Discrete quasirandomness: questions and applications

Jim Propp
(U. Wisconsin)

March 25, 2006

(these slides are available at
`www.math.wisc.edu/~propp/Erdos.pdf`)

Introduction

In my title and in this talk, "quasirandom" means quasirandom in the sense introduced by Niederreiter et al. in the late '70s (aka "subrandom"), *not* quasirandom in the sense introduced by Chung, Graham, and Wilson in the late '80s.

Informally, quasirandom means "like random, but less lumpy".

Continuous quasirandomness:

Example: the van der Corput sequence

$$(1/2, 1/4, 3/4, 1/8, 5/8, \ldots)$$

(whose nth term is $a/2+b/4+c/8+\ldots$ where $n = a + 2b + 4c + \ldots$), used as a substitute for a sequence of i.i.d. random numbers chosen uniformly from $[0, 1]$.

Desideratum: Each subinterval of $[0, 1]$ gets hit about as often as any other (cf. a truly random sequence whose histogram wouldn't be as flat).

Discrete quasirandomness:

Example: the period-8 de Bruijn sequence ..., 0, 0, 0, 1, 1, 1, 0, 1, ... as a substitute for a sequence of i.i.d. random bits.

Desideratum: Each short bit-pattern occurs as often as any other.

Common theme: reducing discrepancy below what's achieved by random processes

Better title for this talk: "Minimizing discrepancy: questions and applications"

Estimating pi with derandomized random walk

Folk theorem: The probability that a random walker in $\mathbf{Z}^2$ starting at $(0,0)$ will hit $(1,1)$ before returning to $(0,0)$ is $\pi/8$.

This gives a really slow way to approximate π by doing repeated independent rounds of random walk: estimate $\pi/8$ by the rational number K/N, where N is the number of trials and K is the number of successful trials (where success means the walker hit $(1,1)$ before returning to $(0,0)$).

For most N's, the lowest error you can hope for is on the order of $1/N$.

Typically, you should expect errors on the order of $1/\sqrt{N}$.

Derandomize: Whenever the walker leaves a site $\neq (1,1)$, he always leaves in the direction 90 degrees clockwise from the direction he used the last time he left that site.

This is the "rotor-router" mechanism for derandomization; for a Java implementation, see
`www.math.wisc.edu/~propp/rotor-router-1.0/`

The exits from each site satisfy a low discrepancy property: for any two neighbors t, t' of site s, the number of exits from s to t and the number of exits from s to t' differ by at most 1 (cf. "by at most $\sqrt{N}$ on average" for random walk, where N is the number of visits to s so far).

Question 1: How quickly does the success-ratio K_N/N approach $\pi/8$ as $N \to \infty$?

Assessed difficulty: Hard

Partial results:

It's known that K_N/N approaches $\pi/8$ with error $O(1/\log N)$.

It has been observed that for $N \leq 10^4$, the error never exceeds $2/N$. (Recall that for the worst N's, $1/2N$ is the smallest error you can achieve with *any* approximation K/N. In contrast, random walk achieves $O(1/\sqrt{N})$.)

For over half of these 10^4 values of N, the derandomized walk gives the best fraction with denominator N, that is, the unique one that is closest to $\pi/8$.

Using derandomized random walk for aggregation in $\mathbf{Z}^2$

Internal diffusion-limited aggregation: Do random walk from a fixed source until you reach a site that you've never visited before; return immediately to the source and continue walking from there. Repeat ad infinitum.

In $\mathbf{Z}^2$, the set of sites visited by time T, suitably rescaled, converges almost surely to a perfect disk as $T \to \infty$.

The derandomized version of this aggregation process makes 4-colored mandalas Ed Pegg dubbed "Propp circles".

See Michael Kleber's article "Goldbug Variations" in the Winter 2005 issue of The Mathematical Intelligencer.

A conjecture both deep and pro-
found
Is whether the circle is round.
In a paper of Erdös
Written in Kurdish
A counterexample is found.

(Leo Moser)

A conjecture a bit less profound
Is whether Propp circles are round.
The verdict is final
Since Yuval and Lionel
Came up with a two-sided bound[1].

[1] L. Levine and Y. Peres, Spherical asymptotics for the rotor-router model in Z^d, preprint, 2005 (and subsequent refinements, not yet published).

Question 2: How big are the radial fluctuations?

