On irregularities of distribution

of binary sequences

relative to arithmetic progressions

Cécile Dartyge (Institut Élie Cartan, Université Lorraine)

joint work with Katalin Gyarmati and András Sárközy

Measures of pseudorandomness

Let $E_N = (e_1, \ldots, e_N) \in \{-1, 1\}^N$.

Definition 1 *(Mauduit and Sárközy 1996) The well-distribution measure of the sequence* E_N *is*

$$W(E_N) = \max_{a,b,t} |\sum_{j=0}^{t-1} e_{aj+b}|$$

where $a, b, t \in \mathbb{N}$ *and* $1 \leq a \leq a + (t-1)b \leq N$*. For* $k \in \mathbb{N}$, $k \leq N$*, the correlation measure of order* k *of the sequence* E_N *is defined by*

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

where $D = (d_1, \ldots, d_k)$ *with* $0 \leq d_1 < d_2 < \ldots < d_k \leq N - M$*.*

A sequence E_N is said to possess strong pseudorandom properties if $W(E_N)$ and $C_k(E_N)$ are small (at least for small k).

Some Examples

The Legendre symbol

$$E_{p-1} = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \ldots, \left(\frac{p-1}{p}\right)\right).$$

This sequence has good PR properties (Mauduit and Sárközy)

$$W(E_p) \leq 9\sqrt{p}\log p, \; C_k(E_p) \leq 9\sqrt{p}\log p.$$

The Thue-Morse sequence : $T_N = (t_0, t_1, \ldots, t_N)$ with $t_n = (-1)^{s_2(n)}$ where $s_2(n)$ is the sum of the digits of n in basis 2.

Gelfond : $W(T_N) \leq 2(1+\sqrt{3})N^{(\log 3/\log 4)}$ for all $N \in \mathbb{N}$

Mauduit and Sárközy : $C_2(T_N) \geq N/12$ for $N \geq 5$.

Weighted distribution measure

In some applications, we need binary sequences such that their "short" subsequences also satisfy good PR properties.

$$E_N(n, M) = (e_{n+1}, e_{n+2}, \dots, e_{n+M}) \quad \text{for } 0 \le n < n + M \le N.$$

Definition 2 *(Gyarmati, Sárközy, D)*

For $0 \le \alpha \le 1/2$, *the weighted* α*-well-distribution measure of* E_N *is defined by*

$$W_\alpha(E_N) = \max_{0 \le n < n+M} M^{-\alpha} W(E_N(n, M)).$$

Remark : $W_0(E_N) = W(E_N)$.

Irregularity results

Theorem 1 *(Roth 1964) If* $N \in \mathbb{N}$, $E_N \in \{-1,1\}^N$, *then there exist* $a,t,q \in \mathbb{N}$, $1 \le a \le a+(t-1)q \le N$ *and* $q \le \sqrt{N}$ *such that :*

$$\Big|\sum_{j=0}^{t-1} e_{a+jq}\Big| > c_1 N^{1/4},$$

for some absolute constant $c_1 > 0$.

The "$N^{1/4}$" in Theorem 1 is optimal.

Theorem 2 *(Matoušek and Spencer (1996)) There exists a sequence* $E_N \in \{-1,1\}^N$ *such that for all* a,t,q *with* $1 \le a \le a+(t-1)q \le N$, *we have*

$$\Big|\sum_{j=0}^{t-1} e_{a+jq}\Big| < c_2 N^{1/4},$$

with some absolute c_2.

For $0 \le \alpha \le 1/2$, we write

$$m_\alpha(N) = \min_{E_N \in \{-1,1\}^N} W_\alpha(E_N).$$

Theorem 1 implies : $m_\alpha(N) \gg N^{1/4-\alpha}$ for $\alpha \in [0, 1/2]$.

Conjecture 1 *For $0 \le \alpha \le 1/2$, we have :*

$$N^{1/4-\alpha/2} \ll m_\alpha(N) \ll N^{1/4-\alpha/2}$$

The case $\alpha = 0$ is a consequence of Theorem 1 and Theorem 2.

Bounds for random binary sequences

Theorem 3 *(Gyarmati, Sárközy, D)*

Let $\alpha \in [0, 1/2]$*. Then for all* $\varepsilon > 0$*, there exists* $N_0 = N_0(\varepsilon)$*,* $\delta = \delta(\varepsilon)$ *such that if* $N > N_0$ *then for a random sequence* $E_N \in \{-1, 1\}^N$ *(that is chosen with probability* $1/2^N$*), we have*

$$P(\delta N^{1/2-\alpha} < W_\alpha(E_N) < 6N^{1/2-\alpha}\sqrt{\log N}) > 1 - \varepsilon.$$

The case $\alpha = 0$ was done by Cassaigne, Mauduit and Sárközy, and sharpened by Alon, Kohayakawa, Mauduit, Moreira and Rödl, and more recently by Aistleitner.

Proof of the upper bound of $W_\alpha(E_N)$ in Theorem 3

We start by applying

$$P\left(\left(\max_{\dots} "\dots"\right) \geq "\dots."\right) \leq \sum_{\dots} P("\dots" \geq "\dots").$$

We find :

$$P\left(W_\alpha(E_N) > 6\frac{\sqrt{N \log N}}{N^\alpha}\right) \leq \sum_{\substack{0\leq n\leq N-M \\ a+(t-1)b\leq M}} P\left(\left|\sum_{j=0}^{t-1} e_{n+a+jb}\right| > 6\sqrt{N \log N}\left(\frac{M}{N}\right)^\alpha\right).$$

Lemma 1 *(Chernoff's inequality, particular case) Let* $X_1, \dots, X_k$ *be independant random variables with* $P(X_i = 1) - 1/2 = P(X_i = -1)$*. Then for* $A > 0$*, we have*

$$P(|\sum_{i=1}^{k} X_i| \geq A) \leq 2e^{-A^2/2k}.$$

We apply this lemma with $X_i = e_{n+a+(i-1)b}$.

