

**Title:** Implementation of Kolmogorov –Smirnov Test for Uniformity and Independence Testing of generated Random numbers

**Problem Statement:**

Write a program in java or macros in Excel to conduct frequency test & independence tests. Generate 5 sets of random numbers using a random number generator developed in a previous experiment. Each set consisting of 100 random numbers performs each test of each set of random numbers.

**Expected Outcome of Experiment:**

| **Index** | **Outcome** |
| --- | --- |
| CO3 | Generate pseudorandom numbers and perform statistical tests to measure the quality of a pseudorandom number generator. |

**Books/ Journals/ Websites referred:**

1. Jerry Banks, John Carson, Barry Nelson, and David M. Nichol, “Discrete Event System

Simulation”; Fifth Edition, Prentice-Hall.

2. Averill M Law, “System Modeling Analysis”; 4th Edition TMH.

3. Banks C M, Sokolowski J A, “Principles of Modelling and Simulation”, Wiley

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Pre Lab/ Prior Concepts:**

Meaning of Uniformity and Independence Tests for Random Numbers:

**Frequency Tests** A basic test that should always be performed to validate a new generator is the test of uniformity. Two different methods of testing are available. They are the **Kolmogorov-Smirnov** and the **chi-square test**. Both of these tests measure the degree of agreement between the distribution of a sample of generated random numbers and the theoretical uniform distribution. Both tests are on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution*.*

**The Kolmogorov-Smirnov test**: This test compares the continuous cdf, F(X), of the uniform distribution to the empirical cdf, SN(x), of the sample of N observations. By definition,

F(x) = x, 0 <= x <= 1

If the sample from the random-number generator is R1 R2, ,• • •, RN,

Then the empirical cdf, SN(X), is defined by

SN(X) = number of R1 R2, • • •, Rn which are <= x / N

As N becomes larger, SN(X) should become a better approximation to F(X) , provided that the null hypothesis is true.

The **Kolmogorov-Smirnov test** is based on the largest absolute deviation between F(x) and SN(X) over the range of the random variable. That is.it is based on the statistic

D = max | F(x) - SN(x)|

For testing against a uniform cdf, the test procedure follows these steps:

**Step 1.** Rank the data from smallest to largest.

Let R (i) denote the i th smallest observation, so that R (1) <= R (2) <= • • • <= R (N)

**Step 2**. Compute D+ = max **{i/N -** R (i) **}**

1<= **i** <=N

& D- = max {R (i) – (i-1)/N}

**Step3.** Compute D = max (D+, D-).

**Step 4.** Determine the critical value, Da, from Table A.8 for the specified significance level a and the given sample size N.

**Step 5.** If the sample statistic D is greater than the critical value, Da, the null hypothesis

That the data are a sample from a uniform distribution is rejected.

**Implementation details & Example**:

***Code:***

from random import random

value = int(input('Enter the number of random numbers: '))

arr = [random() for \_ in range(int(value))]

arr.sort()

print('Numbers: ',arr)

print()

D1 = max([max(i/len(arr)-arr[i-1],0) for i in range(1,len(arr)+1)])

print('Expected values:\n', [i/len(arr) for i in range(1,len(arr)+1)])

print()

print('D+: ',[max(i/len(arr)-arr[i-1],0) for i in range(1,len(arr)+1)])

print()

D2 = max([max(arr[i-1]-(i-1)/len(arr),0) for i in range(1,len(arr)+1)])

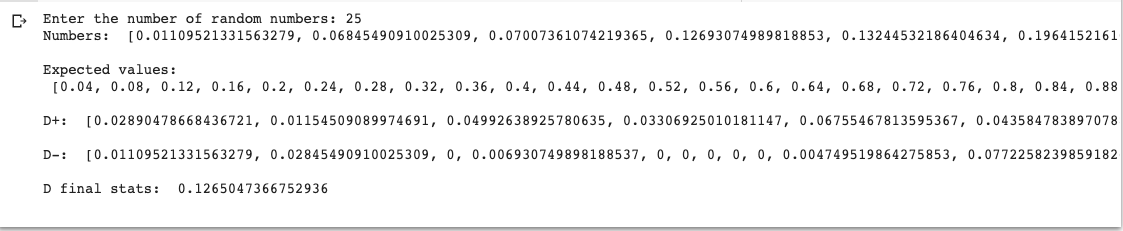
print('D-: ',[max(arr[i-1]-(i-1)/len(arr),0) for i in range(1,len(arr)+1)])

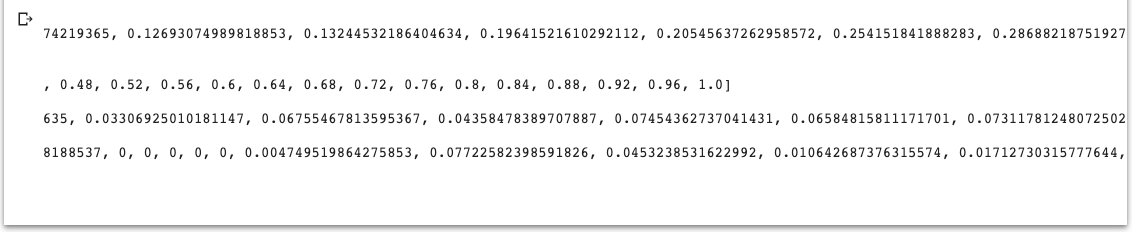
print()

