Lesson 7. Random Number Generation

1 Overview

- How does a computer program sample independent values from the Uniform$[0,1]$ distribution, e.g. the RAND function in Excel?
- It is very difficult to get a computer to do something completely randomly
  - A computer, by design, follows its instructions blindly, and is therefore completely predictable
  - A computer that doesn't do this is broken!
- One approach: **pseudo-random number generators**

2 Pseudo-random number generators (PRNGs)

- "Psuedo" means having a deceptive resemblance
- PRNGs are (deterministic) algorithms that use mathematical formulas or precalculated tables to produce sequences of numbers that appear random
- Some desirable properties of a PRNG:
  1. Efficient: can produce many numbers in a short time
  2. Deterministic: a given sequence of numbers can be reproduced at a later date if the starting point in the sequence is known
     - Useful for comparing different systems
  3. Long cycle: if the PRNG is periodic (generates a sequence that eventually repeats itself), the cycle length should be sufficiently long
     - Modern PRNGs have a period so long that it can be ignored for most practical purposes
  4. (most important) Pass statistical tests for **uniformity** and **independence**
     - These numbers should not be statistically differentiable from a sequence of truly independently sampled values from the Uniform$[0,1]$ distribution
- Some consequences of uniformity and independence:
  - If the interval $[0,1]$ is divided into n subintervals of equal length, and N values are sampled, then the expected number of values in each interval is N/n
  - The probability of observing a value in a particular interval is independent of the previous values observed

3 The linear congruential method

- Produces sequence of integers $X_1, X_2, \ldots$ using the following recursion:

  $$X_{n+1} = (aX_n + c) \bmod m$$

  - The initial value X_0 is called the **seed**
  - The minimum possible value of $X_1, X_2, \ldots$ is 0
  - The maximum possible value of $X_1, X_2, \ldots$ is $m - 1$
- The **stream**, or the sequence of generated pseudo-random numbers is

  $$\frac{X_1}{m}, \frac{X_2}{m}, \ldots$$

- The modulus is often chosen to be a power of 2: binary computations are fast on a computer
- If $c = 0$, this is a **multiplicative congruential method**
- If $c \neq 0$, this is a **mixed congruential method**

Example 1. In Excel, generate 30 pseudo-random numbers using the linear congruential method with a modulus of $2^4 = 16$, a multiplier of 5, an increment of 3, and a seed of 1.

Note: In Excel, MOD(X,m) computes $X \bmod m$.

Example 2. In Excel, generate 30 pseudo-random numbers using the linear congruential method with a modulus of $2^{31} - 1$, a multiplier of 7^5, an increment of 0, and a seed of 123,457.

This generator was used in the IMSL Scientific Subroutine Package in 1978.