Assessed difficulty: Extremely hard

Partial results: The radial fluctuations appear to be on the order of $\log(R)$ or even smaller (perhaps even $O(1)$), but all Levine and Peres can show is that they're $o(R^c)$ for every $c > 1/2$.

See `www.math.wisc.edu/~propp/million.gif`

Beyond rotors

The first part of the talk has been about the underappreciated virtues of periodic sequences as surrogates for random sequences.

But now let's impose more stringent requirements that will rule out periodic sequences.

An Erdös problem, dating back to 1927 or earlier:

Let $f : \mathbf{N} \to \{-1, +1\}$ be an arbitrary 2-coloring of the natural numbers. Is it true that the sup of

$$|\sum_{i=1}^{n} f(id)|,$$

for n and d positive integers, is infinity?

Turning it around:

Does there exist a 2-coloring f such that the sums

$$\sum_{i=1}^{n} f(id),$$

for n and d positive integers, are bounded (in absolute value) by some B?

Still unsolved!

A variant:

Does there exist a 2-coloring f such that for every $d > 0$, the sums

$$\sum_{i=1}^{n} f(c + id),$$

for n and c positive integers, are bounded by some $B(d)$?

I was going to pose this as an open problem here, but David Feldman at UNH showed that the answer is YES.

His example begins $+1, -1, +1, -1, ...$ and has generating function

$$\begin{aligned} 1 - x + x^2 - x^3 ... = \\ (1 - x) \\ \times (1 + x^2 - x^4 - x^6) \\ \times (1 + x^8 + x^{16} - x^{24} - x^{32} - x^{40}) \\ \times \ ... \end{aligned}$$

So instead I'll ask:

Question 3: Does there exist a 2-coloring f such that for every $d > 0$, the sums

$$\sum_{i=1}^{n} f(i)f(i+d),$$

for n a positive integer, are bounded by some $B(d)$?

Assessed difficulty: Moderately easy

Partial results: None.

The Ehrenfeucht-Mycielski sequence

$$0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, ...$$

Rule: to compute a_{n+1} from $a_1 a_2 \ldots a_n$, find the longest suffix $a_m a_{m+1} \ldots a_n$ that occurs earlier in the sequence; if the most recent earlier occurrence is

$$a_{m-d}\ a_{m+1-d}\ \cdots\ a_{n-d}\ ,$$

let a_{n+1} be the complement of the bit a_{n-d+1}.

Question 4: Does the density of 1's converge to 1/2?

Assessed difficulty: Moderately hard

Partial results: See work of Klaus Sutner, `sutner@cs.cmu.edu`.

Derandomized aggregation in the Stern-Brocot tree

Background: If you arrange the dyadic rationals in $(0, 1)$ as vertices of a binary tree rooted at $1/2$, it's natural to do a random walk on this tree, where each arc pointing away from the root (e.g., $1/2 \to 1/4$, $1/2 \to 3/4$, etc.) has probability $1/2$. This walk corresponds to forming a random sum $\frac{1}{2} \pm \frac{1}{4} \pm \frac{1}{8} \pm \ldots$, whose limiting distribution is Lebesgue measure on $(0, 1)$.

If you derandomize this walk with rotors, and do rotor-router aggregation in the binary tree, you pick up the dyadic rationals in the order $1/2, 1/4, 3/4, 1/8$ $5/8, 3/8, 7/8, \ldots$ (the van der Corput sequence) which is very evenly spread.

There's a natural way to arrange *all* the rationals in $(0, 1)$ as vertices of a binary directed tree, rooted at $1/2$.

There's a unique way to assign a probability to each arc so that the random walk "goes to" Lebesgue measure on $(0, 1)$.

You can derandomize this walk with rotors. For one natural way of initializing the rotors, you pick up the rationals in the order $1/2, 1/3, 2/3, 1/4, 3/4, 1/5, 4/5, 2/5, 3/5, ...$ How evenly-spaced is this sequence? Specifically:

Question 5: If we fix n and order the first n terms as $a_1 < a_2 < ... < a_n$, how big is

$$\sum_{i=1}^{n} |a_i - i/n|?$$

Assessed difficulty: Incredibly hard

In particularly the assertion that

$$\sum_{i=1}^{n} |a_i - i/n|$$

is $o(n^c)$ for every $c > 1/2$ is probably equivalent to the Riemann Hypothesis!