Pseudorandom Sequences III: Measures of Pseudorandomness

Arne Winterhof

Austrian Academy of Sciences
Johann Radon Institute for Computational and Applied Mathematics
Linz

Carleton University 2010

Pseudorandom sequences are generated by a deterministic algorithm and 'look random'.

The 'randomness properties' and the corresponding measures depend on the application!

cryptography: unpredictability $\longrightarrow$ linear complexity
numerical integration: uniform distribution $\longrightarrow$ discrepancy
radar: comparison with reflected signal $\longrightarrow$ autocorrelation

Measures for binary sequences

Let $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, 1\}^N$ be a finite binary sequence. Then the well-distribution measure of E_N is defined as

$$W(E_N) = \max_{M,u,v} \left| \sum_{j=0}^{M-1} e_{u+jv} \right|,$$

where the maximum is taken over all M, u, v with $u + (M-1)v \leq N$, and the correlation measure of order k of E_N is defined as

$$C_k(E_N) = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \cdots e_{n+d_k} \right|,$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ and M such that $0 \leq d_1 < d_2 < \ldots < d_k \leq N - M$.

Expected value

Alon et al., 2007:
For a 'truly random' binary sequence E_N of length N the measures $W(E_N)$ and $C_k(E_N)$ (for fixed k) are both $O(N^{1/2}(\log N)^{c(k)})$.

Legendre sequences

$N = p > 2$ prime
$f(X) \in \mathbb{F}_p[X]$ squarefree, $\deg(f) = d \geq 1$

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right), & f(n) \neq 0, \\ 1, & f(n) = 0. \end{cases}$$

Mauduit/Sarközy, 1997:

$$W(E_N) \leq dp^{1/2} \log p$$

$$C_k(E_N) \leq dkp^{1/2} \log p$$

(Polya-Vinogradov method + Weil bound)

Elliptic curve Legendre sequences

$E : y^2 = x^3 + ax + b$, elliptic curve over $\mathbb{F}_p$, $p > 3$
P finite point on E, $f(x, y) \in \mathbb{F}_p[x, y]$ squarefree, $\deg(f) = d \geq 1$

$$e_n = \begin{cases} \left(\frac{f(nP)}{p}\right), & f(nP) \neq 0, \\ 1, & f(nP) = 0. \end{cases}$$

Chen, 2008:

$$W(E_N) \leq dp^{1/2} \log p$$

$$C_k(E_N) \leq dkp^{1/2} \log p$$

Extension to Jacobi symbol (two-prime generator)

$p \neq q$ odd primes, $N = pq$

$$e_n = \begin{cases} \left(\frac{n}{p}\right)\left(\frac{n}{q}\right), & \gcd(n, pq) = 1, \\ 1, & \gcd(n, pq) > 1. \end{cases}$$

Features

- low C_k if $k = 2$ or k is odd
- C_k large if $k > 2$ is even

$$
\begin{aligned}
& \sum_{n=1}^{M} e_n e_{n+p} e_{n+q} e_{n+p+q} \\
\approx\ & \sum_{n=1}^{M} \left(\frac{n}{p}\right)\left(\frac{n}{q}\right)\left(\frac{n+p}{p}\right)\left(\frac{n+p}{q}\right) \\
& \qquad \left(\frac{n+q}{p}\right)\left(\frac{n+q}{q}\right)\left(\frac{n+p+q}{p}\right)\left(\frac{n+p+q}{q}\right) \\
\approx\ & M
\end{aligned}
$$

Measures for quaternary sequences

Let denote by $\mathcal{F}$ the set of the 24 permutations of $\mathcal{E} = \{-1, 1, -i, i\}$. For a quaternary sequence $G_N = (g_1, g_2, \ldots, g_N) \in \mathcal{E}^N$ the well-distribution measure of G_N is defined as

$$\Delta(G_N) = \max_{\varphi, M, u, v} \left| \sum_{j=0}^{M-1} \varphi(g_{u+jv}) \right|, \tag{1}$$

where the maximum is taken over all $\varphi \in \mathcal{F}$ and M, u, v with $u + (M-1)v \le N$, and the correlation measure of order k of G_N is defined as

$$\Gamma_k = \max_{\Phi, M, D} \left| \sum_{n=1}^{M} \varphi_1(g_{n+d_1}) \varphi_2(g_{n+d_2}) \cdots \varphi_k(g_{n+d_k}) \right|,$$

where the maximum is taken over all $\Phi = (\varphi_1, \varphi_2, \ldots, \varphi_k) \in \mathcal{F}^k$, $D = (d_1, d_2, \ldots, d_k)$ and M such that $0 \le d_1 < d_2 < \ldots < d_k \le N - M$.

