Introduction

- Random Numbers (RNs) are a necessary basic ingredient in the simulation of almost all discrete systems.
- Most computer languages have a subroutine, object or function that generates a RN.
- Similarly, simulation languages generate RNs that are used to generate event times and other random variables.
- We will look at the generation of RNs and some randomness tests in this chapter. Next chapter will show how we can use them to generate RVs.
Properties of RNs

- A sequence of RNs, $R_1, R_2, \ldots$, must have two important statistical properties: uniformity and independence.
- Each RN, $R_i$ must be an independent sample drawn from a continuous uniform distribution between 0 and 1.

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

$$E(R) = \int_0^1 x \, dx = \frac{1}{2}$$

$$V(R) = E(R^2) - [E(R)]^2 = \frac{1}{12}$$
Properties of RNs

Some Consequences of Uniformity and Independence

- If the interval \([0, 1]\) is divided into \(n\) classes (sub-intervals) of equal length, the expected number of observations in each interval is \(N/n\), where \(N\) is the total number of observations.
- The probability of observing a value in a particular interval is independent of the previous values drawn.
Generation of Pseudo-RNs

Problems and Errors

- Numbers might not be uniformly distributed.
- Numbers might be discrete-valued.
- The mean / variance of the generated numbers might be too high or too low.
- There might be dependence, such as,
  - autocorrelation
  - numbers successively higher or lower than adjacent numbers
  - several numbers above the mean followed several numbers below the mean
Generation of Pseudo-RNs

Important Considerations

- The routine should be fast.
- The routine should be portable.
- The routine should have a sufficiently long cycle.
- The RNs should be replicable (repeatable).
- Most importantly, the generated RNs should closely approximate the ideal statistical properties of uniformity and independence.
Linear Congruential Method

- The linear congruential method (LCM) produces a sequence of integers, $X_1, X_2, \ldots$ between 0 and $m - 1$ by following a recursive relationship.

$$X_{i+1} = (aX_i + c) \mod m, \quad i = 0, 1, 2, \ldots$$

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

- The initial value $X_0$ is called the seed, $a$ is called the multiplier, $c$ is the increment and $m$ is the modulus.
- If $c = 0$, it is known as the multiplicative congruential method, and if $c \neq 0$, it is called as the mixed congruential method.
Linear Congruential Method

Example

Use the LGM to generate a sequence of RNs with $X_0 = 27$, $a = 17$, $c = 43$ and $m = 100$.

\begin{align*}
X_0 & = 27 \\
X_1 & = (17 \times 27 + 43) \mod 100 = 2 \Rightarrow R_1 = \frac{2}{100} = 0.02 \\
X_2 & = (17 \times 2 + 43) \mod 100 = 77 \Rightarrow R_2 = \frac{77}{100} = 0.77 \\
X_3 & = (17 \times 77 + 43) \mod 100 = 52 \Rightarrow R_3 = \frac{52}{100} = 0.52
\end{align*}
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

Properties to Consider

- For \( m \) a power of 2, say \( m = 2^b \), and \( c \neq 0 \), the longest possible period is \( P = m = 2^b \), which is achieved whenever \( c \) is relatively prime to \( m \) (the greatest common factor of \( c \) and \( m \) is 1) and \( a = 1 + 4k \), where \( k \) is an integer.

- For \( m \) a power of 2, say \( m = 2^b \), and \( c = 0 \), the longest possible period is \( P = m/4 = 2^{b-2} \), which is achieved if the seed \( X_0 \) is odd and if the multiplier \( a \), is given by \( a = 3 + 8k \) or \( a = 5 + 8k \), for some \( k = 0, 1, \ldots \).

