EXACT ORDER OF DISCREPANCY OF NORMAL NUMBERS

10.01.2023, One World Numeration Seminar

Roswitha Hofer on joint work with Gerhard Larcher

JOHANNES KEPLER UNIVERSITY LINZ

Uniform Distribution in $[0, 1)$

We focus on sequences $(x_n)_{n\geq 0}$ in $[0, 1)$.

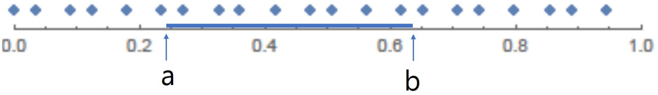


Figure: The first 21 points of $(\{n\varphi\})_{n\geq 0}$.

We say $(x_n)_{n\geq 0}$ is **uniformly distributed** if

$$\lim_{N\to\infty} \frac{1}{N} \underbrace{\#\{0 \leq n < N \ : \ x_n \in [a, b)\}}_{=:A(N,[a,b))} = \lambda([a, b))$$

for all $0 \leq a < b \leq 1$.

The discrepancy D_N ...

... measures how "uniformly" $(x_n)_{n\geq 0}$ is distributed:

$$D_N(x_n) := \sup_{0\leq a<b\leq 1} \left| \frac{1}{N} A(N, [a,b)) - \lambda([a,b)) \right| .$$

Obviously, $(x_n)_{n\geq 0}$ is uniformly distributed if and only if

$D_N \to 0$ as $N \to \infty$.

A regular grid $\left\{ \frac{a}{N} : a \in \{0, 1, \ldots, N-1\} \right\}$, satisfies $D_N = \frac{1}{N}$ or $N D_N = 1$ resp.

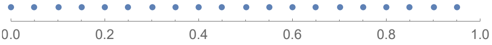


Figure: A regular grid with $N = 20$.

The best possible discrepancy D_N for sequences in N?

W.M. Schmidt [1972] ensured for any sequence $(x_n)_{n\geq 0}$,

$$ND_N(x_n) \gg \log N.$$

We know examples of "**low-discrepancy sequences**" $(x_n)_{n\geq 0}$, i.e., satisfying $ND_N(x_n) \ll \log N$.

How to prove $ND_N(x_n) \ll \log N$? Use e.g. additive properties of $ND_N = \sup_{0\leq a<b\leq 1} |A(N,[a,b)) - N\lambda([a,b))|$.
When we split a pointset of N points into two disjoint sets with N_1 and N_2 points, then

$$ND_N \leq N_1 D_{N_1} + N_2 D_{N_2}.$$

Example of a low-discrepancy sequence

The binary van der Corput sequence $(\mathrm{vdC}_2(n))_{n\geq 0}$:

$$n = n_r 2^r + \cdots + n_1 2 + n_0 \mapsto \frac{n_0}{2} + \frac{n_1}{2^2} + \cdots + \frac{n_r}{2^{r+1}}.$$

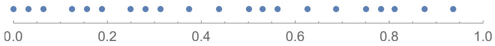


Figure: The first 21 points of $(\mathrm{vdC}_2(n))_{n\geq 0}$.

We know

$$N D_N\big(\mathrm{vdC}_2(n)\big) \ll \log N.$$

The proof uses the binary expansion $N = N_r 2^r + \cdots + N_1 2 + N_0$ with $N_j \in \{0,1\}$ and splits the set $\{x_0, x_1, \ldots, x_{N-1}\}$:

$$\bigcup_{j=0,\, N_j \neq 0}^{r} \{x_{N_r 2^r + \cdots + N_{j+1} 2^{j+1} + i} \,:\, i = 0, 1, \ldots, 2^j - 1\}$$

in $\leq r + 1 \approx \log_2 N$ parts of “almost regular grids” with 2^j points and with $2^j\, D_{2^j} \ll 1$.

Then

$$N D_N \ll \sum_{j=0}^{r} 1 \ll \log N.$$

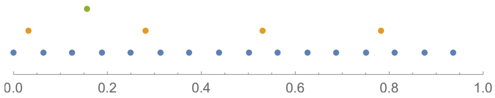


Figure: $N = 21 = 2^4 + 2^2 + 2^0$.