Biquadratic character sequence

Let p be a prime with $p \equiv 1 \bmod 4$ and χ a character of the finite field $\mathbb{F}_p$ of p elements of order 4. Then we define the quaternary sequence $G_{p-1} = (\chi(1), \chi(2), \ldots, \chi(p-1)) \in \mathcal{E}^{p-1}$.
Mauduit/Sarközy, 2002:

$$\Delta(G_{p-1}) = O(p^{1/2} \log p)$$

$$\Gamma_k(G_{p-1}) = O(4^k k p^{1/2} \log p)$$

From quaternary to binary sequences and back

Let $G_N = (g_1, g_2, \ldots, g_N) \in \mathcal{E}^N$ be a quaternary sequence. Then we define two binary sequences $E_N = (e_1, e_2, \ldots, e_N), F_N = (f_1, f_2, \ldots, f_N) \in \{-1, 1\}^N$ by

g_n	(e_n, f_n)
1	$(1, 1)$
-1	$(-1, -1)$
i	$(1, -1)$
$-i$	$(-1, 1)$

$$e_n = \frac{(1-i)g_n + (1+i)\overline{g_n}}{2}, \quad n = 1, 2, \ldots, N,$$

$$f_n = \frac{(1+i)g_n + (1-i)\overline{g_n}}{2} \quad n = 1, 2, \ldots, N.$$

Conversely, let $E_N = (e_1, e_2, \dots, e_N), F_N = (f_1, f_2, \dots, f_N) \in \{-1, 1\}^N$ be two binary sequences. We define a quaternary sequence $G_N = (g_1, g_2, \dots, g_N) \in \mathcal{E}^N$ by

$$g_n = \frac{(1+i)e_n + (1-i)f_n}{2}, \quad n = 1, 2, \dots, N.$$

Well-distribution

Let denote by $E_N F_N$ the product sequence $\{e_1 f_1, e_2 f_2, \ldots, e_N f_N\} \in \{-1, 1\}^N$ of the sequences E_N and F_N.

Theorem

We have the following relations between the well-distributions measures of $E_N, F_N, E_N F_N$ and G_N,

$$\max\{W(E_N), W(F_N)\} \leq \sqrt{2}\Delta(G_N) \quad \textit{and} \quad W(E_N F_N) \leq 3\Delta(G_N).$$

Conversely, we have

$$\Delta(G_N) \leq \sqrt{2}\max\{W(E_N), W(F_N), W(E_N F_N)\}.$$

Cross-correlation measure

We define the crosscorrelation measure $C_k(H_1, \ldots, H_k)$ of k binary sequences $H_1 = (h_{1,1}, h_{2,1}, \ldots, h_{N,1})$, $H_2 = (h_{1,2}, h_{2,2}, \ldots, h_{N,2}), \ldots$, $H_k = (h_{1,k}, h_{2,k}, \ldots, h_{N,k}) \in \{-1, 1\}^N$ by

$$C_k(H_1, \ldots, H_k) = \max_{M,D} \left| \sum_{n=1}^{M} h_{n+d_1,1} h_{n+d_2,2} \cdots h_{n+d_k,k} \right|,$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ and M such that $0 \leq d_1 < d_2 < \ldots < d_k \leq N - M$.

Correlation measure

Theorem

We have

$$\max\{C_k(E_N), C_k(F_N)\} \leq 2^{k/2}\Gamma_k(G_N)$$

and

$$\max C_k(H_1, \ldots, H_k) \leq 3^k \Gamma_k(G_N),$$

where the maximum is taken over all $(H_1, \ldots, H_k) \in \{E_N, F_N, E_N F_N\}^k$.
Conversely, we have

$$\Gamma_k(G_N) \leq 2^{k/2} \max C_k(H_1, \ldots, H_k),$$

where the maximum is taken over all $(H_1, \ldots, H_k) \in \{E_N, F_N, E_N F_N\}^k$.

