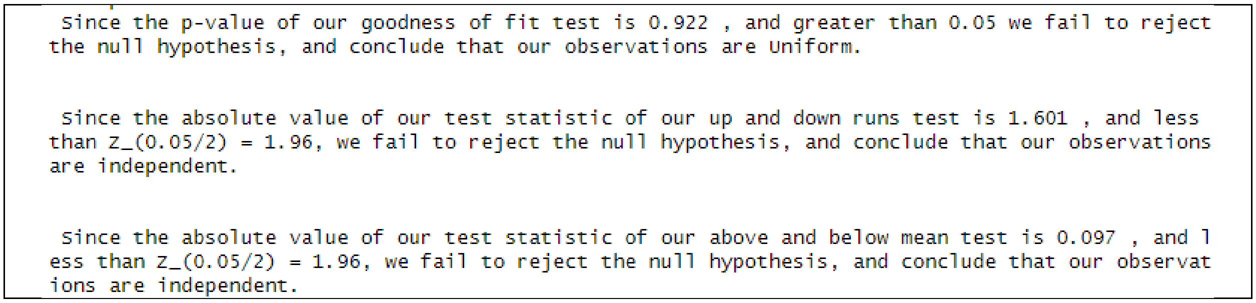
- Row 1: Function call for `taus_gen()`
- Row 2: Histogram of PRN's, histograms of Normal(0,1) deviates
- Row 3: 2-D plot of adjacent PRN's, 3-D plot of twice adjacent PRN's
- Row 4: Means and variances of Normal(0,1) deviates
- Row 5: Goodness of fit and independence test results

A.

taus_gen(N=1000, L=6, Q=7, R=3)

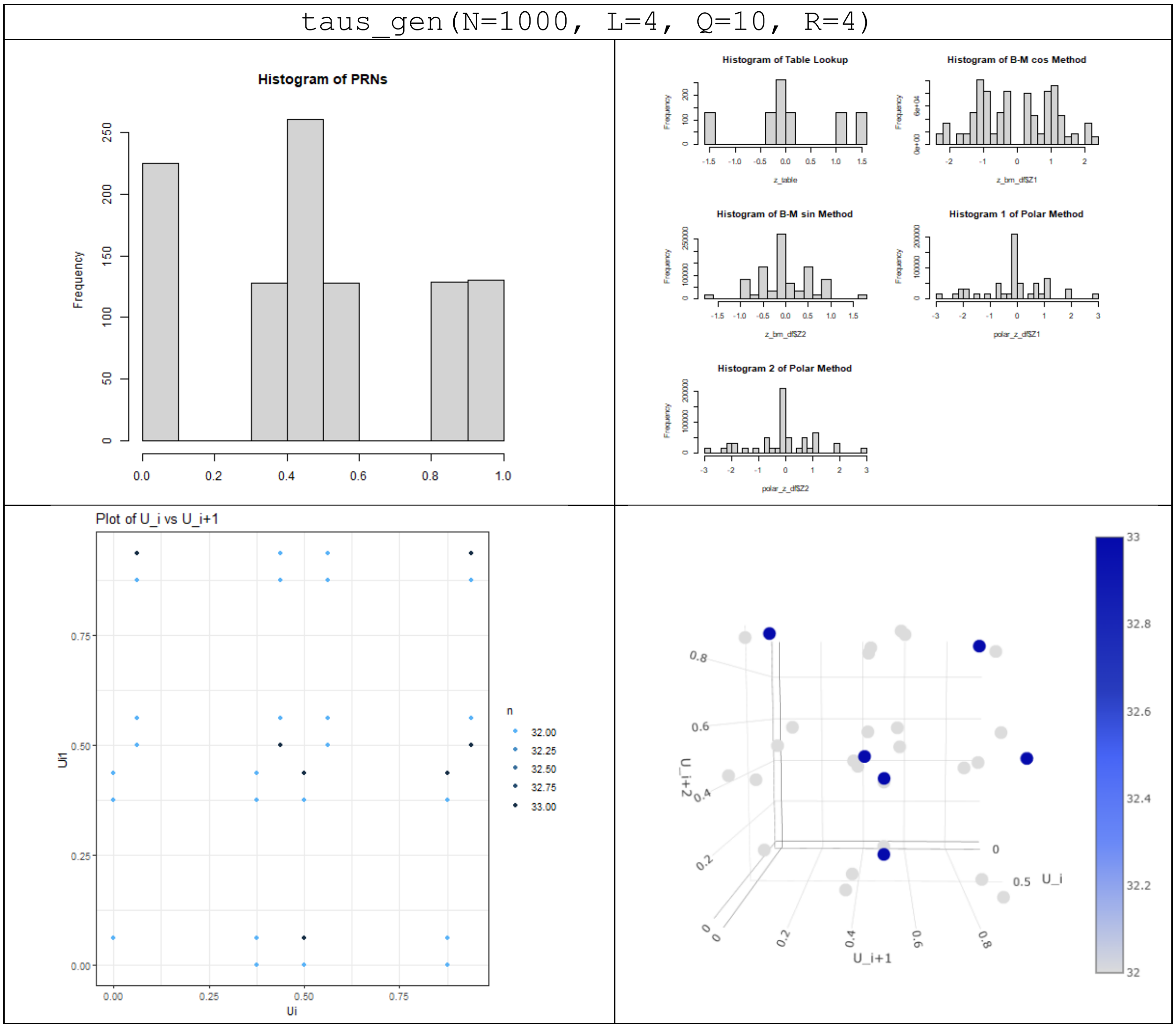


```
Table Lookup:
Mean: -0.0009405902 Variance 0.8980153
Box-Mueller Method:
Mean: -0.006250913 Variance 0.9652291
Mean: 0.0006264497 Variance 0.9768091
Polar Method:
Mean: 0.0005918057 Variance 0.9972024
Mean: 0.0005918057 Variance 0.9972024png
```

