DIGITAL EXPANSIONS, PRIME NUMBERS, AND UNIFORM DISTRIBUTION MODULO 1

Michael Drmota

Institut für Diskrete Mathematik und Geometrie

Technische Universität Wien

michael.drmota@tuwien.ac.at

www.dmg.tuwien.ac.at/drmota/

TU Graz, October 19, 2007

q-Ary Digital Expansion

$q \geq 2$ ... **integer basis** of digital expansion in $\mathbb{N}$

$\mathcal{N} = \{0, 1, \ldots, q-1\}$ ... set of **digits**

$n \in \mathbb{N} \Longrightarrow$

$$\boxed{n = \sum_{j \geq 0} \varepsilon_j(n)\, q^j} \qquad \text{with } \varepsilon_j(n) \in \mathcal{N}.$$

Sum-of-digits function

$$\boxed{s_q(n) = \sum_{j \geq 0} \varepsilon_j(n)}$$

Robert Tichy and Digital Expansions

- More than 25 papers on digital problems.
- Asymptotics for average values $\frac{1}{N}\sum_{n<N} s_q(n)^k$.
- Uniform distribution modulo 1 of $x_n = \alpha s_q(n)$.
- Ergodic properties of $s_q(n)$, odometers.
- Digital expansions based on linear recurrences.
- Digital blocks and digital expansions.
- Fractals and digital expansions.
- Waring's problem with digital restrictions.

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

$$t_n = \begin{cases} 0 & \text{if } s_2(n) \text{ is even,} \\ 1 & \text{if } s_2(n) \text{ is odd.} \end{cases}$$

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

0

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

01

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

0110

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

01101001

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

0110100110010110

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

01101001100101101001011001101001

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

$011010011001011010010110011010011001011001101\cdots$

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

$$011010011001011010010110011010011001011001101\cdots$$

$$t_{2^k+n} = 1 - t_n \quad (0 \le n < 2^k)$$

Thue-Morse Sequence

$$t_n = \frac{1 - (-1)^{s_2(n)}}{2}$$

$$011010011001011010010110011010011001011001101\cdots$$

$$t_{2^k+n} = 1 - t_n \quad (0 \le n < 2^k)$$

or

$$t_{2k} = t_k, \; t_{2k+1} = 1 - t_k$$

Prime Numbers

$$p \text{ is Fermat prime} \iff s_2(p) = 2.$$

$$p \text{ is Mersenne prime} \iff s_2(p) = [\log_2 p].$$

Focus of the Talk

- Distribution properties of $s_q(n)$

- Subsequences of the Thue-Morse sequence

- Gelfond's problems

- Uniform distribution modulo 1

- Relations to primes

Subsequences of the Thue-Morse Sequence

$(n_k)_{k\geq 0}$ increasing sequence of natural numbers

Problem:

$$\#\{k < K : t_{n_k} = 0\} =????$$

Equivalently

$$\#\{k < K : s_2(n_k) \equiv 0 \bmod 2\} =????$$

Examples:

- $n_k = ak + b$
- $n_k = k$-th prime p_k
- $n_k = k^2$ etc.

Linear Subsequences

Gelfond 1967/1968

m, s ... positive integers with $(s, q-1) = 1$.

$$\Longrightarrow \quad \boxed{\#\{n < N : n \equiv \ell \bmod m,\ \ s_q(n) \equiv t \bmod s\} = \frac{N}{ms} + O(N^\lambda)}$$

with $0 < \lambda < 1$.

