Additive congruential pseudo-random number generators

By J. C. P. Miller* and M. J. Prentice†

Pseudo-random number sequences are needed in many calculations. Methods for generating such sequences involving multiplication are familiar; the present paper considers a method of generation involving addition and subtraction only, and discusses the applications of several tests of randomness to the resulting sequencies.

(First received August 1967 and in revised form March 1968)

1. Introduction

The need for, and use of, sequences of pseudo-random numbers on automatic computers for Monte Carlo techniques and other purposes is well established. The various types of sequence needed can be obtained (see e.g. Tocher (1963)) from a uniform distribution on (0, 1). Such a sequence, generated in a digital computer in an exactly reproducible manner and satisfying specified 'tests of randomness' is called a *pseudo-random number sequence*.

Downham and Roberts (1967), among others, have carried out an investigation of such sequences generated by use of multiplication and reduction to some modulus. In particular, several tests of randomness (such as tests of equidistribution and lack of correlation between terms) are examined and discussed. The present paper carries out a similar investigation on sequences produced by addition of integers and reduction to some (prime) modulus, the reduction being by subtraction rather than by division. On many computers it is advantageous to avoid multiplication or division, if used very frequently, in favour of addition or subtraction, in order to save time.

Additive methods for pseudo-random number generation have been tried in the past (e.g. Taussky and Todd, 1956), usually without much success. However, Green, Smith and Klem (1959) were moderately successful with

$$X_j = X_{j-1} + X_{j-n} \quad \text{modulo } 2^r \quad (j > n) \qquad (1)$$
$$u_j = X_j/2^r \quad \text{(Fibonacci-like).}$$

They found that sequences of this type 'passed' most of the tests applied to them, but failed on one of the more sensitive 'run' tests (see § 3) until n was quite large. They claim that the statistical properties of the sequence improved as n increased, and required $n = 16$ (and thus storage of 16 previously generated numbers) for completely satisfactory results. They also noted that taking alternate terms from the sequence generated when $n = 6$ gave quite good results.

In what follows, we describe what we consider to be a relatively successful attempt at generating pseudo-random sequences from the third order recurrence

$$X_j = X_{j-2} + X_{j-3} \quad \text{modulo } p\text{, a prime} \qquad (2)$$
$$u_j = X_j/p.$$

Our reason for using this relation is that (see § 2), provided p can not be written in the form $c^2 + 23d^2$ (c, d, integral), it is sometimes possible to generate periodic sequences with period of order p^2. This means that $p = 2003$, for example, might, and in fact does, provide sequences with period four million approximately, whereas $p = 2347 = 22^2 + 23(9)^2$ gives the much smaller period 2346. Of course, length of period obtained is not the only relevant criterion when looking for a pseudo-random number generator, but it is certainly necessary that the period shall be long compared with the lengths of the sequences likely to be needed.

A number of statistical tests were used on the sequences generated, and on the basis of the results obtained, it appears that the two run tests (see § 3) are the most sensitive (in agreement with Downham and Roberts).

2. Number theory

In this section, we outline properties of sequences (X_j) satisfying $X_j = X_{j-2} + X_{j-3}$ modulo p, with arbitrary (integral) initial values X_0, X_1, X_2. The reason for choosing this sequence for consideration here is that the generation of successive terms is easy—it is difficult to imagine anything easier—comprising

(i) the addition of two given numbers, and
(ii) a very simple reduction, modulo p; we subtract p from the sum, test for sign, and restore p if negative.

This supplies the basic sequence of pseudo-random numbers with no multiplication or division. Subsequent use is then the same as for a sequence of pseudo-random numbers generated by any other method, e.g. by a multiplicative method, though these sometimes yield more immediately a number confined to the range (0, 1). However, this can also be achieved directly with the present method without division, perhaps most simply, if $2^{k+1} > p > 2^k$ by appending to step (ii) a step (iii):

(iii) Test $X_j - 2^k$; if $\geqslant 0$, *do not use* X_j.
If < 0, *use* $X_j/2^k$, in (1, 0], obtained from X_j by shifting.
Continue then to obtain X_{j+1} by step (i), etc.