Special sequences

The Rudin-Shapiro sequence

We consider a trigonometric polynomial

$$P(e^{i\theta}) = \sum_{n=1}^{N} \varepsilon_n e^{2i\pi n\theta}, \qquad \varepsilon_n = \pm 1.$$

Parseval :

$$N = \sum_{n=1}^{N} |\varepsilon_n|^2 = \int_0^1 |P(e^t)|^2 dt \leq \|P\|_\infty^2$$

Does there exists (ε_n) such that $\|P\|_\infty \leq A\sqrt{N}$ for all N? This question was solved independently by Rudin (1958) and Shapiro (1951) : $P_N = \sum_{n=0} r_n X^n$, where $(r_n)_{n\geq 0}$ is the Rudin-Shapiro sequence.

The Rudin-Shapiro sequence

$R_N = (r_0, \ldots, r_{N-1})$, with $r_0 = 1$, $r_{2n} = r_n$ and $r_{2n+1} = -r_n$.

Mauduit and Sárközy : $W(R_N) \leq 2(2+\sqrt{2})\sqrt{N}$ for all $N \in \mathbb{N}$.

Theorem 4 *(K. Gyarmati, A. Sárközy, D)*

$$W_\alpha(R_N) < 40N^{1/2-\alpha}$$

In particular, $W_{1/2}(R_N) < 40$.

Remark : $C_2(R_N) > N/6$ for $N \geq 4$ (Mauduit and Sárközy).

The Legendre symbol

For $p \leq N$ we consider the sequence E_N^p defined by

$$e_n = \begin{cases} \left(\frac{n}{p}\right) & \text{for } (n,p) = 1 \\ 1 & \text{if } p|n \end{cases}$$

Theorem 5 *For every $\alpha \in [0, 1/2]$ there exists $N_0 = N_0(\alpha)$ such that for all $N > N_0$ the exists $p \in \left]\frac{N^{\frac{2(1-\alpha)}{3}}}{2}, N^{\frac{2(1-\alpha)}{3}}\right]$ such that*

$$W_\alpha(E_N^p) < cN^{\frac{1-\alpha}{3}},$$

for some absolute c.

Main ingredient of the proof of Theorem 5

Lemma 2 *(Montgomery and Vaughan)*

There exists an absolute constant c *such that for* $N \in \mathbb{N}$, $N \geq 2$ *there is* $p \in]N/2, N]$ *satisfying for all* $X \in Z$, $Y \in \mathbb{N}$

$$\left|\sum_{n=X+1}^{X+Y} \left(\frac{n}{p}\right)\right| < c\sqrt{p}.$$

For $0 < \alpha \leq 1/2$ we can improve further Theorem 5 by applying Burgess inequality

Theorem 6 *(Gyarmati, Sárközy, D) For all* $0 \leq \alpha \leq 1/2$ *there is* $N_1 = N_1(\alpha)$ *such that for* $N > N_1$ *and* p *satisfying*

$$\frac{1}{2} N^{\frac{8(1-\alpha)}{12-5\alpha}} (\log N)^{-\frac{8\alpha}{12-5\alpha}} < p \leq N^{\frac{8(1-\alpha)}{12-5\alpha}} (\log N)^{-\frac{8\alpha}{12-5\alpha}}$$

and Lemma 2, we have

$$W_\alpha(E_N^p) < cN^{\frac{(1-\alpha)(4-5\alpha)}{12-5\alpha}} (\log N)^{-\frac{8\alpha}{12-5\alpha}}$$

For $\alpha = 1/2$ this gives

$$W_{1/2}(E_N^p) < cN^{3/38} (\log N)^{8/19}$$

Proof of Theorem 6

For $H \in \mathbb{N}$ we define

$$d(p, H) = \max_{X \in \mathbb{Z}} \left| \sum_{n=X+1}^{X+H} \left(\frac{n}{p} \right) \right|.$$

We need to bound

$$\max_{H \le N} H^{-\alpha} d(p, H).$$

Lemma 3 *(Burgess inequality for the Legendre symbol) For p prime, $H, r \in \mathbb{N}$ we have for some absolute constant c*

$$\max_{X \in \mathbb{Z}} \left| \sum_{n=X+1}^{X+H} \left(\frac{n}{p} \right) \right| < c H^{1-\frac{1}{r}} p^{\frac{r+1}{4r^2}} (\log p)^{1/r}.$$

Lower bound for W_α for the Legendre symbol construction

Theorem 7 *(Gyarmati, Sárközy, D) For all* $0 \le \alpha \le 1/2$*, we have*

$$W_\alpha(E_{p-1}^p) > \frac{1}{10} p^{1/2-\alpha}.$$

Main ingredient of the proof of Theorem 7

Lemma 4 *(Winterhof). For any* $\mathcal{D} \subset \mathbb{F}_p$ *and any multiplicative character* $\chi \neq \chi_0$ *modulo* p *we have*

$$\sum_{a \in \mathbb{F}_p} \Big| \sum_{x \in \mathcal{D}} \chi(x+a) \Big|^2 = p|\mathcal{D}| - |\mathcal{D}|^2.$$