In particular:

$$\begin{aligned}\#\{k < K : s_2(ak+b) \equiv 0 \bmod 2\} &= \#\{k < K : t_{ak+b} = 0\} \\ &= \frac{K}{2} + O(K^\lambda)\end{aligned}$$

Linear Subsequences

$q = 2$

Lemma

$$\sum_{n<2^L} x^{s_2(n)} y^n = \prod_{\ell<L} \left(1 + xy^{2^\ell}\right)$$

Corollary 1 $\quad e(x) := e^{2\pi i x}$

$$\begin{aligned} \#\{n < 2^L : n \equiv \ell \bmod m,\ s_q(n) \equiv t \bmod s\} \\ &= \frac{1}{ms} \sum_{i=0}^{m-1} \sum_{j=0}^{s-1} e\left(-\frac{i\ell}{m} - \frac{jt}{s}\right) \prod_{\ell<L} \left(1 + e\left(\frac{i}{s} + \frac{2^\ell j}{m}\right)\right) \\ &= \frac{2^L}{ms} + O(2^{\lambda L}) \end{aligned}$$

Uniform Distribution modulo 1

Corollary 2 α irrational, $h \neq 0$ integer

$$\Longrightarrow \quad \sum_{n<2^L} e(h\,\alpha\,s_2(n)) = (1 + e(h\alpha))^L = o(2^L).$$

With a little bit more effort:

$$\sum_{n<N} e(h\,\alpha\,s_2(n)) = o(N)$$

Weyl's criterion $\Longrightarrow$ $\boxed{\alpha s_2(n) \text{ uniformly distributed modulo 1.}}$

By assuming certain Diophantine approximation properties for α these bounds also imply estimates for the *discrepancy* $D_N(\alpha\,s_2(n))$.

Ergodic Methods

- Compactification of $\mathbb{N}$ by considering the *digit space* $\mathbb{Z}_q = \{0, 1, \ldots, q-1\}^{\mathbb{N}}$.

- Addition extends to a continuous mapping on $\mathbb{Z}_q$, the corresponding action $T : \mathbb{N} \to \mathrm{Aut}(\mathbb{Z}_q)$ is uniquely ergodic.

- The mapping

$$a(z,x) = \begin{cases} \alpha \lim\limits_{w \to x, w \in \mathbb{N}} (s_q(w+z) - s_q(w)) & \text{if the limit exists,} \\ 0 & \text{otherwise} \end{cases}$$

  is a cocycle and the set of essential values of a coincides with

$$\overline{\{a(z,x) : x \in \mathbb{Z}_q,\ \ z \in \mathbb{N}\}}.$$

- Application of the ergodic theorem.

Newman's Phenomenon

Thue-Morse Sequence $t_n = (1 - (-1)^{s_2(n)})/2$

$$011010011001011010010110011010011001011001101\cdots$$

Newman's Phenomenon

Thue-Morse Sequence $t_n = (1 - (-1)^{s_2(n)})/2$

0110100110010110100101100110100110010110011010 $\cdots$

Newman's Phenomenon

Thue-Morse Sequence $t_n = (1 - (-1)^{s_2(n)})/2$

$$0110100110010110100101100110100110010110011010 1\cdots$$

There seem to be more 0's than 1's in the sequence t_{3k}:

$$\#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 0 \bmod 2\}$$
$$> \#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 1 \bmod 2\}$$

Newman's Phenomenon

Thue-Morse Sequence $t_n = (1 - (-1)^{s_2(n)})/2$

$$0110100110010110100101100110100110010110011010 \cdots$$

There seem to be more 0's than 1's in the sequence t_{3k}:

$$\begin{aligned}\#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 0 \bmod 2\} \\ > \#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 1 \bmod 2\}\end{aligned}$$

Note that Gelfond's result implies

$$\begin{aligned}\#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 0 \bmod 2\} \\ \sim \#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 1 \bmod 2\} \\ \sim \frac{N}{6}.\end{aligned}$$

Newman's Phenomenon

Theorem *(Newman 1969, Coquet 1983)*

$$\#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 0 \bmod 2\}$$
$$- \#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 1 \bmod 2\}$$
$$= \sum_{n<N, n\equiv 0 \bmod 3} (-1)^{s_2(n)}$$
$$= \boxed{\Psi(\log_2 n) \cdot N^{\frac{\log 3}{\log 4}} + O(1)},$$

where $\Psi(x)$ *is a* **positive** *continuous periodic function.*

In particular

$$\sum_{n<N, n\equiv 0 \bmod 3} (-1)^{s_2(n)} > 0$$

for all $N > 0$.

