

## Purpose \& Overview

Discuss the generation of random numbers.

- Introduce the subsequent testing for randomness:
$\square$ Frequency test
$\square$ Autocorrelation test.


## Properties of Random Numbers

- Two important statistical properties:
$\square$ Uniformity
$\square$ Independence.
- Random Number, $R_{i}$, must be independently drawn from a uniform distribution with pdf:

$$
\begin{aligned}
& f(x)= \begin{cases}1, & 0 \leq x \leq 1 \\
0, & \text { otherwise }\end{cases} \\
& E(R)=\int_{0}^{1} x d x=\left.\frac{x^{2}}{2}\right|_{0} ^{1}=\frac{1}{2}
\end{aligned}
$$



Figure: pdf for random numbers

## Generation of Pseudo-Random Numbers

- "Pseudo", because generating numbers using a known method removes the potential for true randomness.
- Goal: To produce a sequence of numbers in $[0,1]$ that simulates, or imitates, the ideal properties of random numbers (RN).
- Important considerations in RN routines:
$\square$ Fast
$\square$ Portable to different computers
$\square$ Have sufficiently long cycle
$\square$ Replicable
$\square$ Closely approximate the ideal statistical properties of uniformity and independence.


## Techniques for Generating Random <br> Numbers

- Linear Congruential Method (LCM).
- Combined Linear Congruential Generators (CLCG).
- Random-Number Streams.


## Linear Congruential Method

- To produce a sequence of integers, $X_{1}, X_{2}, \ldots$ between 0 and $m-1$ by following a recursive relationship:

- The selection of the values for $a, c, m$, and $X_{0}$ drastically affects the statistical properties and the cycle length.
- The random integers are being generated [0,m-1], and to convert the integers to random numbers:

$$
R_{i}=\frac{X_{i}}{m}, \quad i=1,2, \ldots
$$

## Example

[LCM]

- Use $X_{0}=27, a=17, c=43$, and $m=100$.
- The $X_{i}$ and $R_{i}$ values are:

$$
\begin{array}{ll}
X_{1}=\left(17^{*} 27+43\right) \bmod 100=502 \bmod 100=2, & R_{1}=0.02 ; \\
X_{2}=\left(17^{*} 2+32\right) \bmod 100=77, & R_{2}=0.77 ; \\
X_{3}=\left(17^{*} 77+32\right) \bmod 100=52, & R_{3}=0.52 ;
\end{array}
$$

## Characteristics of a Good Generator

Maximum Density
$\square$ Such that he values assumed by $R_{i}, i=1,2, \ldots$, leave no large gaps on [0,1]
$\square$ Problem: Instead of continuous, each $R_{i}$ is discrete
Solution: a very large integer for modulus $m$

- Approximation appears to be of little consequence
- Maximum Period

To achieve maximum density and avoid cycling.Achieve by: proper choice of $a, c, m$, and $X_{0}$.

- Most digital computers use a binary representation of numbers

Speed and efficiency are aided by a modulus, $m$, to be (or close to) a power of 2.

## Combined Linear Congruential Generators

- Reason: Longer period generator is needed because of the increasing complexity of stimulated systems.
- Approach: Combine two or more multiplicative congruential generators.
- Let $X_{i, 1}, X_{i, 2}, \ldots, X_{i, k}$, be the $i^{\text {th }}$ output from $k$ different multiplicative congruential generators.
$\square$ The $\mathrm{j}^{\text {th }}$ generator:
- Has prime modulus $m_{j}$ and multiplier $a_{j}$ and period is $m_{j-1}$
- Produces integers $X_{i, j}$ is approx $\sim$ Uniform on integers in [ 1, m-1]
- $W_{i, j}=X_{i, j}-1$ is approx $\sim$ Uniform on integers in [1, $\left.m-2\right]$


## Combined Linear Congruential Generators

[Techniques]
Suggested form:

$$
X_{i}=\left(\sum_{j=1}^{k}(-1)^{j-1} X_{i, j}\right) \bmod m_{1}-1 \quad \text { Hence, } R_{i}= \begin{cases}\frac{X_{i}}{m_{1}}, & X_{i} \succ 0 \\
\frac{m_{1}-1}{m_{1}}, & X_{i}=0 \\
\begin{array}{c}
\text { The coefficient: } \\
\text { Performs the } \\
\text { subtraction } x
\end{array}\end{cases}
$$