It would seem best to choose $p - 2^k = \delta$ small, i.e. p would be a prime just exceeding 2^k; the result is that

* *University Mathematical Laboratory, Corn Exchange St., Cambridge.*
† *I.C.I. Mond Division, The Heath, P.O. Box 14, Runcorn, Cheshire.*

we use numbers in $(2^k, 0]$, easily reduced to $(1, 0]$ by a binary shift rather than by a division.

In the tests below we have, however, preferred to examine the whole range for randomness, by using a division by p. The randomness properties of the partial sequence—a major part of the whole (the exact proportion depending on the choice of p)—may perhaps be expected to be similar to those of the whole.

The sequence is known as the Perron sequence when $(X_0, X_1, X_2) = (3, 0, 2)$ so that $X_n = a^n + b^n + c^n$ modulo p, where a, b, c, are the roots of $z^3 - z - 1 = 0$. These sequences have been rather thoroughly studied, largely because of the simplicity of their generation, and so their structure is known and can be used to help explain statistical results.

Clearly, X_n modulo p is periodic, for the number of triplets $Y_n = (X_n, X_{n+1}, X_{n+2})$ is at most $p^3 - 1$, excluding the triplet $(0, 0, 0)$ which occurs only in the null sequence. In fact, the starting values (X_0, X_1, X_2) must themselves eventually recur.

The polynomial $z^3 - z - 1$ cannot be factorised into real integer factors, but it may be factorisable modulo p, that is, there may exist integers y, w, such that $z^3 - (yp + 1)z - (wp + 1)$ has integer factors (that is, integer zeros).

There are three cases:

$$z^3 - z - 1 = (z - r)(z - s)(z - t) \text{ modulo } p \qquad \text{(A)}$$

$$z^3 - z - 1 = (z - r)(z^2 + rz - m) \text{ modulo } p \qquad \text{(B)}$$

$$z^3 - z - 1 \text{ has no factors modulo } p \qquad \text{(C)}$$

For example,

$$z^3 - z - 1 = (z - 4)(z - 13)(z - 42) \text{ modulo } 59 \qquad \text{(case A)}$$

$$z^3 - z - 1 = (z - 2)(z^2 + 2z + 3) \text{ modulo } 5 \qquad \text{(case B)}$$

and, for

$$p = 3, 13, 29, \ldots, z^3 - z - 1 \text{ has no factors modulo } p. \qquad \text{(case C)}$$

Other examples may be found by studying and factorising $n^3 - n - 1$ for integral n. Putting $z = r$ above, we find that if $z^3 - z - 1$ factorises modulo p, then p is a factor of $r^3 - r - 1$. Thus, if the factors of $n^3 - n - 1$ include p once (as n increases from 1 to $p - 1$), then we have case B, if three times, case A, and if not at all, case C.

Primes of the form $46t + 5, 7, 11, 15, 17, 19, 21, 33, 37, 43, 45$ fall into category B. All others fall into categories A or C. Those in category A are also expressible in the form $c^2 +.23d^2$.

Let $g(z) = \sum_{0}^{\infty} X_r z^r$, then

$$(1 - z^2 - z^3)g(z) = X_0 + X_1 z + (X_2 - X_0)z^2 + \sum_{3}^{\infty}(X_r - X_{r-2} - X_{r-3})z^r.$$

Hence from the recurrence relation (2)

$$(1 - z^2 - z^3)g(z) = X_0 + X_1 z + (X_2 - X_0)z^2 \text{ (modulo } p) = s(z) \text{ (modulo } p\text{), say.}$$

Thus $g(z) = s(z)/F(z)$ is a complete specification of the sequence generated, where

$$F(z) = z^3 f(1/z), \text{ and } f(z) = z^3 - z - 1.$$

The sets of integral zeros of $f(z)$ and $F(z)$ modulo p are isomorphic; all are non-zero and, given r an integral zero of $f(z)$ modulo p, there exists a unique integer R such that $rR = 1$ modulo p and R is a zero of $F(z)$.