We focus on a special type of sequences in $[0, 1)$:

$$(\{2^n\alpha\})_{n\geq 0} \quad \text{with} \quad \alpha \in [0, 1).$$

Why interesting?

We ask, if the binary expansion of α, i.e. $\alpha = 0.\alpha_1\alpha_2\alpha_3\ldots\ldots$ satisfies that each block $(\eta_1, \ldots, \eta_k) \in \{0, 1\}^k$ occurs with frequency $\frac{1}{2^k}$, i.e.

$$\lim_{N\to\infty} \frac{1}{N}\#\Big\{0 \leq n < N : (\alpha_{n+1}, \ldots, \alpha_{n+k}) = (\eta_1, \ldots, \eta_k)\Big\} = \frac{1}{2^k}$$

for all $k \in \mathbb{N}$ and $(\eta_1, \ldots, \eta_k) \in \{0, 1\}^k$.

Such α is called normal in base 2.

It is KNOWN:

- The number $\alpha \in [0, 1)$ is normal in base 2 if and only if $(\{2^n\alpha\})_{n\geq 0}$ is uniformly distributed.
- The quality of normality of a normal number α in base 2 corresponds with the discrepancy D_N of the sequence$(\{2^n\alpha\})_{n\geq 0}$.

It is NOT KNOWN (Korobov [1956]):

- What is the best possible quality of normality?

From Schmidt [1972] we know

$$ND_N \gg \log N.$$

E.g. Schiffer [1986] showed for the Champernowne number

$$ND_N \gg \frac{N}{\log N}.$$

So far the best known examples of α (Levin [1999], Becher & Carton [2019]) satisfy

$$\log N \ll N D_N(\{2^n \alpha\}) \ll \log^2 N.$$

There is a gap of one $\log N$ factor between the lower and the upper bound!

Possible improvements for $\log N \ll ND_N(\{2^n\alpha\}) \ll \log^2 N$?

- What about the known examples of Levin [1999] or Becher & Carton [2019]. What is the correct order of $ND_N(\{2^n\alpha\})$? ... Answered, e.g., for Levin's α, $\log^2 N \ll ND_N(\{2^n\alpha\}) \ll \log^2 N$ (H. & Larcher [2022].)
- Try to find an example of α with better ND_N than $\log^2 N$.
- Try to improve the lower bound of Schmidt for normal numbers:
  The set $\{(\{2^n\alpha\})_{n\geq 0} : \alpha \in [0,1)\}$ is a rather small subset of the set of all sequences $(x_n)_{n\geq 0}$ in $[0,1)$.

The objectives of the rest of the presentation:

- Give the basic ideas of the construction of Levin for the normal number α.
- Point out the main reason, within the construction, why $ND_N(\{2^n\alpha\}) \ll \log^2 N$ holds.
- Work out the essential steps/ideas for proving $ND_N(\{2^n\alpha\}) \gg \log^2 N$.
- Discussing (possible) generalizations.

The ingenious construction of α by Levin

$$0.\underbrace{\alpha_1\alpha_2\ldots\alpha_{2(2^1+1)}}_{\mathcal{A}_1}\underbrace{\alpha_{2(2^1+1)+1}\ldots\alpha_{2(2^1+1)+2(2^2+2)}}_{\mathcal{A}_2}\;\cdots\cdots\;\underbrace{\cdots\cdots\cdots}_{\mathcal{A}_m}\;\cdots\cdots$$

Here the block $\mathcal{A}_m$ for $m \in \mathbb{N}$ consists of $2^{2^m+m} = 2^m \cdot 2^{2^m}$ digits α_i, which are constructed as follows. We write the block $\mathcal{A}_m$ in the form

$$\underbrace{d_0(0)\ldots d_k(0)\ldots d_{2^m-1}(0)}\quad\underbrace{d_0(1)\ldots d_k(1)\ldots d_{2^m-1}(1)}\quad\ldots$$

$$\ldots\quad\underbrace{d_0(n)\ldots d_k(n)\ldots d_{2^m-1}(n)}\quad\ldots$$

$$\ldots\quad\underbrace{d_0(2^{2^m}-1)\ldots d_k(2^{2^m}-1)\ldots d_{2^m-1}(2^{2^m}-1)}.$$

Each n defines a sub-block of length 2^m. The number of sub-blocks (or n) is 2^{2^m}.

