RANDOM, PSEUDORANDOM, AND QUASIRANDOM NUMBERS AND THEIR GENERATION IN SERIAL AND PARALLEL

Prof. Dr. Michael Mascagni
E-mail: mascagni@math.ethz.ch
Home Page:
http://www.cs.fsu.edu/∼mascagni

Seminar für Angewandte Mathematik
Eidgenössische Technische
Hochschule, CH-8092, Zürich **Switzerland**
Department of Computer Science
School of Computational Science
Florida State University
Tallahassee, FL 32306 **USA**

Typeset by $\mathcal{AMS}$-TEX

Monte Carlo Methods: Numerical Experimental that Use Random Numbers

- A Monte Carlo method is any process that consumes random numbers

1. Each calculation is a numerical experiment
   - ▶ Subject to known and unknown sources of error
   - ▶ Should be reproducible by peers
   - ▶ Should be easy to run anew with results that can be combined to reduce the variance
2. Sources of errors must be controllable/isolatable
   - ▶ Programming/science errors under your control
   - ▶ Make possible RNG errors approachable
3. Reproducibility
   - ▶ Must be able to rerun a calculation with the same numbers
   - ▶ Across different machines (modulo arithmetic issues)
   - ▶ Parallel and distributed computers?

What are Random Numbers used for?

1. Random numbers are used extensively in simulation, statistics, and in *Monte Carlo* computations

- Simulation: use random numbers to "randomly pick" event outcomes based on statistical or experiential data
- Statistics: use random numbers to generate data with a particular distribution to calculate statistical properties (when analytic techniques fail)

2. There are many Monte Carlo applications of great interest

- Numerical Quadrature (see below)
- Quantum mechanics: Solving Schrödinger's equation with Green's function Monte Carlo via random walks

What are Random Numbers used for? (Cont.)

- Mathematics: Using the Feynman-Kac/path integral methods to solve partial differential equations with random walks
- Defense: neutronics, nuclear weapons design
- Finance: options, mortgage-backed securities

3. There are many types of random numbers

- "Real" random numbers: uses a "physical source" of randomness
- Pseudorandom numbers: deterministic sequence that passes tests of randomness
- Quasirandom numbers: well distributed (low discrepancy) points

The Classic Monte Carlo Application: Numerical Integration

- Consider computing $I = \int_0^1 f(x)\,dx$
- Conventional quadrature methods:

$$I \approx \sum_{i=1}^{N} w_i f(x_i)$$

A. *Rectangle:* $w_i = \frac{1}{N}$, $x_i = \frac{i}{N}$

B. *Trapezoidal:* $w_i = \frac{2}{N}, w_1 = w_N = \frac{1}{N}$, $x_i = \frac{i}{N}$

1. Monte Carlo quadrature:

$$I \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i), \quad x_i \text{ is } U[0,1],\ i.i.d.$$

2. Big advantage seen in multidimensional integration, consider (s-dimensions):

$$I = \int_{[0,1]^s} f(x_1, \ldots, x_s)\,dx_1 \ldots dx_s$$

The Classic Monte Carlo Application: Numerical Integration (Cont.)

3. Errors are significantly different, with N function evaluations we see the curse of dimensionality:

A. *Product trapezoidal rule: Error* $= O(N^{-2/s})$

B. *Monte Carlo: Error* $= O(N^{-1/2})$ *(indep. of s!!)*

4. Note: the errors are deterministic for the trapezoidal rule whereas the MCM error is a variance bound

5. For $s = 1$, $E[f(x_i)] = I$ when x_i is $U[0,1]$, so $E[\frac{1}{N}\sum_{i=1}^{N} f(x_i)] = I$, and $Var[\frac{1}{N}\sum_{i=1}^{N} f(x_i)] = Var[f(x_i)]/N$. $Var[f(x_i)] = \int_0^1 (f(x) - I)^2\, dx$

Pseudorandom Numbers

- Pseudorandom numbers mimic the properties of "real" random numbers

A. Pass statistical tests

B. Reduce error is $O(N^{-\frac{1}{2}})$ in Monte Carlo

- Some common pseudorandom number generators:

1. Linear congruential: $x_n = ax_{n-1} + c \pmod{m}$

2. Shift register: $y_n = y_{n-s} + y_{n-r} \pmod{2},\ r > s$

3. Additive lagged-Fibonacci:
$z_n = z_{n-s} + z_{n-r} \pmod{2^k},\ r > s$

4. Combined: $w_n = y_n + z_n \pmod{p}$

5. Multiplicative lagged-Fibonacci:
$x_n = x_{n-s} \times x_{n-r} \pmod{2^k},\ r > s$

6. Implicit inversive congruential:
$x_n = a\overline{x_{n-1}} + c \pmod{p}$

7. Explicit inversive congruential: $x_n = a\overline{n} + c \pmod{p}$

Pseudorandom Numbers (Cont.)

- Some properties of pseudorandom number generators, integers: $\{x_n\}$ from modulo m recursion, and $U[0,1], z_n = \frac{x_n}{m}$

A. Should be a purely period sequence (e.g.: DES and IDEA are not provably periodic)

B. Period length: $Per(x_n)$ should be large

C. Cost per bit should be moderate (not cryptography)

D. Should be based on theoretically solid and empirically tested recursions

E. Should be a totally reproducible sequence

Pseudorandom Numbers (Cont.)

- Some common facts (rules of thumb) about pseudorandom number generators:

1. Recursions modulo a power-of-two are cheap, but have simple structure

2. Recursions modulo a prime are more costly, but have higher quality: use Mersenne primes: $2^p - 1$, where p is prime, too

3. Shift-registers (Mersenne Twisters) are efficient and have good quality

4. Lagged-Fibonacci generators are efficient, but have some structural flaws

5. Combining generators is "provably good"

6. Modular inversion is very costly

7. All linear recursions "fall in the planes"

8. Inversive (nonlinear) recursions "fall on hyperbolas"

Periods of Pseudorandom Number Generators (RNGs)

1. Linear congruential: $x_n = ax_{n-1} + c \pmod{m}$, $\mathrm{Per}(x_n) = m - 1, m$ prime, with m a power-of-two, $\mathrm{Per}(x_n) = 2^k$, or $\mathrm{Per}(x_n) = 2^{k-2}$ if $c = 0$
2. Shift register: $y_n = y_{n-s} + y_{n-r} \pmod{2},\ r > s$, $\mathrm{Per}(y_n) = 2^r - 1$
3. Additive lagged-Fibonacci: $z_n = z_{n-s} + z_{n-r} \pmod{2^k},\ r > s$, $\mathrm{Per}(z_n) = 2^r - 1(2^{k-1})$
4. Combined: $w_n = y_n + z_n \pmod{p}$, $\mathrm{Per}(w_n) = \mathrm{lcm}(\mathrm{Per}(y_n), \mathrm{Per}(z_n))$
5. Multiplicative lagged-Fibonacci: $x_n = x_{n-s} \times x_{n-r} \pmod{2^k},\ r > s$,$\mathrm{Per}(x_n) = 2^r - 1(2^{k-3})$
6. Implicit inversive congruential: $x_n = a\overline{x_{n-1}} + c \pmod{p}$, $\mathrm{Per}(x_n) = p$
7. Explicit inversive congruential: $x_n = a\overline{n} + c \pmod{p}$, $\mathrm{Per}(x_n) = p$

Combining RNGs

- There are many methods to combine two streams of random numbers, $\{x_n\}$ and $\{y_n\}$, where the x_n are integers modulo m_x, and y_n's modulo m_y:

1. Addition modulo one: $z_n = \frac{x_n}{m_x} + \frac{y_n}{m_y} \pmod 1$

2. Addition modulo either m_x or m_y

3. Multiplication and reduction modulo either m_x or m_y

4. Exclusive "or-ing"

- Rigorously provable that linear combinations produce combined streams that are "no worse" than the worst
- Tony Warnock: all the above methods seem to do about the same

