

# NAG Library Chapter Introduction

## G05 – Random Number Generators

### Contents

<b>1</b>	<b>Scope of the Chapter</b> .....	2
<b>2</b>	<b>Background to the Problems</b> .....	2
2.1	Pseudorandom Numbers .....	2
2.1.1	NAG Basic Generator .....	2
2.1.2	Wichmann–Hill I Generator .....	3
2.1.3	Wichmann–Hill II Generator .....	3
2.1.4	Mersenne Twister Generator .....	4
2.1.5	ACORN Generator .....	4
2.1.6	L’Ecuyer MRG32k3a Combined Recursive Generator .....	5
2.2	Quasi-random Numbers .....	5
2.3	Scrambled Quasi-random Numbers .....	6
2.4	Non-uniform Random Numbers .....	6
2.5	Copulas .....	6
2.6	Other Random Structures .....	7
2.7	Multiple Streams of Pseudorandom Numbers .....	7
2.7.1	Multiple Streams via Different Initial Values (Seeds) .....	7
2.7.2	Multiple Streams via Different Generators .....	7
2.7.3	Multiple Streams via Skip-ahead .....	7
2.7.4	Multiple Streams via Leap-frog .....	8
2.7.5	Skip-ahead and Leap-frog for a Linear Congruential Generator (LCG): An Example .....	8
2.7.6	Skip-ahead and Leap-frog for the Mersenne Twister: An Example .....	9
<b>3</b>	<b>Recommendations on Choice and Use of Available Routines</b> .....	<b>10</b>
3.1	Pseudorandom Numbers .....	10
3.1.1	Initialization .....	10
3.1.2	Repeated initialization .....	10
3.1.3	Choice of Base Generator .....	10
3.1.4	Choice of Method for Generating Multiple Streams .....	10
3.1.5	Copulas .....	11
3.2	Quasi-random Numbers .....	11
3.3	Programming Advice .....	11
<b>4</b>	<b>Functionality Index</b> .....	<b>11</b>
<b>5</b>	<b>Auxiliary Routines Associated with Library Routine Parameters</b> .....	<b>13</b>
<b>6</b>	<b>Routines Withdrawn or Scheduled for Withdrawal</b> .....	<b>13</b>
<b>7</b>	<b>References</b> .....	<b>15</b>

## 1 Scope of the Chapter

This chapter is concerned with the generation of sequences of independent pseudorandom and quasi-random numbers from various distributions, and the generation of pseudorandom time series from specified time series models.

## 2 Background to the Problems

### 2.1 Pseudorandom Numbers

A sequence of pseudorandom numbers is a sequence of numbers generated in some systematic way such that they are independent and statistically indistinguishable from a truly random sequence. A pseudorandom number generator (PRNG) is a mathematical algorithm that, given an initial state, produces a sequence of pseudorandom numbers. A PRNG has several advantages over a true random number generator in that the generated sequence is repeatable, has known mathematical properties and can be implemented without needing any specialist hardware. Many books on statistics and computer science have good introductions to PRNGs, for example Knuth (1981) or Banks (1998).

PRNGs can be split into base generators, and distributional generators. Within the context of this document a base generator is defined as a PRNG that produces a sequence (or stream) of variates (or values) uniformly distributed over the interval  $(0, 1)$ . Depending on the algorithm being considered, this interval may be open, closed or half-closed. A distribution generator is a routine that takes variates generated from a base generator and transforms them into variates from a specified distribution, for example a uniform, Gaussian (Normal) or gamma distribution.

The period (or cycle length) of a base generator is defined as the maximum number of values that can be generated before the sequence starts to repeat. The initial state of the base generator is often called the seed.

There are six base generators currently available in the NAG Library, these are; a basic linear congruential generator (LCG) (referred to as the NAG basic generator) (see Knuth (1981)), two sets of Wichmann–Hill generators (see Maclaren (1989) and Wichmann and Hill (2006)), the Mersenne Twister (see Matsumoto and Nishimura (1998)), the ACORN generator (see Wikramaratna (1989)) and L'Ecuyer generator (see L'Ecuyer and Simard (2002)).

#### 2.1.1 NAG Basic Generator

The NAG basic generator is a linear congruential generator (LCG) and, like all linear congruential generators, has the form:

$$\begin{aligned}x_i &= a_1 x_{i-1} \bmod m_1, \\ u_i &= \frac{x_i}{m_1},\end{aligned}$$

where the  $u_i$ , for  $i = 1, 2, \dots$ , form the required sequence.

The NAG basic generator uses  $a_1 = 13^{13}$  and  $m_1 = 2^{59}$ , which gives a period of approximately  $2^{57}$ .

This generator has been part of the NAG Library since Mark 6 and as such has been widely used. It suffers from no known problems, other than those due to the lattice structure inherent in all linear congruential generators, and, even though the period is relatively short compared to many of the newer generators, it is sufficiently large for many practical problems.

The performance of the NAG basic generator has been analysed by the Spectral Test, see Section 3.3.4 of Knuth (1981), yielding the following results in the notation of Knuth (1981).

$n$	$\nu_n$	Upper bound for $\nu_n$
2	$3.44 \times 10^8$	$4.08 \times 10^8$
3	$4.29 \times 10^5$	$5.88 \times 10^5$
4	$1.72 \times 10^4$	$2.32 \times 10^4$
5	$1.92 \times 10^3$	$3.33 \times 10^3$
6	593	939
7	198	380
8	108	197

9

67

120

The right-hand column gives an upper bound for the values of  $\nu_n$  attainable by any multiplicative congruential generator working modulo  $2^{59}$ .

An informal interpretation of the quantities  $\nu_n$  is that consecutive  $n$ -tuples are statistically uncorrelated to an accuracy of  $1/\nu_n$ . This is a theoretical result; in practice the degree of randomness is usually much greater than the above figures might support. More details are given in Knuth (1981), and in the references cited therein.

