Pseudo-Random Numbers

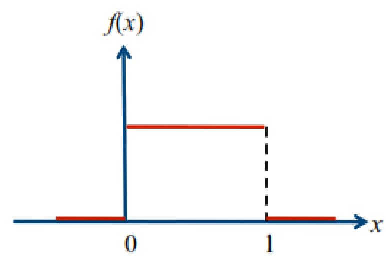
- Important properties of good random number routines:
  - Fast
  - Portable to different computers
  - Have sufficiently long cycle
  - Replicable
    - Verification and debugging
    - Use identical stream of random numbers for different systems
  - Closely approximate the ideal statistical properties of
    - uniformity and
    - independence

Pseudo-Random Numbers: Properties

- Two important statistical properties:
  - Uniformity
  - Independence
- Random number R_i must be independently drawn from a uniform distribution with PDF:

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$E(R) = \int_0^1 x\,dx = \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2}$$



PDF for random numbers

Pseudo-Random Numbers

- Problems when generating pseudo-random numbers
  - The generated numbers might not be uniformly distributed
  - The generated numbers might be discrete-valued instead of continuous-valued
  - The mean of the generated numbers might be too high or too low
  - The variance of the generated numbers might be too high or too low

- There might be dependence:
  - Autocorrelation between numbers
  - Numbers successively higher or lower than adjacent numbers
  - Several numbers above the mean followed by several numbers below the mean

Generating Random Numbers

- Linear Congruential Method (LCM)
- Combined Linear Congruential Generators (CLCG)

Linear Congruential Method

- To produce a sequence of integers $X_1, X_2, \ldots$ between 0 and m-1 by following a recursive relationship:

$$X_{i+1} = (\underbrace{a}_{\text{The multiplier}} X_i + \underbrace{c}_{\text{The increment}}) \bmod \underbrace{m}_{\text{The modulus}}, \quad i = 0,1,2,\ldots$$

- Assumption: $m > 0$ and $a < m$, $c < m$, $X_0 < m$
- The selection of the values for a, c, m, and X_0 drastically affects the statistical properties and the cycle length
- The random integers X_i are being generated in $[0, m\text{-}1]$

Linear Congruential Method

- Convert the integers X_i to random numbers

$$R_i = \frac{X_i}{m}, \quad i = 1,2,\ldots$$

- Note:
  - $X_i \in \{0, 1, ..., m\text{-}1\}$
  - $R_i \in [0, (m\text{-}1)/m]$

Linear Congruential Method: Example

- Use $X_0 = 27$, $a = 17$, $c = 43$, and $m = 100$.
- The X_i and R_i values are:

$$X_1 = (17 \times 27 + 43) \bmod 100 = 502 \bmod 100 = 2 \quad \Rightarrow \quad R_1 = 0.02$$
$$X_2 = (17 \times 2 + 43) \bmod 100 = 77 \quad \Rightarrow \quad R_2 = 0.77$$
$$X_3 = (17 \times 77 + 43) \bmod 100 = 52 \quad \Rightarrow \quad R_3 = 0.52$$
$$X_4 = (17 \times 52 + 43) \bmod 100 = 27 \quad \Rightarrow \quad R_3 = 0.27$$
$$\ldots$$

Linear Congruential Method: Example

- Use $a = 13$, $c = 0$, and $m = 64$
- The period of the generator is very low
- Seed X_0 influences the sequence

i	X_i $X_0=1$	X_i $X_0=2$	X_i $X_0=3$	X_i $X_0=4$
0	1	2	3	4
1	13	26	39	52
2	41	18	59	36
3	21	42	63	20
4	17	34	51	4
5	29	58	23	
6	57	50	43	
7	37	10	47	
8	33	2	35	
9	45		7	
10	9		27	
11	53		31	
12	49		19	
13	61		55	
14	25		11	
15	5		15	
16	1		3	

Linear Congruential Method

Properties to Consider

- Generated numbers must be approximately uniform and independent.
- Moreover, other properties, such as *maximum density* and *maximum period* must be considered.
- By maximum density is meant that the values assumed by $R_i, i = 1, 2, \ldots$, leave no large gaps on $[0, 1]$.
- In many simulation languages, values such as $m = 2^{31} - 1$ and $m = 2^{48}$ are in common use in generators.
- To help achieve maximum density and to avoid cycling, the generator should have the largest possible period.

Linear Congruential Method:

Characteristics of a good Generator

- The LCG has full period if and only if the following three conditions hold (Hull and Dobell, 1962):
  1. The only positive integer that (exactly) divides both m and c is 1
  2. If q is a prime number that divides m, then q divides a-1
  3. If 4 divides m, then 4 divides a-1

Combined Linear Congruential Generators

- Reason: Longer period generator is needed because of the increasing complexity of simulated systems.
- Approach: Combine two or more multiplicative congruential generators.

