Assignment 6

Miguel A. Gomez B.

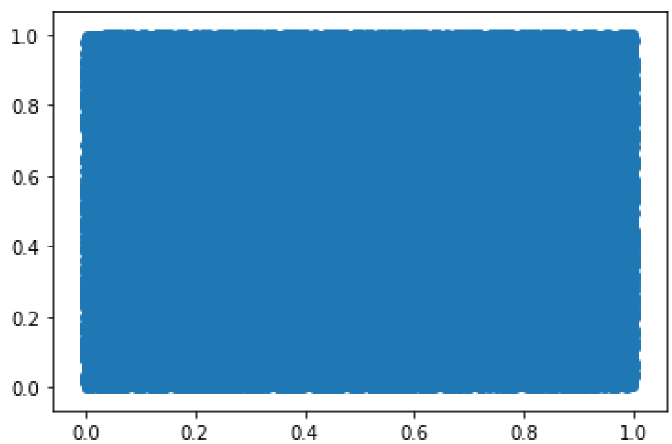
March 14, 2020

1 Testing a Sequence of Numbers.

(50 Points) Together with this assignment you will find a sequence of 100.000 numbers (*'Sequence.txt'*) generated with the brand new *PPP* algorithm. Your task will be to test if the algorithm passes the following tests:

- Spectral tests. (In order to perform this test assume that the algorithm used is *MRG32k3a* by L'Ecuyer).
- χ^2 Test.
- Kolmogorov-Smirnov Test.

Solution Before we proceed with the tests we want to get an idea of how the data is distributed.

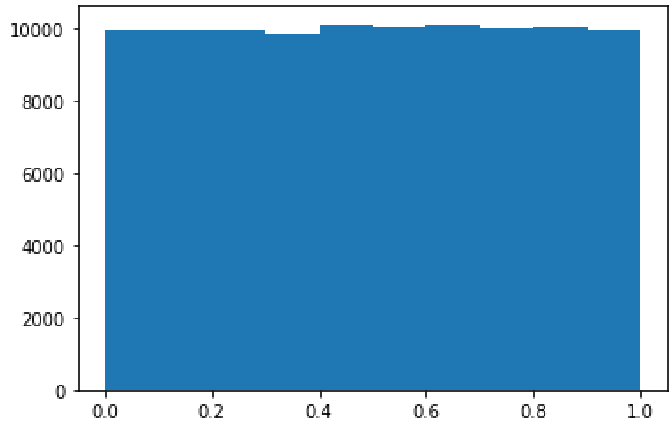


And descriptive statistics gives us an idea about dispersion and uniformity,

Statistic	**Value**
Mean	0.5
Median	0.501
Mode	0.565
Standard Deviation	0.288
Max	0.999
Min	4.243e-05

Table 1: Descriptive Statistics

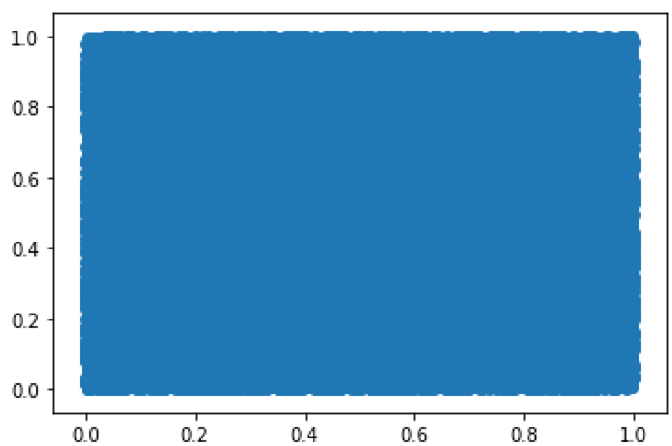
As we can see in the graph it seems uniform, the mean and the median, confirms that, but the standard deviation tells us a different story that it is not visible from the graph, if the data is between 0 and 1 with an average difference of 0.288, it seems suspicious about how the data is distributed, we will a histogram.



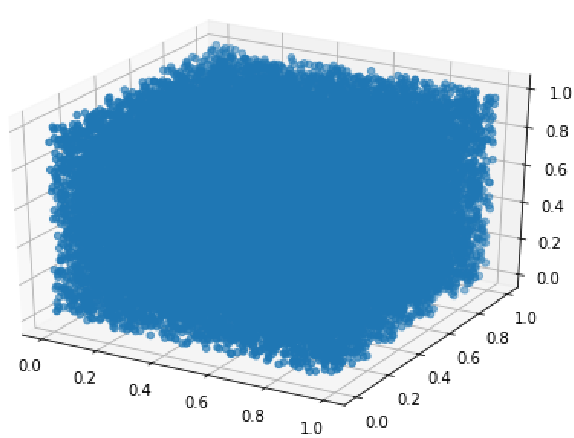
Now we see that there is uniformity between the data,but also we can see differences between intervals, in any case, we see that the data is closely related to a uniform distribution, the tests will determine numerically how uniform these generated values and as a consequence the algorithm that generate them.

