International Journal of Pure and Applied Mathematics
Volume 42 No. 4 2008, 563-568

# GROUP-THEORETIC GENERATION OF NON-UNIFORM PSEUDO-RANDOM SEQUENCES FOR SIMULATION

Marco Pollanen

Department of Mathematics
Trent University
1600 West Bank Drive, Peterborough, Ontario, K9J 7B8, CANADA
e-mail: marcopollanen@trentu.ca

**Abstract:** Many applications involving statistical simulation, such as Monte Carlo methods, require non-uniform random sequences. These are usually created by first generating a uniform sequence and then using techniques such as rejection sampling or transformation. In this paper we introduce a new method to directly generate, without transformation or rejection, some non-uniform pseudo-random sequences. This method is a group-theoretic analogue of linear congruential pseudo-random number generation. We provide examples of such sequences, involving computations in Jacobian groups of plane algebraic curves, that have both good theoretical and statistical properties.

**AMS Subject Classification:** 14H40, 65C10, 65C05, 68W20
**Key Words:** pseudo-random sequences, non-uniform sequences, Jacobian groups, stochastic simulation

## 1. Introduction

Pseudo-random numbers are a critical part of modern computing, especially for use in simulations and cryptography, and consequently there are a myriad of algorithms for generating pseudo-random sequences. Thus far, almost all of the pseudo-random number literature has focused on generating sequences with uniform distribution.

The most popular and well-studied pseudo-random number generator is the linear congruential generator (LCG), proposed by Lehmer in 1948, which is defined by $x_n \equiv ax_{n-1} + b \bmod m$, where $a$, $b$, $m$, and seed $x_0$, are integers. In

Received: August 17, 2007

this paper we study the analagous problem of *multiply sequences*, where $b = 0$, $m = 1$, and $x_0$ is a real number.

Let $S_n$ be a sequence of propositions about the sequence $y_n$. Following the definition by Franklin [3], we define

$$P(S_n) = \lim_{N\to\infty} \frac{1}{N} \sum_{\substack{S_n \text{ is true} \\ 1\le n\le N}} 1\,,$$

when the limit exists.

A sequence $y_n$ is $\infty$-distributed (Knuth [6]) if, for every value of $k$,

$$P(a_1 \le y_n < b_1, \cdots, a_k \le y_{n+k-1} < b_k) = (b_1 - a_1)\cdots(b_k - a_k)$$

for all real $a_j$, $b_j$, with $0 \le a_j < b_j \le 1$ for $1 \le j \le k$.

$\infty$-distributed sequences are of interest in that they automatically pass a large number of asymptotic statistical tests for randomness, including: the frequency test, serial test, gap test, poker test, coupon collector's test, permutation test, run test, maximum-of-$t$ test, collision test, birthday spacings test, serial correlation test, and tests on subsequences (Knuth [6]).

It was shown by Franklin [3] and Pollanen [10] that as $a$ approaches $\infty$, multiple sequences are $\infty$-distributed and thus have good statistical properties. This is not only of theoretical interest, as we will show that we can construct sequences with non-uniform densities, by using multiply sequences over Jacobian groups of certain algebraic curves, that empirically pass many statistical tests for randomness. Similar sequences have also been used by Pollanen [9] for quasi-random number generation.

## 2. Multiply Sequences in Jacobian Groups

Consider a hyperelliptic curve $C$ of genus $g$ defined by $y^2 = f(x)$, where $f(x)$ is a polynomial of degree $2g + 1$ with distinct roots in $\mathbb{C}$. The basic theory of algebraic curves can be found in [2]. Associated with each hyperelliptic curve is a Jacobian group $\mathbb{J}$ with an Abelian structure. An efficient algorithm for adding points in the Jacobian was described by Cantor [1]. Thus, if we let $P_0$ be a point in $\mathbb{J}$ of infinite order, we may define a sequence in $\mathbb{J}$ by $P_{n+1} = aP_n$, $n \ge 0$.