- For \( m \) a prime number and \( c = 0 \), the longest possible period is \( P = m - 1 \), which is achieved whenever the multiplier, \( a \), has the property that the smallest integer \( k \) such that \( a^k - 1 \) is divisible by \( m \) is \( k = m - 1 \).
Linear Congruential Method

Properties to Consider—Example 7.2

Using the multiplicative LCM, find the period of the generator for $a = 13$, $m = 2^6 = 64$ and $X_0 = 1, 2, 3, 4$. The solution is given on the next slide. When the seed is 1 or 3, the sequence has a period of 16. Period lengths of 8 and 4 is achieved when the seed is 2 and 4, respectively. In this example, $m = 2^6 = 64$ and $c = 0$. The maximum period is therefore $P = m/4 = 16$. 
Linear Congruential Method

Properties to Consider-Example 7.2

Table: Periods for Various Seeds

<table>
<thead>
<tr>
<th>i</th>
<th>X_i</th>
<th>X_i</th>
<th>X_i</th>
<th>X_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>26</td>
<td>39</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>18</td>
<td>59</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>42</td>
<td>63</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>34</td>
<td>51</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td>58</td>
<td>23</td>
<td>52</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>50</td>
<td>43</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>37</td>
<td>10</td>
<td>47</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>2</td>
<td>35</td>
<td>4</td>
</tr>
</tbody>
</table>
Linear Congruential Method

Properties to Consider-Example 1

- With $a = 13 = 1 + 4 \times 3$, $c = 3$ is relatively prime to $m = 16$ and $X_0 = 1$, we have the following sequence with a period of $P = m = 2^b = 2^4 = 16$:

<table>
<thead>
<tr>
<th>$i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>14</td>
<td>9</td>
<td>8</td>
<td>11</td>
<td>2</td>
<td>13</td>
<td>12</td>
<td>15</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>
Linear Congruential Method

Properties to Consider-Example 2

With \( a = 3 \), \( c = 0 \), prime number \( m = 17 \) and \( X_0 = 1 \), we have the following sequence with a period of \( P = m - 1 = 16 \) when \( k = 16 \) is the smallest integer such that \( a^k - 1 = 3^{16} - 1 = 43,046,720 \) is divisible by \( k = m - 1 = 16 \):

\[
\begin{array}{cccccccccccccccc}
  i & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\
  X_i & 1 & 3 & 9 & 10 & 13 & 5 & 15 & 11 & 16 & 14 & 8 & 7 & 4 & 12 & 2 & 6 & 1 \\
\end{array}
\]
Combined Linear Congruential Generators

- A RNG with a period of $2^{31} - 1 \approx 2 \times 10^9$ is no longer adequate due to the increasing complexity. So, combine two or more multiplicative congruential generators in such a way that the combined generator has good statistical properties and a longer period.

- If $W_{i_1}, W_{i_2}, \ldots, W_{i_k}$ are any independent, discrete-valued RVs (not necessarily identically distributed), but one of them, say $W_{i_1}$, is uniform on the integers from 0 to $m_1 - 2$, then, the following is uniform on the integers from 0 to $m_1 - 2$.

$$W_i = \left( \sum_{j=1}^{k} W_{ij} \right) \mod (m_1 - 1)$$
Combined Linear Congruential Generators

- Let $X_{i1}, X_{i2}, \ldots X_{ik}$ be the $i$th output from $k$ different multiplicative congruential generators.

$$X_i = \left( \sum_{j=1}^{k} (-1)^{j-1} X_{ij} \right) \mod m_1 - 1$$

$$R_i = \begin{cases} \frac{X_i}{m_1}, & X_i > 0 \\ \frac{X_i}{m_1 - 1}, & X_i = 0 \end{cases}$$

- The maximum period is given by

$$P = \frac{(m_1 - 1)(m_2 - 1) \ldots (m_k - 1)}{2^{k-1}}$$
Combined Linear Congruential Generators

Algorithm by L’Ecuyer (1998)

**Step (1)** Select seed $X_{1,0}$ in the range $[1, 2, 147, 483, 562]$ for the first generator, and seed $X_{2,0}$ in the range $[1, 2, 147, 483, 398]$ for the second. Set $j = 0$.