Idea of proof

$$\left|\sum_{j=0}^{M-1} e_{u+jv}\right| \le \frac{|1-i|}{2}\left|\sum_{j=0}^{M-1} g_{u+jv}\right| + \frac{|1+i|}{2}\left|\sum_{j=0}^{M-1} \overline{g_{u+jv}}\right| \le \sqrt{2}\Delta(G_N)$$

$$\left|\sum_{j=0}^{M-1} f_{u+jv}\right| \le \frac{|1+i|}{2}\left|\sum_{j=0}^{M-1} g_{u+jv}\right| + \frac{|1-i|}{2}\left|\sum_{j=0}^{M-1} \overline{g_{u+jv}}\right| \le \sqrt{2}\Delta(G_N)$$

$$e_n f_n = i(g_n - \phi_1(g_n) - \phi_2(g_n))$$

where ϕ_1 and ϕ_2 are the transpositions $1 \leftrightarrow i$ and $-1 \leftrightarrow i$ of $\mathcal{E}$, resp.

$$\left|\sum_{j=0}^{M-1} e_{u+jv} f_{u+jv}\right| \le \left|\sum_{j=0}^{M-1} g_{u+jv}\right| + \left|\sum_{j=0}^{M-1} \phi_1(g_{u+jv})\right| + \left|\sum_{j=0}^{M-1} \phi_2(g_{u+jv})\right|$$

$$\le 3\Delta(G_n)$$

$$\phi_1(g_n) = \frac{i+1}{2} e_n + \frac{i-1}{2} e_n f_n \quad \text{and} \quad \phi_2(g_n) = \frac{1-i}{2} f_n + \frac{i+1}{2} e_n f_n$$

$$\begin{aligned} \left| \sum_{j=0}^{M-1} g_{u+jv} \right| &= \left| \frac{1+i}{2} \sum_{j=0}^{M-1} e_{n+jv} + \frac{1-i}{2} \sum_{j=0}^{M-1} f_{n+jv} \right| \\ &\leq \frac{1}{\sqrt{2}} \left(W(E_N) + W(F_N) \right) \end{aligned}$$

$$\left| \sum_{j=0}^{M-1} \phi_1(g_{u+jv}) \right| \leq \frac{1}{\sqrt{2}} (W(E_N) + W(E_N F_N)),$$

$$\left| \sum_{j=0}^{M-1} \phi_2(g_{u+jv}) \right| \leq \frac{1}{\sqrt{2}} (W(F_N) + W(E_N F_N))$$

From quaternary to binary sequences

$$G_{p-1} = (\chi(1), \chi(2), \ldots, \chi(p-1)) \in \mathcal{E}^{p-1}$$

Mauduit/Sarközy, 2002:

$$\begin{aligned} \Delta(G_{p-1}) &= O(p^{1/2} \log p) \\ \Gamma_k(G_{p-1}) &= O(4^k k p^{1/2} \log p) \end{aligned}$$

Theorems 1 and 2:

$$\begin{aligned} \max\{W(E_N), W(F_N), W(E_N F_N)\} &= O(p^{1/2} \log p), \\ \max\{C_k(E_N), C_k(F_N\} &= O(2^{k/2} 4^k k p^{1/2} \log p), \\ \max C_k(H_1, \ldots, H_k) &= O(12^k k p^{1/2} \log p), \end{aligned}$$

where the maximum is taken over all $(H_1, \ldots, H_k) \in \{E_N, F_N, E_N F_N\}^k$.

From binary to quaternary sequences

Let $E_p = (e_1, e_2, \ldots, e_p)$ and $F_p = (f_1, f_2, \ldots, f_p)$ be defined by

$$e_n = \begin{cases} \left(\frac{g_1(n)}{p}\right), & g_1(n) \neq 0, \\ 1, & g_1(n) = 0, \end{cases} \quad \text{and} \quad f_n = \begin{cases} \left(\frac{g_2(n)}{p}\right), & g_2(n) \neq 0, \\ 1, & g_2(n) = 0, \end{cases}$$

where $g_1(X), g_2(X) \in \mathbb{F}_p[X]$ are squarefree of degrees $1 \leq D_1, D_2 < p$ with $\gcd(g_1, g_2) = 1$.
Goubin et al. 2004:

$$\begin{aligned} W(E_p) &= O(D_1 p^{1/2} \log p), \\ W(F_p) &= O(D_2 p^{1/2} \log p), \\ W(E_p F_p) &= O((D_1 + D_2) p^{1/2} \log p). \end{aligned}$$

Theorem 1 implies

$$\Delta(G_p) = O((D_1 + D_2) p^{1/2} \log p).$$

$g_1(X) = X$ and $g_2(X) = X + 1$: $C_2(F_p, E_p)$ is large. (Choose $d_1 = 0$ and $d_2 = 1$.)