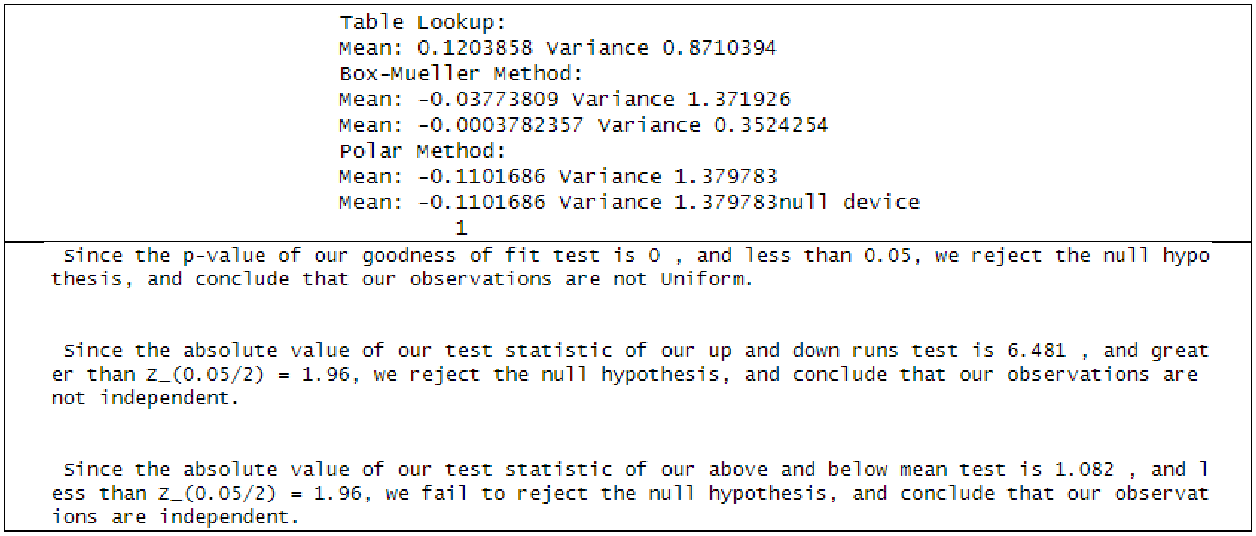


```
 Since the p-value of our goodness of fit test is 0.922 , and greater than 0.05 we fail to reject
the null hypothesis, and conclude that our observations are Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 1.601 , and less
than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observations
are independent.


 Since the absolute value of our test statistic of our above and below mean test is 0.097 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

B.



taus_gen(N=1000, L=4, Q=10, R=4)

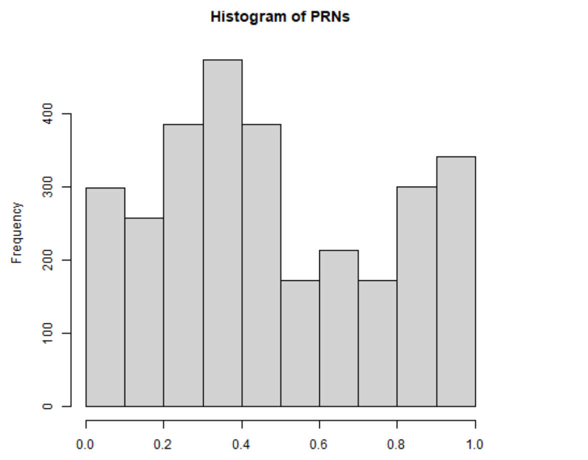


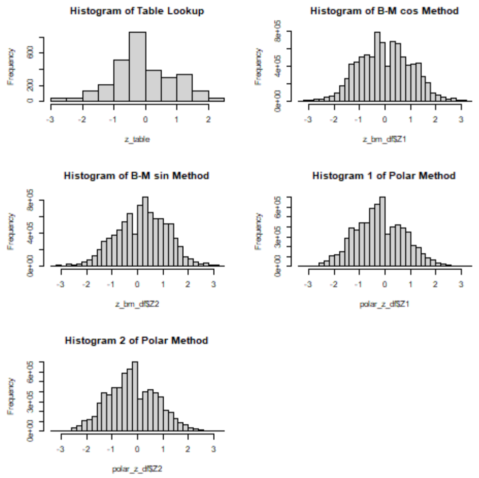
```
Table Lookup:
Mean: 0.1203858 Variance 0.8710394
Box-Mueller Method:
Mean: -0.03773809 Variance 1.371926
Mean: -0.0003782357 Variance 0.3524254
Polar Method:
Mean: -0.1101686 Variance 1.379783
Mean: -0.1101686 Variance 1.379783null device
          1
```

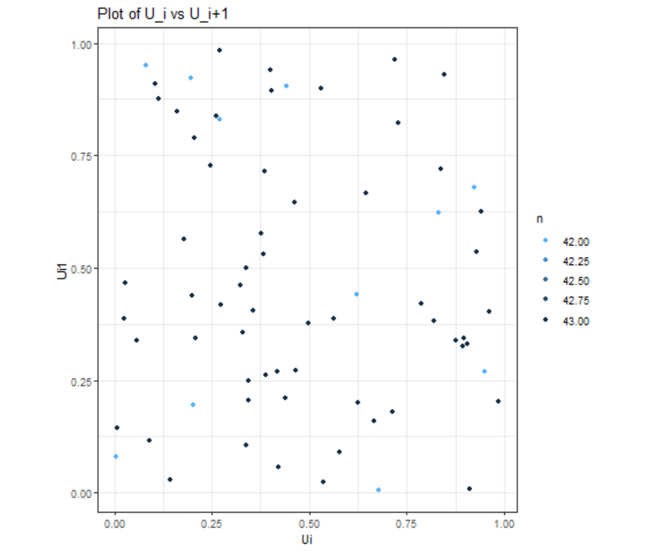
```
 Since the p-value of our goodness of fit test is 0 , and less than 0.05, we reject the null hypo
thesis, and conclude that our observations are not Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 6.481 , and great
er than Z_(0.05/2) = 1.96, we reject the null hypothesis, and conclude that our observations are
not independent.


 Since the absolute value of our test statistic of our above and below mean test is 1.082 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

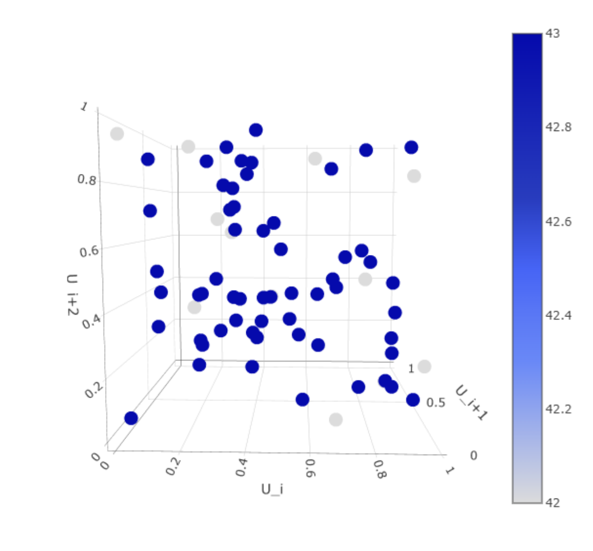
C.

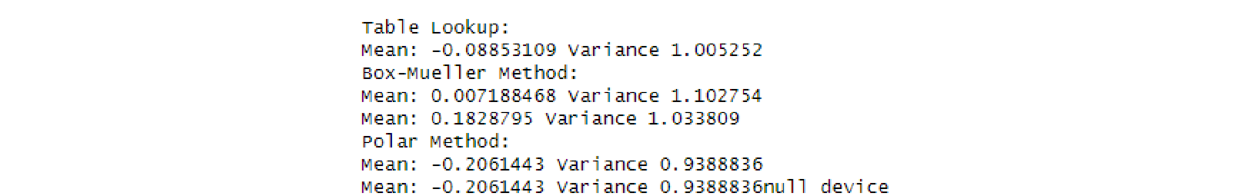
taus_gen(N=3000, L=150, Q=1500, R=500)



