Chapter 7
Random-Number Generation

Random Number Generators

- Simulation programs can get input data from an outside source (trace driven simulation).
- The usefulness of these programs is limited by amount of available data
  - What if more data needed?
  - What if the model changed?
  - What if the input data set is small or unavailable?
- A random number generator address all problems
  - It produces real values between 0.0 and 1.0
  - The output can be converted to other random variate

Generation of Pseudo-Random Numbers

- Algorithmic generators are widely accepted because they meet all of the following criteria:
  - Randomness
    output passes all reasonable statistical tests of randomness
  - controllability
    able to reproduce output, if desired
  - portability
    able to produce the same output on a wide variety of computer systems
  - efficiency
    fast, minimal computer resource requirements
  - documentation - theoretically analyzed and extensively tested

Algorithmic Generators

- An ideal random number generator produces output such that each value in the interval

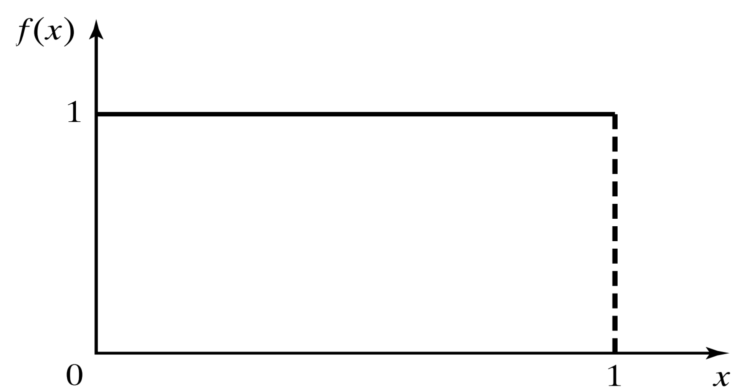
  $0.0 < u < 1.0$ is equally likely to occur.
- A good random number generator produces output that is (almost) statistically indistinguishable from an ideal generator.
- We will construct a good random number generator satisfying all our criteria.

Properties of Random Numbers

- Two important statistical properties
  uniformity and independence
- Random number, RN must be independently drawn from a uniform distribution with pdf:

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

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



Conceptual Model

- Choose a large positive integer m.
  This defines the set $X_m = \{1, 2, \ldots, m - 1\}$
- Each time a random number *u* is needed, draw an integer x "at random" from the set and let
  $u = x/m$
- Each draw simulates a sample of an independent identically distributed sequence of Uniform(0, 1)
- The possible values are $1/m, 2/m, \ldots (m - 1)/m$.
- It is important that m should be large so that the possible values are densely distributed between 0.0 and 1.0

Linear Congruential Generators

- Initially proposed by Lehmer in 1951
- Lehmer discovered the residues of successive powers of a number have good randomness properties
- Produce a sequence of integers, $x_1, x_2, \ldots$ between 0 and m-1 according to the following recursive relationship:

$$x_{i+1} = (ax_i + c) \bmod m, \; i=1, 2, \ldots$$

Pseudo-Random Generator

$$x_{i+1} = (ax_i + c) \bmod m, \; i=1, 2, \ldots$$

- x_0 – seed
- a – constant multiplier
- c – increment
- m – modulus
- The selection of a ,c, and m affect the period and statistical properties. m should be large.

Example 1

- x_0 = 27, a=17, c=43, m=100

Then: X_{100} = {1, 2, …, 99}

x_1=(a*x_0+c) mod m = (17*27+43) mod 100 =2

x_2=(a*x_1+c) mod m = (17*2+43) mod 100 =77

x_3 =(a*x_2+c) mod m= (17*77+43) mod 100 =52

x_4= …

The sequence of x_i: 27, 2, 77, 52, …,

Random numbers: x_0/m, x_1/m, x_2/m, x_3/m, …

0.27, 0.02, 0.77, 0.52, …,

Multiplicative Congruential Generators

- A special case of linear generators is c=0
- Example 2: m=13, a=6, x_0=1, the sequence is

  1, 6, 10, 8, 9, 2, 12, 7, 3, 5, 4, 11, 1, . . .
- Example 3: m = 13 and a = 7 with x_0 = 1, the sequence

  1, 7, 10, 5, 9, 11, 12, 6, 3, 8, 4, 2, 1 . . .
- Because of the 12, 6, 3 and 8, 4, 2, 1 patterns, this sequence appears “less random”
- Example 4: If m = 13 and a = 5 then

  1, 5, 12, 8, 1, . . . or 2, 10, 11, 3, 2, . . . or 4, 7, 9, 6, 4, . . .
- This less-than-full-period behavior is obviously undesirable

Central Issues

- Let $g(x) = ax \bmod m$. For a chosen (a,m) pair, does the function $g(\cdot)$ generate a full-period sequence?
- If a full period sequence is generated, how random does the sequence appear to be?
- Can ($ax \bmod m$) be evaluated efficiently and correctly? Integer overflow can occur when computing ax.