$$g_1(X) = X \quad \text{and} \quad g_2(X) = (X+1)(X+2)(X+2^2)\cdots(X+2^{D-1})$$

with $2k - 1 \leq D < \frac{\log p}{\log 2}$.
We have to show that

$$G(X) = (X+d_1)^{\varepsilon_1} g_2(X+d_1)^{\delta_1} \cdots (X+d_k)^{\varepsilon_k} g_2(X+d_k)^{\delta_k}$$

with $\varepsilon_j, \delta_j \in \{0,1\}$ and $\varepsilon_j + \delta_j \geq 1$ is not a square.
Goubin et al. 2004:

$$\max_{H_1, H_2, \ldots, H_k} C_k(H_1, H_2, \ldots, H_k) = O(Dkp^{1/2} \log p).$$

Theorem 2:

$$\Gamma_k(G_p) = O(Dk2^{k/2}p^{1/2} \log p).$$

Open Problem: Extension: from m-ary sequences to k-ary sequences?

From sequences over $\mathbb{F}_p$ to binary sequences

(s_n) sequence over $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$:
Method 1: Legendre symbol

$$e_n = \left(\frac{s_n}{p}\right), \quad s_n \neq 0.$$

Method 2:

$$e_n = \begin{cases} 1, & \text{if } 0 \leq s_n/p < \frac{1}{2}, \\ -1, & \text{if } \frac{1}{2} \leq s_n/p < 1, \end{cases} \qquad n \geq 0.$$

Discrepancy and correlation measure

Mauduit/Niederreiter/Sarközy, 2007: The correlation measure of order k for (e_n) can be estimated in terms of the discrepancy of the k-dimensional vector sequence modulo p with arbitrary lags $(x_{n+d_0}, x_{n+d_1}, \ldots, x_{n+d_{k-1}})$.

Unfortunately, good discrepancy bounds with arbitrary lags are known only for very few sequences:
Explicit congruential generators: $y_n = f(n) \in \mathbb{F}_p$
A modified recursive inversive generator (Niederreiter/Rivat 2007).
Fermat quotients
Open problem: Find more!

Fermat quotients modulo p

p prime, $\gcd(u,p)=1$

$$q_p(u) \equiv \frac{u^{p-1}-1}{p} \pmod{p}, \qquad 0 \le q_p(u) \le p-1,$$

$$q_p(kp)=0, \qquad k \in \mathbb{Z}.$$

$q_p(u)$ is p^2-periodic

$$q_p(u+vp) = q_p(u) - vu^{-1}, \quad \gcd(u,p) = 1$$

Proof:

$$(u+vp)^{p-1} - 1 \equiv (u^{p-1} - 1) + (p-1)vu^{p-2}p \bmod p^2$$

$$q_p(u+vp) \equiv \frac{u^{p-1}-1}{p} + (p-1)vu^{p-2} \equiv q_p(u) - vu^{-1} \bmod p$$

Multiplicative character sums

Theorem

Let $\psi_1, \ldots, \psi_\ell$ be nontrivial multiplicative characters modulo p. Then we have

$$\sum_{u=0}^{N-1} \psi_1(q_p(u+d_1)) \cdots \psi_\ell(q_p(u+d_\ell)) \ll \max\left\{\frac{\ell N}{p^{1/3}}, \ell p^{3/2} \log p\right\}$$

for any integers $0 \le d_1 < \ldots < d_\ell \le p^2 - 1$ and $1 \le N \le p^2$.

Main tool besides Weil bound:
Heath-Brown: The number of $0 \le v < p$ with

$$q_p(v+d_1)(v+d_1) \equiv q_p(v+d_2)(v+d_2) \bmod p$$

is $O(p^{2/3})$.