Suppose $F(z) = \prod_{i=1}^{h} F_i(z)$ modulo p, $1 \leqslant h \leqslant 3$, where the $F_i(z)$ are irreducible modulo p. Then $h = 1, 2, 3$, correspond to categories C, B, A respectively, and $s(z)/F(z)$ is expressible in partial fractions as $\sum_{i=1}^{h} s_i(z)/F_i(z) = \sum_{i=1}^{h} g_i(z)$, where $\deg(s_i) < \deg(F_i)$ since $\deg(s) \leqslant 2 < 3 = \deg(F)$. For every p except $p = 23$, which does not concern us here, the F_i are distinct and the above formula is valid.

Each $g_i(z)$ generates sequences all of the same length (with period e_i say), which is a factor of $p^{k_i} - 1$ (we exclude $s(z) = 0$ identically) where $k_i = \deg(F_i)$, and so the period of non-zero sequences generated by $g(z)$ is $e = \text{lcm}\ (e_1, \ldots, e_h)$ provided $s(z)/F(z)$ is in its lowest terms. Details and proofs are given in Selmer (1966) and Peterson (1961).

Consider category C, for which $h = 1$, $\deg(F) = 3$, and e divides $p^3 - 1$.

In fact, e divides $p^2 + p + 1$:

$$(f(z))^p = (z^3 - z - 1)^p = (z^{3p} - z^p - 1) \text{ (mod } p)$$

since all coefficients other than cross terms (which contain p as a factor and hence vanish modulo p) are 1 or -1 and $1^p = 1$, $(-1)^p = -1$ modulo p, by Fermat's theorem. Hence $f(z)$ and $f(z^p)$ have the same zeros. Hence if r is a zero of $f(z)$ then so is r^p (and r^{p^2}). Now $X_n = r^n + r^{np} + r^{np^2}$ modulo p satisfies the recurrence, and since $r^{1+p+p^2} = 1$ (product of zeros) modulo p, e divides $p^2 + p + 1$, since all sequences have the same period when F is irreducible.

For category B, $h = 2$ and F_1 gives sequences with period e_1 dividing $(p - 1)$, while F_2 gives sequences with period e_2 dividing $(p^2 - 1)$. Hence F can give sequences with period $e = \text{lcm}\ (e_1, e_2)$ dividing $(p^2 - 1)$, or e_1, or e_2. However, sequences generated need *not* be all of the same length, for if the starting triplet is chosen so that $s(z)$ has either F_1 or F_2 as a factor then by cancellation, F becomes either F_2 or F_1 respectively and the sequences generated have periods e_2 or e_1 respectively. In all other cases, $g(z)$ is in its lowest terms and the period is $e = \text{lcm}\ (e_1, e_2)$.

For category A, there are 7 possible periods (which may be distinct): e_1, e_2, e_3, lcm (e_1, e_2), lcm (e_2, e_3), etc. All divide $(p - 1)$. This is true provided the F_i are distinct, which is always the case if $p \neq 23$.

For $p = 23$, $f(z) = (z - 3)(z - 10)^2$. $X_r = 10^r$ has period 22 modulo 23, but here the general term is $A3^r + (B + Cr)10^r$ and $r10^r$ has period $22.23 = 506$.