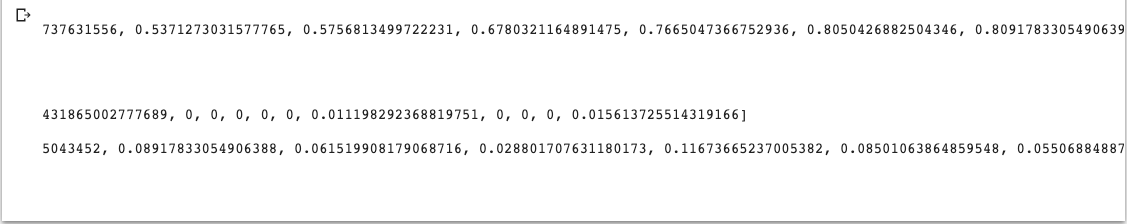
D = max(D1,D2)

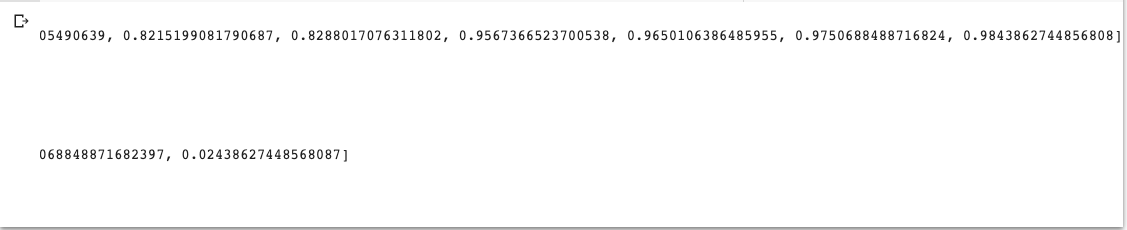
print('D final stats: ',D)

***Output:***









**Conclusion:**

We successfully understood and performed the Implementation of Kolmogorov –Smirnov Test for Uniformity and Independence Testing of generated Random numbers in python.

**Post lab questions:**

1. Explain Gap & Poker test with the help of example.

* The poker test for independence is based on the frequency in which certain digits are repeated in a series of numbers.
* For example 0.255, 0.577, 0.331, 0.414, 0.828, 0.909, 0.303, 0.001... In each case, a pair of like digits appears in the number.
* In a three digit number, there are only three possibilities.
  + The individual digits can be all different. Case 1.
  + The individual digits can all be the same. Case 2.
  + There can be one pair of like digits. Case 3.
* P(case 1) = P(second differ from the first) \* P(third differ from the first and second) = 0.9 \* 0.8 = 0.72  
  P(case 2) = P(second the same as the first) \* P(third same as the first) = 0.1 \* 0.1 = 0.01 P(case 3) = 1 - 0.72 - 0.01 = 0.27

**Gap Test**

* The gap test is used to determine the significance of the interval between recurrence of the same digit.
* A gap of length *x* occurs between the recurrence of some digit.
* See the example on page 313 where the digit 3 is underlined. There are a total of eighteen 3's in the list. Thus only 17 gaps can occur.
* The probability of a particular gap length can be determined by a Bernoulli trail.  
  \begin{displaymath}P ({\rm gap ~of} ~ n) = P(x \ne 3) P(x \ne 3) ... P(x \ne 3) P (x = 3) \end{displaymath}  
  If we are only concerned with digits between 0 and 9, then  
  \begin{displaymath}P ({\rm gap ~of} ~ n) = 0.9^{n} 0.1 \end{displaymath}
* The theoretical frequency distribution for randomly ordered digits is given by  
    
  \begin{displaymath}P (gap \le x) = F(x) = 0.1 \sum_{n=0}^x (0.9)^n = 1 - 0.9^{x+1} \end{displaymath}
* Steps involved in the test.  
  **Step 1.**Specify the cdf for the theoretical frequency distribution given by Equation (8.14) based on the selected class interval width (See Table 8.6 for an example).  
  **Step 2.**Arrange the observed sample of gaps in a cumulative distribution with these same classes.  
  **Step 3.**Find *D*, the maximum deviation between *F(x)* and $S_N(x)$ as in Equation 8.3 (on page 299).  
  **Step 4.**
* Determine the critical value, $D_\alpha$, from Table A.8 for the specified value of $\alpha$ and the sample size $N$.  
  **Step 5.**If the calculated value of *D* is greater than the tabulated value of $D_\alpha$, the null hypothesis of independence is rejected.