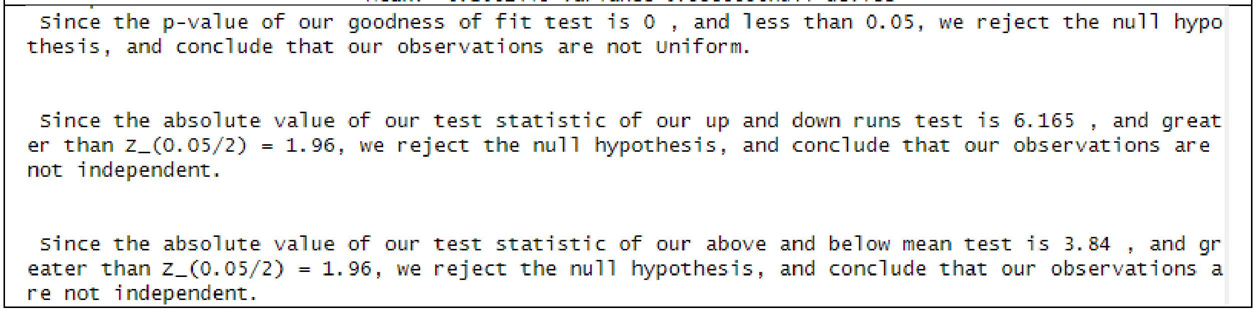








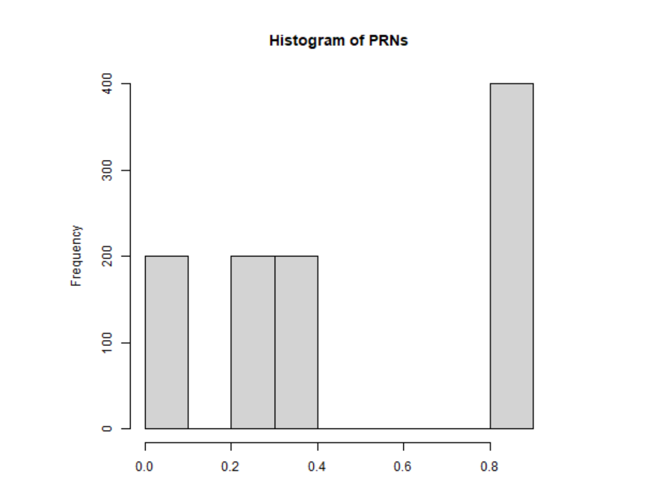
```
Table Lookup:
Mean: -0.08853109 Variance 1.005252
Box-Mueller Method:
Mean: 0.007188468 Variance 1.102754
Mean: 0.1828795 Variance 1.033809
Polar Method:
Mean: -0.2061443 Variance 0.9388836
Mean: -0.2061443 Variance 0.9388836null device
```

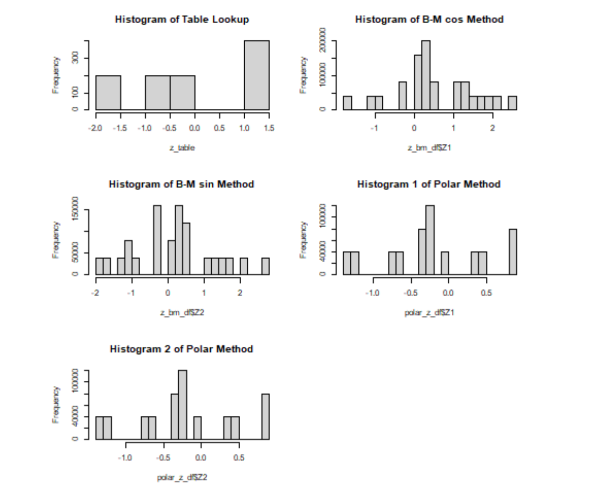


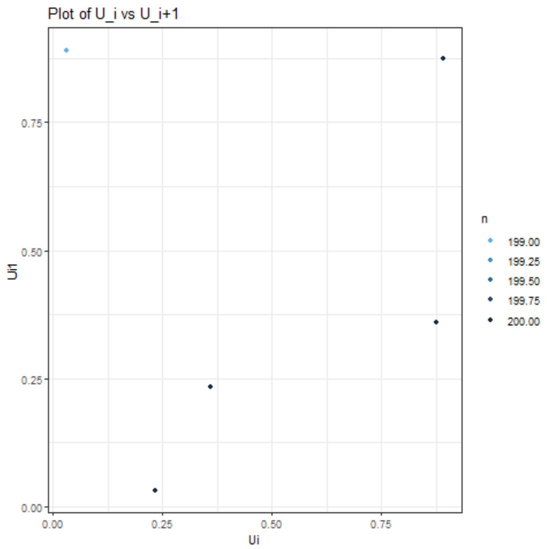
```
 Since the p-value of our goodness of fit test is 0 , and less than 0.05, we reject the null hypo
thesis, and conclude that our observations are not Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 6.165 , and great
er than Z_(0.05/2) = 1.96, we reject the null hypothesis, and conclude that our observations are
not independent.


 Since the absolute value of our test statistic of our above and below mean test is 3.84 , and gr
eater than Z_(0.05/2) = 1.96, we reject the null hypothesis, and conclude that our observations a
re not independent.
```

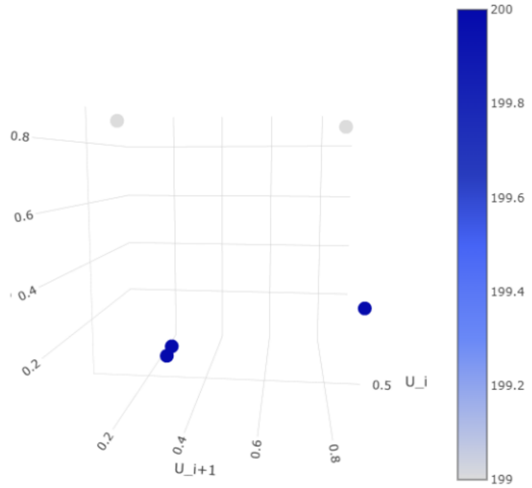
D.

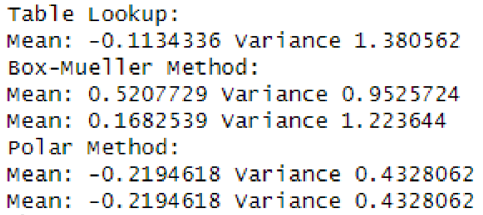
taus_gen(N=1000, L=6, Q=10, R=5)