In hoping for reasonable 'randomness' in the sequences generated, we must presumably restrict ourselves to runs of length less than the period, which suggests rejecting generators of type A, if we are using moderately small p. Some periods are known of full length $p^2 - 1$ and $p^2 + p + 1$, e.g. 199 and 101 respectively. For moderately small p, such as 2003 it is possible to obtain very long sequences before repetition occurs, although even then the sequence may not be statistically satisfactory. When using very large primes, such as 67 101 323, it is quite possibly irrelevant whether or not $p = c^2 + 23d^2$.

3. Statistical tests

(a) *Uniformity*

The range (0, 1) was divided up into 100 disjoint equal intervals and a sample of 10000 pseudo-random numbers was used. A count was made of the number of numbers falling into each cell, and a standard Chi-square test of goodness of fit applied. It should be noted that this test should not be applied blindly when p is only moderately large and just greater than a multiple of 100. For then the expected number in each cell is not the same for each cell when the integers obtained are in fact uniformly distributed on $(0, (p - 1))$. For example (an extreme case which would never arise in practice), for $p = 101$, the expected number in the first cell is double that in any other, as the integers 0 and 1 contribute to it while only one integer contributes to each of the other cells.

(b) *Runs above and below the median*

Given a sequence of numbers $X_1, \ldots, X_N$, the subsequence $X_{i-1}, \ldots, X_{i+r}$, is said to form a run of length r above the median if $X_{i-1} < \frac{1}{2}$, $X_i > \frac{1}{2}, \ldots$, $X_{i+r-1} > \frac{1}{2}$, $X_{i+r} < \frac{1}{2}$. A run below the median is similarly defined, and end runs are defined by $i = 1$ or $i + r = N + 1$.

The expected number of runs of length r, not distinguishing between runs above and below, is approximately given by

$$E(r) = (N - r + 3)2^{-(r+1)},$$

and the expected number of length r or greater by

$$E'(r) = (N - r + 2)2^{-r}.$$

In all cases, $N = 10000$ numbers were used and a Chi-square goodness of fit test of the number of runs of lengths 1 to 9 inclusive and 10 or greater was performed.

(c) *Up and down runs* (see also Downham and Roberts (1967))

Given a sequence of numbers $X_1, \ldots, X_N$, the subsequence $X_{i-1}, \ldots, X_{i+r+1}$, is said to form a run of length r 'up' if

$$X_{i-1} > X_i, X_i < X_{i+1} < X_{i+2} \ldots < X_{i+r} > X_{i+r+1}.$$

Runs down are similarly defined and end runs are defined by $i = 1$ or $i = N - r$. In this case, using analogous notation to test (b),

$$E(r) = 2\{N(r^2 + 3r + 1) - (r^3 + 3r^2 - r - 4)\}/(r + 3)!$$

and

$$E'(r) = 2\{N(r + 1) - (r^2 + r - 1)\}/(r + 2)!$$

very nearly, not distinguishing between runs up and down. In all cases, a sequence of $N = 10000$ numbers was used. The number of runs of length 1 to 5 inclusive and 6 or greater were counted and a Chi-square goodness of fit test performed.

(d) *Serial lag*

To test the serial lag properties of the sequence (for lag q, $1 \leqslant q \leqslant 6$ usually) a sample of $2000 + q$ pseudo-random numbers was used. The interval (0, 1) was divided up into 10 equal disjoint intervals. Let f_i be the number of numbers X_r in the range $\left(\frac{i-1}{10}, \frac{i}{10}\right)$, and $g_{ij}(q)$ be the number of ordered pairs (X_r, X_{r+q}) such that X_r contributes to f_i and X_{r+q} to f_j.

Let

$$S_1^2 = \frac{1}{200}\sum_{i=1}^{10}(f_i - 200)^2,$$

and $$S_2^2(q) = \frac{1}{20}\sum_{i=1}^{10}\sum_{j=1}^{10}(g_{ij}(q) - 20)^2.$$

Then $S_3^2 = S_2^2(q) - S_1^2$ is asymptotically χ^2_{90} for a truly random sequence (see Good, 1953) and so S_3^2 was calculated and compared with the distribution of χ^2_{90}.