Splitting RNGs for Use In Parallel

- We consider splitting a single PRNG:
  - Assume $\{x_n\}$ has $Per(x_n)$
  - Has the fast-leap ahead property: leaping L ahead costs no more than generating $O(\log_2(L))$ numbers
- Then we associate a single block of length L to each parallel subsequence:

1. Blocking:
- First block: $\{x_0, x_1, \ldots, x_{L-1}\}$
- Second block: $\{x_L, x_{L+1}, \ldots, x_{2L-1}\}$
- ith block: $\{x_{(i-1)L}, x_{(i-1)L+1}, \ldots, x_{iL-1}\}$

2. The Leap Frog Technique: define the leap ahead of $\ell = \lfloor \frac{Per(x_i)}{L} \rfloor$:
- First block: $\{x_0, x_\ell, x_{2\ell}, \ldots, x_{(L-1)\ell}\}$
- Second block: $\{x_1, x_{1+\ell}, x_{1+2\ell}, \ldots, x_{1+(L-1)\ell}\}$
- ith block: $\{x_i, x_{i+\ell}, x_{i+2\ell}, \ldots, x_{i+(L-1)\ell}\}$

Splitting RNGs for Use In Parallel (Cont.)

3. The Lehmer Tree, designed for splitting LCGs:

- Define a right and left generator: $R(x)$ and $L(x)$
- The right generator is used within a process
- The left generator is used to spawn a new PRNG stream
- Note: $L(x) = R^W(x)$ for some W for **all** x for an LCG
- Thus, spawning is just jumping a fixed, W, amount in the sequence

4. Recursive Halving Leap-Ahead, use fixed points or fixed leap aheads:

- First split leap ahead: $\lfloor \frac{Per(x_i)}{2} \rfloor$
- ith split leap ahead: $\lfloor \frac{Per(x_i)}{2^{l+1}} \rfloor$
- This permits effective user of all remaining numbers in $\{x_n\}$ without the need for *a priori* bounds on the stream length L

Generic Problems with Splitting RNGs for Use In Parallel

1. Splitting for parallelization is not scalable:

- ◇ It usually costs $O(\log_2(Per(x_i)))$ bit operations to generate a random number
- ◇ For parallel use, a given computation that requires L random numbers per process with P processes must have $Per(x_i) = O((LP)^e)$
- ◇ Rule of thumb: never use more than $\sqrt{Per(x_i)}$ of a sequence $\rightarrow e = 2$
- ◇ Thus cost per random number is not constant with number of processors!!

2. Correlations within sequences are generic!!

- ◇ Certain offsets within any modular recursion will lead to extremely high correlations
- ◇ Splitting in any way converts auto-correlations to cross-correlations between sequences
- ◇ Therefore, splitting generically leads to inter-processor correlations in PRNGs

New Results in Parallel RNGs: Using Distinct Parameterized Streams in Parallel

1. Default generator: additive lagged-Fibonacci, $x_n = x_{n-s} + x_{n-r} \pmod{2^k},\ r > s$

- Very efficient: 1 add & pointer update/number
- Good empirical quality
- Very easy to produce distinct parallel streams

2. Alternative generator #1: prime modulus LCG, $x_n = ax_{n-1} + c \pmod{m}$

- Choice: Prime modulus (quality considerations)
- Parameterize the multiplier
- Less efficient than lagged-Fibonacci
- Provably good quality
- Multiprecise arithmetic in initialization

New Results in PRNGs: Using Distinct Parameterized Streams in Parallel (Cont.)

3. Alternative generator #2: power-of-two modulus LCG, $x_n = ax_{n-1} + c \pmod{2^k}$

- Choice: Power-of-two modulus (efficiency considerations)
- Parameterize the prime additive constant
- Less efficient than lagged-Fibonacci
- Provably good quality
- Must compute as many primes as streams

Parameterization Based on Seeding

◇ Consider the Lagged-Fibonacci generator:

$$x_n = x_{n-5} + x_{n-17} \pmod{2^{32}} \text{ or in general:}$$
$$x_n = x_{n-s} + x_{n-r} \pmod{2^k},\ r > s$$

◇ The seed is 17 32-bit integers; 544 bits, longest possible period for this linear generator is $2^{17\times 32} - 1 = 2^{544} - 1$