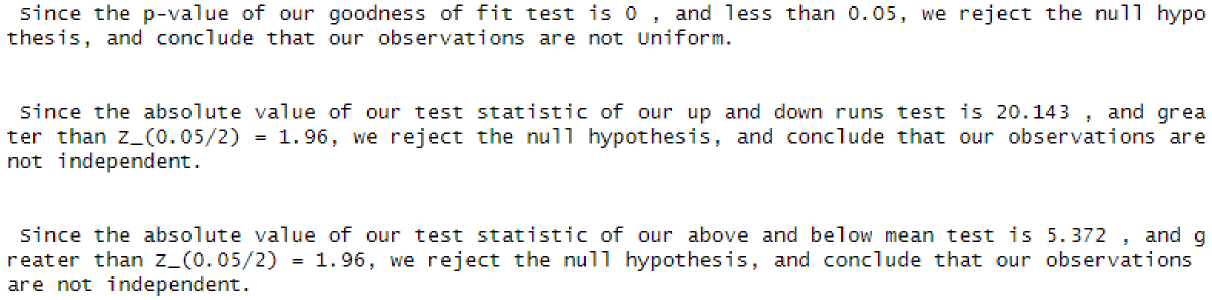








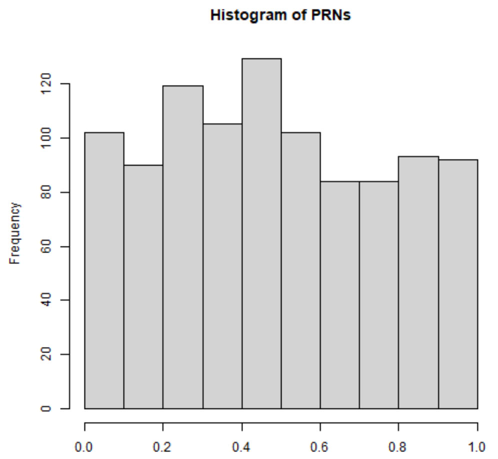
```
Table Lookup:
Mean: -0.1134336 Variance 1.380562
Box-Mueller Method:
Mean: 0.5207729 Variance 0.9525724
Mean: 0.1682539 Variance 1.223644
Polar Method:
Mean: -0.2194618 Variance 0.4328062
Mean: -0.2194618 Variance 0.4328062
```

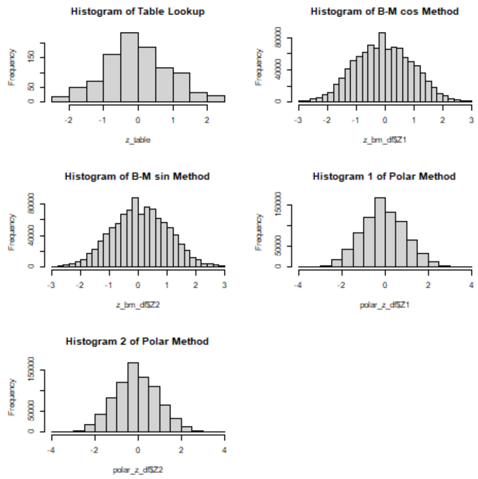


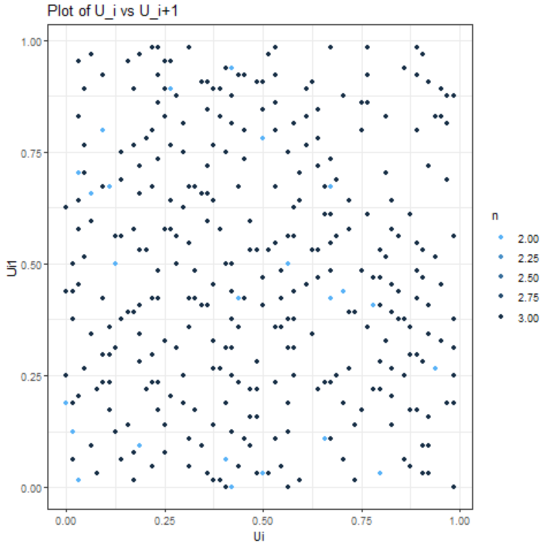
```
 Since the p-value of our goodness of fit test is 0 , and less than 0.05, we reject the null hypo
thesis, and conclude that our observations are not Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 20.143 , and grea
ter than Z_(0.05/2) = 1.96, we reject the null hypothesis, and conclude that our observations are
not independent.


 Since the absolute value of our test statistic of our above and below mean test is 5.372 , and g
reater than Z_(0.05/2) = 1.96, we reject the null hypothesis, and conclude that our observations
are not independent.
```

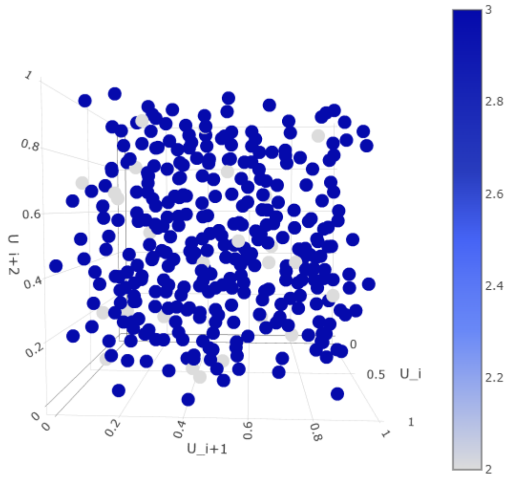
E.

```
taus_gen(N=1000, L=6, Q=10, R=7)
```







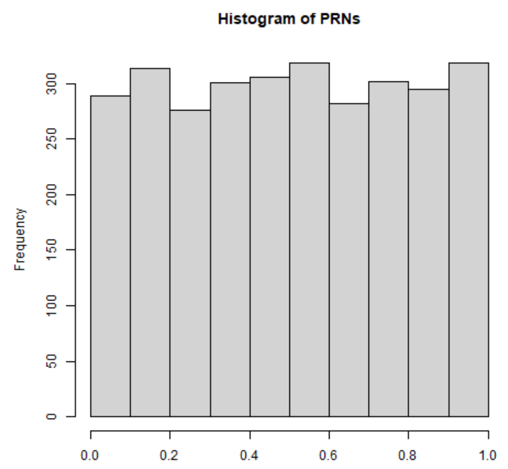


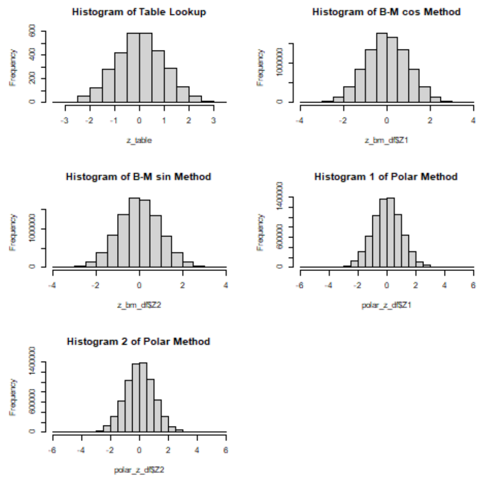
```
Table Lookup:
Mean: -0.02823068 Variance 0.8813078
Box-Mueller Method:
Mean: -0.04497671 Variance 1.00978
Mean: 0.06826471 Variance 0.959513
Polar Method:
Mean: -0.07289046 Variance 1.056986
Mean: -0.07289046 Variance 1.056986null device
```