(e) *Poker test*

A run of 2000 sets of first digits of 5 consecutive numbers in the sequences was examined and classified according to the type of poker hand to which they corresponded, as shown in **Table 1.**

Table 1

EVENT	PROBABILITY
'bust'	0·3024
'one pair	0·5040
'two pairs'	0·1080
'three of a kind'	0·0720
'full house'	0·0090
'four or more of a kind'	0·0046

Again, a standard Chi-square test of goodness of fit was performed.

(f) *Powers*

One method of estimating $\int_0^1 f(x)dx$ by a Monte Carlo method would be to evaluate $\frac{1}{n}\sum_1^n f(X_i)$ where the X_i are (supposedly) uniformly and independently distributed on (0, 1). Given a pseudo-random number generator, it is reasonable to ask whether sequences generated from it will satisfactorily 'integrate' polynomials (in one variable) by this method, since, if so, then they will also 'integrate' most 'smooth' functions (with a Taylor expansion) satisfactorily.

For a random variable X, uniformly distributed on (0, 1),

$$E(X^j) = 1/(j+1), \text{ and var } (X^j) = j^2/(2j+1)(j+1)^2. \qquad (j > 0)$$

We did not persevere with this test for long since it was not very sensitive, and our main concern was to try to find sequences that satisfy the most sensitive tests we could find. It was hoped that the powers test might provide a sequence of graded tests distinct from those given by correlation and run tests; this is, however, more likely to be useful for, say, a sequence of random normal deviates, where sample variance is perhaps more variable than it is in a more or less uniform distribution over a finite interval, as in our sequences.

4. Results

A few results were obtained with Lehmer's Multiplicative Congruence (using a prime modulus; see also Downham and Roberts, 1967). Under suitable conditions on k (where the generator is $X_{r+1} = kX_r$ modulo p), satisfactory results can be obtained from all the tests we used.

A few runs with the second order Fibonacci sequence were also made, and in agreement with Taussky and Todd (1956), the results were extremely unsatisfactory. For example, the number of up or down runs of length 1 was extremely low, while the number of up or down runs of length 4 or greater was very high. Attempts to obtain more satisfactory sequences were made by using only every other, then every third, and finally every fourth number in the sequence. In every case, the value of χ^2 obtained in both run tests was highly significant.

The uniformity, serial lag, powers, and poker tests were not particularly sensitive, whereas the two run tests gave very large values of χ^2 at the least provocation. Incidentally, Green, Smith, and Klem (1959) also found that the median runs test and a generalisation of it (also testing goodness of fit of runs above and below the point q, $q = 0{\cdot}75$, $0{\cdot}25$, etc.) were the most sensitive, and on the basis of results obtained, felt it necessary to go as far as the sixth order for a satisfactory generator. They did not use the up and down run test at all.

After using all six tests on a few generators, it was found that no generator 'passed' the run tests and yet failed the others, and hence, unless a generator 'passed' the run tests, it was considered a waste of computer time running the others.

Run test results from the basic Perron recurrence $X_j = X_{j-2} + X_{j-3}$ modulo p were not at all good. There were far too many up or down runs of length 3 or greater, and *never* any median runs of length 4. It was relatively easy to prove that in fact a median run of exactly length 4 is impossible, i.e.

if $a < \frac{1}{2}, b > \frac{1}{2}, c > \frac{1}{2}, (a+b)_1 > \frac{1}{2}, (b+c)_1 > \frac{1}{2}$

then $(a+b+c)_1 > \frac{1}{2}$

(and its dual), where $(X)_1$ denotes the fractional part of X.

