CS 812
Lecture 41
Friday 5/1/20

Fine Structure of Pseudo-Random Sequences

Reference: G. Marsaglia, Proc. Nat. Acad. Sci. (USA), 1969

See Also:

G. Marsaglia, Numerische Mathematik, 1970

Knuth, ACP v. 2, Section 3.3.4.

Overview

Consider a multiplicative pseudo-random sequence defined by the iteration

$$r_i = kr_{i-1} \bmod m, \qquad i = 1, 2, 3, \ldots \ .$$

It is common to normalize this and set $u_i = r_i/m$, so that $0 \leq u_i < 1$.

Marsaglia's main result is that successive length n "windows" of the normalized sequence

$$u_1, u_2, u_3, \ldots$$

lie on a small number of parallel hyperplanes. Thus, the pseudo-random sequence has a structure that disinguishes it from a truly random sequence.

This is another nice application of lattices. We will not use the LLL algorithm but rather a famous "existence" result of Minkowski.

Geometry of Numbers

Theorem: Let L be a full-rank lattice in $\mathbf{R}^n$. Let $C \subset \mathbf{R}^n$ be convex and symmetric around the origin. If the volume of C exceeds $2^n \det(L)$, then C contains a lattice point different from 0.

Convex means that if $a, b \in C$, then the line segment linking a with b lies within C.

The symmetry condition means that $x \in C$ iff $-x \in C$.

Example: take $n = 2$, and let $L = \mathbf{Z}^2$. Let C be the disk of radius r centered at the origin. The area of C is πr^2, and this exceeds 4 when $r > 2/\sqrt{\pi} = 1.128\ldots$. In this case there is room to spare, and $(1, 0) \in C$.

It will be enough to prove the theorem for the case $L = \mathbf{Z}^n$, which has determinant 1.

Lemma (Blichfeldt's principle): Let $D \subset \mathbf{R}^n$ have volume > 1. Then there are two distinct points of D whose coordinates differ by integers.

Proof of the lemma: Partition $\mathbf{R}^n$ into axis-aligned cubes with unit-length edges. Precisely, if $z = (z_1, \ldots, z_n) \in \mathbf{Z}^n$, the z-th cube B_z is given by $z_i \leq x_i < z_i + 1$, $i = 1 \ldots, n$. Let

$$D = \bigcup D_z \qquad \text{(disjoint)},$$

where $D_z = D \cap B_z$. Note that $\sum_z \text{Vol}(D_z) = \text{Vol}(D) > 1$. To derive a contradiction, suppose that all $D_z - z$ are disjoint. Since all $D_z - z \subset B_0$, we'd have

$$\sum_z \text{Vol}(D_z - z) \leq \text{Vol}(B_0) = 1.$$

So there is an overlap, say $x - z = y - z'$, making $x - y \in \mathbf{Z}^n$. (Note that $z - z' \neq 0$, so $x \neq y$.)

Proof of the theorem: Let $D = \frac{1}{2}C$. By hypothesis, $\text{Vol}(D) > 1$. So there are $v_1 \neq v_2 \in D$ with $v_1 - v_2 \in \mathbf{Z}^n$. Then,

$$0 \neq v_1 - v_2 = \frac{(2v_1) + (-2v_2)}{2} \in C,$$

by convexity and symmetry.

I learned this proof from a web site by Alexander Gorodnik of U. Bristol.

Marsaglia's Theorem (The Random Numbers Lie Mainly in the Planes)

Let $f(r) = kr$, a self-map on $\mathbf{Z}_m$, and let $u_1, u_2, \ldots$ be the normalized version of the sequence generated by iterating f.

The successive "windows" of the sequence are

$$\begin{aligned} \pi_1 &= u_1 \ldots u_n \\ \pi_2 &= u_2 \ldots u_{n+1} \\ &\cdots \end{aligned}$$

Note that the state evolution for the unnormalized windows is basically a shift register

$$[\,] \quad \leftarrow \quad [\,] \quad \leftarrow \quad \cdots \quad \leftarrow [\,] \quad \supset$$

where the loop on the right (sorry no arrowhead yet) indicates multiplication by k in $\mathbf{Z}_m$.

Theorem: Let $c = (c_1, \ldots, c_n) \in \mathbf{Z}^n$ satisfy

$$c_1 + c_2 k + \cdots c_n k^{n-1} \equiv 0 \pmod{m}.$$

Then all windows $\pi_1, \pi_2, \ldots$ lie on the hyperplanes defined by

$$cx = 0, \pm 1, \pm 2, \ldots \ .$$

We can pick such a c so that all windows lie within $\leq (n!m)^{1/n}$ of these parallel hyperplanes.

Before proving this, let's think about the bound a little. Stirling's formula is

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

So

$$(n!)^{\frac{1}{n}} \sim 2\pi^{\frac{1}{2n}} e^{\frac{\log n}{2n}} \frac{n}{e} \sim \frac{n}{e},$$

since the first two factors tend to 1. So for m fixed and n large, $(n!m)^{\frac{1}{n}} = O(n)$.

Proof (following Marsaglia): The j-th window is

$$\pi_j = (\frac{sk^j}{m} + t_1, \ldots, \frac{sk^{j+n-1}}{m} + t_n),$$

where the t_i's are integers. So

$$\begin{aligned} c\pi_j &= c_1\frac{sk^j}{m} + \cdots + c_n\frac{sk^{j+n-1}}{m} + [\text{integer}] \\ &= \frac{sk^{j-1}}{m}(c_1k + \cdots + c_nk^n) + [\text{integer}] \\ &\in \mathbf{Z}. \end{aligned}$$

The number of hyperplanes $cx = [\text{integer}]$ intersecting the unit box $[0,1)^n$ is $\leq \sum |c_i|$. (Proof: We can assume the integers are all nonnegative; if not, replace c by $-c$. Then for $0 \leq x_i < 1$, $\sum c_i x_i \leq \sum |c_i|$.) Now we want a "short" vector c in the lattice L defined by

$$c_1 + c_2k + \cdots + c_nk^{n-1} \equiv 0 \pmod{m}.$$

Please check that this is indeed a lattice.

It will suffice to find such a vector in the lattice L_k generated by the rows of

$$\begin{pmatrix} m & 0 & 0 & 0 & \cdots & 0 & 0 \\ -k & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -k & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -k & 1 & \cdots & 0 & 0 \\ & & & & \vdots & & \\ 0 & 0 & 0 & 0 & \cdots & -k & 1 \end{pmatrix}$$

Note this matrix has determinant m.

We know $L_k \subset L$. Marsaglia claims they are equal. Try to prove that.

Now we apply the Minkowski theorem to the lattice L_k and the convex body C defined by $||x||_1 \leq R$.

Volume computation: the unit simplex has volume $1/n!$, so $||x||_1 \leq 1$ has volume $2^n/n!$. By scaling, the volume of C is $(2R)^n/n!$.

The Minkowski theorem applies as soon as

$$\frac{(2R)^n}{n!} > 2^n m,$$

that is, when $R > (n!m)^{1/n}$.

To get the conclusion of the theorem, consider the short nonzero vectors $c, c', c'', \ldots$ produced by applying the theorem succesively to a sequence of R's descending monotonically to the limit $(n!m)^{1/n}$. Since there are a finite number of possible vectors, there must be one appearing infinitely often in the sequence, and that one will have 1-norm $\leq (n!m)^{\frac{1}{n}}$, as claimed.