The core of the construction is the Pascal-matrix $\pmod 2$

$$P = \left(\binom{i+j}{j}\right)_{i,j\geq 0}, \qquad P(m) = \left(\binom{i+j}{j} \pmod 2\right)_{0\leq i,j<2^m}$$

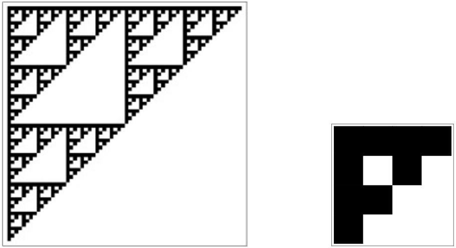


Figure: The $2^6 \times 2^6$ matrix $P(6)$ and the $2^2 \times 2^2$ matrix $P(2)$.

We construct $\mathcal{A}_2$ of length $2^2 2^{2^2} = 4 \cdot 16$:

$$\mathcal{A}_2 = \overbrace{0,0,0,0}^{d_0(0),d_1(0),d_2(0),d_3(0)}, \overbrace{1,1,1,1}, \overbrace{1,0,1,0}, \overbrace{0,1,0,1}, \overbrace{1,1,0,0},$$
$$\overbrace{0,0,1,1}, \overbrace{0,1,1,0}, \overbrace{1,0,0,1}, \overbrace{1,0,0,0}, \overbrace{0,1,1,1}, \overbrace{0,0,1,0}, \overbrace{1,1,0,1},$$
$$\overbrace{0,1,0,0}, \overbrace{1,0,1,1}, \overbrace{1,1,1,0}, \overbrace{0,0,0,1}^{d_0(15),d_1(15),d_2(15),d_3(15)}$$

The $d_k(n)$ for $k \in \{0,1,2,3\}$, $n \in \{0,\dots,15\}$ are given by:

$$\underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}}_{=P(2)} \begin{pmatrix} n_0 \\ n_1 \\ n_2 \\ n_3 \end{pmatrix} = \begin{pmatrix} d_0(n) \\ d_1(n) \\ d_2(n) \\ d_3(n) \end{pmatrix} \pmod 2.$$

How to see $ND_N(\{2^n\alpha\}) \ll \log^2 N$?

Note that $\{2^0\alpha\} = \alpha =$

$$0.\underbrace{\alpha_1\alpha_2\ldots\alpha_{2(2^1+1)}}_{\mathcal{A}_1}\underbrace{\alpha_{2(2^1+1)+1}\ldots\alpha_{2(2^1+1)+2(2^2+2)}}_{\mathcal{A}_2}\cdots\underbrace{\ldots\ldots\ldots}_{\mathcal{A}_m}\cdots$$

$\{2^1\alpha\} = 0.\alpha_2\alpha_3\alpha_4\ldots$,

$$\{2^{2(2^1+1)}\alpha\} = 0.\underbrace{\alpha_{2(2^1+1)+1}\ldots\alpha_{2(2^1+1)+2(2^2+2)}}_{\mathcal{A}_2}\cdots\underbrace{\ldots\ldots\ldots}_{\mathcal{A}_m}\cdots$$

- Write
  $N = 2^{(1+2^1)} + 2^{(2+2^2)} + \cdots + 2^{(m-1+2^{m-1})} + N' = n_m + N'$
  with $1 \le N' \le 2^m 2^{2^m}$ and $N' \approx 2^m N''$, $1 \le N'' \le 2^{2^m}$.
- Write $x_{n_m+2^m n+k}$ as $x_{n,k}$, $k \in \{0,\ldots,2^m-1\}$,
  $n \in \{0,\ldots,N''-1\}$. Then

$$x_{n,k} = 0.\alpha_{n_m+2^m n+k+1}, \alpha_{n_m+2^m n+k+2}, \ldots = 0.d_k(n)d_{k+1}(n)\ldots.$$