Let ψ be the quadratic character (Legendre-symbol) of $\mathbb{F}_p$.
$q_p(u+vp) = q_p(u) - vu^{-1}$ and $x = u + vp$:

$$\sum_{x=0}^{N-1} \psi(q_p(u))\psi(q_p(u+d))$$

$$\ll \quad p + \sum_{\substack{u=1 \\ u\neq p-d}}^{p-1} \left| \sum_{v=0}^{\lfloor N/p \rfloor} \psi((q_p(u) - vu^{-1})(q_p(u+d) - v(u+d)^{-1})) \right|$$

If $q_p(u)u = q_p(u+d)(u+d) \in \mathbb{F}_p$ we use the trivial bound N/p for the inner sum and otherwise the Polya-Vinogradov-Weil bound $p^{1/2} \log p$.
Heath-Brown (using Stepanov-Schmidt method): $O(p^{2/3})$ bad u

$$\ll p^{2/3} N/p + pp^{1/2} \log p$$

Correlation measure

ψ be the quadratic character of $\mathbb{F}_p$ (Legendre symbol)
$e_n = \psi(q_p(n))$, $\gcd(n, p) = 1$

$$C_k(E_N) \ll kp^{5/3}$$

Some open problems

1. Polynomial analogs:

$P \in \mathbb{F}_q[X]$ irreducible over $\mathbb{F}_q$ of degree d

$$\frac{f^{q^d-1}-1}{P} \bmod P, \quad f \in \mathbb{F}_q[X], \;\; \gcd(f,P)=1$$

2. $f(X) = f_0(X) + f_1(X)p \in \mathbb{Z}[X]$:

$$\frac{f(u)-f_0(u)}{p} \bmod p$$

3. $q_p(f(u))$, $f \in \mathbb{F}_p[X]$
4. Matrix analogs of Fermat quotients.
5. Boolean functions with Fermat quotients.

Vectors of consecutive Fermat-quotients

Ostafe/Shparlinski, 2010:

$$\Gamma = \left\{ \left(\frac{q_p(n)}{p}, \frac{q_p(n+1)}{p}, \ldots, \frac{q_p(n+s-1)}{p} \right)_{n=1}^{N} \right\}$$

Exponential sums: $\psi(z) = \exp(2\pi i z/p)$

$$\Sigma_N^{(s)}(\mathbf{a}) = \sum_{u=1}^{N} \psi\left(\sum_{j=0}^{s-1} a_j q_p(u+j) \right)$$

$$\max_{\gcd(a_0,\ldots,a_{s-1},p)=1} \left| \Sigma_N^{(s)}(\mathbf{a}) \right| \ll sp \log p \quad \text{for } 1 \le N \le p^2$$

Vectors of Fermat-quotients with arbitrary lags

Chen/Ostafe/W., 2010:

$$\Gamma = \left\{ \left(\frac{q_p(n+d_0)}{p}, \frac{q_p(n+d_1)}{p}, \ldots, \frac{q_p(n+d_{s-1})}{p} \right)_{n=1}^{N} \right\}$$

for any integers with $0 \le d_0 < d_1 < \cdots < d_{s-1} < p^2$.

Same method and result if $d_l \not\equiv d_h \pmod{p}$ for $0 \le l < h < s$.

A bad example

$s = 3$: $d_0 = 0$, $d_1 = p$, $d_2 = 2p$, $a_0 = 1$, $a_1 = -2$, $a_2 = 1$, $\gcd(u, p) = 1$

$$\begin{aligned} & a_0 q_p(u + d_0) + a_1 q_p(u + d_1) + a_2 q_p(u + d_2) \\ = \;& q_p(u) - 2q_p(u + p) + q_p(u + 2p) \\ = \;& q_p(u) - 2(q_p(u) - u^{-1}) + (q_p(u) - 2u^{-1}) = 0 \end{aligned}$$

$$\sum_{u=1}^{N} \psi \left(\sum_{j=0}^{2} a_j q_p(u + d_j) \right) = N$$

Theorem

For $s \geq 1$ *and* $D = (d_0, d_1, \ldots, d_{s-1})$ *with* $0 \leq d_0 < d_1 < \cdots < d_{s-1} < p^2$ *such that no triple* (d_l, d_h, d_t) *satisfies* $d_l \equiv d_h \equiv d_t \pmod{p}$ *for* $0 \leq l < h < t < s$*, we have*

$$\max_{\gcd(a_0,\ldots,a_{s-1},p)=1} \left| \Sigma_N^{(s)}(\mathbf{a}) \right| \ll s \max\{p \log p, Np^{-1/2}\} \quad 1 \leq N \leq p^2.$$

If $s = 2$ *or* $d_{s-1} < p$*, the stronger bound* $sp \log p$ *holds.*

$$\Sigma_N^{(s)}(\mathbf{a}) = \sum_{u=1}^{N} \psi\left(\sum_{j=0}^{s-1} a_j q_p(u + d_j)\right)$$

Lattice test with arbitrary lags

Again only very few examples of good sequences are known!
Problem: Find more!

Thank you for your attention.