Proof

$0 \leqslant a < \frac{1}{2}$ and $\frac{1}{2} < b < 1$, hence $\frac{1}{2} < (a+b) < 1\frac{1}{2}$.
But $(a+b)_1 > \frac{1}{2}$, and hence $\frac{1}{2} < (a+b) < 1$.
Also $\frac{1}{2} < b < 1$ and $\frac{1}{2} < c < 1$, and so $1 < (b+c) < 2$.
But $(b+c)_1 > \frac{1}{2}$. Hence $1\frac{1}{2} < b + c < 2$.
$\frac{1}{2} < (a+b) < 1$ and $\frac{1}{2} < c < 1$, so $1 < (a+b+c) < 2$.
$1\frac{1}{2} < (b+c) < 2$ and $0 < a < \frac{1}{2}$, so $1\frac{1}{2} < (a+b+c) < 2\frac{1}{2}$.
Hence $1\frac{1}{2} < (a+b+c) < 2$, i.e. $(a+b+c)_1 > \frac{1}{2}$
q.e.d.

Taking every other number in the sequence, and also every third, still gave very unsatisfactory run test results. However, when sampling every fourth number, extremely good results were obtained for quite small p (such as 2003), provided p was not of the form $c^2 + 23d^2$; see **Table 2.**

Table 2

MEDIAN RUN LENGTH	EXPECTATION		NUMBER OBTAINED	
		$p = 2003$	$p = 2347$	$p = 5237$
1	2500·5	2401	2457	2661
2	1250·1	1247	1167	1295
3	625·0	631	745	667
4	312·5	312	302	290
5	156·2	171	124	162
6	78·1	81	111	55
7	39·0	36	50	9
8	19·5	19	16	31
9	9·8	16	0	15
10 or more	9·8	13	0	0
	χ^2_9	11·0	73·4*	65·8*
UP-DOWN				
1	4166·7	4095	4156	4043
2	1833·1	1857	1844	1868
3	527·7	545	580	583
4	115·0	118	82	90
5	20·3	14	17	22
6 or more	3·5	2	0	0
	χ^2_5	4·8	19·8*	19·2*

2347 and 5237 both fall into category A, whereas 2003, in category C, produces sequences with periods of length 4014013, and much more satisfactory run test results.

Generators of this type (i.e. every fourth number from the generalisation of the Perron sequence) were then tested with the uniformity, serial lag, powers and poker tests, and, as expected were satisfactory provided p was not of the form $c^2 + 23d^2$. When p was of this form, all results were unsatisfactory. One possible reason is, of course, that the cycle length is much shorter ($p - 1$ or less). Another is that the actual integers obtained from the recurrence are by no means uniformly distributed; see, for example, **Table 3**.

Table 3

$p = 59$ $(x_0, x_1, x_2) = (16, 50, 43)$ gives a cycle of length 58:

16	50	43	7	34	50	41	25	32	7	57
39	5	37	44	42	22	27	5	49	32	54
22	27	17	49	44	7	34	51	41	26	33
8	0	41	8	41	49	49	31	39	21	11
1	32	12	33	44	45	18	30	4	48	34
52	23	27								

The integer 49 occurs 4 times whereas twenty-three positive integers <59 do not occur at all. When p is not in category A of § 2, this phenomenon is not as marked, as satisfactory results for all tests can be obtained. Thus, with the above restriction on p, it appears at first sight that the sequence generated by taking every fourth number from the generalised Perron sequence is as good as the multiplicative method. About 25 different values of p, ranging from 2003 up to 67101323, were found to give satisfactory results to all our tests. However, the limitation with any additive method is that if the generator is used to produce 'random' points in k-space (in the obvious way), then when k is at all large, results of Monte Carlo integrations (for instance) will quite possibly be very inaccurate because the generator is really only producing points in a subspace of much lower dimension (in our case, of dimension 3). Greenberger (1965) describes another hazard with additive generators, observed with the aid of an oscilloscope.

