Fast and Reliable Random Number Generators for Scientific Computing (extended abstract)

Richard P. Brent[1]

Oxford University Computing Laboratory,
Wolfson Building, Parks Road,
Oxford OX1 3QD, UK

random@rpbrent.co.uk

Abstract. Fast and reliable pseudo-random number generators are required for simulation and other applications in Scientific Computing. We outline the requirements for good uniform random number generators, and describe a class of generators having very fast vector/parallel implementations with excellent statistical properties. We also discuss the problem of initialising random number generators, and consider how to combine two or more generators to give a better (though usually slower) generator.

1 Introduction

Monte Carlo and quasi-Monte Carlo methods are of great importance in simulation [20], computational finance, numerical integration, computational physics [7, 13], etc. Due to Moore's Law and increases in parallelism, the statistical quality of random number generators is becoming even more important than in the past. A program running on a supercomputer might use 10^9 random numbers per second over a period of many hours (or even months in some cases), so 10^{16} or more random numbers might contribute to the result. Small correlations or other deficiencies in the random number generator could easily lead to spurious effects and invalidate the results of the computation, see e.g. [7, 19].

Applications require random numbers with various distributions (e.g. normal, exponential, Poisson, ...) but the algorithms used to generate these random numbers almost invariably require a good uniform random number generator. In this paper we consider only the generation of uniformly distributed numbers. Usually we are concerned with *real* numbers u_n that are intended to be uniformly distributed on the interval $[0, 1)$. Sometimes it is convenient to consider *integers* U_n in some range $0 \le U_n < m$. In this case we require $u_n = U_n/m$ to be (approximately) uniformly distributed.

Pseudo-random numbers generated in a deterministic fashion on a digital computer can not be truly random. What is required is that finite segments

This work was supported in part by EPSRC grant GR/N35366. rpb217a

of the sequence $(u_0, u_1, \cdots)$ behave in a manner indistinguishable from a truly random sequence. In practice, this means that they pass all statistical tests that are relevant to the problem at hand. Since the problems to which a library routine will be applied are not known in advance, random number generators in subroutine libraries should pass a number of stringent statistical tests (and not fail any) before being released for general use.

Random numbers generated by physical sources are available [1]. However, there are problems in generating such numbers sufficiently fast, and experience with them is insufficient to be confident of their statistical properties. Thus, for the present, we recommend treating such physical sources of random numbers with caution. They can be used to initialise (and perhaps periodically reinitialise) deterministic generators, and can be combined with deterministic generators by the algorithms considered in §4. For the moment we restrict our attention to deterministic random number generators.

A sequence $(u_0, u_1, \cdots)$ depending on a finite state must eventually be periodic, i.e. there is a positive integer ρ such that $u_{n+\rho} = u_n$ for all sufficiently large n. The minimal such ρ is called the *period*.

In §2 we consider desiderata for random number generators. Then, in §3, we describe one popular class of random number generators. In §4 we discuss how to combine two or more generators to give a (hopefully) better generator. Finally, in §5 we briefly mention implementations.

2 Requirements for good random number generators

Requirements for a good pseudo-random number generator have been discussed in many surveys, e.g. [2, 3, 5, 10, 14]. Here we summarize and comment briefly on the most important requirements.

2.1 Uniformity

The sequence of random numbers should pass statistical tests for uniformity of distribution. This is usually easy for deterministic generators implemented in software. For physical/hardware generators, the well-known technique of Von Neumann, or similar but more efficient techniques [9], can be used to extract uniform bits from a sequence of independent but possibly biased bits.

2.2 Independence

Subsequences of the full sequence $(u_0, u_1, \cdots)$ should be independent. Random numbers are often used to sample a d-dimensional space, so the sequence of d-tuples $(u_{dn}, u_{dn+1}, \ldots, u_{dn+d-1})$ should be uniformly distributed in the d-dimensional cube $[0, 1]^d$ for all "small" values of d (certainly for all $d \leq 6$). For random number generators on parallel machines, the sequences generated on each processor should be independent.

2.3 Long period

As mentioned above, a simulation might use 10^{16} random numbers. In such a case the period ρ must exceed 10^{16}. For many generators there are strong correlations between $u_0, u_1, \cdots$ and $u_m, u_{m+1}, \cdots$, where $m = \rho/2$ (and similarly for other simple fractions of the period). Thus, in practice the period should be *much* larger than the number of random numbers that will ever be used. A good rule of thumb is to use at most $\sqrt{\rho}$ numbers.