The selection of m

- On a 32-bit computer system, $2^{31} - 1$ is the largest possible positive integer, and it is prime.
- On a 64-bit computer system, $2^{63} - 1$ is the largest possible positive integer, but it is not prime.

Random-Number Generation Library

- The C library <stdlib.h> contains: rand( )
  range: {0, 1, 2, …, m-1}

where m-1 required at least 2^{15}.

u= (double) rand( ) / RAND_MAX

Theorem 1

- If the sequence x0, x1, x2, . . . is produced by a Lehmer generator with multiplier *a* and modulus *m* then

$$x_i = a^i x_0 \bmod m$$

- It is a bad idea to compute xi by first computing a^i.
- Theorem 1 has significant theoretical value.

Theorem 2

- If $x_0 \in Xm$ and the sequence $x_0, x_1, x_2 \ldots$ is produced by a multiplicative generator with multiplier a and (prime) modulus m then there is a positive integer p with $p \leq m - 1$ such that

  $x_0, x_1, x_2 \ldots x_{p-1}$ are all different and

  $$x_{i+p} = x_i, \quad i = 0, 1, 2, \ldots$$

- That is, the sequence is periodic with fundamental period p.
- In addition $(m - 1) \bmod p = 0$,

  thus, p is a divisor of m-1.

The Periodical Property

- If we pick any initial seed $x_0 \in$ Xm and generate the sequence

  $x_0, x_1, x_2, \ldots$ then x_0 will occur again.
- Further x_0 will reappear at index p that is either $m - 1$ or a divisor of $m - 1$.
- The pattern will repeat forever.