Davis and Rabinowitz (see Tocher, 1963) claim to have obtained satisfactory results from simulations of finding the volume of k-dimensional hypercubes by randomising between three second order additive generators. Accordingly, to show the limitations of our third-order method, simulations of finding the volume of a hypercube of side $2^{-1/k}$ $(2 \leqslant k \leqslant 9)$ in k-dimensional space were run, and an attempt to estimate the volume of a 100-dimensional hypercube of side $0{\cdot}99$ was made (volume $= \left(1 - \frac{1}{100}\right)^{100} \approx 1/e$). Estimates of the volume of the first 'quadrant' of a hypersphere of unit radius in k-space $(2 \leqslant k \leqslant 6)$ were also made. In both types of simulation, every term in the sequence was used, and thus successive x_i values (for example) are generated by taking every kth member in the sequence, *but* the x_i, $1 \leqslant i \leqslant k$ determining $(x_1, \ldots, x_k,)$ in k-space are by no means independent, and in fact $(x_1, \ldots, x_k,)$ is in a 3-dimensional subspace of $(0, 1)^k$.

Hypercube

(Number of points inside out of 10000)

k	RUN 1	RUN 2	EXPECTATION	S.D.
2	5048	5029	5000	50
3	4918	4918	5000	50
4	4942	4973	5000	50
5	4909	4964	5000	50
6	4791	4882	5000	50
7	4848	4823	5000	50
8	4818	4866	5000	50
9	4805	4868	5000	50
100	3562		3679	48

The deviation from expectation is negative and significant for $k \geqslant 5$, and the effect appears to increase with k, although it is not as large as one might expect. The simulation with $k = 100$ was repeated with the Cambridge University Library Routine for pseudo-random numbers (Lehmer's Method with prime $(2^{31} - 1)$ modulus). Of 10000 points, 3648 were inside the hypercube, which is quite satisfactory.

Hypersphere (the volume of a hypersphere of unit radius is $\pi^{k/2}/(\frac{1}{2}k)!$)

(Number of points inside out of 10000)

k	EXPECTATION	S.D.	RUN 1	RUN 2
2	7854·0	41	7922	7792
3	5236·0	50	5254	5189
4	3084·0	46	2974	3066
5	1644·9	37	1471	1399
6	807·5	27	721	678

Again, a downward drift (relative to expectation) is noticeable as k increases. The results are not acceptable for $k \geqslant 5$, and only one is reasonable for $k = 4$.

Tests were also made on sequences generated by taking every fifth, sixth, seventh, eighth and ninth numbers from the original recurrence. All were satisfactory apart from the run test results when taking every seventh number; there were too few long median runs and too many of lengths 3, 4 and 5, while the up and down run test gave too many of length 1 and too few of length 4. We were unable to find a satisfactory explanation of this.

5. Conclusions

It appears that satisfactory pseudo-random numbers can be generated from the additive congruence $X_j = X_{j-2} + X_{j-3}$ modulo p provided only every fourth term is used and p is not of the form $c^2 + 23d^2$. If pseudo-random numbers are required for four or more independent purposes, then every term in the sequence can be used (if seven sequences are required, reject every eighth term), but caution is needed when evaluating integrals by Monte Carlo methods, as this additive method should certainly not be used in more than three dimensions.

Even when p is not of the form $c^2 + 23d^2$, it is advisable to check that the period obtained is long enough. For example, $p = 151$ gives a period of $1093 = (p^2 + p + 1)/21$, and $p = 157$ gives a period of $4108 = (p^2 - 1)/6 = 26(p + 1)$, so it appears that quite possibly periods considerably shorter than p^2 are common. However, supposing the word length for integer addresses on a given (binary) machine is n, (i.e. integers $\leqslant 2^n - 1$ are allowed), then it is quite likely that there exists a prime p in the range $2^{n-1/2}$ to 2^n, satisfying all the conditions for category C of § 2 and such that $p^2 + p + 1$ is prime. For example, it was relatively easy to find $p = 2957$ and 3137 to satisfy all these conditions, and they must produce sequences of period 8 746 807 and 9 843 907 respectively, as these two numbers are themselves primes. For the Titan computer at Cambridge, $n = 20$ and since (Hardy and Wright, 1960) there are of the order of 20000 primes between 750000 and 1 048576 it appears almost certain that such a p exists, which will produce sequences with period at least 500000 million. Actually finding such a p is not an insurmountable problem; only the lack of a suitable program prevented us from finding one.