- Let $X_{i,1}$, $X_{i,2}$, ..., $X_{i,k}$ be the i-th output from k different multiplicative congruential generators.
  - The j-th generator $X_{\bullet,j}$:

$$X_{i+1,j} = (a_j X_i + c_j) \bmod m_j$$

  - has prime modulus m_j, multiplier a_j, and period m_j-1
  - produces integers $X_{i,j}$ approx ~ Uniform on $[0, m_j - 1]$
  - $W_{i,j} = X_{i,j}$ - 1 is approx ~ Uniform on integers on $[0, m_j - 2]$

Combined Linear Congruential Generators

- Suggested form:

$$X_i = \left(\sum_{j=1}^{k} (-1)^{j-1} X_{i,j} \right) \bmod m_1 - 1 \qquad \text{Hence, } R_i = \begin{cases} \dfrac{X_i}{m_1}, & X_i > 0 \\ \dfrac{m_1 - 1}{m_1}, & X_i = 0 \end{cases}$$

- The maximum possible period is: $P = \dfrac{(m_1 - 1)(m_2 - 1)...(m_k - 1)}{2^{k-1}}$

Combined Linear Congruential Generators

- Example: For 32-bit computers, combining $k = 2$ generators with $m_1 = 2147483563$, $a_1 = 40014$, $m_2 = 2147483399$ and $a_2 = 40692$.
  The algorithm becomes:
  Step 1: Select seeds
  $X_{0,1}$ in the range $[1, 2147483562]$ for the 1st generator
  $X_{0,2}$ in the range $[1, 2147483398]$ for the 2nd generator
  Step 2: For each individual generator,
  $X_{i+1,1} = 40014 \times X_{i,1} \bmod 2147483563$
  $X_{i+1,2} = 40692 \times X_{i,2} \bmod 2147483399$
  Step 3: $X_{i+1} = (X_{i+1,1} - X_{i+1,2}) \bmod 2147483562$
  Step 4: Return

$$R_{i+1} = \begin{cases} \dfrac{X_{i+1}}{2147483563}, & X_{i+1} > 0 \\ \dfrac{2147483562}{2147483563}, & X_{i+1} = 0 \end{cases}$$

  Step 5: Set $i = i+1$, go back to step 2.
  - Combined generator has period: $(m_1 - 1)(m_2 - 1)/2 \sim 2 \text{ x } 10^{18}$

Tests for Random Numbers

- Two categories:
  - Testing for **uniformity**:

    H_0: $R_i \sim U[0,1]$

    H_1: $R_i \nsim U[0,1]$
    - Failure to reject the null hypothesis, H_0, means that evidence of non-uniformity has not been detected.
  - Testing for **independence**:

    H_0: $R_i \sim$ independent

    H_1: $R_i \nsim$ independent
    - Failure to reject the null hypothesis, H_0, means that evidence of dependence has not been detected.

- Level of significance α, the probability of rejecting H_0 when it is true:

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

Tests for Random Numbers

- When to use these tests:
  - If a well-known simulation language or random-number generator is used, it is probably unnecessary to test
  - If the generator is not explicitly known or documented, e.g., spreadsheet programs, symbolic/numerical calculators, tests should be applied to many sample numbers.

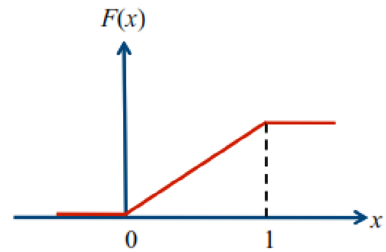
- Types of tests:
  - Theoretical tests: evaluate the choices of m, a, and c without actually generating any numbers
  - Empirical tests: applied to actual sequences of numbers produced.
    - Our emphasis.

Test for Random Numbers

1. *Frequency test*. Uses the Kolmogorov–Smirnov or the chi-square test to compare the distribution of the set of numbers generated to a uniform distribution.
2. *Autocorrelation test*. Tests the correlation between numbers and compares the sample correlation to the desired correlation, zero.

