Chapter 4:
(0,1) Random Number Generation

Refer to Text Book:

- Refer to Reading Files

Review Last Lecture

- **Idea of Random Number Generators**

  Why we need Random Generators in simulation

- **Pseudo-Random Numbers**
  - Linear congruential generator (LCG)
  - Definitions: Seed, Period, Cycle , Streems
  - Conditions for LCG Full Cycle
  - Numerical Examples
- **Random Streams**
  - What is random streams?
  - How to get multiple random streams?
  - Why do need to change random streams?

Today's Lecture Plan

- **Testing outputs of Random Number Generators**

  Why do we the test?

- **Statistical Tests of (0,1) Random Numbers**
- **P-Value**
- **Chi-Squared Goodness-of-Fit Test**
  - Main idea of the test
  - Hypotheses of the test
  - The decisions of the test
  - Numerical Example
- **Kolmogorov–Smirnov Test (K-S test)**
  - Main idea of the test
  - Hypotheses of the test
  - The decisions of the test
  - Numerical Example

Testing (0,1) Random Numbers

- Any random number generator must produce sequences of numbers that appear to be independent and identically distributed (IID) $U(0, 1)$ random variables

- The hypothesis that a sample from the generator is IID $U(0, 1)$ must be made and the hypothesis must be statistically tested.

Testing (0,1) Random Numbers

Distributional Tests

Given that we have a sequence of numbers, null hypothesis is that the generated numbers are uniformly distributed on the interval 0 to 1 and the alternative hypothesis that the generated numbers are not uniformly distributed on the interval 0 to 1

$$H_0: U_i \sim U(0,1) \text{ versus } H_1: U_i \text{ not } U(0,1)$$

Testing (0,1) Random Numbers

Distributional Tests

$$\mathbf{H_0: U_i \sim U(0,1) \text{ versus } H_1: U_i \text{ not } U(0,1)}$$

We can do the test by one of two ways

- pre-specifying the Type 1 error (the confidence level α) and comparing to a critical value (Value to Value comparison)
- pre-specifying the Type 1 error (the confidence level α) and comparing to a p-value (Probability to Probability comparison)

$\boldsymbol{\alpha} = P(\text{Type 1 error}) = P(\text{rejecting the } \mathbf{H_0} \text{ when it is true})$

Testing (0,1) Random Numbers

***p*-value:**

- The *p*-value for a statistical test is the smallest α level at which the observed test statistic is significant.
- The smaller the *p*-value, the more the result can be considered statistically significant. Thus, the *p*-value can be compared to the desired significance level, α.
- The testing criterion is
  - If the *p*-value $> \alpha$, then do not reject H_0
  - If the *p*-value $\leq \alpha$, then reject H_0.

Chi-Squared Goodness-of-Fit Test

Main idea

If observation are taken from U[0,1] then on average number of data in each interval is equal.

- The chi-square test divides the range of the data into, *k*, intervals.
- Count the number in each interval
- the number of observations that fall in each interval is close the expected number

Chi-Squared Goodness-of-Fit Test

Test Steps:

1. Divide the interval (0, 1) into k equally spaced classes so that $\Delta b = b_j - b_{j-1}$ resulting in $p_j = \frac{1}{k}$ for $j = 1, 2, \ldots, k$. This results in the expected number in each interval being $np_j = n \times \frac{1}{k} = \frac{n}{k}$.
2. As a practical rule, the expected number in each interval np_j should be at least 5. Thus, in this case, $\frac{n}{k} \geq 5$ or $n \geq 5k$ or $k \leq \frac{n}{5}$. Thus, for a given value of n, you should choose the number of intervals $k \leq \frac{n}{5}$.
3. Since the parameters of the distribution are known $a = 0$ and $b = 1$, then $s = 0$. Therefore, we reject $H_0 : U_i \sim U(0, 1)$ if $\chi_0^2 > \chi_{\alpha,k-1}^2$ or if the p-value is less than α.

$$\chi_0^2 = \frac{k}{n} \sum_{j=1}^{k} \left(c_j - \frac{n}{k} \right)^2$$

we reject $H_0 : U_i \sim U(0, 1)$ if $\chi_0^2 > \chi_{\alpha,k-1}^2$ or if the p-value is less than α

Chi-Squared Goodness-of-Fit Test

Example:

Suppose we have 100 observations from a pseudorandom number generator. Perform a chi-squared test that the numbers appear $U(0, 1)$.