- Maximal period is $\text{Per}(x_n) = (2^{17} - 1) \times 2^{31}$

◦ Period is maximal $\iff$ at least one of the 17 32-bit integers is odd

- This seeding failure results in only even "random numbers"

◦ Are $(2^{17} - 1) \times 2^{31\times 17}$ seeds with full period

◦ Thus there are the following number of full-period equivalence classes (ECs):

$$E = \frac{(2^{17} - 1) \times 2^{31\times 17}}{(2^{17} - 1) \times 2^{31}} = 2^{31\times 16} = 2^{496}$$

The Equivalence Class Structure

◇ With the "standard" b_0:

m.s.b				l.s.b.	
b_{k-1}	b_{k-2}	...	b_1	b_0	
□	□	...	0	0	x_{r-1}
0	□	...	□	0	x_{r-2}
⋮	⋮	⋮	⋮	⋮	
□	0	...	□	0	x_1
□	□	...	□	1	x_0

◇ Or can choose a special b_0:

m.s.b				l.s.b.	
b_{k-1}	b_{k-2}	...	b_1	b_0	
□	□	...	□	b_{0n-1}	x_{r-1}
□	□	...	□	b_{0n-2}	x_{r-2}
⋮	⋮	⋮	⋮	⋮	
□	□	...	□	b_{01}	x_1
0	0	...	0	b_{00}	x_0

Parameterization of Prime Modulus LCGs

- ◇ Consider only $x_n = ax_{n-1} \pmod{m}$, with m prime has maximal period when a is a primitive root modulo m
- ◇ If α and a are primitive roots modulo m then $\exists\ l$ s.t. $\gcd(l, m-1) = 1 \& \alpha \equiv a^l \pmod{m}$
- ◇ If $m = 2^{2^n} + 1$ (Fermat prime) then all odd powers of α are primitive elements also
- ◇ If $m = 2q+1$ with q also prime (Sophie-Germain prime) then all odd powers (save the qth) of α are primitive elements
- ○ Consider $x_n = ax_{n-1} \pmod{m}$ and $y_n = a^l y_{n-1} \pmod{m}$ and define the full-period exponential-sum cross-correlation between then as:

$$C(j,l) = \sum_{n=0}^{m-1} e^{\frac{2\pi i}{m}(x_n - y_{n-j})}$$

then the Riemann hypothesis over finite-fields implies $|C(j,l)| \leq (l-1)\sqrt{m}$

Parameterization of Prime Modulus LCGs (Cont.)

- Mersenne modulus: relatively easy to do modular multiplication
- With Mersenne prime modulus, $m = 2^p - 1$ must compute $\phi_{m-1}^{-1}(k)$, the kth number relatively prime to $m - 1$
- Can compute $\phi_{m-1}(x)$ with a variant of the Meissel-Lehmer algorithm fairly quickly:
  - Use partial sieve functions to trade off memory for more than 2^j operations, $j = \#$ of factors of $m - 1$
  - Have fast implementation for $p = 31, 61, 127, 521, 607$

Parameterization of Power-of-Two Modulus LCGs

- $x_n = ax_{n-1} + c_i \pmod{2^k}$, here the c_i's are distinct primes
- Can prove (Percus and Kalos) that streams have good spectral test properties among themselves
- Best to choose $c_i \approx \sqrt{2^k} = 2^{k/2}$
- Must enumerate the primes, uniquely, not necessarily exhaustively to get a unique parameterization
- Note: in $0 \leq i < m$ there are $\approx \frac{m}{\log m}$ primes via the prime number theorem, thus if $m \approx 2^k$ streams are required, then must exhaust all the primes modulo $\approx 2^{k+\log k} = 2^k k = m \log m$
- Must compute distinct primes on the fly either with table or something like Meissel-Lehmer algorithm

Quality Issues in Serial and Parallel PRNGs

◇ Empirical tests (more later)

◇ Provable measures of quality:

1. Full- and partial-period discrepancy (Niederreiter) test equidistribution of overlapping k-tuples
2. Also full- ($k = \mathrm{Per}(x_n)$) and partial-period exponential sums:

$$C(j,k) = \sum_{n=0}^{k-1} e^{\frac{2\pi i}{m}(x_n - x_{n-j})}$$

◦ For LCGs and SRGs full-period and partial-period results are similar

- ▶ $|C(\cdot, \mathrm{Per}(x_n))| < O(\sqrt{\mathrm{Per}(x_n)})$
- ▶ $|C(\cdot, j)| < O(\sqrt{\mathrm{Per}(x_n)})$

• Additive lagged-Fibonacci generators have poor provable results, yet empirical evidence suggests $|C(\cdot, \mathrm{Per}(x_n))| < O(\sqrt{\mathrm{Per}(x_n)})$

Parallel Neutronics: A Difficult Example

1. The structure of parallel neutronics

- Use a parallel queue to hold unfinished work
- Each processor follows a distinct neutron
- Fission event places a new neutron(s) in queue with initial conditions

2. Problems and solutions

- Reproducibility: each neutron is queued with a new generator (and with the next generator)
- Using the binary tree mapping prevents generator reuse, even with extensive branching
- A global seed reorders the generators to obtain a statistically significant new but reproducible result

Many Parameterized Streams Facilitate Implementation and Use

1. Advantages of using parameterized generators
   - Each unique parameter value gives an "independent" stream
   - Each stream is uniquely numbered
   - Numbering allows for absolute reproducibility, even with queuing on MIMD platforms
   - Effective serial implementation + enumeration yield a portable scalable implementation
   - Provides theoretical testing basis
2. Implementation details
   - Generators mapped canonically to a binary tree
   - Extended seed data structure contains current seed and next generator
   - Spawning uses new next generator as starting point: assures no reuse of generators
3. All these ideas in the **S**calable **P**arallel **R**andom **N**umber **G**enerators (`SPRNG`) library

Many Different Generators and A Unified Interface

1. Advantages of having more than one generator
   - An application exists that stumbles on a given generator
   - Generators based on different recursions allow comparison to rule out spurious results
   - Makes the generators real experimental tools
2. Two interfaces to the `SPRNG` library: simple and default
   - Initialization returns a pointer to the generator state: `init_SPRNG()`
   - Single call for new random number: `SPRNG()`
   - Generator type chosen with parameters in `init_SPRNG()`
   - Makes changing generator very easy
   - Can use more than one generator *type* in code
   - Parallel structure is extensible to new generators through dummy routines

Implementation, Testing, and Extensibility

1. Implementational details (some yet to be done)

- Common parallel interface/design (C, FORTRAN, C++, MPI)
- HPF: not by `SPRNG` group but Portland Group
- Distribution of public domain software through GAMS, netlib, National High Performance Software Exchange, and collaboration

2. Target architectures

- IBM SP
- SGI systems
- Convex/HP SPPs
- Clusters
- Cray machines
- Distributed machines (heterogeneous)
- Computational Grid

Implementation, Testing, and Extensibility (Cont.)

3. Testing Routines

- Standard randomness tests (e.g. Marsaglia's DIEHARD suite & Knuth tests)
- Parallel randomness tests: includes processor-wise versions of serial randomness tests and (linear) exponential sums
- Application (physics) motivated tests: Ising model computations, first-passage computations, etc.

4. Extensibility

- Many researchers have own favorite PRNG (fanatical sects exist)
- User supplied generator must have an explicit parameterization, but splitting is possible
- Uses library's existing parallel infrastructure
- Testing routines are crucial to ensure correct implementation and overall suitability

Quasirandom Numbers

- Many problems require uniformity, not randomness (independence): use a deterministic sequence with small *star discrepancy*: so-called "quasirandom" numbers
- **Definition**: The *star discrepancy* D_N^* of $x_1, \ldots, x_N \in \overline{J}$ (measure of uniformity):

$$D_N^* = D_N^*(x_1, \ldots, x_N) =$$

$$\sup_{0 \le u \le 1} \left| \frac{1}{N} \sum_{n=1}^{N} \chi_{[0,u)}(x_n) - u \right|,$$

where χ is the characteristic function

- **Theorem** (Koksma, 1942): if $f(x)$ has bounded variation $V(f)$ on $[0,1]$ and $x_1, \ldots, x_N \in [0,1]$ with star discrepancy D_N^*, then:

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(x_n) - \int_0^1 f(x)\, dx \right| \le V(f) D_N^*$$

◇ Note: Many different types of discrepancies are definable

Some Types of Quasirandom Numbers (Cont.)