Note that the achievable accuracy drops rapidly as the number of dimensions increases. This is a property of all multiplicative congruential generators and is the reason why very long periods are needed even for samples of only a few random numbers.

### 2.1.2 Wichmann–Hill I Generator

This series of Wichmann–Hill base generators (see Maclaren (1989)) use a combination of four linear congruential generators and has the form:

$$\begin{aligned} w_i &= a_1 w_{i-1} \bmod m_1 \\ x_i &= a_2 x_{i-1} \bmod m_2 \\ y_i &= a_3 y_{i-1} \bmod m_3 \\ z_i &= a_4 z_{i-1} \bmod m_4 \\ u_i &= \left( \frac{w_i}{m_1} + \frac{x_i}{m_2} + \frac{y_i}{m_3} + \frac{z_i}{m_4} \right) \bmod 1, \end{aligned} \quad (1)$$

where the  $u_i$ , for  $i = 1, 2, \dots$ , form the required sequence. The NAG Library implementation includes 273 sets of parameters,  $a_j, m_j$ , for  $j = 1, 2, 3, 4$ , to choose from.

The constants  $a_i$  are in the range 112 to 127 and the constants  $m_j$  are prime numbers in the range 16718909 to 16776971, which are close to  $2^{24} = 16777216$ . These constants have been chosen so that each of the resulting 273 generators are essentially independent, all calculations can be carried out in 32-bit integer arithmetic and the generators give good results with the spectral test, see Knuth (1981) and Maclaren (1989). The period of each of these generators would be at least  $2^{92}$  if it were not for common factors between  $(m_1 - 1)$ ,  $(m_2 - 1)$ ,  $(m_3 - 1)$  and  $(m_4 - 1)$ . However, each generator should still have a period of at least  $2^{80}$ . Further discussion of the properties of these generators is given in Maclaren (1989).

### 2.1.3 Wichmann–Hill II Generator

This Wichmann–Hill base generator (see Wichmann and Hill (2006)) is of the same form as that described in Section 2.1.2, i.e., a combination of four linear congruential generators. In this case  $a_1 = 11600$ ,  $m_1 = 2147483579$ ,  $a_2 = 47003$ ,  $m_2 = 2147483543$ ,  $a_3 = 23000$ ,  $m_3 = 2147483423$ ,  $a_4 = 33000$ ,  $m_4 = 2147483123$ .

Unlike in the original Wichmann–Hill generator, these values are too large to carry out the calculations detailed in (1) using 32-bit integer arithmetic, however, if

$$w_i = 11600 w_{i-1} \bmod 2147483579$$

then setting

$$W_i = 11600(w_{i-1} \bmod 185127) - 10379(w_{i-1}/185127)$$

gives

$$w_i = \begin{cases} W_i & \text{if } W_i \geq 0 \\ 2147483579 + W_i & \text{otherwise} \end{cases}$$

and  $W_i$  can be calculated in 32-bit integer arithmetic. Similar expressions exist for  $x_i$ ,  $y_i$  and  $z_i$ . The period of this generator is approximately  $2^{121}$ .

Further details of implementing this algorithm and its properties are given in Wichmann and Hill (2006). This paper also gives some useful guidelines on testing PRNGs.

### 2.1.4 Mersenne Twister Generator

The Mersenne Twister (see Matsumoto and Nishimura (1998)) is a twisted generalized feedback shift register generator. The algorithm underlying the Mersenne Twister is as follows:

- (i) Set some arbitrary initial values  $x_1, x_2, \dots, x_r$ , each consisting of  $w$  bits.
- (ii) Letting

$$A = \begin{pmatrix} 0 & I_{w-1} \\ a_w & a_{w-1} \cdots a_1 \end{pmatrix},$$

where  $I_{w-1}$  is the  $(w-1) \times (w-1)$  identity matrix and each of the  $a_i, i = 1$  to  $w$  take a value of either 0 or 1 (i.e., they can be represented as bits). Define

$$x_{i+r} = \left( x_{i+s} \oplus \left( x_i^{(\omega:(l+1))} | x_{i+1}^{(l:1)} \right) A \right),$$

where  $x_i^{(\omega:(l+1))} | x_{i+1}^{(l:1)}$  indicates the concatenation of the most significant (upper)  $w-l$  bits of  $x_i$  and the least significant (lower)  $l$  bits of  $x_{i+1}$ .

- (iii) Perform the following operations sequentially:

$$\begin{aligned} z &= x_{i+r} \oplus (x_{i+r} \ggg t_1) \\ z &= z \oplus ((z \lll t_2) \text{ AND } m_1) \\ z &= z \oplus ((z \lll t_3) \text{ AND } m_2) \\ z &= z \oplus (z \ggg t_4) \\ u_{i+r} &= z / (2^w - 1), \end{aligned}$$

where  $t_1, t_2, t_3$  and  $t_4$  are integers and  $m_1$  and  $m_2$  are bit-masks and ' $\ggg t$ ' and ' $\lll t$ ' represent a  $t$  bit shift right and left respectively,  $\oplus$  is bit-wise exclusively or (xor) operation and 'AND' is a bit-wise and operation.

The  $u_{i+r}$ , for  $i = 1, 2, \dots$ , form the required sequence. The supplied implementation of the Mersenne Twister uses the following values for the algorithmic constants:

$$\begin{aligned} w &= 32 \\ a &= 0x9908b0df \\ l &= 31 \\ r &= 624 \\ s &= 397 \\ t_1 &= 11 \\ t_2 &= 7 \\ t_3 &= 15 \\ t_4 &= 18 \\ m_1 &= 0x9d2c5680 \\ m_2 &= 0xefc60000 \end{aligned}$$

where the notation  $0xDD\dots$  indicates the bit pattern of the integer whose hexadecimal representation is  $DD\dots$ .