JOHANNES KEPLER UNIVERSITY LINZ

- Estimate $N''D_{N''}$ for $x_{n,k}$ with $n \in \{0, \dots, N''-1\}$ and fixed $k \in \{0, ..., 2^m - 1\}$.
- Write $N'' = N''_{2^m} 2^{2^m} + \cdots + N''_1 2^1 + N''_0 2^0$ with $N''_j \in \{0, 1\}$. Write $\{0, \dots, N''-1\}$ as a union of $\ll \log N$ subsets:

$$\bigcup_{j=0,\, N''_j \neq 0}^{2^m} \{N''_{2^m} 2^{2^m} + \cdots + N''_{j+1} 2^{j+1} + i \, : \, i = 0, 1, \dots, 2^j - 1\}.$$

- Focus on

$$x_{n,k} = 0.d_k(n) d_{k+1}(n) \dots \in \left[\frac{c}{2^j}, \frac{c+1}{2^j}\right)$$

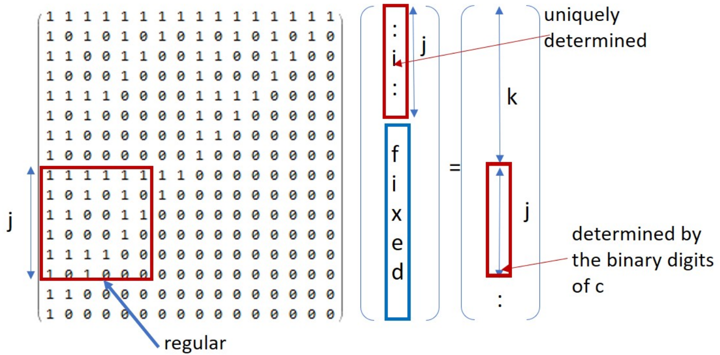
with $c \in \{0, \dots, 2^j - 1\}$ and with fixed $k \in \{0, 1, \dots, 2^m - 1\}$, that is

$$d_k(n) d_{k+1}(n) \dots = c_{j-1} \dots c_1 c_0 \dots$$

where $c = c_{j-1} 2^{j-1} + \cdots + c_1 2 + c_0$ and $n \in \{N''_{2^m} 2^{2^m} + \cdots + N''_{j+1} 2^{j+1} + i \, : \, i = 0, 1, \dots, 2^j - 1\}$.

Translate $d_k(n)d_{k+1}(n)\ldots = c_{j-1}\ldots c_1c_0\ldots$ into linear-algebra

Suppose only $k + j \leq 2^m$ then
$x_{n,k} = 0.d_k(n)d_{k+1}(n)\ldots d_{k+j-1}(n)\ \ldots = c_{j-1}\ldots c_1c_0\ldots$. Note the first j digits stem from n.



We summarize:

For fixed $k \in \{0, \ldots, 2^m - 1\}$ and $c \in \{0, 1, \ldots, 2^j - 1\}$ one value of n in $\{N''_{2^m} 2^{2^m} + \cdots + N''_{j+1} 2^{j+1} + i \,:\, i = 0, 1, \ldots, 2^j - 1\}$ yields $x_{n,k} \in [\frac{c}{2^j}, \frac{c+1}{2^j})$[1]. Those 2^j points form an "almost regular grid" amongst $[\frac{c}{2^j}, \frac{c+1}{2^j}),\ c \in \{0, 1, \ldots, 2^j - 1\}$. Thus $2^j D_{2^j} \ll 1$. Altogether we obtain:

$$N D_N \leq \overbrace{n_m D_{n_m}}^{\ll \log N,\ \text{less relevant}} + \overbrace{2^m}^{\#\, k's} N'' D_{N''} \ll \overbrace{2^m}^{\#\, k's} \sum_{j=0, N''_j \neq 0}^{2^m} 1.$$

Since $2^m \approx \log N$ we arrive at

$$N D_N \ll (2^m)^2 \ll \log^2 N.$$

[1] with exceptions, but those exceptions can be controlled.