2.4 Ability to skip ahead

If a simulation is to be run on a machine with several processors, or if a large simulation is to be performed on several independent machines, it is essential to ensure that the sequences of random numbers used by each processor are disjoint. Two methods of subdivision are commonly used. Suppose, for example, that we require 4 disjoint subsequences for a machine with 4 processors. One processor could use the subsequence $(u_0, u_4, u_8, \cdots)$, another the subsequence $(u_1, u_5, u_9, \cdots)$, etc. For efficiency each processor should be able to "skip over" the terms that it does not require.

Alternatively, processor j could use the subsequence $(u_{m_j}, u_{m_j+1}, \cdots)$, where the indices m_0, m_1, m_2, m_3 are sufficiently widely separated that the (finite) subsequences do not overlap, but this requires some efficient method of generating u_m for large m without generating all the intermediate values $u_1, \ldots, u_{m-1}$.

2.5 Proper initialization

The initialization of random number generators, especially those with a large amount of state information, is an important and often neglected topic. In some applications only a short sequence of random numbers is used after each initialization of the generator, so it is important that short sequences produced with different seeds are uncorrelated.

For example, suppose that a random number generator with seed s produces a sequence $(u_1^{(s)}, u_2^{(s)}, u_3^{(s)}, \ldots)$. If we use m different seeds $s_1, s_2, \ldots, s_m$ and generate n numbers from each seed, we get an $m \times n$ array U with elements $U_{i,j} = u_j^{(s_i)}$. We do not insist that the seeds are random – they could for example be consecutive integers.

Testing packages such as Marsaglia's *Diehard* [15] typically test a 1-D array of random numbers. We can generate a 1-D array by concatenating the rows (or columns) of U. Irrespective of how this is done, we would hope that the random numbers would pass the standard statistical tests. However, many current generators fail because they were intended for the case $m = 1$ (or small) and n large [8, 11]. The other extreme is m large and $n = 1$. In this case we expect $u_1^{(s)}$ to behave like a pseudo-random function of s.

2.6 Unpredictability

In cryptographic applications, it is not sufficient for the sequence to pass standard statistical tests for randomness; it also needs to be *unpredictable* in the sense that there is no efficient deterministic algorithm for predicting u_n (with probability of success significantly greater than 0.5) from $(u_0, u_1, \ldots, u_{n-1})$, unless n is so large that the prediction is infeasible.

At first sight it appears that unpredictability is not required in scientific applications. However, if a random number generator is predictable then we can always devise a statistical test (albeit an artificial one) that the generator will fail. Thus, it seems a wise precaution to use an unpredictable generator if the cost of doing so is not too high. We discuss techniques for this in §4.

Strictly speaking, unpredictability implies uniformity, independence, and a (very) long period. However, it seems worthwhile to state these simpler requirements separately.

2.7 Efficiency

It should be possible to implement the method efficiently so that only a few arithmetic operations are required to generate each random number, all vector/parallel capabilities of the machine are used, and overheads such as those for subroutine calls are minimal. Of course, efficiency tends to conflict with other requirements such as unpredictability, so a tradeoff is often involved.

2.8 Repeatability

For testing and development it is useful to be able to repeat a run with *exactly* the same sequence of random numbers as was used in an earlier run. This is usually easy if the sequence is restarted from the beginning (u_0). It may not be so easy if the sequence is to be restarted from some other value, say u_m for a large integer m, because this requires saving the state information associated with the random number generator.

2.9 Portability

Again, for testing and development purposes, it is useful to be able to generate *exactly* the same sequence of random numbers on two different machines, possibly with different wordlengths.

3 Generalized Fibonacci generators

Given a circular buffer of length r words (or bits), we can generate pseudo-random numbers from a linear or nonlinear recurrence

$$u_n = f(u_{n-1}, u_{n-2}, \ldots, u_{n-r}) \; .$$

For speed it is desirable that $f(u_{n-1}, u_{n-2}, \ldots, u_{n-r})$ depends explicitly on only a small number of its r arguments. An important case is the class of "generalized Fibonacci" or "lagged Fibonacci" random number generators [10].

