Systems Simulation
Chapter 7: Random-Number Generation

Fatih Cavdur
fatihcavdur@uludag.edu.tr

April 22, 2014

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.

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

$$E(R) = \int_0^1 r dr = \frac{1}{2}$$

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

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 $m = 100$.

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

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

1. 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.
2. 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 $k = 0, 1, \ldots$.
3. 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 1

Using the multiplicative LCM, find the period of the generator for $a = 13, m = 2^6 = 64$ and $X_0 = 1, 2, 3, 4$. 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 max period is then $P = m/4 = 16$.

Table : Periods for Various Seeds

i	X_i	X_i	X_i	X_i
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	52
6	57	50	43	36
7	37	10	47	20
8	33	2	35	4

Linear Congruential Method

Properties to Consider-Example 2

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

Table : Max Period

i	X_i	i	X_i
1	0	9	8
2	3	10	11
3	10	11	2
4	5	12	13
5	4	13	12
6	7	14	15
7	14	15	6
8	9	16	1

Linear Congruential Method

Properties to Consider-Example 3

With $a = 3$, $c = 0$, prime number $m = 17$ and $X_0 = 1$, we have the following sequence with the max period of $P = m - 1 = 16$ when $k = 16$ is the smallest integer such that $a^k - 1 = 3^{16} - 1$ (which equals to 43,046,720) is divisible by $k = m - 1 = 16$ (verify that for $k < 16$, $a^k - 1$ is not divisible by $k = m - 1$):

Table : Max Period

i	X_i	i	X_i
1	3	9	14
2	9	10	8
3	10	11	7
4	13	12	4
5	5	13	12
6	15	14	2
7	11	15	6
8	16	16	1

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 $W_{i1}, W_{i2}, \ldots, W_{ik}$ are any independent, discrete-valued RVs (not necessarily identically distributed), but one of them, say W_{i1}, 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 ith output from k different multiplicative congruential generators.

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

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

- 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.

$$\begin{aligned} X_{1,j+1} &= 40,014 X_{1,j} \mod 2,147,483,563 \\ X_{2,j+1} &= 40,692 X_{2,j} \mod 2,147,483,399 \end{aligned}$$

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 $i = 1, 2, \ldots, \lfloor P/b \rfloor$.
- 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 o compare the distribution of the set of numbers generated to a uniform distribution.
- Autocorrelation Test: Tests the correlation between numbers and compares the sample compares the sample correlation to the expected correlation, zero.

Tests for RNs

- In testing for uniformity, the hypotheses are as follows:

$$\begin{aligned} H_0 &: R_i \sim U[0,1] \\ H_1 &: R_i \nsim U[0,1] \end{aligned}$$

- In testing for uniformity, the hypotheses are as follows:

$$\begin{aligned} H_0 &: R_i \sim \text{ independently} \\ H_1 &: R_i \nsim \text{ independently} \end{aligned}$$

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 ith smallest observation.

Step (2) Compute

$$\begin{aligned} D^+ &= \max\left\{\frac{i}{N} - R_{(i)}\right\} \\ D^- &= \max\left\{R_{(i)} - \frac{i-1}{N}\right\} \end{aligned}$$

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

$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
i/N	0.20	0.40	0.60	0.80	1.00
$i/N - R_{(i)}$	0.15	0.26	0.16	-	0.07
$R_{(i)} - (i-1)/N$	0.05	-	0.04	0.21	0.13

Frequency Tests

K-S Test Example

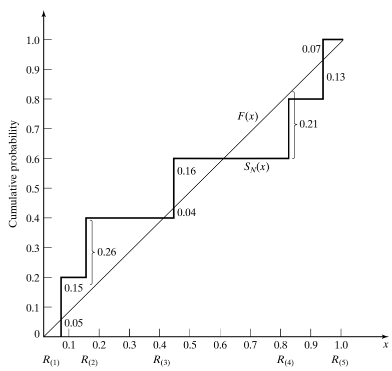

Figure : Comparison of $F(x)$ and $S_N(x)$

Frequency Tests

Chi-Square (C-S) Test

- The C-S test uses the sample statistic

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

- 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 χ_0^2 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 $\chi_0^2 = 3.4 < \chi_{0.05,9}^2 = 16.9$, the null hypothesis is not rejected.

Table : Calculations for C-S Test

Interval	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
1	8	10	-2	4	0.4
2	8	10	-2	4	0.4
3	10	10	0	0	0.0
...	...	...	...	...	...
8	14	10	4	16	1.6
9	10	10	0	0	0.0
10	11	10	1	1	0.0
	100	100	0		3.4

Autocorrelation Tests

- 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 ith number.
- Thus, the autocorrelation ρ_{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,

$$\begin{aligned} H_0 &: \rho_{im} = 0 \\ H_1 &: \rho_{im} \neq 0 \end{aligned}$$

Autocorrelation Tests

- The distribution of the estimator $\widehat{\rho}_{im}$ 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 $-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$.

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

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

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

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,

$$\begin{aligned}
\widehat{\rho}_{im} &= \frac{1}{M+1}\left(\sum_{k=0}^{M}\left[R_{i+km}\right]\left[R_{i+(k+1)m}\right]\right) - 0.25 \\
&= \frac{1}{4+1}(.23(.28) + .28(.33) + .33(.27) + .27(.05) + .05(.36)) - 0.25 \\
&= -0.1945
\end{aligned}$$

Autocorrelation Tests

Autocorrelation Test Example (Example 7.8 in DESS)

$$\sigma_{\widehat{\rho}_{im}} = \frac{\sqrt{13M+7}}{12(M+1)} = \frac{\sqrt{13(4)+7}}{12(4+1)} = 0.1280$$

$$Z_0 = \frac{\widehat{\rho}_{im}}{\sigma_{\widehat{\rho}_{im}}} = -\frac{0.1945}{0.1280} = -1.516$$

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.