**Step (2)** Evaluate each individual generator.

\[
X_{1,j+1} = 40,014X_{1,j} \mod 2,147,483,563
\]
\[
X_{2,j+1} = 40,692X_{2,j} \mod 2,147,483,399
\]

**Step (3)** Set

\[
X_{j+1} = (X_{1,j+1} - X_{2,j+1}) \mod 2,147,483,562
\]
Combined Linear Congruential Generators

Algorithm by L’Ecuyer (1998)

Step (4) Return

\[ R_{j+1} = \begin{cases} \frac{X_{j+1}}{2,147,483,563}, & X_{j+1} > 0 \\ \frac{2,147,483,562}{2,147,483,563}, & X_{j+1} = 0 \end{cases} \]

Step (5) Set \( j = j + 1 \) and go to step 2.
RN Streams

- The seed for a LCG is the integer value $X_0$ that initializes the RN sequence.
- Any value in the sequence $X_0, X_1, \ldots, X_P$ could be used to “seed” the generator.
- A RN stream is a convenient way to refer to a starting seed taken from the sequence.
- Typically these starting seeds are far apart in the sequence. If the streams are $b$ values apart, then, stream $i$ could be defined by starting seed $S_i = X_{b(i-1)}$, for $i = 1, 2, \ldots, \lceil P/b \rceil$.
- Values of $b = 100,000$ were common in older generators, but values as large as $b = 10^{37}$ are in use in modern combined LCGs.
Tests for RNs

To check on whether the desirable properties of uniformity and independence, a number of tests can be performed.

The tests can be placed in two categories, according to the properties of interest: uniformity and independence.

- 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.

- Autocorrelation Test: Tests the correlation between numbers and compares the sample correlation to the expected correlation, zero.
Tests for RNs

In testing for uniformity, the hypotheses are as follows:

\[ H_0 : R_i \sim U[0, 1] \]
\[ H_1 : R_i \not\sim U[0, 1] \]

In testing for uniformity, the hypotheses are as follows:

\[ H_0 : R_i \sim \text{independently} \]
\[ H_1 : R_i \not\sim \text{independently} \]
Frequency Tests

Kolmogorov-Smirnov (K-S) Test

- This test compared the continuous CDF, $F(x)$, of the uniform distribution with the empirical CDF, $S_N(x)$. We have
  \[ F(x) = x, \quad 0 \leq x \leq 1 \]

- The empirical CDF $S_N(x)$ defined by
  \[ S_N(x) = \frac{\text{number of } R_1, R_2, \ldots, R_N \text{ which are } \leq x}{N} \]

- K-S test is based on the largest absolute deviation between
  \[ D = \max |F(x) - S_N(x)| \]
Frequency Tests

K-S Test

Step (1) Rank the data from smallest to largest. Let $R(i)$, denote the $i$th smallest observation.

Step (2) Compute

$$D^+ = \max \left\{ \frac{i}{N} - R(i) \right\}$$

$$D^- = \max \left\{ R(i) - \frac{i - 1}{N} \right\}$$
Frequency Tests

K-S Test

Step (3) Compute $D = \max(D^+, D^-)$

Step (4) Locate in Table A.8 the critical value $D_{\alpha, N}$.

Step (5) If $D > D_{\alpha, N}$, the null hypothesis is rejected. If $D \leq D_{\alpha, N}$, conclude that no difference has been detected between the distributions.
Frequency Tests

K-S Test Example

- Suppose that we have five numbers, 0.44, 0.81, 0.14, 0.05 and 0.93. Perform a test for uniformity using the K-S test with the significance level of $\alpha = 0.05$.
- We must first rank the numbers from smallest to largest. The calculations are seen in the table on the next slide.
- The computations for $D^+$ and $D^-$ are shown as $i/N - R(i)$ and $R(i) - (i - 1)/N$, respectively.
- We see that $D^+ = 0.26$, $D^- = 0.21$, $D = 0.26$ and $D_{\alpha,N} = 0.565$. Since $D < D_{\alpha,N}$, the hypothesis that the distribution is uniform distribution is not rejected.
Frequency Tests

K-S Test Example

Table: Calculations for K-S Test

<table>
<thead>
<tr>
<th>$R(i)$</th>
<th>0.05</th>
<th>0.14</th>
<th>0.44</th>
<th>0.81</th>
<th>0.93</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i/N$</td>
<td>0.20</td>
<td>0.40</td>
<td>0.60</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td>$i/N - R(i)$</td>
<td>0.15</td>
<td>0.26</td>
<td>0.16</td>
<td>-</td>
<td>0.07</td>
</tr>
<tr>
<td>$R(i) - (i - 1)/N$</td>
<td>0.05</td>
<td>-</td>
<td>0.04</td>
<td>0.21</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Frequency Tests