This algorithm has a period length of approximately  $2^{19,937} - 1$  and has been shown to be uniformly distributed in 623 dimensions (see Matsumoto and Nishimura (1998)).

### 2.1.5 ACORN Generator

The ACORN generator is a special case of a multiple recursive generator (see Wikramaratna (1989) and Wikramaratna (2007)). The algorithm underlying ACORN is as follows:

- (i) Choose an integer value  $k \geq 1$ .
- (ii) Choose an integer value  $M$ , and an integer seed  $Y_0^{(0)}$ , such that  $0 < Y_0^{(0)} < M$  and  $Y_0^{(0)}$  and  $M$  are relatively prime.
- (iii) Choose an arbitrary set of  $k$  initial integer values,  $Y_0^{(1)}, Y_0^{(2)}, \dots, Y_0^{(k)}$ , such that  $0 \leq Y_0^{(m)} < M$ , for all  $m = 1, 2, \dots, k$ .

(iv) Perform the following sequentially:

$$Y_i^{(m)} = \left( Y_i^{(m-1)} + Y_{i-1}^{(m)} \right) \bmod M$$

for  $m = 1, 2, \dots, k$ .

(v) Set  $u_i = Y_i^{(k)}/M$ .

The  $u_i$ , for  $i = 1, 2, \dots$ , then form a pseudorandom sequence, with  $u_i \in [0, 1)$ , for all  $i$ .

Although you can choose any value for  $k$ ,  $M$ ,  $Y_0^{(0)}$  and the  $Y_0^{(m)}$ , within the constraints mentioned in (i) to (iii) above, it is recommended that  $k \geq 10$ ,  $M$  is chosen to be a large power of two with  $M \geq 2^{60}$  and  $Y_0^{(0)}$  is chosen to be odd.

The period of the ACORN generator, with the modulus  $M$  equal to a power of two, and an odd value for  $Y_0^{(0)}$  has been shown to be an integer multiple of  $M$  (see Wikramaratna (1992)). Therefore, increasing  $M$  will give a series with a longer period.

### 2.1.6 L'Ecuyer MRG32k3a Combined Recursive Generator

The base generator L'Ecuyer MRG32k3a (see L'Ecuyer and Simard (2002)) combines two multiple recursive generators:

$$\begin{aligned} x_i &= a_{11}x_{i-1} + a_{12}x_{i-2} + a_{13}x_{i-3} \bmod m_1 \\ y_i &= a_{21}y_{i-1} + a_{22}y_{i-2} + a_{23}y_{i-3} \bmod m_2 \\ z_i &= x_i - y_i \bmod m_1 \\ u_i &= (z_i + 1)/d \end{aligned}$$

where  $a_{11} = 0$ ,  $a_{12} = 1403580$ ,  $a_{13} = -810728$ ,  $m_1 = 2^{32} - 209$ ,  $a_{21} = 527612$ ,  $a_{22} = 0$ ,  $a_{23} = -1370589$ ,  $m_2 = 2^{32} - 22853$ , and  $u_i, i = 1, 2, \dots$  form the required sequence. If  $d = m_1$  then  $u_i \in (0, 1]$  else if  $d = m_1 + 1$  then  $u_i \in (0, 1)$ . Combining the two multiple recursive generators (MRG) results in sequences with better statistical properties in high dimensions and longer periods compared with those generated from a single MRG. The combined generator described above has a period length of approximately  $2^{191}$ .

## 2.2 Quasi-random Numbers

Low discrepancy (quasi-random) sequences are used in numerical integration, simulation and optimization. Like pseudorandom numbers they are uniformly distributed but they are not statistically independent, rather they are designed to give more even distribution in multidimensional space (uniformity). Therefore they are often more efficient than pseudorandom numbers in multidimensional Monte-Carlo methods.

The quasi-random number generators implemented in this chapter generate a set of points  $x^1, x^2, \dots, x^N$  with high uniformity in the  $S$ -dimensional unit cube  $I^S = [0, 1]^S$ . One measure of the uniformity is the discrepancy which is defined as follows:

Given a set of points  $x^1, x^2, \dots, x^N \in I^S$  and a subset  $G \subset I^S$ , define the counting function  $S_N(G)$  as the number of points  $x^i \in G$ . For each  $x = (x_1, x_2, \dots, x_S) \in I^S$ , let  $G_x$  be the rectangular  $S$ -dimensional region

$$G_x = [0, x_1) \times [0, x_2) \times \dots \times [0, x_S)$$

with volume  $x_1, x_2, \dots, x_S$ . Then the discrepancy of the points  $x^1, x^2, \dots, x^N$  is

$$D_N^*(x^1, x^2, \dots, x^N) = \sup_{x \in I^S} \left| S_N(G_x) - N \prod_{k=1}^S x_k \right|.$$

The discrepancy of the first  $N$  terms of such a sequence has the form

$$D_N^*(x^1, x^2, \dots, x^N) \leq C_S (\log N)^S + O\left((\log N)^{S-1}\right) \quad \text{for all } N \geq 2.$$

The principal aim in the construction of low-discrepancy sequences is to find sequences of points in  $I^S$  with a bound of this form where the constant  $C_S$  is as small as possible.

Three types of low-discrepancy sequences are supplied in this library, these are due to Sobol, Faure and Niederreiter. Two sets of Sobol sequences are supplied, the first is based on work of Joe and Kuo (2008) and the second on the work of Bratley and Fox (1988). More information on quasi-random number generation and the Sobol, Faure and Niederreiter sequences in particular can be found in Bratley and Fox (1988) and Fox (1986).

### 2.3 Scrambled Quasi-random Numbers

Scrambled quasi-random sequences are an extension of standard quasi-random sequences that attempt to eliminate the bias inherent in a quasi-random sequence whilst retaining the low-discrepancy properties. The use of a scrambled sequence allows error estimation of Monte–Carlo results by performing a number of iterates and computing the variance of the results.

