THE SUM OF DIGITS OF PRIMES

Michael Drmota

joint work with **Christian Mauduit** and **Joël Rivat**

Institute of Discrete Mathematics and Geometry

Vienna University of Technology

michael.drmota@tuwien.ac.at

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

Seminari combinatòria, teoria de grafs i aplicacions, UPC Barcelona, May 21, 2009

Binary Representation of Primes

2 extremal cases

$$p = 2^k + 1 \qquad \text{Fermat prime } (k = 2^m)$$

$$p = 2^k + 2^{k-1} + \cdots + 2 + 1 \qquad \text{Mersenne prime } (k + 1 \in \mathbb{P})$$

Question of Ben Green

Given k, does there exist a prime p and $0 = j_1 < j_2 < \cdots < j_k$ with

$$p = 2^{j_1} + 2^{j_2} + \cdots + 2^{j_k} \text{ ??}$$

Summary

- Thue–Morse sequence
- Gelfond's theorem on linear subsequences
- Gelfond's problems
- Exponential sum estimates
- A global central limit theorem
- A local central limit theorem
- Proof methods

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

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

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

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

$$011010011001011010010110011010011001011001101\cdots$$

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

or

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

Frequency of Letters

$$\begin{aligned}\#\{n < N : t_n = 0\} &= \#\{n < N : t_n = 1\} + O(1) \\ &= \frac{N}{2} + O(1)\end{aligned}$$

equivalently

$$\begin{aligned}\#\{n < N : s_2(n) \equiv 0 \bmod 2\} &= \#\{n < N : s_2(n) \equiv 1 \bmod 2\} + O(1) \\ &= \frac{N}{2} + O(1)\end{aligned}$$

Subsequences of the Thue-Morse Sequence

$(n_k)_{k\ge 0}$ 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 $e(x) := e^{2\pi i x}$

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

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* $D_N(\alpha\,s_2(n))$.

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

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

2. $(m, q-1) = 1$: 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: $q_1, q_2, \ldots, q_d \geq 2$, $(q_i, q_j) = 1$ for $i \neq j$, $\alpha_1, \ldots, \alpha_d$ irrational:

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

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

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

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

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

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

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 $\Longrightarrow$

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

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

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

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

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

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

Drmota, Rivat, Stoll 2008

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 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

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

Global Results

Central limit theorem for $s_q(n)$:

$$\frac{1}{N} \cdot \#\{n \le N : s_q(n) \le \mu_q \log_q N + y\sqrt{\sigma_q^2 \log_q N}\} = \Phi(y) + O\left(\frac{1}{\sqrt{\log N}}\right),$$

where

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2}\, dt$$

distribution function of the normal distribution

and

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

Remark $(q = 2)$

$$\frac{1}{2^L} \sum_{n<2^L} e^{its_2(n)} = \left(\frac{1+e^{it}}{2}\right)^L \quad \Longrightarrow \quad \text{CLT}$$

Global Results

Central limit theorem for $s_q(p)$: Bassily, Katai

$$\boxed{\frac{1}{\pi(N)} \cdot \#\{p \le N : s_q(p) \le \mu_q \log_q N + y\sqrt{\sigma_q^2 \log_q N}\} = \Phi(y) + o(1)}.$$

Central limit theorem for $s_q(n^2)$: Bassily, Katai

$$\boxed{\frac{1}{N} \cdot \#\{n \le N : s_q(n^2) \le 2\mu_q \log_q N + y\sqrt{2\sigma_q^2 \log_q N}\} = \Phi(y) + o(1)}$$

Remark. The results also hold for $s_q(P(n))$ and $s_q(P(p))$ if $P(x)$ in a non-negative integer polynomial.

Local Results for all n

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

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 $\frac{1}{N} \sum_{n<N} s_q(n) \sim \mu_q \log_q N$.