To prove $ND_N \gg \log^2 N$ we define for $m \in \mathbb{N}$

- $n_m < N := N_m < n_{m+1}$
- and an interval $J := J_m \subseteq [0,1)$.
- For m large enough we find a $\overline{c} > 0$ such that

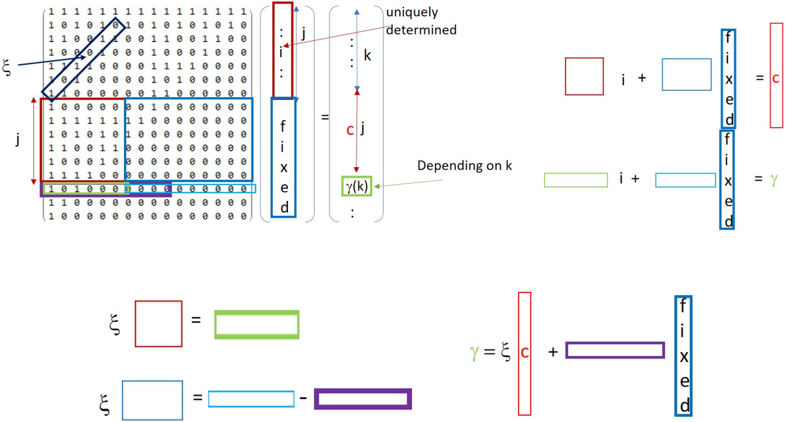
$$ND_N \geq A(N,J) - N\lambda(J) \geq \overline{c} \cdot (\log N)^2 .$$

- We reach this with $N = N_m = n_m + 2^m \sum_{l=0}^{M} N_l$ and $M + 1 \approx c' \log N$;

$$A(N,J) - N\lambda(J) \geq -\underbrace{n_m D_{n_m}}_{\ll \log N} + \sum_{l=0}^{M} \underbrace{\sum_{k=0}^{2^m-1} \big(A(N_l,J) - N_l\lambda(J)\big)}_{\text{a suitable surplus of points}} .$$

- As tools serve specific features of the Pascal-matrix $P(m)$.

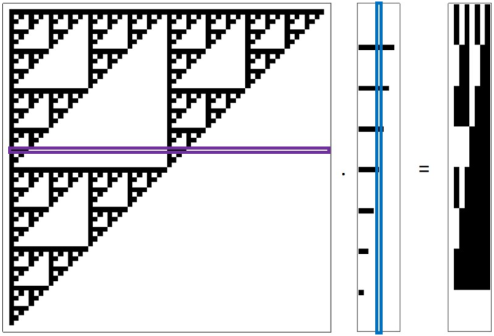
$x_{n,k} \in \left[\frac{c+\gamma/2}{2^j}, \frac{c+(\gamma+1)/2}{2^j}\right)$ **with** $k+j+1 \le 2^m$ **?**



Need many k's with equal values for "violet part · fixed vector"

For our choice of $N_m = n_m + 2^m 2^{w_0} \sum_{l=0}^{M} 2^{-8l}$ the fixed vector for $l = 0, \ldots, M$ has a special form $(00000000)(10000000)\ldots(10000000)00....$.

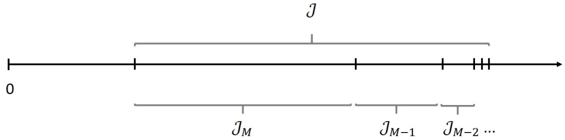
Again we use specific properties of the Pascal-matrix:



"violet part · fixed vector" is more often black than white. This relative surplus grows with m.

We return to obtaining $ND_N \gg \log^2 N$

With $N = n_m + 2^m 2^{w_0} \sum_{l=0}^{M} 2^{-8l} = n_m + 2^m \sum_{l=0}^{M} N_l$ and $M + 1 \approx c' \log N$



$$ND_N \geq A(N, J) - N\lambda(J) \geq -\underbrace{n_m D_{n_m}}_{\ll \log N} +$$

$$+\sum_{l=0}^{M} \underbrace{\sum_{k=0}^{2^m-1} \big(A(N_l, J) - N_l\lambda(J)\big)}_{\text{surplus of points}} \geq (M+1)c'' 2^m \geq \overline{c} \cdot (\log N)^2 \,.$$

Further normal numbers with $ND_N \ll \log^2 N$

The essential properties of the Pascal-matrix mod 2, guaranteeing $ND_N \ll \log^2 N$, can be identified in an appropriate form in a **modified Pascal-matrix mod 2** (Becher & Carton [2019]) and **mod p** (H. & Larcher [2022]).