This implementation of scrambled quasi-random sequences is based on TOMS algorithm 823 and details can be found in the accompanying paper, Hong and Hickernell (2003). Three methods of scrambling are supplied; the first a restricted form of Owen’s scrambling (Owen (1995)), the second based on the method of Faure and Tezuka (2000) and the last method combines the first two.

Scrambled versions of both Sobol sequences and the Niederreiter sequence can be obtained.

The efficiency of a simulation exercise may often be increased by the use of variance reduction methods (see Morgan (1984)). It is also worth considering whether a simulation is the best approach to solving the problem. For example, low-dimensional integrals are usually more efficiently calculated by routines in Chapter D01 rather than by Monte–Carlo integration.

### 2.4 Non-uniform Random Numbers

Random numbers from other distributions may be obtained from the uniform random numbers by the use of transformations and rejection techniques, and for discrete distributions, by table based methods.

#### (a) Transformation Methods

For a continuous random variable, if the cumulative distribution function (CDF) is  $F(x)$  then for a uniform  $(0, 1)$  random variate  $u$ ,  $y = F^{-1}(u)$  will have CDF  $F(x)$ . This method is only efficient in a few simple cases such as the exponential distribution with mean  $\mu$ , in which case  $F^{-1}(u) = -\mu \log u$ . Other transformations are based on the joint distribution of several random variables. In the bivariate case, if  $v$  and  $w$  are random variates there may be a function  $g$  such that  $y = g(v, w)$  has the required distribution; for example, the Student’s  $t$ -distribution with  $n$  degrees of freedom in which  $v$  has a Normal distribution,  $w$  has a gamma distribution and  $g(v, w) = v\sqrt{n/w}$ .

#### (b) Rejection Methods

Rejection techniques are based on the ability to easily generate random numbers from a distribution (called the envelope) similar to the distribution required. The value from the envelope distribution is then accepted as a random number from the required distribution with a certain probability; otherwise, it is rejected and a new number is generated from the envelope distribution.

#### (c) Table Search Methods

For discrete distributions, if the cumulative probabilities,  $P_i = \text{Prob}(x \leq i)$ , are stored in a table then, given  $u$  from a uniform  $(0, 1)$  distribution, the table is searched for  $i$  such that  $P_{i-1} < u \leq P_i$ . The returned value  $i$  will have the required distribution. The table searching can be made faster by means of an index, see Ripley (1987). The effort required to set up the table and its index may be considerable, but the methods are very efficient when many values are needed from the same distribution.

### 2.5 Copulas

A copula is a function that links the univariate marginal distributions with their multivariate distribution. Sklar’s theorem (see Sklar (1973)) states that if  $f$  is an  $m$ -dimensional distribution function with

continuous margins  $f_1, f_2, \dots, f_m$ , then  $f$  has a unique copula representation,  $c$ , such that

$$f(x_1, x_2, \dots, x_m) = c(f_1(x_1), f_2(x_2), \dots, f_m(x_m))$$

The copula,  $c$ , is a multivariate uniform distribution whose dependence structure is defined by the dependence structure of the multivariate distribution  $f$ , with

$$c(u_1, u_2, \dots, u_m) = f(f_1^{-1}(u_1), f_2^{-1}(u_2), \dots, f_m^{-1}(u_m))$$

where  $u_i \in [0, 1]$ . This relationship can be used to simulate variates from distributions defined by the dependence structure of one distribution and each of the marginal distributions given by another. For additional information see Nelsen (1998) or Boye (Unpublished manuscript) and the references therein.

## 2.6 Other Random Structures

In addition to random numbers from various distributions, random compound structures can be generated. These include random time series, random matrices and random samples.

## 2.7 Multiple Streams of Pseudorandom Numbers

It is often advantageous to be able to generate variates from multiple, independent, streams (or sequences) of random variates. For example when running a simulation in parallel on several processors. There are four ways of generating multiple streams using the routines available in this chapter:

- (i) using different initial values (seeds);
- (ii) using different generators;
- (iii) skip ahead (also called block-splitting);
- (iv) leap-frogging.

### 2.7.1 Multiple Streams via Different Initial Values (Seeds)

A different sequence of variates can be generated from the same base generator by initializing the generator using a different set of seeds. The statistical properties of the base generators are only guaranteed within, not between sequences. For example, two sequences generated from two different starting points may overlap if these initial values are not far enough apart. The potential for overlapping sequences is reduced if the period of the generator being used is large. In general, of the four methods for creating multiple streams described here, this is the least satisfactory.

The one exception to this is the Wichmann–Hill II generator. The Wichmann and Hill (2006) paper describes a method of generating blocks of variates, with lengths up to  $2^{90}$ , by fixing the first three seed values of the generator ( $w_0$ ,  $x_0$  and  $y_0$ ), and setting  $z_0$  to a different value for each stream required. This is similar to the skip-ahead method described in Section 2.7.3, in that the full sequence of the Wichmann–Hill II generator is split into a number of different blocks, in this case with a fixed length of  $2^{90}$ . But without the computationally intensive initialization usually required for the skip-ahead method.

### 2.7.2 Multiple Streams via Different Generators

Independent sequences of variates can be generated using a different base generator for each sequence. For example, sequence 1 can be generated using the NAG basic generator, sequence 2 using Mersenne Twister, sequence 3 the ACORN generator and sequence 4 using L’Ecuyer generator. The Wichmann–Hill I generator implemented in this chapter is, in fact, a series of 273 independent generators. The particular sub-generator to use is selected using the SUBID variable. Therefore, in total, 277 independent streams can be generated with each using a different generator (273 Wichmann–Hill I generators, and 4 additional base generators).