Newman's Phenomenon

Extensions

Grabner 1993

$$\sum_{n<N, n\equiv 0 \bmod Q} (-1)^{s_2(n)} > 0 \quad \text{for } \boxed{Q = 3^k 5^\ell}.$$

Grabner, Herendi and Tichy 1997

$$\sum_{n<N, n\equiv 0 \bmod \boxed{17}} (-1)^{s_2(n)} > 0$$

Newman's Phenomenon

Further Extensions Drmota, Leinfellner, Skałba 1999, 2000

$$\sum_{n<N, n\equiv 0 \bmod Q} (-1)^{s_2(n)} > 0$$

for $\boxed{Q \equiv 0 \bmod 3,\ Q = 2^K + 1,\ Q = (2^{4K-1} + 1)/3}$.

A negative statement: Drmota, Skałba 2000

$$\#\left\{\text{primes } p \le x : \sum_{n<N, n\equiv 0 \bmod p} (-1)^{s_2(n)} > 0\right\} = o\left(\frac{x}{\log x}\right)$$

Newman's phenomenon occurs only for *few primes*.

Gelfond's Problems

Gelfond 1967/1968

1. $q_1, q_2, \ldots, q_d \geq 2$, $(q_i, q_j) = 1$ for $i \neq j$, $(m_j, q_j - 1) = 1$:

$$\boxed{\#\{n < N : s_{q_j}(n) \equiv \ell_j \bmod m_j,\ 1 \leq j \leq d\} = \frac{N}{m_1 \cdots m_d} + O(N^{1-\eta})}$$

2. $(m, q-1) = 1$:

$$\boxed{\#\{\text{primes } p < N : s_q(p) \equiv \ell \bmod m\} = \frac{\pi(N)}{m} + O(N^{1-\eta})}$$

3. $(m, q-1) = 1$, $P(x) \in \mathbb{N}[x]$:

$$\boxed{\#\{n < N : s_q(P(n)) \equiv \ell \bmod m\} = \frac{N}{m} + O(N^{1-\eta})}$$

Gelfond's Problems

Gelfond 1967/1968

1. $q_1, q_2, \ldots, q_d \ge 2$, $(q_i, q_j) = 1$ for $i \ne j$, $(m_j, q_j - 1) = 1$: Kim 1999

$$\#\{n < N : s_{q_j}(n) \equiv \ell_j \bmod m_j,\ 1 \le j \le d\} = \frac{N}{m_1 \cdots m_d} + O(N^{1-\eta})$$

2. $(m, q-1) = 1$: Mauduit, Rivat 2005+

$$\#\{\text{primes } p < N : s_q(p) \equiv \ell \bmod m\} = \frac{\pi(N)}{m} + O(N^{1-\eta})$$

3. $(m, q-1) = 1$, $P(x) \in \mathbb{N}[x]$: Mauduit, Rivat 2007+ for $P(x) = x^2$

$$\#\{n < N : s_q(n^2) \equiv \ell \bmod m\} = \frac{N}{m} + O(N^{1-\eta})$$

Gelfond's 1^{st} Problem

Besineau 1972: solution without error terms

Kim 1999: bounds on exponential sums:
($e(x) = e^{2\pi i x}$, $\|x\| = \min_{k \in \mathbb{Z}} |x - k|$ ... distance to the nearest integer)

$$\left| \frac{1}{N} \sum_{n<N} e(\alpha_1 s_{q_1}(n) + \alpha_2 s_{q_2}(n) + \cdots + \alpha_d s_{q_d}(n)) \right|$$
$$\ll \exp \left(-\eta \log N \sum_{j=1}^{d} \|(q_j - 1)\alpha_j\|^2 \right),$$

($\alpha_j \in \mathbb{Q}$: Kim, $\alpha_j \in \mathbb{R}$: Drmota, Larcher)