0.971	0.668	0.742	0.171	0.350	0.931	0.803	0.848	0.160	0.085
0.687	0.799	0.530	0.933	0.105	0.783	0.828	0.177	0.535	0.601
0.314	0.345	0.034	0.472	0.607	0.501	0.818	0.506	0.407	0.675
0.752	0.771	0.006	0.749	0.116	0.849	0.016	0.605	0.920	0.856
0.830	0.746	0.531	0.686	0.254	0.139	0.911	0.493	0.684	0.938
0.040	0.798	0.845	0.461	0.385	0.099	0.724	0.636	0.846	0.897
0.468	0.339	0.079	0.902	0.866	0.054	0.265	0.586	0.638	0.869
0.951	0.842	0.241	0.251	0.548	0.952	0.017	0.544	0.316	0.710
0.074	0.730	0.285	0.940	0.214	0.679	0.087	0.700	0.332	0.610
0.061	0.164	0.775	0.015	0.224	0.474	0.521	0.777	0.764	0.144

Chi-Squared Goodness-of-Fit Test

Example:

Since we have $n = 100$, the number of intervals should be less than or equal to 20. Let us choose $k = 10$. This means that $p_j = 0.1$ for all j.

Start Interval

End Interval

Diff. between #End and #Start

j	b_{j-1}	b_j	p_j	c_j	np_j	$\frac{(c_j - np_j)^2}{np_j}$
1	0	0.1	0.1	13	10	0.9
2	0.1	0.2	0.1	8	10	0.4
3	0.2	0.3	0.1	7	10	0.9
4	0.3	0.4	0.1	7	10	0.9
5	0.4	0.5	0.1	6	10	1.6
6	0.5	0.6	0.1	9	10	0.1
7	0.6	0.7	0.1	13	10	0.9
8	0.7	0.8	0.1	14	10	1.6
9	0.8	0.9	0.1	13	10	0.9
10	0.9	1	0.1	10	10	0

Chi-Squared Goodness-of-Fit Test

Example:

Summing the last column yields:

$$\chi_0^2 = \sum_{j=1}^{k} \frac{(c_j - np_j)^2}{np_j} = 8.2$$

Computing the p-value for $k - s - 1 = 10 - 0 - 1 = 9$ degrees of freedom, yields $\mathrm{P}\{\chi_9^2 > 8.2\} = 0.514$. Thus, given such a high p-value, we would not reject the hypothesis that the observed data is $U(0,1)$.

Chi-Squared Goodness-of-Fit Test

Exercise :

Given the following data, test the hypothesis that the data appears U(0, 1) versus that it is not U(0, 1) using the Chi-Square test at the $\alpha = 0.05$ significance level.

Use $k = 6$ intervals

n	U_n
1	0.651
2	0.222
3	0.836
4	0.991
5	0.210
6	0.308
7	0.400
8	0.031
9	0.552
10	0.049
11	0.268
12	0.660
13	0.849
14	0.438
15	0.883
16	0.833
17	0.417
18	0.156
19	0.897
20	0.597

Kolmogorov–Smirnov Test (K-S test)

Main idea

If observation are taken from U[0,1] then the graph of the empirical distribution of the observation follows the CDF of U[0,1] distribution.

(K-S) Test ***does not depend on specifying intervals*** for tabulating the test statistic.