- The maximum possible period is:

$$
P=\frac{\left(m_{1}-1\right)\left(m_{2}-1\right) \ldots\left(m_{k}-1\right)}{2^{k-1}}
$$

## Combined Linear Congruential Generators

■ Example: For 32-bit computers, L'Ecuyer [1988] suggests combining $k=2$ generators with $m_{1}=2,147,483,563, a_{1}=40,014, m_{2}=$ $2,147,483,399$ and $a_{2}=20,692$. The algorithm becomes:

Step 1: Select seeds

- $X_{1,0}$ in the range $[1,2,147,483,562]$ for the $1^{\text {st }}$ generator
- $X_{2,0}$ in the range [1, 2,147,483,398] for the $2^{\text {nd }}$ generator.

Step 2: For each individual generator,
$X_{1, j+1}=40,014 X_{1, j} \bmod 2,147,483,563$
$X_{2, j+1}=40,692 X_{1, j} \bmod 2,147,483,399$.
Step 3: $X_{j+1}=\left(X_{1, j+1}-X_{2, j+1}\right) \bmod 2,147,483,562$.
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 $j=j+1$, go back to step 2.
$\square$ Combined generator has period: $\left(m_{1}-1\right)\left(m_{2}-1\right) / 2 \sim 2 \times 10^{18}$

## Random-Numbers Streams

- The seed for a linear congruential random-number generator:
$\square$ Is the integer value $X_{0}$ that initializes the random-number sequence.
$\square$ Any value in the sequence can be used to "seed" the generator.
- A random-number stream:
$\square$ Refers to a starting seed taken from the sequence $X_{0}, X_{1}, \ldots, X_{P}$.
$\square$ If the streams are $b$ values apart, then stream $i$ could defined by starting seed: $\quad S_{i}=X_{b(i-1)}$

Older generators: $b=10^{5}$; Newer generators: $b=10^{37}$.

- A single random-number generator with $k$ streams can act like $k$ distinct virtual random-number generators
- To compare two or more alternative systems.
$\square$ Advantageous to dedicate portions of the pseudo-random number sequence to the same purpose in each of the simulated systems.


## Tests for Random Numbers

Two categories:
$\square$ Testing for uniformity:

$$
\begin{array}{ll}
H_{0}: & R_{i} \sim U[0,1] \\
H_{1}: & R_{i} \sim U[0,1]
\end{array}
$$

- Failure to reject the null hypothesis, $\mathrm{H}_{0}$, means that evidence of non-uniformity has not been detected.
$\square$ Testing for independence:
$H_{0}: R_{i} \sim$ independently
$H_{1}: R_{i} \uparrow$ independently
- Failure to reject the null hypothesis, $\mathrm{H}_{0}$, means that evidence of dependence has not been detected.
- Level of significance $\alpha$, the probability of rejecting $\mathrm{H}_{0}$ when it is true:

$$
\alpha=P\left(\text { reject } H_{0} \mid H_{0} \text { is true }\right)
$$

## Tests for Random Numbers

- When to use these tests:
$\square$ If a well-known simulation languages or random-number generators is used, it is probably unnecessary to test
$\square$ 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:
$\square$ Theoretical tests: evaluate the choices of $\mathrm{m}, \mathrm{a}$, and c without actually generating any numbers
$\square$ Empirical tests: applied to actual sequences of numbers produced. Our emphasis.
- Test of uniformity
- Two different methods:
$\square$ Kolmogorov-Smirnov test
$\square$ Chi-square test


## 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.
$\square$ We know:

$$
F(x)=x, \quad 0 \leq x \leq 1
$$If the sample from the RN generator is $R_{1}, R_{2}, \ldots, R_{N}$, then the empirical cdf, $S_{N}(x)$ is:

$$
S_{N}(x)=\frac{\text { number of } R_{1}, R_{2}, \ldots, R_{n} \text { which are } \leq x}{N}
$$

- Based on the statistic: $D=\max \left|F(x)-S_{N}(x)\right|$
$\square$ Sampling distribution of $D$ is known (a function of $N$, tabulated in Table A.8.)
- A more powerful test, recommended.