- ◇ Must choose point sets (finite #) or sequences (infinite #) with small D_N^*
- ◇ Often used is the *van der Corput sequence* in base b: $x_n = \Phi_b(n-1), n = 1, 2, \ldots,$ where for $b \in \mathbb{Z}, b \geq 2$:

$$\Phi_b\left(\sum_{j=0}^{\infty} a_j b^j\right) = \sum_{j=0}^{\infty} a_j b^{-j-1} \quad \text{with}$$

$$a_j \in \{0, 1, \ldots, b-1\}$$

  For v.d. Corput sequ. $ND_N^* \leq \frac{\log N}{3 \log 2} + O(1)$
- ◇ With $b = 2$, we get $\{\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8} \ldots\}$
- ◇ With $b = 3$, we get $\{\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9} \ldots\}$

Some Types of Quasirandom Numbers (Cont.)

- Other small D_N^* points sets and sequences:

1. Halton sequence: $\mathbf{x}_n = (\Phi_{b_1}(n-1), \ldots, \Phi_{b_s}(n-1))$, $n = 1, 2, \ldots$, $D_N^* = O\left(N^{-1}(\log N)^s\right)$ if $b_1, \ldots, b_s$ pairwise relatively prime

2. Hammersley point set: $\mathbf{x}_n = \left(\frac{n-1}{N}, \Phi_{b_1}(n-1), \ldots, \Phi_{b_{s-1}}(n-1)\right)$, $n = 1, 2, \ldots, N$, $D_N^* = O\left(N^{-1}(\log N)^{s-1}\right)$ if $b_1, \ldots, b_{s-1}$ are pairwise relatively prime

3. Ergodic dynamics: $\mathbf{x}_n = \{n\alpha\}$, where $\alpha = (\alpha_1, \ldots, \alpha_s)$ is irrational and $\alpha_1, \ldots, \alpha_s$ are linearly independent over the rationals then for almost all $\alpha \in \mathbb{R}^s$, $D_N^* = O(N^{-1}(\log N)^{s+1+\epsilon})$ for all $\epsilon > 0$

4. Other methods of generation

- ◇ Method of good lattice points (Sloan and Joe)
- ◇ Soboĺ sequences
- ◇ Faure sequences
- ◇ Niederreiter sequences

Some Types of QRNs (Cont.)

1. Another interpretation of the v.d. Corput sequence:

- ◇ Define the ith ℓ-bit "direction number" as: $v_i = 2^i$ (think of this as a bit vector)
- ◇ Think of $n-1$ via its base-2 representation $n-1 = b_{\ell-1}b_{\ell-2}\ldots b_1 b_0$
- ◇ Thus we have