Marsaglia [14] considers generators $F(r, s, \theta)$ that satisfy

$$U_n = U_{n-r} \; \theta \; U_{n-s} \mod m$$

for fixed "lags" r and s ($r > s > 0$) and $n \geq r$. Here m is a modulus (typically 2^w if w is the wordlength in bits), and θ is some binary operator, e.g. addition, subtraction, multiplication or "exclusive or". We abbreviate these operators by $+, -, *$ and $\oplus$ respectively. Generators using $\oplus$ are also called "linear feedback shift register" (LFSR) generators or "Tausworthe" generators. Usually U_n is normalised to give a floating-point number $u_n = U_n/m \in [0, 1)$.

It is possible to choose lags r, s so that the period ρ of the generalized Fibonacci generators $F(r, s, +)$ is a large prime p or a small multiple of such a prime. Typically, the period of the least-significant bit is p; because carries propagate from the least-significant bit into higher-order bits, the overall period is usually $2^{w-1}\rho$ for wordlength w. For example, [3, Table 1] gives several pairs (r, s) with $r > 10^6$. (The notation in [3] is different: $r + \delta$ corresponds to our r.)

There are several ways to improve the performance of generalized Fibonacci generators on statistical tests such as the Birthday Spacings and Generalized Triple tests [14]. The simplest is to include small odd integer multipliers α and β in the generalized Fibonacci recurrence, e.g.

$$U_n = \alpha U_{n-r} + \beta U_{n-s} \mod m \; .$$

Other ways to improve statistical properties (at the expense of speed) are to include more terms in the linear recurrence [12], to discard some members of the sequence [13], or to combine two or three generators in various ways (see §4).

With suitable choice of lags (r, s), the generalised Fibonacci generators satisfy the requirements of uniformity, independence, long period, efficiency, and ability to skip ahead. They do *not* satisfy the requirement for unpredictability. In the following section we show how to overcome this difficulty.

4 Improving generators

In this section we consider how generators that suffer some defects can be improved.

4.1 Improving a generator by "decimation"

If $(x_0, x_1, \ldots)$ is generated by a 3-term recurrence, we can obtain a (hopefully better) sequence $(y_0, y_1, \ldots)$ by defining $y_j = x_{jp}$, where $p > 1$ is a suitable constant. In other words, use every p-th number and discard the others.

Consider the case $F(r, s, \oplus)$ with $w = 1$ (LFSR) and $p = 3$. (If $p = 2$, the y_j satisfy the same 3-term recurrence as the x_j.)

Using generating functions, it is easy to show that the y_j satisfy a 5-term recurrence. For example, if $x_n = x_{n-1} \oplus x_{n-127}$, then $y_n = y_{n-1} \oplus y_{n-43} \oplus y_{n-85} \oplus y_{n-127}$. A more elementary approach is given in [21].

A possible improvement over simple decimation is decimation by blocks [13].

Better than regular decimation is "irregular decimation" (§4.4).

4.2 Combining generators by addition

We can combine some number K of generalized Fibonacci generators by addition (mod 2^w). If each component generator is defined by a primitive trinomial $T_k(x) = x^{r_k} + x^{s_k} + 1$, with distinct prime degrees r_k, then the combined generator has period at least $2^{w-1} \prod_{k=1}^{K} (2^{r_k} - 1)$ and satisfies a 3^K-term linear recurrence.

Because the speed of the combined generator decreases like $1/K$, we would probably take $K \leq 3$ in practice. The case $K = 2$ seems to be better (and more efficient) than "decimation" with $p = 3$.

4.3 Combining by shuffling

Suppose we have two pseudo-random sequences $X = (x_0, x_1, \ldots)$ and $Y = (y_0, y_1, \ldots)$. We can use a buffer V of size B say, fill the buffer using the sequence X, then use the sequence Y to generate indices into the buffer. If the index is j then the random number generator returns $V[j]$ and replaces $V[j]$ by the next number in the X sequence [10, Algorithm M].

In other words, we use one generator to shuffle the output of another generator. This seems to be as good (and about as fast) as combining two generators by addition. B should not be too small.

4.4 Combining by shrinking

Coppersmith *et al* [6] suggested using one sequence to "shrink" another sequence.

Suppose we have two pseudo-random sequences $(x_0, x_1, \ldots)$ and $(y_0, y_1, \ldots)$, where $y_i \in \mathrm{GF}(2)$. Suppose $y_i = 1$ for $i = i_0, i_1, \ldots$ Define a sequence $(z_0, z_1, \ldots)$ to be the subsequence $(x_{i_0}, x_{i_1}, \ldots)$ of $(x_0, x_1, \ldots)$. In other words, one sequence of bits (y_i) is used to decide whether to "accept" or "reject" elements of another sequence (x_i). This is sometimes called "irregular decimation" (compare §4.1).