Kolmogorov–Smirnov Test (K-S test)

Steps

Suppose we have a sample of data x_i for $i = 1, 2, \dots, n$

- Sort the data to obtain $x_{(i)}$ for $i = 1, 2, \dots, n$, with $x_{(1)}$ is the smallest, $x_{(2)}$ is the 2^{nd} smallest,… , $x_{(n)}$ is the largest value.
- Empirical distribution function is the proportion of the data values that are less than or equal to the i^{th} order statistic $x_{(i)}$ for each $i = 1, 2, \dots, n$,

Kolmogorov–Smirnov Test (K-S test)

Steps

- From $x_{(i)}$, $i = 1, 2, \ldots, n$, the empirical distribution gives :

$$\Pr\{X \leq x_{(1)}\} = \frac{1}{n}$$

$$\Pr\{X \leq x_{(2)}\} = \frac{2}{n}$$

$$\Pr\{X \leq x_{(3)}\} = \frac{3}{n}$$

$$\ldots$$

$$\Pr\{X \leq x_{(n)}\} = \frac{n}{n} = 1$$

Kolmogorov–Smirnov Test (K-S test)

Steps

Finally, Compute the values:

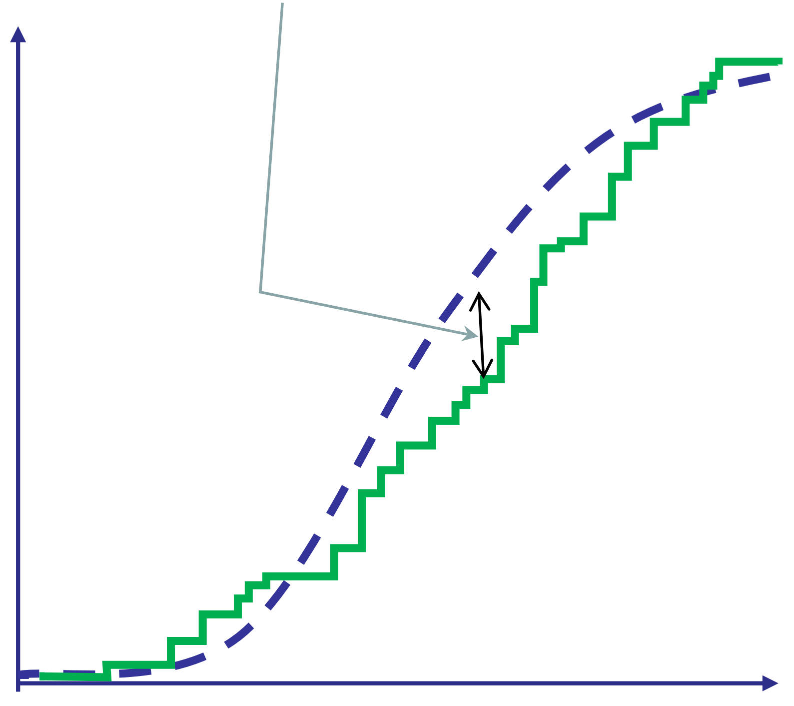
D_n : is the maximum vertical distance between the empirical curve and the theoretical curve

$$D_n^+ = \max_{1\le i\le n}\left\{\frac{i}{n} - \hat{F}(x_{(i)})\right\}$$

$$= \max_{1\le i\le n}\left\{\frac{i}{n} - x_{(i)}\right\}$$

$$D_n^- = \max_{1\le i\le n}\left\{\hat{F}(x_{(i)}) - \frac{i-1}{n}\right\}$$

$$= \max_{1\le i\le n}\left\{x_{(i)} - \frac{i-1}{n}\right\}$$



$$\boldsymbol{D_0 = \max\{\, D^- ,\, D^+ \,\}}$$

Kolmogorov–Smirnov Test (K-S test)

Test Decision

- Reject the null hypothesis (H_0 : x ~ U[0,1]) if D_n is greater than the critical value D_α , where α is the significance level.

- The K-S test statistic, D_n, represents the largest vertical distance between the suggested distribution and the empirical distribution

Kolmogorov–Smirnov Test (K-S test)

TABLE B 7 Kolmogorov–Smirnov Test Critical Values

n	$D_{0.1}$	$D_{0.05}$	$D_{0.01}$
10	0.36866	0.40925	0.48893
11	0.35242	0.39122	0.46770
12	0.33815	0.37543	0.44905
13	0.32549	0.36143	0.43247
14	0.31417	0.34890	0.41762
15	0.30397	0.33760	0.40420
16	0.29472	0.32733	0.39201
17	0.28627	0.31796	0.38086
18	0.27851	0.30936	0.37062
19	0.27136	0.30143	0.36117
20	0.26473	0.29408	0.35241
25	0.23768	0.26404	0.31657
30	0.21756	0.24170	0.28987
35	0.20185	0.22425	0.26897
Over 35	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

Kolmogorov–Smirnov Test (K-S test)

Example:

Consider the same data

Suppose we have 100 observations from a pseudorandom number generator. Perform a chi-squared test that the numbers appear $U(0, 1)$.