Kolmogorov-Smirnov Test

- Compares the continuous CDF, $F(x)$, of the uniform distribution with the empirical CDF, $S_N(x)$, of the N sample observations.
  - We know: $F(x) = x, \ \ 0 \le x \le 1$
  - If the sample from the RNG is $R_1, R_2, \ldots, R_N$, then the empirical CDF, $S_N(x)$ is:



$$S_N(x) = \frac{\text{Number of } R_i \text{ where } R_i \le x}{N}$$

- Based on the statistic: $D = max \mid F(x) - S_N(x) \mid$
  - Sampling distribution of D is known

Kolmogorov-Smirnov Test

- The test consists of the following steps
  - **Step 1:** Rank the data from smallest to largest $R_{(1)} \le R_{(2)} \le \ldots \le R_{(N)}$
  - **Step 2:** Compute

$$D^+ = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_{(i)} \right\}$$

$$D^- = \max_{1 \le i \le N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

  - **Step 3:** Compute $D = \max(D^+, D^-)$
  - **Step 4:** Get D_α for the significance level α
  - **Step 5:** If $D \le D_\alpha$ accept, otherwise reject H_0

Kolmogorov-Smirnov Critical Values

Degrees of Freedom (N)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27
Over 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Kolmogorov-Smirnov Test

- Example: Suppose N=5 numbers: 0.44, 0.81, 0.14, 0.05, 0.93.

	i	1	2	3	4	5
Step 1:	$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
	i/N	0.20	0.40	0.60	0.80	1.00
Step 2:	$i/N - R_{(i)}$	0.15	0.26	0.16	-	0.07
	$R_{(i)} - (i\text{-}1)/N$	0.05	-	0.04	0.21	0.13

Arrange $R_{(i)}$ from smallest to largest

$D^{+} = max\{i/N - R_{(i)}\}$

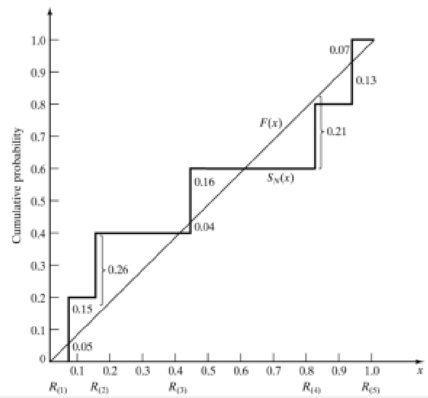
$D^{-} = max\{R_{(i)} - (i\text{-}1)/N\}$

Step 3: $D = \max(D^{+}, D^{-}) = 0.26$

Step 4: For $\alpha = 0.05$,

$D_{\alpha} = 0.565 > D = 0.26$

Hence, H_0 is not rejected.



Chi-square Test

- Chi-square test uses the sample statistic:

n is the # of classes

O_i is the observed # in the i-th class

E_i is the expected # in the i-th class

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

  - Approximately the chi-square distribution with n-1 degrees of freedom
  - For the uniform distribution, E_i, the expected number in each class is:

$$E_i = \frac{N}{n}, \quad \text{where N is the total number of observations}$$

- Valid only for large samples, e.g., $N \geq 50$

Chi-square Test: Example

- Example with 100 numbers from [0,1], α=0.05
- 10 intervals
- $\chi^2_{0.05,9} = 16.9$
- Accept, since
  - $X^2_0 = 11.2 < \chi^2_{0.05,9}$

Interval	Upper Limit	O_i	E_i	O_i-E_i	(O_i-E_i)^2	(O_i-E_i)^2/E_i
1	0.1	10	10	0	0	0
2	0.2	9	10	-1	1	0.1
3	0.3	5	10	-5	25	2.5
4	0.4	6	10	-4	16	1.6
5	0.5	16	10	6	36	3.6
6	0.6	13	10	3	9	0.9
7	0.7	10	10	0	0	0
8	0.8	7	10	-3	9	0.9
9	0.9	10	10	0	0	0
10	1.0	14	10	4	16	1.6
Sum		**100**	**100**	**0**	**0**	**11.2**

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

X^2_0=11.2

Practice Questions

15. Consider the multiplicative congruential generator under the following circumstances:

    (a) $X_0 = 7, a = 11, m = 16$
    (b) $X_0 = 8, a = 11, m = 16$
    (c) $X_0 = 7, a = 7, m = 16$
    (d) $X_0 = 8, a = 7, m = 16$

    Generate enough values in each case to complete a cycle. What inferences can be drawn? Is maximum period achieved?

23. Solution to Exercise 23:

	Case (a)	Case (b)	Case (c)	Case (d)
i	X_i	X_i	X_i	X_i
0	7	8	7	8
1	13	8	1	8
2	15		7	8
3	5			
4	7			

Inferences:
Maximum period, $p = 4$, occurs when X_0 is odd and $a = 3 + 8k$ where $k = 1$. Even seeds have the minimal possible period regardless of a.

4. Use the linear congruential method to generate a sequence of three two-digit random integers and corresponding random numbers. Let $X_0 = 27$, $a = 8$, $c = 47$, and $m = 100$.