## Kolmogorov-Smirnov Test

- Example: Suppose 5 generated numbers are $0.44,0.81,0.14$, $0.05,0.93$.

Step 1:

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

Step 3: $D=\max \left(D^{+}, D^{-}\right)=0.26$
Step 4: For $\alpha=0.05$,

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

Hence, $H_{0}$ is not rejected.


## Chi-square test

- Chi-square test uses the sample statistic:


Approximately the chi-square distribution with $n-1$ degrees of freedom (where the critical values are tabulated in Table A.6)
$\square$ For the uniform distribution, $E_{i}$, the expected number in the each class is:

$$
E_{i}=\frac{N}{n}, \quad \text { where } \mathrm{N} \text { is the total \# of observation }
$$

- Valid only for large samples, e.g. $\mathrm{N}>=50$


## Tests for Autocorrelation

- Testing the autocorrelation between every m numbers ( m is a.k.a. the lag), starting with the $i^{\text {th }}$ number
$\square$ The autocorrelation $\rho_{i m}$ between numbers: $R_{i j}, R_{i+m}, R_{i+2 m}$ $R_{i+(M+1) m}$
$M$ is the largest integer such that $i+(M+1) m \leq N$
- Hypothesis:

$$
\begin{array}{ll}
H_{0}: & \rho_{i m}=0, \quad \text { if numbers are independent } \\
H_{1}: & \rho_{i m} \neq 0, \quad \text { if numbers are dependent }
\end{array}
$$

- If the values are uncorrelated:

For large values of M , the distribution of the estimator of $\rho_{i m}$, denoted $\hat{\rho}_{i m}$ is approximately normal.

Test statistics is:

$$
Z_{0}=\frac{\hat{\rho}_{i m}}{\hat{\sigma}_{\hat{\rho}_{i m}}}
$$

$\square Z_{0}$ is distributed normally with mean $=0$ and variance $=1$, and:

$$
\begin{aligned}
& \hat{\rho}_{i m}=\frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+k m} R_{i+(k+1) m}\right]-0.25 \\
& \hat{\sigma}_{\rho_{i m}}=\frac{\sqrt{13 M+7}}{12(M+1)}
\end{aligned}
$$

- If $\rho_{\text {im }}>0$, the subsequence has positive autocorrelation
$\square$ High random numbers tend to be followed by high ones, and vice versa.
- If $\rho_{i m}<0$, the subsequence has negative autocorrelation
$\square$ Low random numbers tend to be followed by high ones, and vice versa.


## Example

- Test whether the $3^{\text {rd }}, 8^{\text {th }}, 13^{\text {th }}$, and so on, for the following output on P. 265.
$\square$ Hence, $\alpha=0.05, i=3, m=5, N=30$, and $\mathrm{M}=4$

$$
\begin{aligned}
\hat{\rho}_{35} & =\frac{1}{4+1}\left[\begin{array}{l}
(0.23)(0.28)+(0.25)(0.33)+(0.33)(0.27) \\
+(0.28)(0.05)+(0.05)(0.36)
\end{array}\right]-0.25 \\
& =-0.1945 \\
\hat{\sigma}_{\rho_{35}} & =\frac{\sqrt{13(4)+7}}{12(4+1)}=0.128 \\
Z_{0} & =-\frac{0.1945}{0.1280}=-1.516
\end{aligned}
$$

From Table A.3, $z_{0.025}=1.96$. Hence, the hypothesis is not rejected.

## Shortcomings

- The test is not very sensitive for small values of M, particularly when the numbers being tests are on the low side.
- Problem when "fishing" for autocorrelation by performing numerous tests:
$\square$ If $\alpha=0.05$, there is a probability of 0.05 of rejecting a true hypothesis.
If 10 independence sequences are examined,
- The probability of finding no significant autocorrelation, by chance alone, is $0.95^{10}=0.60$.
- Hence, the probability of detecting significant autocorrelation when it does not exist $=40 \%$


## Summary

- In this chapter, we described:
$\square$ Generation of random numbers
$\square$ Testing for uniformity and independence
- Caution:
$\square$ Even with generators that have been used for years, some of which still in used, are found to be inadequate.
$\square$ This chapter provides only the basic
$\square$ Also, even if generated numbers pass all the tests, some underlying pattern might have gone undetected.