Full Period Multipliers

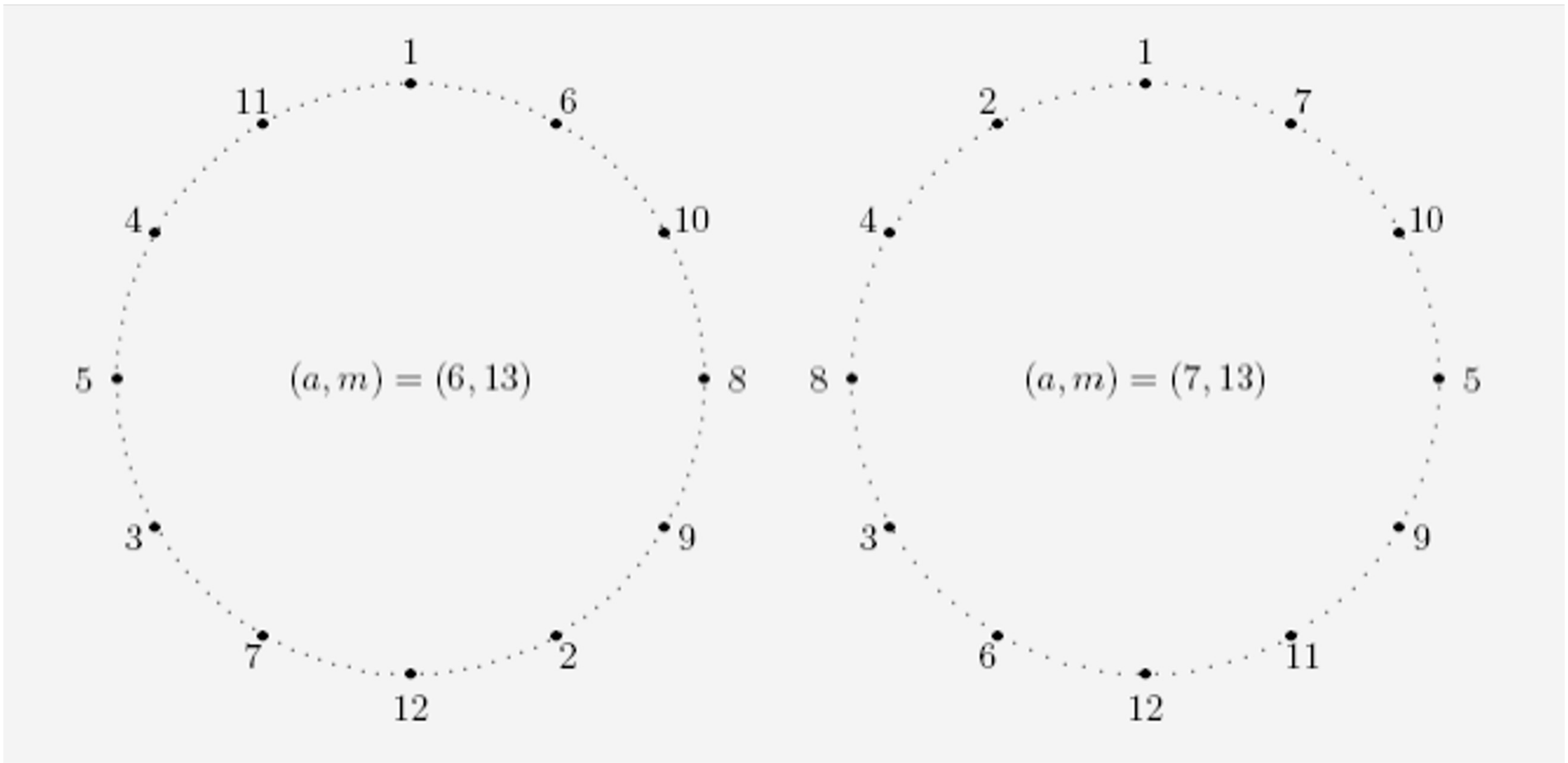
- Definition: The sequence

  $$x_0, x_1, x_2, \ldots$$

  produced by a multiplicative congruential generator with modulus m and multiplier a has a full period if and only if the fundamental period p is m-1. If the sequence has a full period, then a is said to be a full-period multiplier relative to m.
- We are interested in choosing full-period multipliers where $p = m - 1$.

Example 5

- Full-period multipliers generate a virtual circular list with $m - 1$ distinct elements.



Finding Full Period Multipliers

// the following algorithm uses $x_0=1$ as the seed ($x_1=a$). New value
// of x is recursively generated until the initial seed reappears.

```
p = 1;
x = a;
while (x != 1)     /* until x=1, it is the end if the period */
{
   p++;
   x = (a * x)% m;  /* beware of a * x overflow */
}
If (p == m − 1)
           /* a is a full-period multiplier */
else
         /* a is not a full-period multiplier */
```

Frequency of Full-Period Multipliers

- Given a prime modulus *m*, how many corresponding full-period multipliers are there?
- **Theorem 3:** If m is prime and $p_1, p_2, \ldots, p_r$ are the (unique) prime factors of $m - 1$. Then the number of full-period multipliers is:

$$\frac{(p_1 - 1)(p_2 - 1)\ldots(p_r - 1)}{p_1 p_2 \cdots p_r}(m-1).$$

- **Example 6:** If m = 13 then m − 1 = 12 = 2^2 *3. Therefore, there are $\frac{(2-1)(3-1)}{2 \times 3}(13-1) = 4$ full-period multipliers (2, 6, 7, and 11).

Example 7

- If $m = 2^{31} - 1 = 2147483647$ then since the prime decomposition of $m - 1$ is

$$m - 1 = 2^{31} - 2 = 2 \cdot 3^2 \cdot 7 \cdot 11 \cdot 31 \cdot 151 \cdot 331$$

- the number of full period multipliers is

$$\frac{(1 \cdot 2 \cdot 6 \cdot 10 \cdot 30 \cdot 150 \cdot 330)\,(2 \cdot 3^2 \cdot 7 \cdot 11 \cdot 31 \cdot 151 \cdot 331)}{2 \cdot 3 \cdot 7 \cdot 11 \cdot 31 \cdot 151 \cdot 331}$$

$= 534600000$

- Therefore, approximately 25% of the multipliers are full-period.

Relative Prime

- Definition: two positive integers *a* and *b* are relative prime if they have no common prime divisors.
- How to test *a* and *b* are relative prime?

$$\gcd(a, b) = 1$$

gcd(a, b) returns the greatest common divisor of a and b.