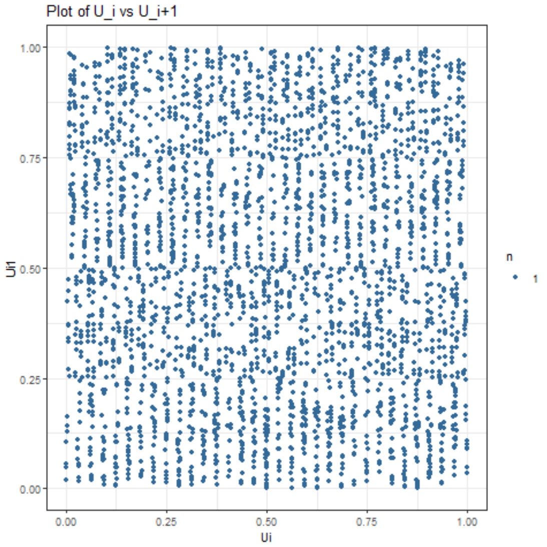
```
 Since the p-value of our goodness of fit test is 0.007 , and less than 0.05, we reject the null
hypothesis, and conclude that our observations are not Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 0.25 , and less t
han Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observations
are independent.


 Since the absolute value of our test statistic of our above and below mean test is 0.964 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

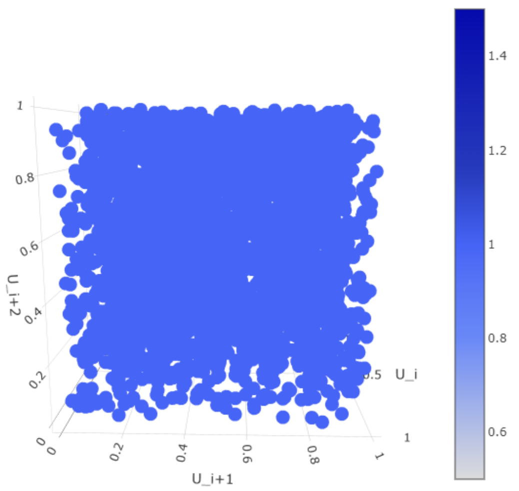
F.

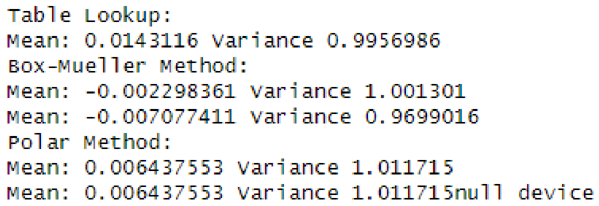
taus_gen(N=3000, L=20, Q=15, R=14)



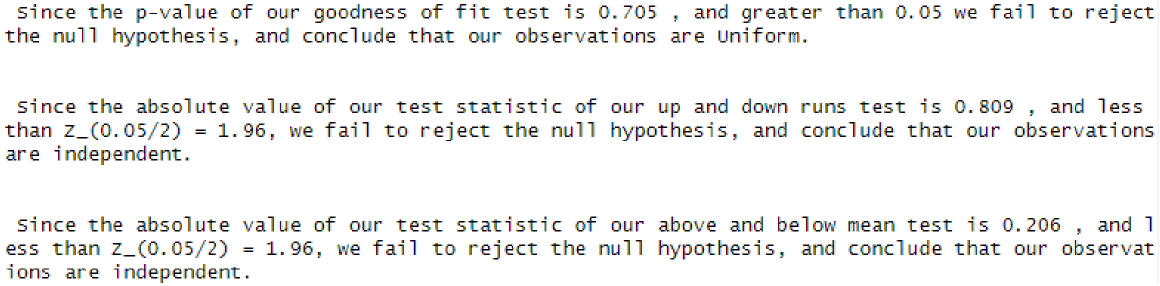








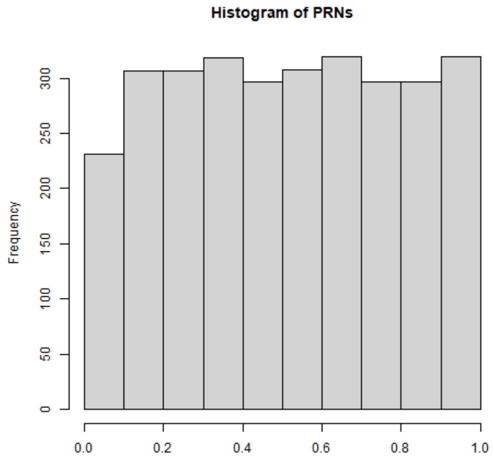
```
Table Lookup:
Mean: 0.0143116 Variance 0.9956986
Box-Mueller Method:
Mean: -0.002298361 Variance 1.001301
Mean: -0.007077411 Variance 0.9699016
Polar Method:
Mean: 0.006437553 Variance 1.011715
Mean: 0.006437553 Variance 1.011715null device
```

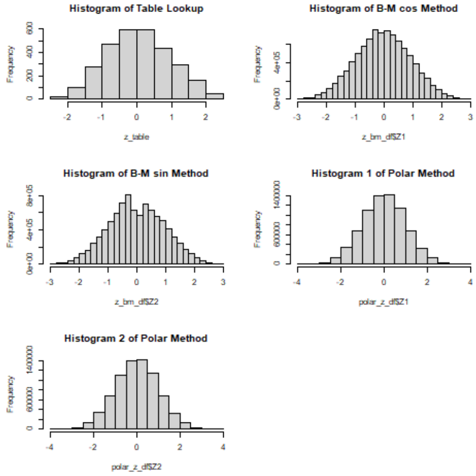


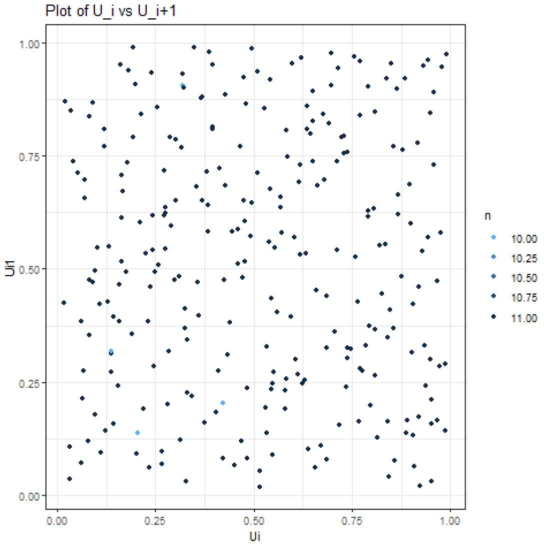
```
 Since the p-value of our goodness of fit test is 0.705 , and greater than 0.05 we fail to reject
the null hypothesis, and conclude that our observations are Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 0.809 , and less
than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observations
are independent.


 Since the absolute value of our test statistic of our above and below mean test is 0.206 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

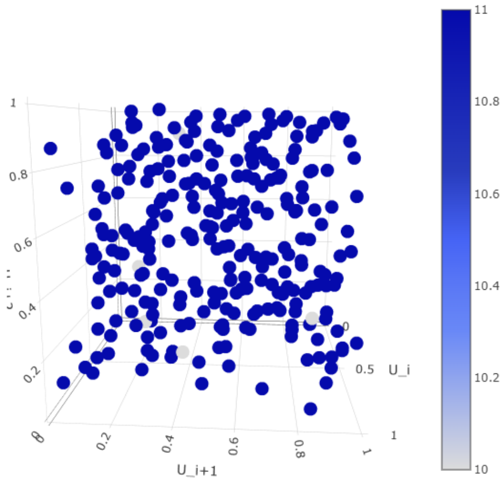
G.

```
taus_gen(N=3000, L=23, Q=17, R=16)
```