K-S Test Example

Figure: Comparison of $F(x)$ and $S_N(x)$
Frequency Tests

Chi-Square (C-S) Test

- The C-S test uses the sample statistic

\[
\chi^2_0 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}
\]

- \(O_i\) and \(E_i\) are the observed and expected number in class \(i\). For equally spaced classes,

\[
E_i = \frac{N}{n}
\]

- It can be shown that \(\chi^2_0\) is approximately chi-squared distributed with \(n - 1\) degrees of freedom.
Frequency Tests

C-S Test Example (Example 7.7 in DESS)

Considering the given data the following computations are done. Since $\chi^2_0 = 3.4 < \chi^2_{0.05,9} = 16.9$, the null hypothesis is not rejected.

Table: Calculations for C-S Test

<table>
<thead>
<tr>
<th>Interval</th>
<th>$O_i$</th>
<th>$E_i$</th>
<th>$O_i - E_i$</th>
<th>$(O_i - E_i)^2$</th>
<th>$\frac{(O_i - E_i)^2}{E_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>10</td>
<td>-2</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>10</td>
<td>-2</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>10</td>
<td>4</td>
<td>16</td>
<td>1.6</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>3.4</td>
<td></td>
</tr>
</tbody>
</table>


The tests for autocorrelation are concerned with the dependence between numbers in a sequence.

We will consider a test for autocorrelation. It requires the computation of autocorrelation between every $m$ numbers ($m$ is the lag), starting with the $i$th number.

Thus, the autocorrelation $\rho_{im}$ between the following numbers would be of interest: $R_i, R_{i+m}, R_{i+2m}, \ldots, R_{i+(M+1)m}$.

The value $M$ is largest integer st $i + (M + 1)m \leq N$, where $N$ is the total number of values in the sequence. We have,

$$
H_0 : \rho_{im} = 0 \\
H_1 : \rho_{im} \neq 0
$$
Autocorrelation Tests

The distribution of the estimator $\hat{\rho}_{im}$ is approximately normal if the data are uncorrelated. We have the standard normal test statistic of $Z_0$ and do not reject $H_0$ if $-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$.

$$Z_0 = \frac{\hat{\rho}_{im}}{\sigma_{\hat{\rho}_{im}}}$$

$$\hat{\rho}_{im} = \frac{1}{M + 1} \left( \sum_{k=0}^{M} [R_{i+km}] [R_{i+(k+1)m}] \right) - 0.25$$

$$\sigma_{\hat{\rho}_{im}} = \frac{\sqrt{13M + 7}}{12(M + 1)}$$
Autocorrelation Tests

Autocorrelation Test Example (Example 7.8 in DESS)

Considering the data in the text, we test for whether the 3rd, 8th, 13th and so on, numbers are autocorrelated using $\alpha = 0.05$. Here, $i = 3$, $m = 5$, $N = 30$ and $M = 4$ (largest integer st $3 + (M + 1)5 \leq 30$). Then,

$$
\hat{\rho}_{im} = \frac{1}{M+1} \left( \sum_{k=0}^{M} [R_{i+km}] [R_{i+(k+1)m}] \right) - 0.25
$$

$$
= \frac{1}{4+1} \left( .23(.28) + .28(.33) + .33(.27) + .27(.05) + .05(.36) \right) - 0.25
$$

$$
= -0.1945
$$
Autocorrelation Tests

Autocorrelation Test Example (Example 7.8 in DESS)

\[
\hat{\rho}_{im} = \frac{\sqrt{13M + 7}}{12(M + 1)} = \frac{\sqrt{13(4) + 7}}{12(4 + 1)} = 0.1280
\]

\[
Z_0 = \frac{\hat{\rho}_{im}}{\sigma_{\hat{\rho}_{im}}} = -\frac{0.1945}{0.1280} = -1.516
\]

Since \(-z_{0.025} = -1.96 \leq Z_0 \leq 1.96 = z_{0.025}\), we cannot reject the null hypothesis.
Summary

- Reading HW: Chapter 7.
- Chapter 7 Exercises.