So one obtains similar to Levin's construction a normal number α in base p with $ND_N(\{p^n\alpha\}) \ll \log^2 N$.

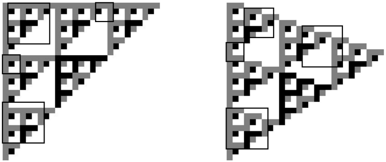
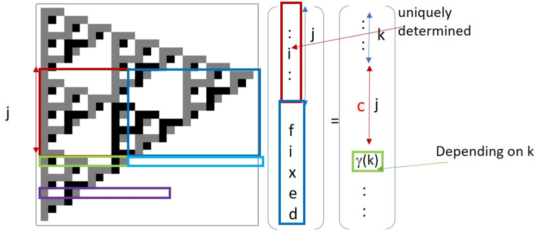
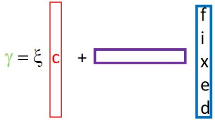


Figure: All the sketched, square submatrices are regular. The right shows a modified Pascal-matrix.

Can we prove also $ND_N \gg \log^2 N$?

Some important properties in our method of proof might be re-identified in a **modified Pascal-matrix mod p.**





Can we prove also $ND_N \gg \log^2 N$?

But a comparable pattern as for the unmodified case $p = 2$ is clear only for the unmodified case for general p.

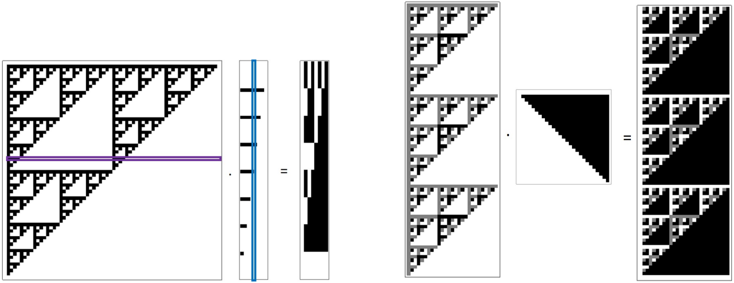


Figure: For general p we again obtain many black entries for "violet part $\cdot$ fixed vector" .

The main problem for a modified case is:

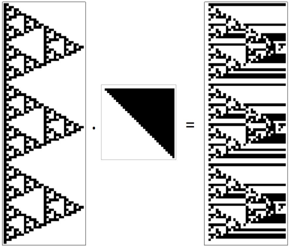


Figure: How to obtain many equal values for “violet part · fixed vector”?.

This might be feasible for a single example of such α. But a general “recipe” for all such α seems to be unfeasible.

We summarize:

- **We know:** The exact order of ND_N is $\log^2 N$ for Levin's number based on the Pascal-matrix mod p (H. & Larcher 2022).
- **Open:** What is the correct order of ND_N for normal numbers that are obtained from nested perfect necklaces in the sense of Becher & Carton [2019] for $p = 2$ or from affine necklaces for $p \in \mathbb{P}$ in the sense of H. & Larcher [2022].

Thank you for the attention!

1. M.B. Levin. *On the discrepancy estimate of normal numbers.* Acta Arithmetica, v. 88, no. 2., pp. 99-111, 1999.
2. V. Becher and O. Carton. *Normal numbers and nested perfect necklaces.* Journal of Complexity, 54: 101403, 2019.
3. R. Hofer and G. Larcher. *The exact order of the discrepancy of Levin's normal number in base* 2*.* To appear in Journal de Théorie des Nombres de Bordeaux. `https://arxiv.org/abs/2205.01566`
4. R. Hofer and G. Larcher. *Discrepancy bounds for normal numbers generated by necklaces in arbitrary base.* Submitted 2022. `https://arxiv.org/abs/2211.04212`