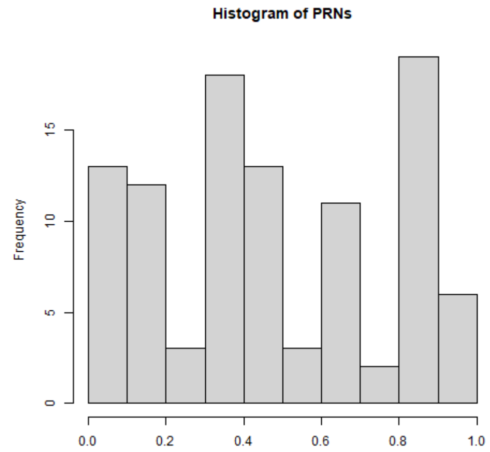


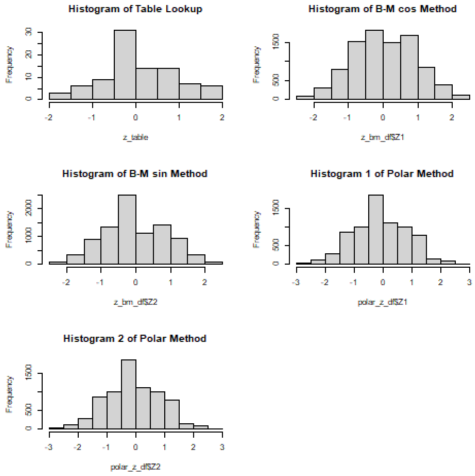
```
Table Lookup:
Mean: 0.05424044 Variance 0.8471281
Box-Mueller Method:
Mean: -0.03132781 Variance 0.8814452
Mean: -0.004637075 Variance 0.9280633
Polar Method:
Mean: 0.008244445 Variance 1.008405
Mean: 0.008244445 Variance 1.008405null device
```

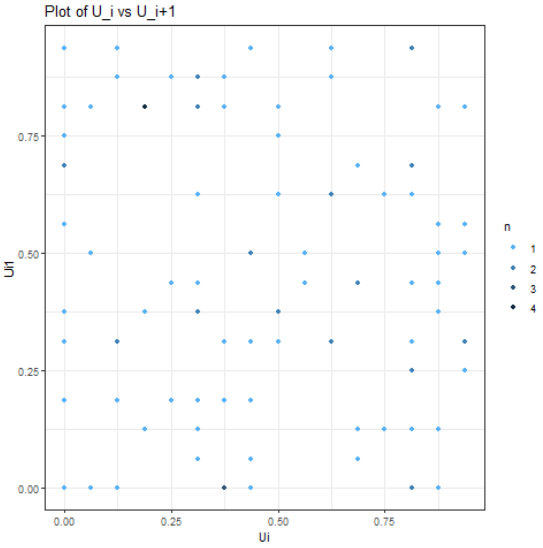
```
 Since the p-value of our goodness of fit test is 0.08 , and greater than 0.05 we fail to reject
the null hypothesis, and conclude that our observations are Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 4.75 , and greate
r than Z_(0.05/2) = 1.96, we reject the null hypothesis, and conclude that our observations are n
ot independent.


 Since the absolute value of our test statistic of our above and below mean test is 1.446 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

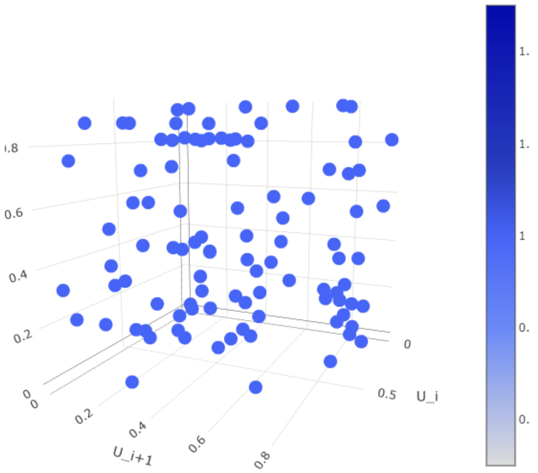
H.

```
taus_gen(N=100, L=4, Q=11, R=3)
```







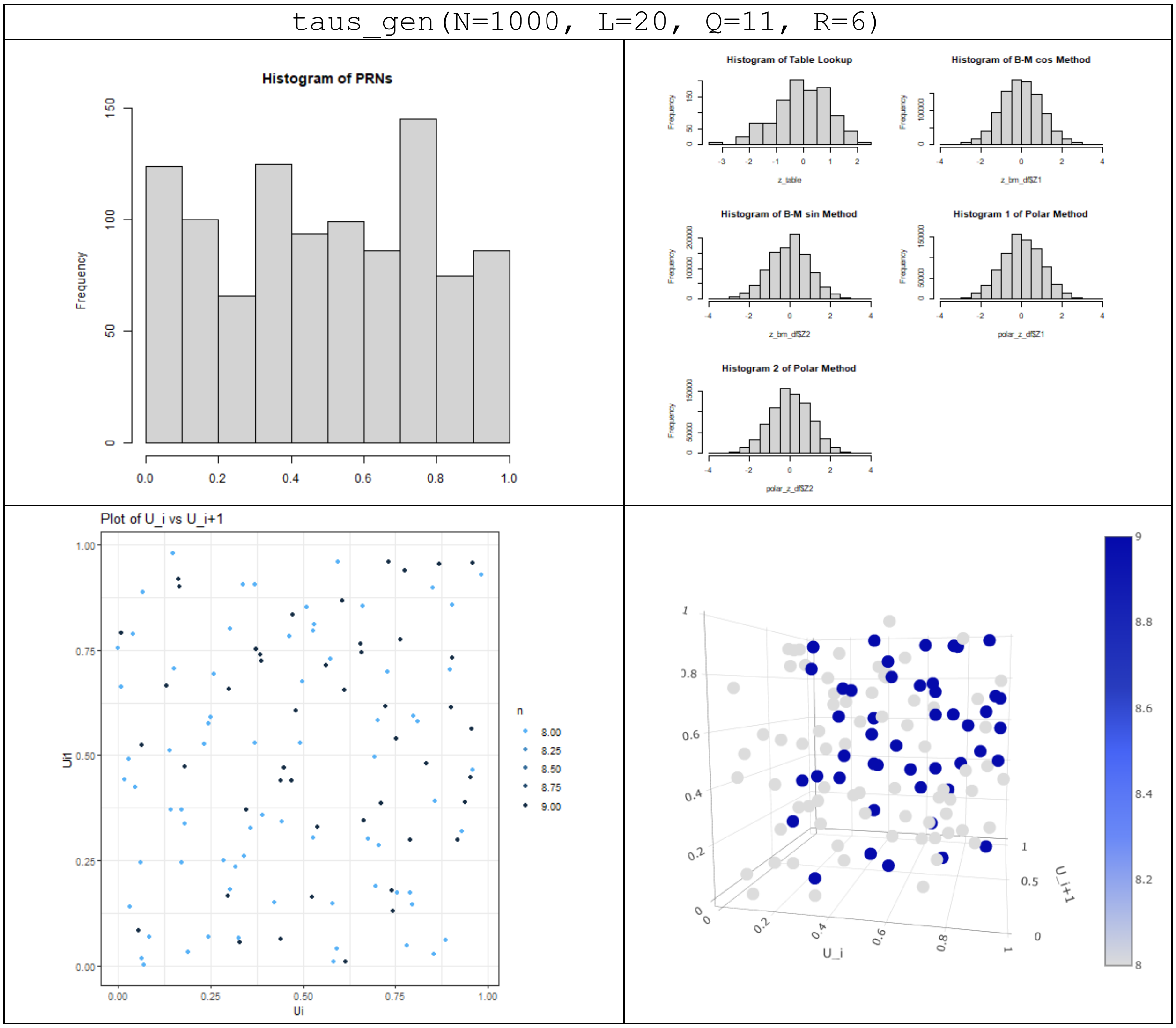


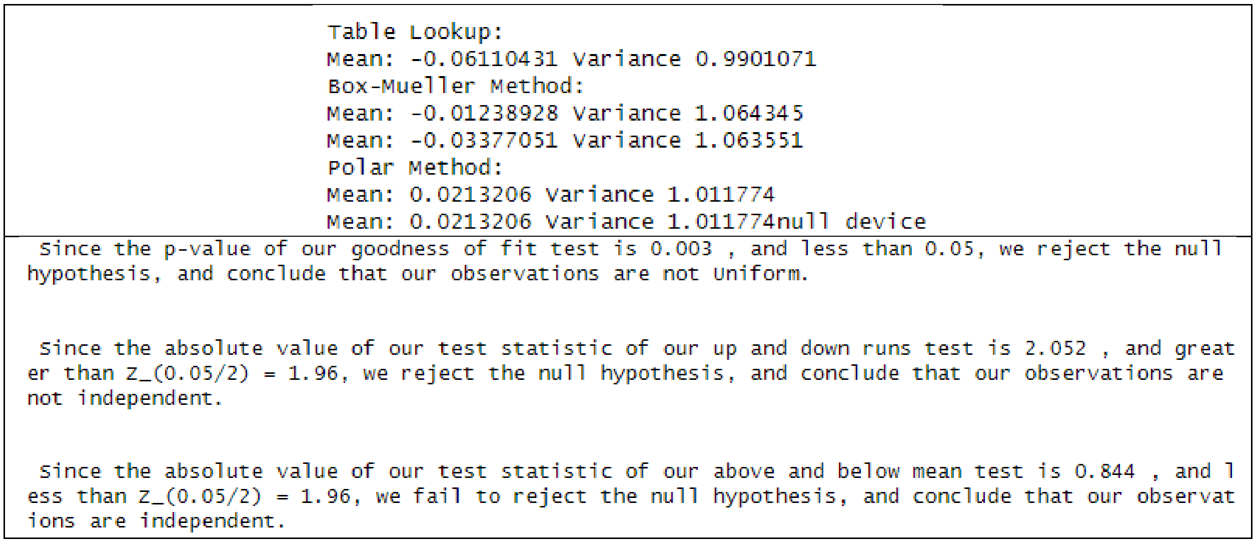
```
Table Lookup:
Mean: -0.1134336 Variance 1.380562
Box-Mueller Method:
Mean: 0.5207729 Variance 0.9525724
Mean: 0.1682539 Variance 1.223644
Polar Method:
Mean: -0.2194618 Variance 0.4328062
Mean: -0.2194618 Variance 0.4328062
```