It seems reasonable to suppose that the time lost in storage and retrieval of three previously generated numbers (as opposed to one) will on many machines be more than offset by the time gained through using addition and subtraction rather than multiplication and division.

Acknowledgements

M. J. Prentice is indebted to I.C.I. Mond Division for financial support while working on this project. We are grateful to members of the Cambridge University Statistical Laboratory, especially Mr. A. D. McLaren, for suggestions and comments.

References

Downham, D. Y., and Roberts, F. D. K. (1967). Multiplicative congruential pseudo-random number generators, *Computer Journal*, Vol. 10, pp. 74–77.

Good, I. J. (1953). The Serial Test for Sampling Numbers and other Tests for Randomness, *Proc. Camb. Phil. Soc.*, Vol. 49, pp. 276–284.

Green, B. F., Smith, J. E. K., and Klem, L. (1959). Empirical Tests of an Additive Random Number Generator, *JACM*, Vol. 6, pp. 527–537.

Greenberger, M. (1965). Method in Randomness, *Comm. ACM*, Vol. 8, pp. 177–179.

Hardy, G. H., and Wright, E. M. (1960). *An Introduction to the Theory of Numbers*, Oxford University Press (4th Edition).

Naylor, T. H., Balintfy, J. L., Burdick, D. S. and King Chu (1966). *Computer Simulation Techniques*, Wiley.

Peterson, W. W. (1961). *Error Correcting Codes*, M.I.T. Press, Cambridge, Mass.

Prentice, M. J. (1967). Applied Field Notebook, Diploma in Mathematical Statistics, Cambridge.

Selmer, E. S. (1966). *Linear Recurrence Relations over Finite Fields*, Department of Mathematics, University of Bergen, Norway.

Taussky, O., and Todd, J. (1956). Generation and Testing of Pseudo-Random Numbers (in *Symposium on Monte Carlo Methods*, Ed. H. A. Meyer, Wiley, pp. 15–28).

Tocher, K. D. (1963). *The Art of Simulation*, English Universities Press.

Book Review

Lectures on Advanced Numerical Analysis, by Fritz John, 1967; 179 pages. (London: *Thomas Nelson*, 60s.)

As the author explains in a preface, this book consists essentially of the lecture notes of a course of lectures given in 1956–57, with revisions to the wording and the list of references, but no significant changes to the content to bring the material up to date. This explains, in some measure, the curious selection of numerical methods considered, no doubt also directed by the personal preferences of the author, and the historical nature of some parts of the book.

The first chapter treats the solution of linear simultaneous equations, principally by iterative and gradient methods, but includes also a discussion of matrix norms and estimates for the norm of the inverse matrix. This is followed by a chapter on the roots of polynomials largely devoted to theorems on bounds for these roots, but describing also the direct methods of Bernoulli and Graeffe, and the iterative method of Newton. These two chapters together comprise over one-third of the book.

The determination of eigenvalues of a matrix, polynomial and trigonometric approximation, and the solution of ordinary differential equations are next discussed very briefly, and the remainder of the book is devoted to the solution of partial differential equations in fair detail, especially of equations of the parabolic and hyperbolic type. Here besides the normal difference methods, the methods of Friedrich and Courant, Isaacson and Rees for the solution of hyperbolic equations as systems of first order equations are described.

Although the book cannot be recommended as a basic textbook on advanced numerical analysis today, it does contain much interesting material and would form interesting background reading to a course.

D. C. Gilles (Glasgow)