### 2.7.3 Multiple Streams via Skip-ahead

Independent sequences of variates can be generated from a single base generator through the use of block-splitting, or skipping-ahead. This method consists of splitting the sequence into  $k$  non-overlapping blocks, each of length  $n$ , where  $n$  is no smaller than the maximum number of variates required from any of the sequences. For example,

$$\frac{x_1, x_2, \dots, x_n}{\text{block 1}}, \frac{x_{n+1}, x_{n+2}, \dots, x_{2n}}{\text{block 2}}, \frac{x_{2n+1}, x_{2n+2}, \dots, x_{3n}}{\text{block 3}}, \text{etc.}$$

where  $x_1, x_2, \dots$  is the sequence produced by the generator of interest. Each of the  $k$  blocks provide an independent sequence.

The skip-ahead algorithm therefore requires the sequence to be advanced a large number of places, as to generate values from say, block  $b$ , you must skip over the  $(b-1)n$  values in the first  $b-1$  blocks. Due to their form this can be done efficiently for linear congruential generators and multiple congruential generators. A skip-ahead algorithm is also provided for the Mersenne Twister generator.

Although skip-ahead requires some additional computation at the initialization stage (to ‘fast forward’ the sequence) no additional computation is required at the generation stage.

This method of producing multiple streams can also be used for the Sobol and Niederreiter quasi-random number generator via the parameter ISKIP in G05YLF.

### 2.7.4 Multiple Streams via Leap-frog

Independent sequences of variates can also be generated from a single base generator through the use of leap-frogging. This method involves splitting the sequence from a single generator into  $k$  disjoint subsequences. For example:

$$\begin{aligned} \text{Subsequence 1: } & x_1, x_{k+1}, x_{2k+1}, \dots \\ \text{Subsequence 2: } & x_2, x_{k+2}, x_{2k+2}, \dots \\ & \vdots \\ \text{Subsequence } k: & x_k, x_{2k}, x_{3k}, \dots \end{aligned}$$

where  $x_1, x_2, \dots$  is the sequence produced by the generator of interest. Each of the  $k$  subsequences then provides an independent stream of variates.

The leap-frog algorithm therefore requires the generation of every  $k$ th variate from the base generator. Due to their form this can be done efficiently for linear congruential generators and multiple congruential generators. A leap-frog algorithm is provided for the NAG Basic generator, both the Wichmann–Hill I and Wichmann–Hill II generators and L’Ecuyer generator.

It is known that, dependent on the number of streams required, leap-frogging can lead to sequences with poor statistical properties, especially when applied to linear congruential generators. In addition, leap-frogging can increase the time required to generate each variate. Therefore leap-frogging should be avoided unless absolutely necessary.

### 2.7.5 Skip-ahead and Leap-frog for a Linear Congruential Generator (LCG): An Example

As an illustrative example, a brief description of the algebra behind the implementation of the leap-frog and skip-ahead algorithms for a linear congruential generator is given. A linear congruential generator has the form  $x_{i+1} = a_1 x_i \bmod m_1$ . The recursive nature of a linear congruential generator means that

$$\begin{aligned} x_{i+v} &= a_1 x_{i+v-1} \bmod m_1 \\ &= a_1 (a_1 x_{i+v-2} \bmod m_1) \bmod m_1 \\ &= a_1^2 x_{i+v-2} \bmod m_1 \\ &= a_1^v x_i \bmod m_1. \end{aligned}$$

The sequence can therefore be quickly advanced  $v$  places by multiplying the current state ( $x_i$ ) by  $a_1^v \bmod m_1$ , hence skipping the sequence ahead. Leap-frogging can be implemented by using  $a_1^k$ , where  $k$  is the number of streams required, in place of  $a_1$  in the standard linear congruential generator recursive formula, in order to advance  $k$  places, rather than one, at each iteration.

In a linear congruential generator the multiplier  $a_1$  is constructed so that the generator has good statistical properties in, for example, the spectral test. When using leap-frogging to construct multiple streams this multiplier is replaced with  $a_1^k$ , and there is no guarantee that this new multiplier will have suitable properties especially as the value of  $k$  depends on the number of streams required and so is likely to change depending on the application. This problem can be emphasised by the lattice structure of linear



congruential generators. Similarly, the value of  $a_1$  is often chosen such that the computation  $a_1 x_i \bmod m_1$  can be performed efficiently. When  $a_1$  is replaced by  $a_1^k$ , this is often no longer the case.

Note that, due to rounding, when using a distributional generator, a sequence generated using leap-frogging and a sequence constructed by taking every  $k$  value from a set of variates generated without leap-frogging may differ slightly. These differences should only affect the least significant digit.

### 2.7.6 Skip-ahead and Leap-frog for the Mersenne Twister: An Example

Skipping ahead with the Mersenne Twister generator is based on the definition of a  $k \times k$  (where  $k = 19937$ ) transition matrix,  $A$ , over the finite field  $\mathbb{F}_2$  (with elements 0 and 1). Multiplying  $A$  by the current state  $x_n$ , represented as a vector of bits, produces the next state vector  $x_{n+1}$ :

$$x_{n+1} = Ax_n.$$

Thus, skipping ahead  $v$  places in a sequence is equivalent to multiplying by  $A^v$ :

$$x_{n+v} = A^v x_n.$$

Since calculating  $A^v$  by a standard square and multiply algorithm is  $O(k^3 \log v)$  and requires over 47MB of memory (see Haramoto *et al.* (2008)), an indirect calculation is performed which relies on a property of the characteristic polynomial  $p(z)$  of  $A$ , namely that  $p(A) = 0$ . We then define

$$g(z) = z^v \bmod p(z) = a_{k-1} z^{k-1} + \dots + a_1 z + a_0,$$

and observe that

$$g(z) = z^v + q(z)p(z)$$

for a polynomial  $q(z)$ . Since  $p(A) = 0$ , we have that  $g(A) = A^v$  and

$$A^v x_n = (a_{k-1} A^{k-1} + \dots + a_1 A + a_0 I) x_n.$$

This polynomial evaluation can be performed using Horner's method:

$$A^v x_n = A(\dots A(A(Aa_{k-1}x_n + a_{k-2}x_n) + a_{k-3}x_n) + \dots + a_1x_n) + a_0x_n,$$

which reduces the problem to advancing the generator  $k - 1$  places from state  $x_n$  and adding (where addition is as defined over  $\mathbb{F}_2$ ) the intermediate states for which  $a_i$  is non-zero.