```
 Since the p-value of our goodness of fit test is 0.215 , and greater than 0.05 we fail to reject
the null hypothesis, and conclude that our observations are Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 0.319 , and less
than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observations
are independent.


 Since the absolute value of our test statistic of our above and below mean test is 0.065 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

I.

taus_gen(N=1000, L=20, Q=11, R=6)

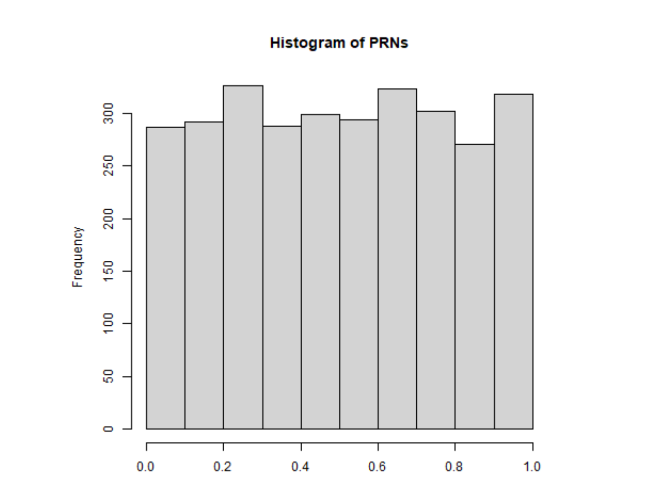


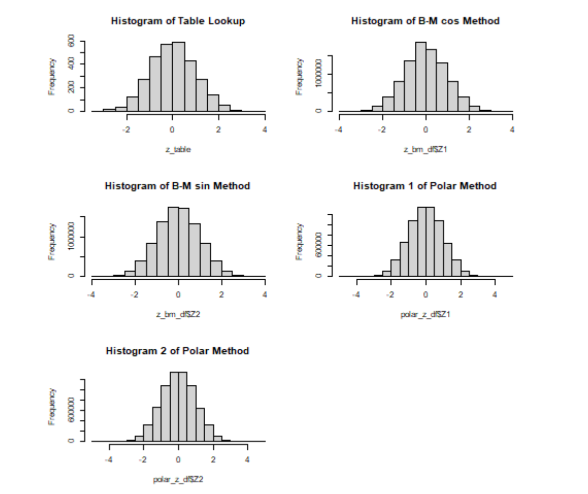


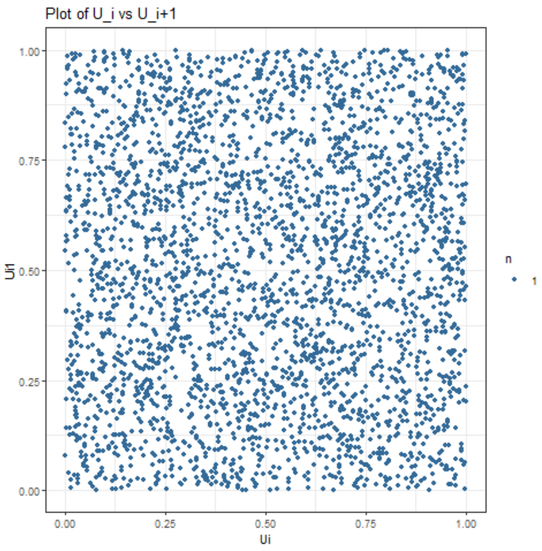
```
                          Table Lookup:
                          Mean: -0.06110431 Variance 0.9901071
                          Box-Mueller Method:
                          Mean: -0.01238928 Variance 1.064345
                          Mean: -0.03377051 Variance 1.063551
                          Polar Method:
                          Mean: 0.0213206 Variance 1.011774
                          Mean: 0.0213206 Variance 1.011774null device
 Since the p-value of our goodness of fit test is 0.003 , and less than 0.05, we reject the null
hypothesis, and conclude that our observations are not Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 2.052 , and great
er than Z_(0.05/2) = 1.96, we reject the null hypothesis, and conclude that our observations are
not independent.


 Since the absolute value of our test statistic of our above and below mean test is 0.844 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

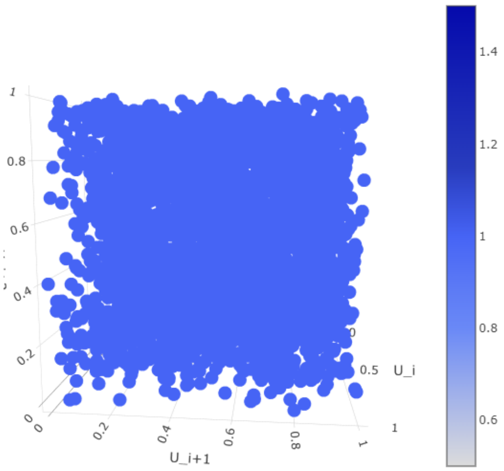
J.

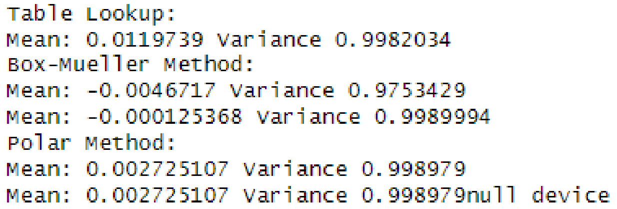
taus_gen(N=3000, L=20, Q=15, R=7)



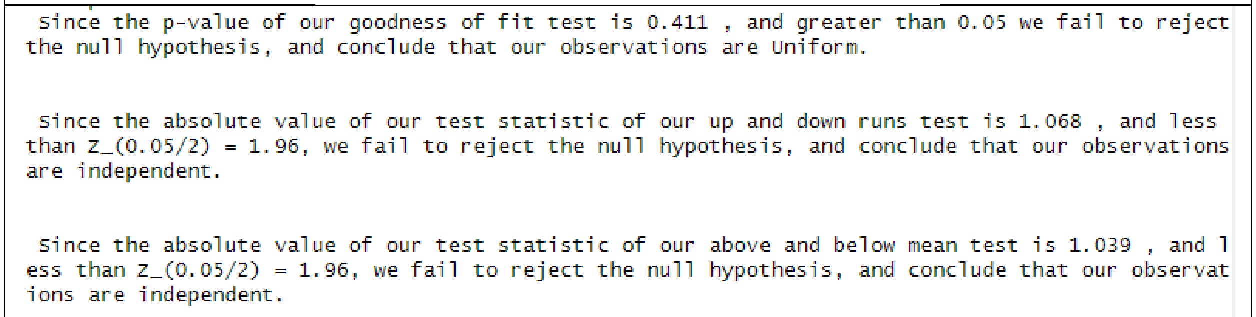








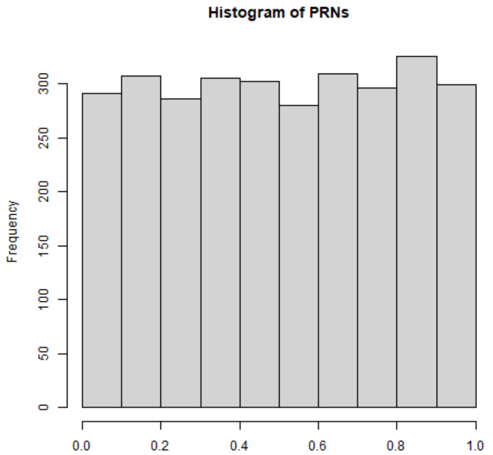
```