Gelfond's 1^{st} Problem

Applications of Kim's method

Drmota, Larcher 2001: $q_1, q_2, \ldots, q_d \ge 2$, $(q_i, q_j) = 1$ for $i \ne j$, $\alpha_1, \ldots, \alpha_d$ irrational:

$$(\alpha_1 s_{q_1}(n), \ldots, \alpha_d s_{q_d}(n))_{n \ge 0} \in \mathbb{R}^d \quad \text{uniformly distributed mod 1.}$$

Thuswaldner, Tichy 2005: $q_1, q_2, \ldots, q_d \ge 2$, $(q_i, q_j) = 1$ for $i \ne j$.

For $d > 2^k$ the number of representations of

$$N = x_1^k + \cdots + x_d^k \quad \text{with} \quad s_{q_j}(x_j) \equiv \ell_j \bmod m_j, 1 \le j \le d$$

is asymptotically given by

$$\frac{\mathfrak{S}(N)}{m_1 \cdots m_d} \frac{\Gamma\left(1 + \frac{1}{k}\right)^d}{\Gamma\left(\frac{d}{k}\right)} N^{\frac{d}{k}-1} + O\left(N^{\frac{d}{k}-1-\eta}\right).$$

Gelfond's 2^{nd} Problem

Mauduit, Rivat 2005+: α real number

$$\left| \frac{1}{N} \sum_{n<N} \Lambda(n) e(\alpha s_q(n)) \right| \ll \exp\left(-\eta \log N \|(q-1)\alpha\|^2 \right),$$

Corollary

$$\left| \frac{1}{\pi(N)} \sum_{p<N} e(\alpha s_q(p)) \right| \ll \exp\left(-\eta \log N \|(q-1)\alpha\|^2 \right),$$

Applications

- α irrational $\Longrightarrow$

$$(\alpha s_q(p))_{p \text{ prime}} \quad \text{is uniformly distributed mod 1.}$$

Gelfond's 2^{nd} Problem

Applications (cont.)

- Set $\alpha = j/m$ + discrete Fourier analysis $\Longrightarrow$

$$\#\{\text{primes } p < N : s_q(p) \equiv \ell \bmod m\} = \frac{\pi(N)}{m} + O(N^{1-\eta}).$$

- t_n ... Thue-Morse sequence $\Longrightarrow$

$$\#\{\text{primes } p < N : t_p = 0\} = \frac{\pi(N)}{2} + O(N^{1-\eta}).$$

Gelfond's 2^{nd} Problem

Gaussian primes

$q = -a + i$ ... basis for digital expansion in $\mathbb{Z}[i]$ ($a \in \{1, 2, \ldots\}$)

$\mathcal{N} = \{0, 1, \ldots, a^2\}$ ... digit set

$z \in \mathbb{Z}[i] \Longrightarrow \boxed{z = \sum_{j \ge 0} \varepsilon_j(z)\, q^j}$ with $\varepsilon_j(z) \in \mathcal{N}$

$s_q(z) = \sum_{j \ge 0} \varepsilon_j(z)$ ... sum-of-digits function

Drmota, Rivat, Stoll 2007+

Suppose that $a \ge 28$ such that $q = -a + i$ is prime,
i.e. $a \in \{36, 40, 54, 56, 66, 74, 84, 90, 94, \ldots\}$. Then

$$\frac{1}{N/\log N} \sum_{|z|^2 \le N,\, z \text{ prime}} e(\alpha s_q(z)) \ll \exp\left(-\eta \log N \|(a^2 + 2a + 2)\alpha\|^2\right).$$

Gelfond's 2^{nd} Problem

Drmota, Mauduit, Rivat 2007+: $(k, q-1) = 1$

$$\#\{\text{primes } p < N : s_q(p) = k\}$$
$$= \frac{q-1}{\varphi(q-1)} \frac{\pi(N)}{\sqrt{2\pi\sigma_q^2 \log_q N}} \left(\exp\left(-\frac{(k-\mu_q \log_q N)^2}{2\sigma_q^2 \log_q N} \right) + O((\log N)^{-\frac{1}{2}+\varepsilon}) \right),$$

where

$$\mu_q := \frac{q-1}{2}, \quad \sigma_q^2 := \frac{q^2-1}{12}.$$

Remark: This asymptotic expansion is only significant if

$$\left| k - \mu_q \log_q N \right| \le C (\log N)^{\frac{1}{2}}$$

Note that $\frac{1}{\pi(N)} \sum_{p<N} s_q(p) \sim \mu_q \log_q N$.