Examples: gcd(12, 30)=6, gcd(12, 24)=12
gcd(10, 18)=2, gcd(10, 21)=1

Example 8

- If m = 13 then we know from Example 6 there are 4 full period multipliers.

- From Example 5, a = 6 is one. Then, since 1, 5, 7, and 11 are relatively prime to 13-1,

  $6^1 \bmod 13 = 6$, $6^5 \bmod 13 = 2$

  $6^7 \bmod 13 = 7$, $6^{11} \bmod 13 = 11$

- Equivalently, if we knew a = 2 is a full-period multiplier,

  $2^1 \bmod 13 = 2$, $2^5 \bmod 13 = 6$

  $2^7 \bmod 13 = 11$, $2^{11} \bmod 13 = 7$

Finding All Full-Period Multipliers

- Once one full-period multiplier has been found, then all others can be found by the following Algorithm

```
i = 1;
x = a;  /* assume a is a full-period multiplier
while (x != 1)
{
    if (gcd(i, m − 1) == 1)
                /* a^i mod m is a full-period multiplier*/
    i++
    x = (a * x) % m;      /* beware a * x overflow */
}
```

Empirical Tests of Randomness

Testing for Randomness

- The output of pseudorandom number generators must be tested for uniformity and independence.

- There are several statistical tests but none of them are powerful enough to guarantee perfect randomness.

Empirical Test of Randomness

An empirical test of randomness is a statistical test of the hypothesis that repeated calls to a random number generator will produce an iid (independent, identically distributed) sample from a Uniform(0,1) distribution.

Three Steps

1. Generate a sample by repeated call to the generator.
2. Compute a test statistic whose statistical distribution is known when the random numbers are truly iid Uniform(0,1) random variates.
3. Assess the likelihood of the observed (computed from step 2) value of the test statistic relative to the theoretical distribution from which it is assumed to have been drawn.

Frequency Test

- Frequency test is applied to check the uniformity of a set of random numbers.
- Measure the observed frequencies and theoretical frequencies by means of a chi-squared test.
- A chi-squared test measures the degree of fit between the observed (actual) and the expected (theoretical) probability distribution.

Frequency Test Steps

1. Suppose $r_1, r_2, \ldots, r_n$ is a sequence of n pseudo-random numbers generated over the interval (0, 1) which is divided into s subintervals;

2. Let

   $f_o(j)$ = number of observations in
   subinterval j ,

   $f_e(j)$ = number of expected observations in
   subinterval j = n/s;

Frequency Test Steps - cont.

3. The chi-squared statistic is computed as:

$$\chi^2 = \sum_{j=1}^{s} \frac{[f_o(j) - f_e(j)]^2}{f_e(j)} = \frac{s}{n} \sum_{j=1}^{s} \left(f_o(j) - \frac{n}{s} \right)^2$$

4. Compare the obtained χ^2 value against a theoretical value in a table based on level of significance α and the degree of freedom (s – 1): $\chi^2_{\alpha,(s-1)}$

Frequency Test Steps - cont.

5. If $\chi^2 < \chi^2_{\alpha,(s-1)}$, then we accept the hypothesis that the numbers $r_1, r_2, \ldots, r_n$ are uniformly distributed.

Example: Assume we have $n = 100$ observations divided into $s = 10$ subintervals over (0, 1). The output of a random number generator is found to have the distribution:

Example - cont.

Interval	f_e	f_o
$0 - 0.1$	10	10
$0.1 - 0.2$	10	9
$0.2 - 0.3$	10	9
$0.3 - 0.4$	10	16
$0.4 - 0.5$	10	8
$0.5 - 0.6$	10	11
$0.6 - 0.7$	10	8
$0.7 - 0.8$	10	9
$0.8 - 0.9$	10	16
$0.9 - 1.0$	10	4

Applying the χ^2 formula:

$$\chi^2 = \frac{1}{10}\sum_{j=1}^{10}\left(f_o(j) - 10\right)^2 = 12$$

Example - cont.

- From the Chi-square table:

$$\alpha = 95\%, \ \ \chi^2_{0.95,9} = 16.9,$$

Thus we accept the hypothesis that the sample is from a population with uniform distribution.

- A small value of χ^2 implies a good fit between experiment and prediction.

Serial Test

- The serial test is applied to check randomness or independence of successive pseudo-random numbers.
- It is an extension of the chi-square test to higher dimensions.
- The first member of the sequence is compared with the second, the third, the fourth, etc. and checked for frequency derivations from theoretical values.