To determine the distribution our sequence defines, we need to take another view of the Jacobian, namely the Jacobian variety. Using a basis of first order differentials of the first kind,

$$\left\{\frac{dx}{y}, \frac{xdx}{y}, \cdots, \frac{x^{g-1}dx}{y}\right\},$$

and the associated lattice of their periods $\Lambda \subset \mathbb{C}^g$, we define a natural embedding $L : C \to \mathbb{C}^g/\Lambda$ by

$$P \to \left(\int_\infty^P \frac{dx}{y}, \int_\infty^P \frac{xdx}{y}, \dots, \int_\infty^P \frac{x^{g-1}dx}{y}\right).$$

Obviously $L(\infty) = (0, \dots, 0)$. Now, given two points P and Q in $\mathbb{J}$, the Abel-Jacobi Theorem guarantees that:

$$\left(\int_\infty^P \frac{dx}{y}, \dots, \int_\infty^P \frac{x^{g-1}dx}{y}\right) + \left(\int_\infty^Q \frac{dx}{y}, \dots, \int_\infty^Q \frac{x^{g-1}dx}{y}\right)$$
$$= -\left(\int_\infty^{P\oplus Q} \frac{dx}{y}, \dots, \int_\infty^{P\oplus Q} \frac{x^{g-1}dx}{y}\right) \bmod \Lambda.$$

As there is a natural bijection between $\mathbb{C}^g/\Lambda$ and $\mathbb{R}^{2g}/\mathbb{Z}^{2g}$ given by linear transformation, the sequence $P_{n+1} = aP_n$ has a natural identification to $\mathbf{y}_n = c^n\mathbf{y}_0 \bmod \mathbb{Z}^{2g}$ for $\mathbf{y}_0 \in \mathbb{R}^{2g}/\mathbb{Z}^{2g}$ and $c$ some integer. It was shown by Franklin [4] that, for almost all choices of $y_0$, the sequence defined as $y_n$ is uniformly distributed in $\mathbb{R}^{2g}$. Thus, the Jacobian determinant of our map $L$ can be used to calculate the density to which our sequence of points $P_n$ converges.

## 3. Empirical Results

In the case when $g = 1$, we have an elliptic curve $y^2 = x^3 + ax^2 + bx + c$. The points of the curve coincide with the Jacobian group. Accordingly, we can derive a group law for addition on the curve (see Lang [7]). Given a point $P = [x, y]$ on the elliptic curve, the duplication formula defines the $x$-coordinate of $2P$ as

$$x(2P) = \frac{x^4 - 2bx^2 - 8cx + b^2 - 4ac}{4(x^3 + ax^2 + bx + c)}.$$

Note that, although the use of elliptic curves over finite fields has recently been used by Hess and Shparlinski [5] for uniform random number generation, our method is fundamentally different and focuses on non-uniform distributions.

For our tests, a sequence was calculated by taking the $x$-coordinates of the sequence of points $P_n = 2^{101}P_{n-1}$ on an elliptic curve, with initial point $P_0$ having the $x$-coordinate equal to 100 (the $y$-coordinate was not relevant, as

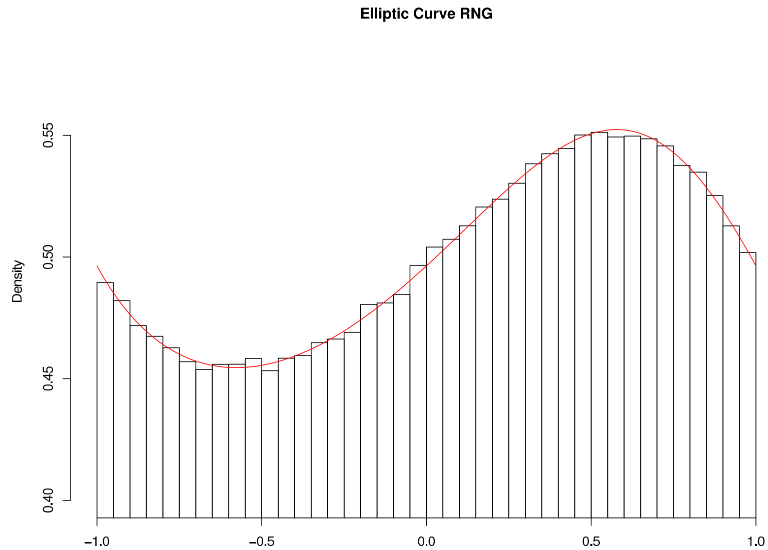


Figure 1: Comparison of pseudo-random versus theoretical density for the elliptic curve $y^2 = x^3 - x$ with 3524322 samples

