A General-Purpose Uniform Random Number Package

E. Jack Chen
https://orcid.org/0000-0002-3038-8048
BASF Corporation, USA

INTRODUCTION

As computer capacities and simulation (e.g., Monte Carlo method) technologies advance, simulation has become the method of choice for modeling and analysis. The fundamental advantage of simulation is that it can tolerate far less restrictive modeling assumptions, leading to an underlying model that is more reflective of reality and thus more valid, leading to better decisions (Lucas et al., 2015). Simulation studies are typically preceded by transforming in a more or less complicated way of a sequence of numbers between 0 and 1 produced by a pseudorandom generator into an observation of the measure of interest. Random numbers are a fundamental resource in science and technology. A facility for generating sequences of pseudorandom numbers is a fundamental part of computer simulation systems. Furthermore, random number generators also play an important role in cryptography and in the blockchain ecosystem. For example, common cryptosystems employ keys that must be generated in a random fashion. A collection of random variables $x_1,x_2,\ldots,x_n$ is a random sample if they are independent and identically distributed. All samples of the sequence are generated independently of each other, and the value of the next sample in the sequence cannot be predicted, regardless of how many samples have already been produced.

True random numbers cannot be produced by a deterministic algorithm, and hence, random numbers generated by using a recursive equation are referred to as pseudorandom numbers. Pseudorandom numbers seem random, but in fact, they are not random and perfectly predictable. The purpose of these algorithms is to produce sequences of numbers whose behavior is difficult to distinguish from that of their "truly random" counterparts. The deterministic nature of these techniques is important because it can be reproduced effectively in computations. A facility for generating sequences of pseudorandom numbers is a fundamental part of computer simulation systems. Usually, in practice, such a facility produces a deterministic sequence of values, but externally these values should appear to be drawn independently from a uniform distribution between 0 and 1, i.e., they are independent and statistically indistinguishable from a truly random sequence. Furthermore, multiple independent streams of random numbers are often required in simulation studies, for instance, to facilitate synchronization for variance-reduction purposes, and for making independent replications.

A random number generator (RNG) is an algorithm that starting from an initial seed (or seeds), produces a stream of numbers that behaves as if it were a random sample when analyzed using statistical tests, see L'Ecuyer and Simard (2007) and Rukhin et al. (2010). The RNG is closely related to the Deterministic Random Bit Generators (DRBGs). See L'Ecuyer (1990, 2013) and references therein for more information on RNGs. We describe a portable set of software utilities for uniform random-number generation. It provides for multiple generators (streams) running simultaneously, and each generator (stream) has its sequence of numbers partitioned into many long disjoint contiguous substreams, see L'Ecuyer et al. (2002). Simple procedure calls allow the user to make any generator "jump" ahead/back

DOI: 10.4018/978-1-7998-3479-3.ch013

v steps (random numbers). This is more flexible than provides a pre-computed list of seeds that are predetermined steps apart. Implementation issues are discussed. The basic underlying generator CMRG (Combined Multiple Recursive Generator) combines two multiple recursive random number generators with a period length of approximately 2^{191} ($\approx 3.1\times10^{57}$), good speed, and excellent theoretical properties, e.g., the lattice structure, see Kroese et al. (2011) for a list of desired properties.

BACKGROUND

There are several methods for generating the random numbers, of which the most popular are the congruential methods (mixed, multiplicative, and additive). These congruential methods are fast, convenient, and have good enough statistical properties when their parameters are chosen carefully. The (mixed) linear congruential generators (LCGs) are defined by

$$x_i = \left(ax_{i-1} + c\right) \text{MOD } m, u_i = \frac{x_i}{m}, x_0 \in \left\{1,\ldots,m-1\right\}, i > 0.$$

where m (the modulus) is a positive integer (usually a very large primary number), a (the multiplier) $\in\{0,1,\ldots,m-1\}$ and c (the increment) is a nonnegative integer. This mathematical notation signifies that x_i is the remainder of $(ax_{i-1}+c)$ divided by m.. Hence, $x_i\in\{0,1,\ldots,m-1\}$. Thus, random variable u_i is a uniform 0, 1 variable. Note that

$$x_{i+v} = \left(a^v x_i + \frac{c\left(a^v - 1\right)}{a-1} \right) \text{MOD } m.$$

Hence, every x_i is completely determined by *m, a, c,* and x_0. The sequence x_i repeats once it returns to a previously visited value. The period of a generator is the length of a generated steam before it begins to repeat. If $u_0=u_p$ (where $p>0$), then the length p is called the period. The longest possible period for a LCG is m, i.e., m represents the desired number of different values that could be generated for the random numbers. Hence, the modulus m is often taken as a large prime number close to the largest integer directly representable on the computer (i.e., equal or near $2^{31}-1$ for 32- bit computers). If $p=m$, we say that the generator has full period. The required conditions on how to choose *m, a,* and *c* so that the corresponding LCG will have full period are known, see Knuth (1997) or Law (2014).

LCGs are sensitive with respect to the parameters, especially the value of a. When $c>0$, the LCGs are called mixed LCGs. When $c=0$,

$$x_i = ax_{i-1} \text{ MOD } m, u_i = \frac{x_i}{m}, x_0 \in \left\{1,\ldots,m-1\right\}, i > 0.$$

These LCGs are called multiplicative LCGs. Most existing implementations of LCGs are multiplicative LCGs, because in general the value of c does not have a large impact of the quality of an LCG. Note that if $x_i=0$, then all subsequent x_i are identically 0. Thus, the longest possible period for a multiplicative