Combining two sequences by shrinking is slower than combining the sequences by $\oplus$, but is less amenable to analysis based on linear algebra or generating functions, so is preferable in applications where the sequence needs to be unpredictable. In the final version of the paper we shall consider whether x_i should be a single bit (as originally proposed) or whether it is safe to take x_i as a byte or word (faster but with a possible loss of unpredictability).

5 Implementations

In the final version of the paper we shall comment on some implementations of random number generators.

References

1. Anonymous, *Random number generation and testing*, NIST, December 2000. `http://csrc.nist.gov/rng/` .
2. R. P. Brent, Random number generation and simulation on vector and parallel computers, *LNCS* **1470**, Springer-Verlag, Berlin, 1998, 1–20.
3. R. P. Brent and P. Zimmermann, Random number generators with period divisible by a Mersenne prime, *LNCS* **2667**, Springer-Verlag, Berlin, 2003, 1–10. Preprint available at `http://www.comlab.ox.ac.uk/oucl/work/richard.brent/pub/pub211.html` .
4. R. P. Brent and P. Zimmermann, Algorithms for finding almost irreducible and almost primitive trinomials, in *Primes and Misdemeanours: Lectures in Honour of the Sixtieth Birthday of Hugh Cowie Williams*, Fields Institute, Toronto, to appear. Preprint available at `···/pub212.html` .
5. P. D. Coddington, Random number generators for parallel computers, *The NHSE Review* **2** (1996). `http://nhse.cs.rice.edu/NHSEreview/RNG/PRNGreview.ps` .
6. D. Coppersmith, H. Krawczyk and Y. Mansour, The shrinking generator, *Advances in Crpyptology – CRYPTO'93, LNCS* **773**, Springer-Verlag, Berlin, 1994, 22–39.
7. A. M. Ferrenberg, D. P. Landau and Y. J. Wong, Monte Carlo simulations: hidden errors from "good" random number generators, *Phys. Review Letters* **69** (1992), 3382–3384.
8. P. Gimeno, *Problem with* `ran_array`, personal communication, 10 Sept. 2001.
9. A. Juels, M. Jakobsson, E. Shriver and B. K. Hillyer, How to turn loaded dice into fair coins, *IEEE Trans. on Information Theory* **46**, 2000, 911–921.
10. D. E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms* (third edition), Addison-Wesley, Menlo Park, CA, 1998.
11. D. E. Knuth, *A better random number generator*, January 2002, `http://www-cs-faculty.stanford.edu/~knuth/news02.html` .
12. T. Kumada, H. Leeb, Y. Kurita and M. Matsumoto, New primitive t-nomials ($t = 3, 5$) over GF(2) whose degree is a Mersenne exponent, *Math. Comp.* **69** (2000), 811–814. Corrigenda: *ibid* **71** (2002), 1337–1338.
13. M. Lüscher, A portable high-quality random number generator for lattice field theory simulations, *Computer Physics Communications* **79** (1994), 100–110.
14. G. Marsaglia, A current view of random number generators, in *Computer Science and Statistics: The Interface*, Elsevier Science Publishers B. V.,1985, 3–10.
15. G. Marsaglia, *Diehard*, 1995. Available from `http://stat.fsu.edu/~geo/` .
16. M. Mascagni, M. L. Robinson, D. V. Pryor and S. A. Cuccaro, Parallel pseudorandom number generation using additive lagged-Fibonacci recursions, *Lecture Notes in Statistics* **106**, Springer-Verlag, Berlin, 1995, 263–277.
17. A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, New York, 1997. `http://cacr.math.uwaterloo.ca/hac/` .
18. W. P. Petersen, Lagged Fibonacci series random number generators for the NEC SX-3, *Internat. J. High Speed Computing* **6** (1994), 387–398.
19. L. N. Shchur, J. R. Heringa and H. W. J. Blöte, Simulation of a directed random-walk model: the effect of pseudo-random-number correlations, *Physica A* **241** (1997), 579.
20. I. Vattulainen, T. Ala-Nissila and K. Kankaala, Physical tests for random numbers in simulations, *Phys. Review Letters* **73** (1994), 2513–2516.
21. R. M. Ziff, Four-tap shift-register-sequence random-number generators, *Computers in Physics* **12** (1998), 385–392.