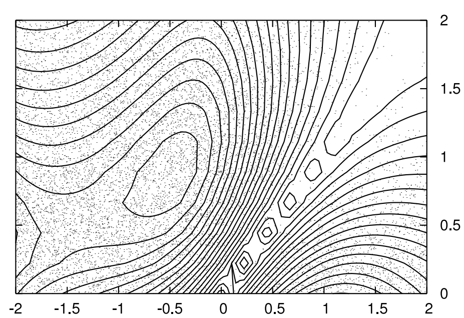


Figure 2: Scatter plot of density generated from the hyperelliptic curve $y^2 = x^5 + x^4 - 2x^3 + 7x^2 - x + 4$ versus contour plot of bivariate density proportional to $\frac{|x-y|}{\sqrt{(x^5+x^4-2x^3+7x^2-x+4)(y^5+y^4-2y^3+7y^2-y+4)}}$ on $[-2, 2] \times [0, 2]$

all our operations were duplications). Thus, the algorithm used was a simple execution of the following recursion (101 iterations per sample):

$$x \to \frac{x^4 - 2bx^2 - 8cx + b^2 - 4ac}{4(x^3 + ax^2 + bx + c)},$$

which theoretically leads to a density proportional to

$$1/\sqrt{x^3 + ax^2 + bx + c}.$$

The curve used for the arithmetic was $y^2 = x^3 - x$, and so the expected density was proportional to $\frac{1}{\sqrt{x^3-x}}$. Note that, as $x_0$ was rational, the entire sequence was rational (although we would expect rounding errors to propagate). As there is an isomorphism between the elliptic curve and a fundamental parallelogram in the complex plane, the asymptotic results for multiply sequences in Section 2 apply. Thus, for a multiplier as large as $2^{101}$, we anticipate excellent statistical results. Some theoretical statistical bounds are given by Pollanen [10].

We performed an empirical test in which a total of 3,524,322 samples were generated on the interval $[-1, 1]$ using double precision arithmetic with the GNU *C* compiler. Normalized elliptic logarithms were used to convert the samples into 24-bit "uniform" samples in $[0, 1)$. The resulting sequence of bits was subjected to Marsaglia's *DIEHARD* testing standard [8], which even some supposedly good pseudo-random number generators can fail. With a threshold of $\alpha = 0.001$ for the resulting $p$-values, the sequence passed each of the *DIEHARD* tests. A histogram of the actual density plotted against the theoretical density is presented in figure 1.

In Figure 2, we plotted a scatter plot of 10,000 samples of a bivariate density generated from the curve $y^2 = x^5 + x^4 - 2x^3 + 7x^2 - x + 4$, with an overlayed contour plot of the theoretical density. This method shows promise for generating classes of multivariate pseudo-random sequences related to hyperelliptic curves, and perhaps could be generalized to larger classes of distributions by employing more general plane algebraic curves (see Volcheck [11]).

## References

[1] D.G. Cantor, Computing in the Jacobian of a hyper-elliptic curve, *Math. Comp.*, **48** (1987), 95-101.

[2] C. Chevalley, *Introduction to the Theory of Algebraic Functions of One Variable*, AMS Surveys VI, New York (1951).

[3] J.N. Franklin, Deterministic simulation of random processes, *Math. of Comp.*, **17** (1963), 28-59.

[4] J.N. Franklin, Equidistribution of matrix-power residues modulo One, *Math. of Comp.*, **18** (1964), 560-568.

[5] F. Hess, I. Shparlinski, On the linear complexity of multidimensional distribution of congruential generators over elliptic curves, *Design. Code. Cryptogr.*, **35** (2005), 111-117.

[6] D.E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*, Volume 2, Addison-Wesley, New York (1998).

[7] S. Lang, *Elliptic Curves: Diophantine Analysis*, Springer-Verlag, New York (1978).

[8] G. Marsaglia, Diehard: A battery of tests for randomness, http://www.stat.fsu.edu/pub/diehard/ (1996).

[9] M. Pollanen, Formal group laws and non-uniform quasi-random sequences, *Int. J. Pure Appl. Math.*, **37** (2007), 79-100.

[10] M. Pollanen, Bounds on equipartition tests for multiply sequences, *Preprint* (2007).

[11] E.J. Volcheck, Computing in the Jacobian of a plane algebraic curve, In: *Proceedings of the First Algorithmic Number Theory Symposium*, Cornell University (1994).