There are therefore two stages to skipping the Mersenne Twister ahead  $v$  places:

- (i) Calculate the coefficients of the polynomial  $g(z) = z^v \bmod p(z)$ ;
- (ii) advance the sequence  $k - 1$  places from the starting state and add the intermediate states that correspond to non-zero coefficients in the polynomial calculated in the first step.

The resulting state is that for position  $v$  in the sequence.

The cost of calculating the polynomial is  $O(k^2 \log v)$  and the cost of applying it to state is constant. Skip ahead functionality is typically used in order to generate  $n$  independent pseudorandom number streams (e.g., for separate threads of computation). There are two options for generating the  $n$  states:

- (i) On the master thread calculate the polynomial for a skip ahead distance of  $v$  and apply this polynomial to state  $n$  times, after each iteration  $j$  saving the current state for later usage by thread  $j$ .
- (ii) Have each thread  $j$  independently and in parallel with other threads calculate the polynomial for a distance of  $(j + 1)v$  and apply to the original state.

Since  $\lim_{v \rightarrow \infty} \log v = \log nv$ , then for large  $v$  the cost of generating the polynomial for a skip ahead distance of  $nv$  (i.e., the calculation performed by thread  $n - 1$  in option (ii) above) is approximately the same as generating that for a distance of  $v$  (i.e., the calculation performed by thread 0). However, only one application to state need be made per thread, and if  $n$  is sufficiently large the cost of applying the polynomial to state becomes the dominant cost in option (i), in which case it is desirable to use option (ii). Tests have shown that as a guideline it becomes worthwhile to switch from option (i) to option (ii) for approximately  $n > 30$ .

Leap frog calculations with the Mersenne Twister are performed by computing the sequence fully up to the required size and discarding the redundant numbers for a given stream.

## 3 Recommendations on Choice and Use of Available Routines

### 3.1 Pseudorandom Numbers

Prior to generating any pseudorandom variates the base generator being used must be initialized. Once initialized, a distributional generator can be called to obtain the variates required. No interfaces have been supplied for direct access to the base generators. If a sequence of random variates from a uniform distribution on the open interval  $(0, 1)$ , is required, then the uniform distribution routine (G05SAF) should be called.

#### 3.1.1 Initialization

Prior to generating any variates the base generator must be initialized. Two utility routines are provided for this, G05KFF and G05KGF, both of which allow any of the base generators to be chosen.

G05KFF selects and initializes a base generator to a repeatable (when executed serially) state: two calls of G05KFF with the same argument-values will result in the same subsequent sequences of random numbers (when both generated serially).

G05KGF selects and initializes a base generator to a non-repeatable state in such a way that different calls of G05KGF, either in the same run or different runs of the program, will almost certainly result in different subsequent sequences of random numbers.

No utilities for saving, retrieving or copying the current state of a generator have been provided. All of the information on the current state of a generator (or stream, if multiple streams are being used) is stored in the integer array STATE and as such this array can be treated as any other integer array, allowing for easy copying, restoring, etc.

#### 3.1.2 Repeated initialization

As mentioned in Section 2.7.1, it is important to note that the statistical properties of pseudorandom numbers are only guaranteed within sequences and not between sequences produced by the same generator. Repeated initialization will thus render the numbers obtained less rather than more independent. In a simple case there should be only one call to G05KFF or G05KGF and this call should be before any call to an actual generation routine.

#### 3.1.3 Choice of Base Generator

If a single sequence is required then it is recommended that the Mersenne Twister is used as the base generator (GENID = 3). This generator is fast, has an extremely long period and has been shown to perform well on various test suites, see Matsumoto and Nishimura (1998), L'Ecuyer and Simard (2002) and Wichmann and Hill (2006) for example.

When choosing a base generator, the period of the chosen generator should be borne in mind. A good rule of thumb is never to use more numbers than the square root of the period in any one experiment as the statistical properties are impaired. For closely related reasons, breaking numbers down into their bit patterns and using individual bits may also cause trouble.

#### 3.1.4 Choice of Method for Generating Multiple Streams

If the Wichmann–Hill II base generator is being used, and a period of  $2^{90}$  is sufficient, then the method described in Section 2.7.1 can be used. If a different generator is used, or a longer period length is required then generating multiple streams by altering the initial values should be avoided.

Using a different generator works well if less than 277 streams are required.

Of the remaining two methods, both skip-ahead and leap-frogging use the sequence from a single generator, both guarantee that the different sequences will not overlap and both can be scaled to an arbitrary number of streams. Leap-frogging requires no *a-priori* knowledge about the number of variates being generated, whereas skip-ahead requires you to know (approximately) the maximum number of

variates required from each stream. Skip-ahead requires no *a-priori* information on the number of streams required. In contrast leap-frogging requires you to know the maximum number of streams required, prior to generating the first value. Of these two, if possible, skip-ahead should be used in preference to leap-frogging. Both methods required additional computation compared with generating a single sequence, but for skip-ahead this computation occurs only at initialization. For leap-frogging additional computation is required both at initialization and during the generation of the variates. In addition, as mentioned in Section 2.7.4, using leap-frogging can, in some instances, change the statistical properties of the sequences being generated.

### 3.1.5 Copulas

After calling G05RCF or G05RDF the G01F routines in Chapter G01 can be used to convert the uniform marginal distributors into a different form as required.