It **does not apply** for $k = 2$ or $k = [\log_2 p]$ (if $q = 2$).

Gelfond's 2^{nd} Problem

Lemma 1 *For every fixed integer* $q \geq 2$ *there exist two constants* $c_1 > 0$, $c_2 > 0$ *such that for every* k *with* $(k, q-1) = 1$

$$\sum_{p \leq N \;\; p \equiv k \bmod q-1} e(\alpha s_q(p)) \ll (\log N)^3 N^{1-c_1 \|(q-1)\alpha\|^2}$$

uniformly for real α *with* $\|(q-1)\alpha\| \geq c_2 (\log N)^{-\frac{1}{2}}$.

Remark. This is a refined version of the previous estimate.

Gelfond's 2^{nd} Problem

Lemma 2 *Suppose that* $0 < \nu < \frac{1}{2}$ *and* $0 < \eta < \frac{\nu}{2}$. *Then for every* k *with* $(k, q-1) = 1$ *we have*

$$\sum_{p \le N,\ p \equiv k \bmod q-1} e(\alpha s_q(p)) = \frac{\pi(N)}{\varphi(q-1)} e(\alpha \mu_q \log_q N) \times \left(e^{-2\pi^2\alpha^2\sigma_q^2 \log_q N} \left(1 + O\left(|\alpha|\right)\right) + O\left(|\alpha| \left(\log N\right)^{\nu}\right)\right)$$

uniformly for real α *with* $|\alpha| \le (\log N)^{\eta - \frac{1}{2}}$.

Remark. This is a refined version of a *central limit theorem* of $s_q(p)$:

$$\boxed{\#\{p \le N : s_q(p) \le \mu_q \log_q N + y\sqrt{\sigma_q^2 \log_q N}\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2}\, dt}.$$

Proof

Set

$$S(\alpha) := \sum_{p\le N} e(\alpha s_q(p)) \quad \text{and} \quad S_k(\alpha) := \sum_{p\le N,\ p\equiv k \bmod q-1} e(\alpha s_q(p)).$$

Then by using $s_q(p) \equiv p \bmod (q-1)$ we get

$$\begin{aligned}\#\{p \le N : s_q(p) = k\} &= \int_{-\frac{1}{2(q-1)}}^{1-\frac{1}{2(q-1)}} S(\alpha) e(-\alpha k)\, d\alpha \\ &= (q-1) \int_{-\frac{1}{2(q-1)}}^{\frac{1}{2(q-1)}} \left(\sum_{p\le N,\ p\equiv k \bmod q-1} e(\alpha s_q(p)) \right) e(-\alpha k)\, d\alpha \\ &= (q-1) \int_{-\frac{1}{2(q-1)}}^{\frac{1}{2(q-1)}} S_k(\alpha)\, e(-\alpha k)\, d\alpha.\end{aligned}$$

Proof

We split up the integral

$$\int_{-\frac{1}{2(q-1)}}^{\frac{1}{2(q-1)}} = \int_{|\alpha|\le(\log N)^{\eta-1/2}} + \int_{(\log N)^{\eta-1/2}<|\alpha|\le 1/(2(q-1))}$$