$$\Phi_2(n-1) = 2^{-\ell} \bigoplus_{i=0,\ b_i=1}^{i=\ell-1} v_i$$

2. The Soboĺ sequence works the same!!

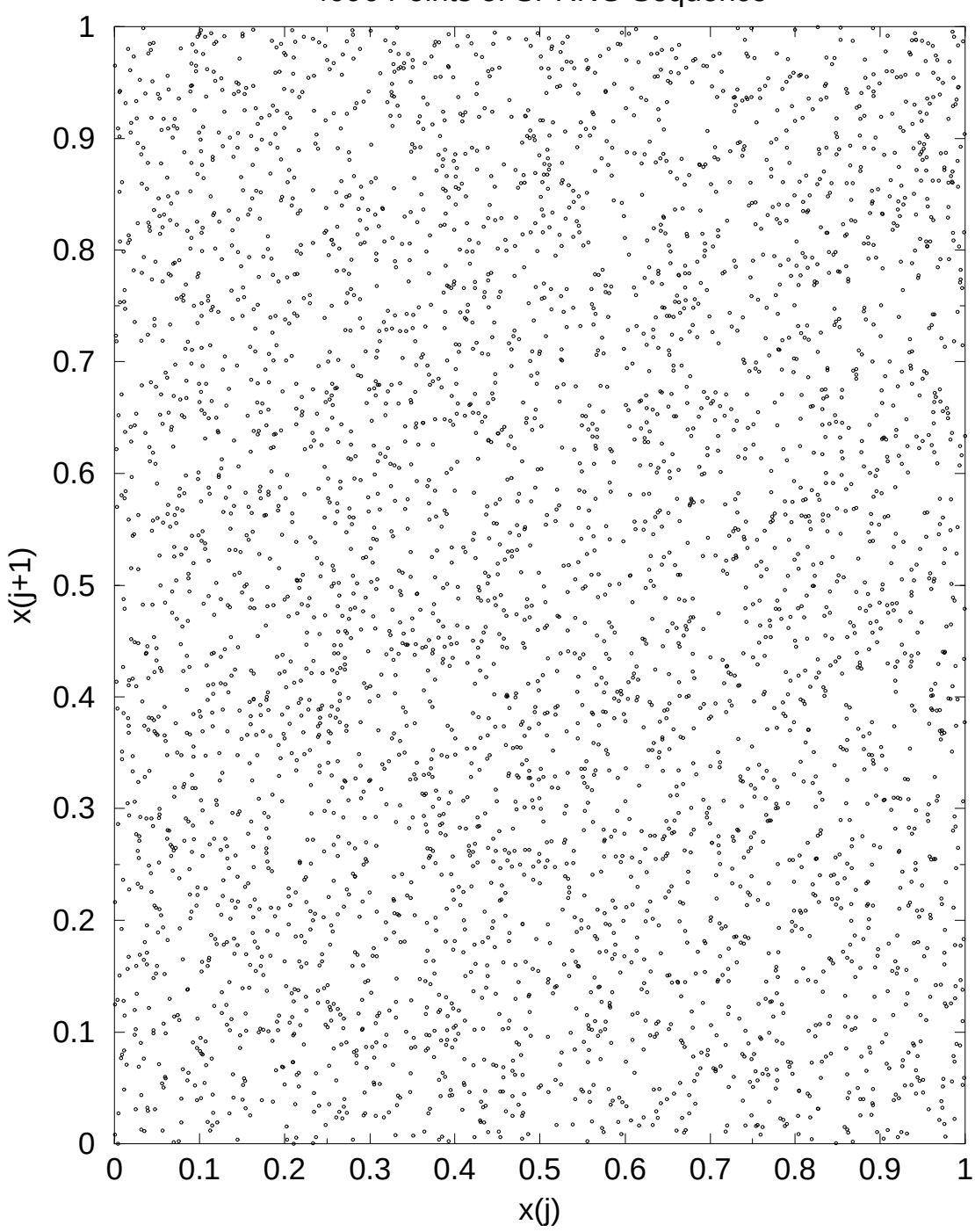
- ◇ Use recursions with a primitive binary polynomial define the (dense) v_i
- ◇ The Soboĺ sequence is defined as:

$$s_n = 2^{-\ell} \bigoplus_{i=0,\ b_i=1}^{i=\ell-1} v_i$$

- ◇ For speed of implementation, we use Gray-code ordering

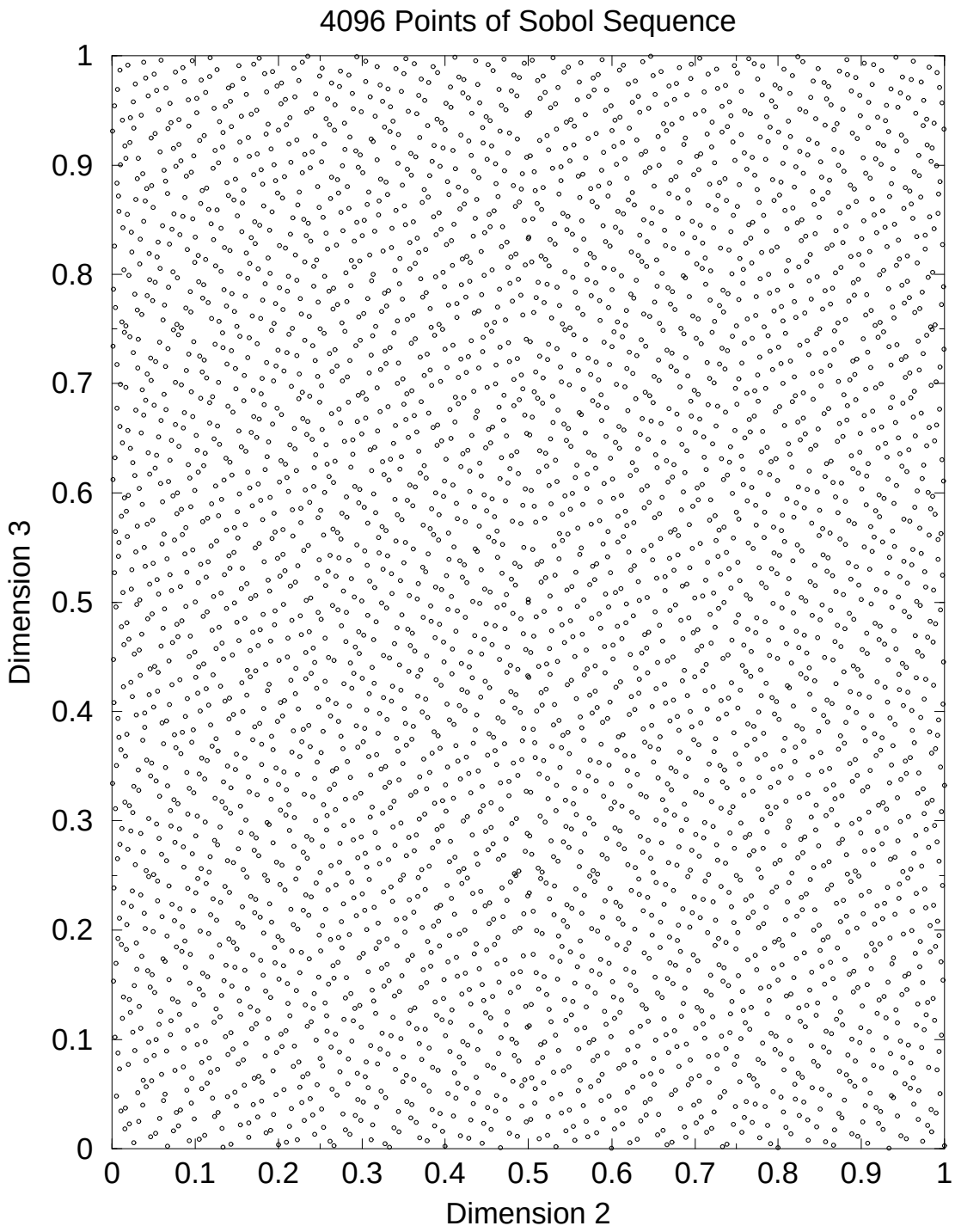
A Picture is Worth a Thousand Words: 4096 Pseudorandom Pairs

SPRNG Sequence



A Picture is Worth a Thousand Words: 4096 Quasirandom Pairs

2−D Projection of Sobol' Sequence



Possible Student Projects on Random Numbers

1. `SPRNG` available at: `http://sprng.cs.fsu.edu`
   - New generators: Well, Mersenne Twister, LCGs, etc.
   - Spawn-intensive generators
   - Generators with small memory footprints
   - More comprehensive testing suite
   - C++ implementation
   - Adding a quasirandom number sequence
2. Basic Random Number Research
   - Improved methods for initializing shift-register sequences
   - Fast $GF(2)$ library for polynomial computations
   - Optimal quasirandom numbers
   - Comparison of quasirandom numbers with sparse grids
   - Improved theoretical tests