1.1 Spectral Test (Not finished).

1.1.1 2D



1.2 3D



1.3 χ^2 Test.

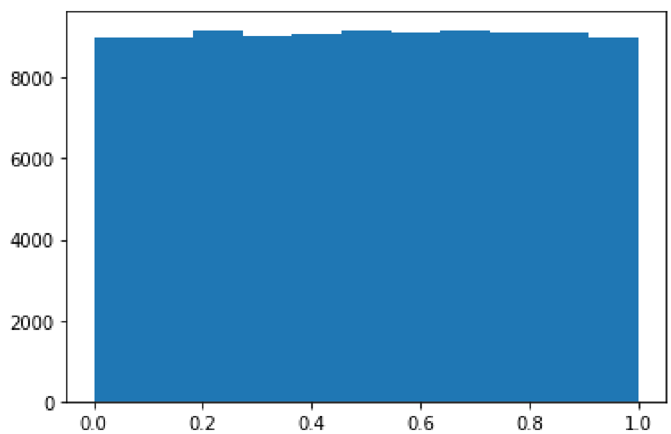
On the χ^2 test, we compute χ^2 with the following equation

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \tag{1}$$

where n is the number of intervals, O_i is the observed value and E_i is the expected value. Then we lookup on the χ^2 table for an $\alpha = 0.05$(generally) and an $n-1$ degrees of freedom. And formulate a hypothesis(*Null Hypothesis*): There is no difference between our computed distribution (the generated numbers) and the theoretical distribution. If the sum is less than $\chi^2_{[1-\alpha;n-1]}$, then the hypothesis previous hypothesis cannot be rejected at the level of significance α:

$$\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} < \chi^2_{[1-\alpha;n-1]}$$

Using the information computed from the following histogram



we choose our $n = 11$ and to formulate an expected value for every bin, we state that the expected value is $\frac{100000/11}{\approx}9090.90$, because we are assuming that the values are related to a uniform distribution.

The following results resume our computation for χ^2

Interval	E_i	O_i	$\frac{(O_i-E_i)^2}{E_i}$
1	9090.90	9007	0.77448091
2	9090.90	8997	0.97008091
3	9090.90	9157	0.48048091
4	9090.90	9041	0.27400091
5	9090.90	9066	0.06825091
6	9090.90	9163	0.57168091
7	9090.90	9126	0.13545091
8	9090.90	9181	0.89280091
9	9090.90	9113	0.05368091
10	9090.90	9143	0.29848091
11	9090.90	9006	0.79305091
χ^2			5.312

Table 2: Computation of χ^2

Now we compare our χ^2 compared with the value on the table, for a $\nu = 11 - 1 = 10$, and $p = 1 - 0.05$ is 18.31, as we can see $5.312 < 18.31$ so the hypothesis cannot be rejected at least with a level of significance of 0.05.

	$p = .01$	$p = .05$	$p = .25$	$p = .50$	$p = .75$	$p = .95$	$p = .99$
$\nu = 1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$\nu = 2$	0.02010	0.1026	0.5753	1.386	2.773	5.991	9.210
$\nu = 3$	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$\nu = 4$	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
$\nu = 5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$\nu = 6$	0.8720	1.635	3.455	5.348	7.841	12.59	16.81
$\nu = 7$	1.239	2.167	4.255	6.346	9.037	14.07	18.48
$\nu = 8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09
$\nu = 9$	2.088	3.325	5.899	8.343	11.39	16.92	21.67
$\nu = 10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$\nu = 11$	3.053	4.575	7.584	10.34	13.70	19.68	24.73
$\nu = 12$	3.571	5.226	8.438	11.34	14.84	21.03	26.22
$\nu = 15$	5.229	7.261	11.04	14.34	18.25	25.00	30.58
$\nu = 20$	8.260	10.85	15.45	19.34	23.83	31.41	37.57
$\nu = 30$	14.95	18.49	24.48	29.34	34.80	43.77	50.89
$\nu = 50$	29.71	34.76	42.94	49.33	56.33	67.50	76.15