Serial Test - cont.

- This is repeated for all other numbers in the sequence.

- The simplest form of the test is applied to pairs of consecutive numbers in the sequence.

Serial Test Steps

1. Suppose $r_1, r_2, \ldots, r_{2n}$ is a sequence of 2n pseudo- random numbers generated in the range [1, m – 1],

   $$r_i \in \{1, 2, \ldots, m-1\}.$$

2. There are total n pairs of random numbers:

   (r_1, r_2) (r_3, r_4) (r_5, r_6) . . . (r_i , r_{i+1}) . . . (r_{2n-1}, r_{2n})

Serial Test Steps - cont.

3. Every pair (r_i , r_{i+1}) will select one of the $(m-1)^2$ squares:

	1	2	…	m-1
1				
2				
…				
m-1				

Serial Test Steps - cont.

4. If pairs of random numbers are independent and uniformly distributed,

$$\Pr[(r_i, r_i+1) \text{ selecting each square}] = 1/(m-1)^2$$

5. Applying the chi-square test with $(m-1)^2$ intervals (one for each square) and degree of freedom $(m-1)^2 - 1$.

Remark on Serial Test

The method only test independence within each pair of random numbers. It can be generated to triples, quadruples, etc., but number of intervals becomes very large.

Gap Test

- Gap test is based on the distribution of the length of the gaps between consecutive occurrence of random numbers in a specified interval.
- It is defined by two real-valued parameters a and b chosen so that $0 \le a < b \le 1$. These two parameters define a subinterval $(a, b) \subseteq (0, 1)$ with length $\delta = b - a$.

Length of Gaps

- Assume a sequence of random numbers:

  0.12 , 0.43 , 0.75 , 0.60 , 0.88 , 0.32 ,

  0.71 , 0.02 , 0.92 , 0.74 , 0.15 , ...

- If we use the interval (0, 0.5), we will have

  0.12 , 0.43 , 0.75 , 0.60 , 0.88 , 0.32 ,

  0.71 , 0.02 , 0.92 , 0.74 , 0.15 , ...

Gap - Geometric Distribution

$$\Pr[\text{Gap length} = 0] = \delta$$
$$\Pr[\text{Gap length} = 1] = \delta(1 - \delta)$$
$$\Pr[\text{Gap length} = 2] = \delta(1 - \delta)^2$$
$$\ldots = \ldots$$
$$\Pr[\text{Gap length} = i] = \delta(1 - \delta)^{i}$$

- To carry out the gap test, collect gap lengths from sequence of random numbers and use chi-square test to check if sample of lengths is from above distribution.

Gap Test - Example

0.12 , 0.97 , 0.34 , 0.76 , 0.09 , 0.22 , 0.65 , 0.87 ,
0.43 , 0.13 , 0.37 , 0.58 , 0.88 , 0.96 , 0.24 , 0.66 ,
0.47 , 0.27 , 0.18 , 0.06 , 0.25 , 0.64 , 0.75 , 0.11 ,
0.97 , 0.33 , 0.48 , 0.55 , 0.17 , 0.91 , 0.61 , 0.29 ,
0.36 , 0.01 , 0.41 , 0.10 , 0.82 , 0.98 , 0.71 , 0.35 ,
0.62 , 0.02 , 0.80 , 0.70 , 0.33 , 0.22 , 0.78 , 0.39 ,
0.51 , 0.30 .

- If the interval is (0.1, 0.4), there are total s = 19 gaps.

Distribution of Gap Length

Gap Length i	$f_e = \delta(1-\delta)^i * 19$	f_o
0	5.7	4
1	3.99	5
2	2.79	6
3	1.96	3
4	1.37	1
5	0.96	0
> 5	2.23	0

Applying the χ^2 Test

$$\begin{aligned}
\chi^2 &= \sum_{j=1}^{s} \frac{[f_o(j) - f_e(j)]^2}{f_e(j)} \\
&= \frac{(-1.7)^2}{5.7} + \frac{1.01^2}{3.99} + \frac{3.21^2}{2.79} + \frac{1.04^2}{1.96} + \frac{0.37^2}{1.37} \\
&\quad + \frac{0.96^2}{0.96} + \frac{2.23^2}{2.23} \\
&= 8.30.
\end{aligned}$$

From the table, we find that at

$\alpha = 95\%$, $\chi^2_{0.95,18} = 30.1$.

Hence we accept the hypothesis that the random numbers are independent.