Q1. The following output are obtained from running CSC446_A3_Q1.ipynb in Jupyter Hub

(using random.seed(1) and random.random())
The list of 10 random numbers
[0.13436424411240122,
 0.8474337369372327,
 0.763774618976614,
 0.2550690257394217,
 0.49543508709194095,
 0.4494910647887381,
 0.651592972722763,
 0.7887233511355132,
 0.0938595867742349,
 0.02834747652200631]

From reading the documentation of Python standard libraries, we can know that the random() module uses **Mersenne Twister (MT)**.

use Kolmogorv-Smirnov method with a 0.05 level of significance for uniformity test

1. Rank the data from the smallest to largest ($R_{(i)}$)

   ten_random =
   [0.02834747652200631, 0.0938595867742349, 0.13436424411240122, 0.2550690257394217, 0.4494910647887381, 0.49543508709194095, 0.651592972722763, 0.763774618976614, 0.7887233511355132, 0.8474337369372327]
2. Produce a list for i/N (N = sample size)

   freq_list = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
3. Produce a list for D^+ (i/N - $R_{(i)}$, negative values are ignored)

   d_plus_list =
   [0.0716525234779937, 0.10614041322576512, 0.16563575588759877, 0.14493097426057833, 0.050508935211261874, 0.10456491290805903, 0.048407027277237, 0.036225381023386016, 0.11127664886448685, 0.15256626306276733]
4. Produce a list for D^- ($R_{(i)}$ - (i-1)/N, negative values are ignored)

   d_minus_list =
   [0.02834747652200631, 0.049491064788738104, 0.051592972722762975, 0.06377461897661407]

5. Locate the critical value using level of significance and degree of freedom (sample size) and compare it with sample statistics D ($D = max(D^-, D^+)$, D_{alpha} = 0.410)

   d_max = 0.16563575588759877
   d_critical = 0.410
   d_max < d_critical, therefore H_0, the null hypothesis, is not rejected. There are no significant difference between the sample distribution and theoretical distribution

Chi-square test on 2000 random numbers

1. Decide/count the number of classes n and determine the expected number in each class E_i

   n=20, [0, 0.05] [0.05, 0.01].....[0.95 1.0], E_i = 2000/20 = 100
2. Count the observed number in each class O_i

   [112, 88, 97, 89, 96, 102, 109, 103, 100, 81, 121, 107, 104, 92, 101, 84, 107, 109, 88, 110]
   (from O_1 to O_{20})
3. Calculate x_0 ($(O_i - E_i)^2 / E_i$)
   x_0 = 20.9

4. Locate the critical value from the table using (sample size-1) and level of significance and compare it with x_0 for the final conclusion
   x_critical = 30.1
   x_critical > x_0 therefore H_0, the null hypothesis, is not rejected. There are no significant difference between the sample distribution and theoretical distribution

Kolmogorv-Smirnov method with a = 0.01

1. Locate the critical value using level of significance and degree of freedom (sample size) and compare it with sample statistics D ($D = max(D^-, D^+)$, D_{alpha} = 0.490)

   d_max = 0.16563575588759877
   d_critical = 0.490
   d_max < d_critical, therefore H_0, the null hypothesis, is not rejected. There are no significant difference between the sample distribution and theoretical distribution

Chi-square test with a = 0.01

1. Locate the critical value from the table using (sample size-1) and level of significance and compare it with x_0 for the final conclusion

   x_critical = 36.2
   x_0 = 20.9
   x_critical > x_0 therefore H_0, the null hypothesis, is not rejected. There are no significant difference between the sample distribution and theoretical distribution

Q2. The output can be re-generated from running CSC446_A3_Q2.ipynb

O_i list = [94, 102, 97, 94, 90, 103, 106, 102, 102, 110]
E_i list = [100, 100, 100, 100, 100, 100, 100, 100, 100, 100]
X_0 = 3.3800000000000003
Critical value = 16.9

H_0, the null hypothesis, is not rejected since x_0 < critical value. There are no significant difference between the sample distribution and theoretical distribution

One potential logical issue with this algorithm is that the first 2 random numbers are obtained from other random number generators and the rest of the numbers are dependent on it. If different random number generation techniques are used, the uniformity test may vary. In addition, we cannot use this for the cryptography testing due to the fact every 3rd random number can be obtained when its previous 2 random numbers are known.

Q3. The output can be re-generated from running CSC446_A3_Q3.ipynb

a)
   Maximum period P is 4 and a = 3 + 8k where k=1
   List of xi = [7, 13, 15, 5, 7]
   List of ri = [0.4375, 0.8125, 0.9375, 0.3125, 0.4375]
   We can see that it starts to cycle after every 4th random number is generated which matches the theoretical maximum period. However, the gap is relatively large.
b) Maximum period P is 4 and a = 3 + 8k where k=1
   However the seed or x_0 is an even number
   List of xi = [8, 8, 8, 8, 8]
   List of ri = [0.5, 0.5, 0.5, 0.5, 0.5]
   Since the seed is an even number, the multiplier and modular have lost their purpose to generate pseudo-random numbers. All random numbers are identical to each other, so maximum period is not achieved. It's not a viable random number generator.
c) Maximum period P is 4 and a = 7 which cannot be obtained from a=3+8k or a=5+8k where k is a set which contains 0 and all positive integers

List of xi = [7, 1, 7, 1, 7]
List of ri = [0.4375, 0.0625, 0.4375, 0.0625, 0.4375]
Although the maximum period is 4, the random numbers start to cycle after every 2 random numbers are generated, which is different from the maximum period. Only 2 unique numbers are generated. Cycling happens very often therefore this is not a viable random number generator.

d) Maximum period P is 4 and a = 7 which cannot be obtained from a=3+8k or a=5+8k where k is a set which contains 0 and all positive integers
In addition, the seed x0 is an even number
List of xi = [8, 8, 8, 8, 8]
List of ri = [0.5, 0.5, 0.5, 0.5, 0.5]
From the result, it is noticeable that all random numbers are the same, which violates the purpose of using a random number generator. Maximum period is not achieved.

Q4. For cdf F(x)

$$\int_{-\infty}^{x} e^{2t}dt = ½ * e^{2x}$$

$$\int_{-\infty}^{0} e^{2t}dt + \int_{0}^{x} e^{-2t}dt = 1 - ½ * (e^{-2x})$$

F(x) =

$½ * e^{2x}$	$-\infty < x <= 0$
$1 - ½ * e^{-2x}$	$0 < x < \infty$

$X_i = F^{-1}(R_i)$

$½ * \ln(2R)$	$0 < R <= 0.5$
$-½ * \ln(2-2R)$	$0.5 < R < 1$

Q5. The output can be re-generated from running CSC446_A3_Q5.ipynb
ρ_{15} = -0.0471803333333333296
σ_{15} = 0.1178511301977579
$Z_0 = \rho_{15} / \sigma_{15}$ = -0.4003384036637002
From the test statistics, we can see that the subsequence has negative correlation. $Z_{0.025}$ = 1.96. Hence, the hypothesis is not rejected.

Q6. Please check the CSC446_A3_Q6.ipynb