4. Solution to Exercise 4:

$X_0 = 27,\ a = 8,\ c = 47,\ m = 100$

$X_1 = (8 \times 27 + 47)\text{mod } 100 = 63, \quad R_1 = 63/100 = .63$

$X_2 = (8 \times 63 + 47)\text{mod } 100 = 51, \quad R_2 = 51/100 = .51$

$X_3 = (8 \times 51 + 47)\text{mod } 100 = 55, \quad R_3 = 55/100 = .55$

7. The sequence of numbers 0.54, 0.73, 0.98, 0.11, and 0.68 has been generated. Use the Kolmogorov–Smirnov test with $\alpha = 0.05$ to learn whether the hypothesis that the numbers are uniformly distributed on the interval [0, 1] can be rejected.

7. Solution to Exercise 7:

$R_{(i)}$	.11	.54	.68	.73	.98
i/N	.20	.40	.60	.80	1.0
$i/N - R_{(i)}$	.09	–	–	.07	.02
$R_{(i)} - (i-1)/N$	.11	.34	.28	.13	.18

$D^+ = \max_{1 \le i \le N}(i/N - R_{(i)}) = .09$

$D^- = \max_{1 \le i \le N}(R_{(i)} - (i-1)/N) = .34$

$D = \max(D^+, D^-) = .34$

The critical value, D_α, obtained from Table A.8 is

$$D_{.05} = .565$$

since $D < D_{.05}$, the hypothesis that there is no difference between the true distribution of $\{R_1, R_2, \ldots, R_5\}$ and the uniform distribution on $[0, 1]$ cannot be rejected on the basis of this test.

8. Reverse the 100 two-digit random numbers in Example 7 to get a new set of random numbers. Thus, the first random number in the new set will be 0.43. Use the chi-square test, with $\alpha = 0.05$, to learn whether the hypothesis that the numbers are uniformly distributed on the interval [0, 1] can be rejected.

0.34	0.90	0.25	0.89	0.87	0.44	0.12	0.21	0.46	0.67
0.83	0.76	0.79	0.64	0.70	0.81	0.94	0.74	0.22	0.74
0.96	0.99	0.77	0.67	0.56	0.41	0.52	0.73	0.99	0.02
0.47	0.30	0.17	0.82	0.56	0.05	0.45	0.31	0.78	0.05
0.79	0.71	0.23	0.19	0.82	0.93	0.65	0.37	0.39	0.42
0.99	0.17	0.99	0.46	0.05	0.66	0.10	0.42	0.18	0.49
0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.30	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.40	0.64	0.40	0.19	0.79	0.62
0.18	0.26	0.97	0.88	0.64	0.47	0.60	0.11	0.29	0.78

8. Let ten intervals be defined each from $(10i-9)$ to $(10i)$ where $i = 1, 2, \ldots, 10$. By counting the numbers that fall within each interval and comparing this to the expected value for each interval, $E_i = 10$, the following table is generated:

Interval	O_i	$(O_i - E_i)^2/E_i$
(01-10)	9	0.1
(11-20)	9	0.1
(21-30)	9	0.1
(31-40)	6	1.6
(41-50)	17	4.9
(51-60)	5	2.5
(61-70)	10	0.0
(71-80)	12	0.4
(81-90)	7	0.9
(91-00)	16	3.6
	100	14.2= χ_0^2

From Table A.6, $\chi^2_{.05,9} = 16.9$. Since $\chi_0^2 < \chi_{.05,9}$, then the null hypothesis of no difference between the sample distribution and the uniform distribution is not rejected.

10. Use the mixed congruential method to generate a sequence of three two-digit random integers and corresponding random numbers with $X_0 = 37$, $a = 7$, $c = 29$, and $m = 100$.

11. Use the mixed congruential method to generate a sequence of three two-digit random integers between 0 and 24 and corresponding random numbers with $X_0 = 13$, $a = 9$, and $c = 35$.

12. Write a computer program that will generate four-digit random numbers, using the multiplicative congruential method. Allow the user to input values of X_0, a, c, and m.

18. Solution to Exercise 18:

$X_1 = [7 \times 37 + 29] \bmod 100 = 88$
$R_1 = .88$
$X_2 = [7 \times 88 + 29] \bmod 100 = 45$
$R_2 = .45$
$X_3 = [7 \times 45 + 29] \bmod 100 = 44$
$R_3 = .44$

19. Use $m = 25$

$X_1 = [9 \times 13 + 35] \bmod 25 = 2$

$X_2 = [9 \times 2 + 35] \bmod 25 = 3$
$X_3 = [9 \times 3 + 35] \bmod 25 = 12$

20. Test the following sequence of numbers for uniformity and independence, using procedures you learned in this chapter: 0.594, 0.928, 0.515, 0.055, 0.507, 0.351, 0.262, 0.797, 0.788, 0.442, 0.097, 0.798, 0.227, 0.127, 0.474, 0.825, 0.007, 0.182, 0.929, 0.852.