## 3.2 Quasi-random Numbers

Prior to generating any quasi-random variates the generator being used must be initialized via G05YLF or G05YNF. Of these, G05YLF can be used to initialize a standard Sobol, Faure or Niederreiter sequence and G05YNF can be used to initialize a scrambled Sobol or Niederreiter sequence.

Due to the random nature of the scrambling, prior to calling the initialization routine G05YNF one of the pseudorandom initialization routines, G05KFF or G05KGF, must be called.

Once a quasi-random generator has been initialized, using either G05YLF or G05YNF, one of three generation routines can be called to generate uniformly distributed sequences (G05YMF), Normally distributed sequences (G05YJF) or sequences with a log-normal distribution (G05YKF). For example, for a repeatable sequence of scrambled quasi-random variates from the Normal distribution, G05KFF must be called first (to initialize a pseudorandom generator), followed by G05YNF (to initialize a scrambled quasi-random generator) and then G05YJF can be called to generate the sequence from the required distribution.

Sequences from other distributions can be obtained by calling the ‘deviate’ routines supplied in Chapter G01 on the results from G05YMF. However, care should be taken when doing this as some of these ‘deviate’ routines are only accurate up to a limited number of significant figures which may effect the statistical properties of the resulting sequence of variates.

## 3.3 Programming Advice

Take care when programming calls to those routines in this chapter which are functions. The reason is that different calls with the same parameters are intended to give different results.

For example, if you wish to assign to Z the difference between two successive random numbers generated by G05KAF, beware of writing

```
Z = G05KAF(IGEN, ISEED) - G05KAF(IGEN, ISEED)
```

It is quite legitimate for a Fortran compiler to compile zero, one or two calls to G05KAF; if two calls, they may be in either order (if zero or one calls are compiled, Z would be set to zero). A safe method to program this would be

```
X = G05KAF(IGEN, ISEED)
Y = G05KAF(IGEN, ISEED)
Z = X-Y
```

## 4 Functionality Index

Generating samples, matrices and tables,	
random correlation matrix.....	G05PYF
random orthogonal matrix.....	G05PXF
random permutation of an integer vector .....	G05NCF
random sample from an integer vector	
unequal weights, without replacement .....	G05NEF
unweighted, without replacement .....	G05NDF
random table .....	G05PZF

Generation of time series,	
asymmetric GARCH Type II.....	G05PEF
asymmetric GJR GARCH.....	G05PFF
EGARCH.....	G05PGF
exponential smoothing.....	G05PMF
type I AGARCH.....	G05PDF
univariate ARMA.....	G05PHF
vector ARMA.....	G05PJF
Pseudorandom numbers,	
array of variates from multivariate distributions,	
Dirichlet distribution.....	G05SEF
multinomial distribution.....	G05TGF
Normal distribution.....	G05RYF
Student's $t$ distribution.....	G05RZF
copulas	
Clayton/Cook–Johnson copula (bivariate).....	G05REF
Clayton/Cook–Johnson copula (multivariate).....	G05RHF
Frank copula (bivariate).....	G05RFF
Frank copula (multivariate).....	G05RJF
Gaussian copula.....	G05RDF
Gumbel–Hougaard copula.....	G05RKF
Plackett copula.....	G05RGF
Student's $t$ copula.....	G05RCF
initialize generator,	
multiple streams,	
leap-frog.....	G05KHF
skip-ahead.....	G05KJF
skip-ahead (power of 2).....	G05KKF
nonrepeatable sequence.....	G05KGF
repeatable sequence.....	G05KFF
vector of variates from discrete univariate distributions,	
binomial distribution.....	G05TAF
geometric distribution.....	G05TCF
hypergeometric distribution.....	G05TEF
logarithmic distribution.....	G05TFF
logical value .TRUE. or .FALSE.....	G05TBF
negative binomial distribution.....	G05THF
Poisson distribution.....	G05TJF
uniform distribution.....	G05TLF
user-supplied distribution.....	G05TDF
variate array from discrete distributions with array of parameters,	
Poisson distribution with varying mean.....	G05TKF
vectors of variates from continuous univariate distributions,	
beta distribution.....	G05SBF
Cauchy distribution.....	G05SCF
exponential mix distribution.....	G05SGF
$F$ -distribution.....	G05SHF
gamma distribution.....	G05SJF
logistic distribution.....	G05SLF
log-normal distribution.....	G05SMF
negative exponential distribution.....	G05SFF
Normal distribution.....	G05SKF
real number from the continuous uniform distribution.....	G05SAF
Student's $t$ -distribution.....	G05SNF
triangular distribution.....	G05SPF
uniform distribution.....	G05SQF
von Mises distribution.....	G05SRF

Weibull distribution .....	G05SSF
$\chi^2$ square distribution.....	G05SDF
Quasi-random numbers, array of variates from univariate distributions, log-normal distribution .....	G05YKF
Normal distribution.....	G05YJF
uniform distribution.....	G05YMF
initialize generator, scrambled Sobol or Niederreiter.....	G05YNF
Sobol, Niederreiter or Faure .....	G05YLF

## 5 Auxiliary Routines Associated with Library Routine Parameters

None.