As we can see, our computed value lies between $p = .05$ and

1.4 Kolmogorov-Smirnov Test.

On this test we will compare the PDF (Probability Density Function), $F(X)$, against an empirical PDF uniform distribution, $S_n(X)$ from a sample of N observations, then, by definition,

$$F(x) = x, 0 \leq x \leq 1$$

$$S_n(x) = \frac{R_1, R_2, R_3, \ldots, R_n \leq x}{N} \tag{2}$$

where R_n is an observation. The test also requires us to sort the observations and compute D^+ and D^-,

$$D^+ = \max_{1 \leq i \leq N} \left[\frac{i}{n} - R_i \right], D^- = \max_{1 \leq i \leq N} \left[R_i - \frac{i-1}{n} \right],$$

Then we compute D as

$$D = \max(D^+, D^-),$$

and same as before we find our correspondent value on the KS table, with the expected level of significance(as before we will use $\alpha = 0.05$), if our value is less that the value founded, we accept the null hypothesis(The data may be was generated from an uniform distribution).

Same as before we will use the bin values obtained from the previous histogram. And therefore we compute D^+ and D^- with $F(x_i)$ defined as:

$$F(x_i) = \frac{1}{b-a}, a \leq x \leq b,$$

with $a = 0$ and $b = 1$. Then $F(x) = x$. Now sorting the values, the following table resumes our computations:

i	$F(x_i)$	i/n	$i/n - F(x_i)$	$F(x_i) - (i-1)/n$
1	4.24310000e-05	0.08333333	8.32909023e-02	4.24310000e-05
2	9.09467555e-02	0.16666667	7.57199112e-02	7.61342212e-03
3	1.81851080e-01	0.25	6.81489201e-02	1.51844132e-02
4	2.72755404e-01	0.33333333	6.05779290e-02	2.27554044e-02
5	3.63659729e-01	0.41666667	5.30069378e-02	3.03263955e-02
6	4.54564053e-01	0.5	4.54359467e-02	3.78973866e-02
7	5.45468378e-01	0.58333333	3.78649556e-02	4.54683777e-02
8	6.36372702e-01	0.66666667	3.02939645e-02	5.30393688e-02
9	7.27277027e-01	0.75	2.27229734e-02	6.06103600e-02
10	8.18181351e-01	0.83333333	1.51519822e-02	6.81813511e-02
11	9.09085676e-01	0.91666667	7.58099112e-03	7.57523422e-02
12	9.99990000e-01	1.	1.00000000e-05	8.33233333e-02
			$D^+ = 0.08329090233333333$	$D^- = 0.08332333333333342$
$D = \max(D^+, D^-) = 0.08332333333333342$				

Table 3: KS Test computation

Now we compare our result against the KS Table:

	$p = .01$	$p = .05$	$p = .25$	$p = .50$	$p = .75$	$p = .95$	$p = .99$
$n = 1$	0.01000	0.05000	0.2500	0.5000	0.7500	0.9500	0.9900
$n = 2$	0.01400	0.06749	0.2929	0.5176	0.7071	1.0980	1.2728
$n = 3$	0.01699	0.07919	0.3112	0.5147	0.7539	1.1017	1.3589
$n = 4$	0.01943	0.08789	0.3202	0.5110	0.7642	1.1304	1.3777
$n = 5$	0.02152	0.09471	0.3249	0.5245	0.7674	1.1392	1.4024
$n = 6$	0.02336	0.1002	0.3272	0.5319	0.7703	1.1463	1.4144
$n = 7$	0.02501	0.1048	0.3280	0.5364	0.7755	1.1537	1.4246
$n = 8$	0.02650	0.1086	0.3280	0.5392	0.7797	1.1586	1.4327
$n = 9$	0.02786	0.1119	0.3274	0.5411	0.7825	1.1624	1.4388
$n = 10$	0.02912	0.1147	0.3297	0.5426	0.7845	1.1658	1.4440
$n = 11$	0.03028	0.1172	0.3330	0.5439	0.7863	1.1688	1.4484
$n = 12$	0.03137	0.1193	0.3357	0.5453	0.7880	1.1714	1.4521
$n = 15$	0.03424	0.1244	0.3412	0.5500	0.7926	1.1773	1.4606
$n = 20$	0.03807	0.1298	0.3461	0.5547	0.7975	1.1839	1.4698
$n = 30$	0.04354	0.1351	0.3509	0.5605	0.8036	1.1916	1.4801

same as before the test is passed if $D < K_{[1-\alpha=0.95;n=12]}$, some authors like Knuth, even compare D^+ and D^-, and as we see:

$$0.083 < 1.17,$$

Which satisfies both conditions and the test is passed.