Table Lookup:
Mean: 0.0119739 Variance 0.9982034
Box-Mueller Method:
Mean: -0.0046717 Variance 0.9753429
Mean: -0.000125368 Variance 0.9989994
Polar Method:
Mean: 0.002725107 Variance 0.998979
Mean: 0.002725107 Variance 0.998979null device
```

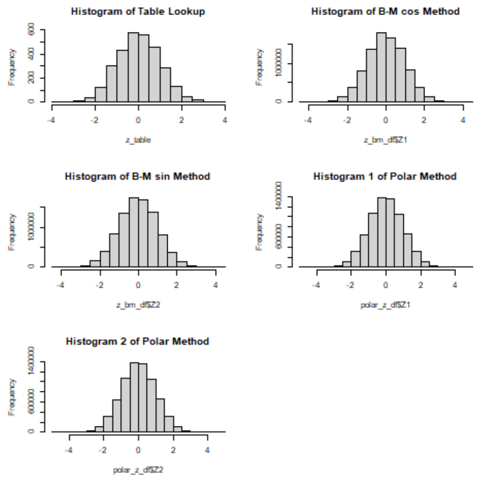


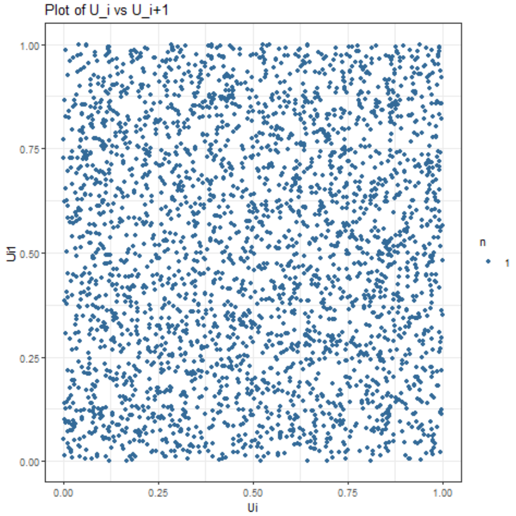
```
 Since the p-value of our goodness of fit test is 0.411 , and greater than 0.05 we fail to reject
the null hypothesis, and conclude that our observations are Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 1.068 , and less
than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observations
are independent.


 Since the absolute value of our test statistic of our above and below mean test is 1.039 , and l
ess than Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observat
ions are independent.
```

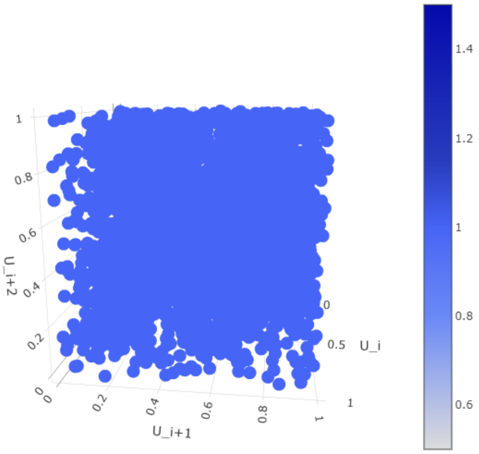
K.

taus_gen(N=3000, L=157, Q=1113, R=536)









```
Table Lookup:
Mean:  0.01852196 Variance 0.9938038
Box-Mueller Method:
Mean:  0.004581701 Variance 0.9798927
Mean:  -0.00629107 Variance 0.9790715
Polar Method:
Mean:  0.003997407 Variance 0.9918115
Mean:  0.003997407 Variance 0.9918115null device
```

```
 Since the p-value of our goodness of fit test is 0.841 , and greater than 0.05 we fail to reject the null
hypothesis, and conclude that our observations are Uniform.


 Since the absolute value of our test statistic of our up and down runs test is 1.761 , and less than Z_
(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observations are independen
t.


 Since the absolute value of our test statistic of our above and below mean test is 1.404 , and less than
Z_(0.05/2) = 1.96, we fail to reject the null hypothesis, and conclude that our observations are independe
nt.
```