## 6 Routines Withdrawn or Scheduled for Withdrawal

Withdrawn Routine	Mark of Withdrawal	Replacement Routine(s)
G05CAF	22	G05SAF
G05CBF	22	G05KFF
G05CCF	22	G05KGF
G05CFF	22	F06DFF
G05CGF	22	F06DFF
G05DAF	22	G05SQF
G05DBF	22	G05SFF
G05DCF	22	G05SLF
G05DDF	22	G05SKF
G05DEF	22	G05SMF
G05DFF	22	G05SCF
G05DGF	16	G05SJF
G05DHF	22	G05SDF
G05DJF	22	G05SNF
G05DKF	22	G05SHF
G05DLF	16	G05SBF
G05DMF	16	G05SBF
G05DPF	22	G05SSF
G05DRF	22	G05TKF
G05DYF	22	G05TLF
G05DZF	22	G05TBF
G05EAF	22	G05RZF
G05EBF	22	G05TLF
G05ECF	22	G05TJF
G05EDF	22	G05TAF
G05EEF	22	G05THF
G05EFF	22	G05TEF
G05EGF	22	G05PHF
G05EHF	22	G05NCF
G05EJF	22	G05NDF
G05EWF	22	G05PHF
G05EXF	22	G05TDF
G05EYF	22	G05TDF
G05EZF	22	G05RZF
G05FAF	22	G05SQF
G05FBF	22	G05SFF
G05FDF	22	G05SKF

G05FEF	22	G05SBF
G05FFF	22	G05SJF
G05FSF	22	G05SRF
G05GAF	22	G05PXF
G05GBF	22	G05PYF
G05HDF	22	G05PJF
G05HKF	24	G05PDF
G05HLF	24	G05PEF
G05HMF	24	G05PFF
G05HNF	24	G05PGF
G05KAF	24	G05SAF
G05KBF	24	G05KFF
G05KCF	24	G05KGF
G05KEF	24	G05TBF
G05LAF	24	G05SKF
G05LBF	24	G05SNF
G05LCF	24	G05SDF
G05LDF	24	G05SHF
G05LEF	24	G05SBF
G05LFF	24	G05SJF
G05LGF	24	G05SQF
G05LHF	24	G05SPF
G05LJF	24	G05SFF
G05LKF	24	G05SMF
G05LLF	24	G05SJF
G05LMF	24	G05SSF
G05LNF	24	G05SLF
G05LPF	24	G05SRF
G05LQF	24	G05SGF
G05LXF	24	G05RYF
G05LYF	24	G05RZF
G05LZF	24	G05RZF
G05MAF	24	G05TLF
G05MBF	24	G05TCF
G05MCF	24	G05THF
G05MDF	24	G05TFF
G05MEF	24	G05TKF
G05MJF	24	G05TAF
G05MKF	24	G05TJF
G05MLF	24	G05TEF
G05MRF	24	G05TGF
G05MZF	24	G05TDF
G05NAF	24	G05NCF
G05NBF	24	G05NDF
G05PAF	24	G05PHF
G05PCF	24	G05PJF
G05QAF	24	G05PXF
G05QBF	24	G05PYF
G05QDF	24	G05PZF
G05RAF	24	G05RDF
G05RBF	24	G05RCF
G05YAF	23	G05YLF and G05YMF
G05YBF	23	G05YLF and either G05YJF or G05YKF
G05YCF	24	G05YLF
G05YDF	24	G05YMF
G05YEF	24	G05YLF
G05YFF	24	G05YMF
G05YGF	24	G05YLF
G05YHF	24	G05YMF

G05ZAF            22            No replacement routine required

## 7 References

Banks J (1998) *Handbook on Simulation* Wiley

Boye E (Unpublished manuscript) Copulas for finance: a reading guide and some applications Financial Econometrics Research Centre, City University Business School, London

Bratley P and Fox B L (1988) Algorithm 659: implementing Sobol's quasirandom sequence generator *ACM Trans. Math. Software* **14** (1) 88–100

Faure H and Tezuka S (2000) Another random scrambling of digital (t,s)-sequences *Monte Carlo and Quasi-Monte Carlo Methods* Springer-Verlag, Berlin, Germany (eds K T Fang, F J Hickernell and H Niederreiter)

Fox B L (1986) Algorithm 647: implementation and relative efficiency of quasirandom sequence generators *ACM Trans. Math. Software* **12** (4) 362–376

Haramoto H, Matsumoto M, Nishimura T, Panneton F and L'Ecuyer P (2008) Efficient jump ahead for F2-linear random number generators *INFORMS J. on Computing* **20** (3) 385–390

Hong H S and Hickernell F J (2003) Algorithm 823: implementing scrambled digital sequences *ACM Trans. Math. Software* **29:2** 95–109

Joe S and Kuo F Y (2008) Constructing Sobol sequences with better two-dimensional projections *SIAM J. Sci. Comput.* **30** 2635–2654

Knuth D E (1981) *The Art of Computer Programming (Volume 2)* (2nd Edition) Addison–Wesley

L'Ecuyer P and Simard R (2002) *TestU01: a software library in ANSI C for empirical testing of random number generators* Departement d'Informatique et de Recherche Operationnelle, Universite de Montreal <http://www.iro.umontreal.ca/~lecuyer>

Maclaren N M (1989) The generation of multiple independent sequences of pseudorandom numbers *Appl. Statist.* **38** 351–359

Matsumoto M and Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudorandom number generator *ACM Transactions on Modelling and Computer Simulations*

Morgan B J T (1984) *Elements of Simulation* Chapman and Hall

Nelsen R B (1998) *An Introduction to Copulas. Lecture Notes in Statistics 139* Springer

Owen A B (1995) Randomly permuted (t,m,s)-nets and (t,s)-sequences *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Lecture Notes in Statistics* **106** Springer-Verlag, New York, NY 299–317 (eds H Niederreiter and P J-S Shiue)

Ripley B D (1987) *Stochastic Simulation* Wiley

Sklar A (1973) Random variables: joint distribution functions and copulas *Kybernetika* **9** 499–460

Wichmann B A and Hill I D (2006) Generating good pseudo-random numbers *Computational Statistics and Data Analysis* **51** 1614–1622

Wikramaratna R S (1989) ACORN - a new method for generating sequences of uniformly distributed pseudo-random numbers *Journal of Computational Physics* **83** 16–31

Wikramaratna R S (1992) Theoretical background for the ACORN random number generator *Report AEA-APS-0244* AEA Technology, Winfrith, Dorest, UK

Wikramaratna R S (2007) The additive congruential random number generator a special case of a multiple recursive generator *Journal of Computational and Applied Mathematics*