From Lemma 1 we get an upper bound for the second integral

$$\int_{(\log N)^{\eta-1/2}<|\alpha|\le 1/(2(q-1))} S_k(\alpha)\, e(-\alpha k)\, d\alpha \ll (\log N)^2\, N\, e^{-c_1 (q-1)^2 (\log N)^{2\eta}}$$

$$\ll \frac{\pi(N)}{\log N}.$$

Proof

Set

$$\alpha := t/(2\pi\sigma_q\sqrt{\log_q N}) \quad \text{and} \quad \Delta_k = \frac{k - \mu_q \log_q N}{\sqrt{\sigma_q^2 \log_q N}}.$$

Then by Lemma 2 we an asymptotic expansion for the first integral:

$$\begin{aligned}
&\int_{|\alpha|\le(\log N)^{\eta-1/2}} S_k(\alpha)e(-\alpha k)\,d\alpha \\
&\quad = \frac{\pi(N)}{\varphi(q-1)} \int_{|\alpha|\le(\log N)^{\eta-1/2}} e(\alpha(\mu_q \log_q N - k))\, e^{-2\pi^2\alpha^2\sigma_q^2 \log_q N} \cdot (1 + O \\
&\quad + O\left(\pi(N) \int_{|\alpha|\le(\log N)^{\eta-1/2}} |\alpha|\,(\log N)^\nu\,d\alpha\right) \\
&\quad = \frac{1}{\varphi(q-1)} \frac{\pi(N)}{2\pi\sigma_q\sqrt{\log_q N}} \int_{-\infty}^{\infty} e^{it\Delta_k - t^2/2}\,dt + O\left(\pi(N)e^{-2\pi^2\sigma_q^2(\log N)^{2\eta}}\right) \\
&\quad + O\left(\frac{\pi(N)}{\log N}\right) + O\left(\frac{\pi(N)}{(\log N)^{1-\nu-2\eta}}\right) \\
&\quad = \frac{1}{\varphi(q-1)} \frac{\pi(N)}{\sqrt{2\pi\sigma_q^2 \log_q N}} \left(e^{-\Delta_k^2/2} + O((\log N)^{-\frac{1}{2}+\nu+2\eta})\right).
\end{aligned}$$

Gelfond's 3^{rd} Problem

Mauduit, Rivat 1995, 2005 $1 \le c \le \frac{7}{5}$:

$$\#\{n < N : s_q([n^c]) \equiv \ell \bmod m\} \sim \frac{N}{m}$$

Dartyge, Tenenbaum 200?: There exists $C > 0$ with

$$\#\{n < N : s_q(n^2) \equiv \ell \bmod m\} \ge C\, N$$

Drmota, Rivat 2005: $s_2^{[<\lambda]}(n) = \sum_{j<\lambda} \epsilon_j(n)$, $s_2^{[\ge\lambda]}(n) = \sum_{j\ge\lambda} \epsilon_j(n)$:

$$\#\{n < 2^L : s_2^{[<L]}(n^2) \equiv 0 \bmod 2\} \sim \frac{2^L}{2},$$

$$\#\{n < 2^L : s_2^{[\ge L]}(n^2) \equiv 0 \bmod 2\} \sim \frac{2^L}{2}.$$

Gelfond's 3^{rd} Problem

Mauduit, Rivat 2007+

$$\boxed{\frac{1}{N}\sum_{n<N} e(\alpha s_q(n^2)) \ll (\log N)^A \exp\left(-\eta \log N \|(q-1)\alpha\|^2\right)}$$

Applications

- $\#\{n < N : s_q(n^2) \equiv \ell \bmod m\} = \dfrac{N}{m} + O(N^{1-\eta})$
- $\#\{n < N : t_{n^2} = 0\} = \dfrac{N}{2} + O(N^{\lambda})$
- $(\alpha\, s_q(n^2))_{n \geq 0}$ is uniformly distributed modulo 1

Gelfond's 3^{rd} Problem

Open problem:

distribution properties of $s_q(n^3)$??

General problem:

Let S be a set of natural numbers that are determined by congruence conditions and bounds on the exponents of the prime factorization.

What is the distribution of $(s_q(n))_{n \in S}$??

Examples: primes, squares, cubes, square-free numbers etc.

Happy Birthday!