0.971	0.668	0.742	0.171	0.350	0.931	0.803	0.848	0.160	0.085
0.687	0.799	0.530	0.933	0.105	0.783	0.828	0.177	0.535	0.601
0.314	0.345	0.034	0.472	0.607	0.501	0.818	0.506	0.407	0.675
0.752	0.771	0.006	0.749	0.116	0.849	0.016	0.605	0.920	0.856
0.830	0.746	0.531	0.686	0.254	0.139	0.911	0.493	0.684	0.938
0.040	0.798	0.845	0.461	0.385	0.099	0.724	0.636	0.846	0.897
0.468	0.339	0.079	0.902	0.866	0.054	0.265	0.586	0.638	0.869
0.951	0.842	0.241	0.251	0.548	0.952	0.017	0.544	0.316	0.710
0.074	0.730	0.285	0.940	0.214	0.679	0.087	0.700	0.332	0.610
0.061	0.164	0.775	0.015	0.224	0.474	0.521	0.777	0.764	0.144

Kolmogorov–Smirnov Test (K-S test)

Example:

Compute the following table

i	$x_{(i)}$	i/n	$\frac{i-1}{n}$	$F(x_{(i)})$	$\frac{i}{n} - F(x_{(i)})$	$F(x_{(i)}) - \frac{i-1}{n}$
1	0.006	0.010	0.000	0.006	0.004	0.006
2	0.015	0.020	0.010	0.015	0.005	0.005
3	0.016	0.030	0.020	0.016	0.014	−0.004
4	0.017	0.040	0.030	0.017	0.023	−0.013
5	0.034	0.050	0.040	0.034	0.016	−0.006
⋮	⋮	⋮	⋮	⋮	⋮	⋮
95	0.933	0.950	0.940	0.933	0.017	−0.007
96	0.938	0.960	0.950	0.938	0.022	−0.012
97	0.940	0.970	0.960	0.940	0.030	−0.020
98	0.951	0.980	0.970	0.951	0.029	−0.019
99	0.952	0.990	0.980	0.952	0.038	−0.028
100	0.971	1.000	0.990	0.971	0.029	−0.019

Kolmogorov–Smirnov Test (K-S test)

Example:

Computing D_n^+ and D_n^- yields

$$D_n^+ = \max_{1\le i\le n}\left\{\frac{i}{n} - x_{(i)}\right\} = 0.108$$

$$D_n^- = \max_{1\le i\le n}\left\{x_{(i)} - \frac{i-1}{n}\right\} = 0.038$$

Thus, we have that

$$D_n = \max\{D_n^+, D_n^-\} = \max\{0.108, 0.038\} = 0.108$$

Referring to Appendix B and using the approximation for sample sizes greater than 35, we have that $D_{0.05} \approx 1.35/\sqrt{n}$. Thus, $D_{0.05} \approx 1.35/\sqrt{100} = 0.135$. Since $D_n < D_{0.05}$, we would not reject the hypothesis that the data is uniformly distributed over the range from 0 to 1.

Chi-Squared Goodness-of-Fit Test

Exercise:

Given the following data, test the hypothesis that the data appears U(0, 1) versus that it is not U(0, 1) using the Kolmogorov–Smirnov Test (K-S) test at the $\alpha = 0.05$ significance level.

n	U_n
1	0.651
2	0.222
3	0.836
4	0.991
5	0.210
6	0.308
7	0.400
8	0.031
9	0.552
10	0.049
11	0.268
12	0.660
13	0.849
14	0.438
15	0.883
16	0.833
17	0.417
18	0.156
19	0.897
20	0.597