Local Results for all n

More precise results (only stated for $q = 2$) Mauduit, Sarközy

$$\#\{n < N : s_q(n) = k\} = F\left(\frac{k}{\log N}, \log_2 N\right) \binom{[\log_2 N]}{k} \left(1 + O\left(\frac{1}{\log N}\right)\right),$$

uniformly for $\varepsilon \log_2 N \le k \le (1-\varepsilon)\log_2 N$, where $F(x,t)$ is analytic in x and periodic in t.

Proof is based on an representation of the form

$$\sum_{n<N} x^{s_2(n)} = F(x, \log_2 N)(1+x)^{\log_2 N}.$$

and a saddle point analysis applied to the integral in Cauchy's formula.

Local Results for primes

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

Local Results for primes

This result **does NOT apply** for $k = 2$ and $k = [\log_2 p]$ (for $q = 2$):

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

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

Local Results for primes

... **but** we have:

$$\boxed{\#\{\text{primes } p < 2^{2k} : s_2(p) = k\} \sim \frac{2^{2k}}{\sqrt{2\pi}\log 2\, k^{\frac{3}{2}}}}$$

Local Results for squares

Drmota, Mauduit, Rivat 2007+:

$$\#\{n < N : s_q(n^2) = k\}$$
$$= \frac{N}{\sqrt{4\pi\sigma_q^2 \log_q N}} \left(\exp\left(-\frac{(k - 2\mu_q \log_q N)^2}{4\sigma_q^2 \log_q N} \right) + O((\log N)^{-\frac{1}{2}+\varepsilon}) \right).$$

Remark: Again this asymptotic expansion is only significant if

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

which is the significant range.

Idea of the proof for primes

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

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

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

Remark. This is a refined version of the previous estimate by Mauduit and Rivat.

Idea of the proof for primes

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* $|\alpha| \le (\log N)^{\eta-\frac{1}{2}}$.

Idea of the proof for primes

Lemma 1 + 2 imply the result

Set

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

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

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

Idea of the proof for primes

We split up the integral

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

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

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

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

Idea of the proof for primes

Set

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

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

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

$$= \frac{\pi(N)}{\varphi(q-1)} \int_{|\alpha| \le (\log N)^{\eta - 1/2}} e(\alpha(\mu_q \log_q N - k))\, e^{-2\pi^2\alpha^2\sigma_q^2 \log_q N} \cdot (1 + O$$

$$+ O\left(\pi(N) \int_{|\alpha| \le (\log N)^{\eta - 1/2}} |\alpha|\, (\log N)^{\nu}\, d\alpha\right)$$

$$= \frac{1}{\varphi(q-1)} \frac{\pi(N)}{2\pi\sigma_q\sqrt{\log_q N}} \boxed{\int_{-\infty}^{\infty} e^{it\Delta_k - t^2/2}\, dt} + O\left(\pi(N) e^{-2\pi^2\sigma_q^2 (\log N)^{2\eta}}\right)$$

$$+ O\left(\frac{\pi(N)}{\log N}\right) + O\left(\frac{\pi(N)}{(\log N)^{1-\nu-2\eta}}\right)$$

$$= \frac{1}{\varphi(q-1)} \frac{\pi(N)}{\sqrt{2\pi\sigma_q^2 \log_q N}} \left(\boxed{e^{-\Delta_k^2/2}} + O((\log N)^{-\frac{1}{2}+\nu+2\eta})\right).$$

Idea of the proof for primes

Proof idea of Lemma 1

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 $\|(q-1)\alpha\| \geq c_2 (\log N)^{-\frac{1}{2}}$.

Idea of the proof for primes

Proof idea of Lemma 1: Vaughan's method

Let $q \geq 2$, $x \geq q^2$, $0 < \beta_1 < 1/3$, $1/2 < \beta_2 < 1$. Let g be an arithmetic function. Suppose that uniformly for all real numbers $M \leq x$ and all complex numbers a_m, b_n such that $|a_m| \leq 1$, $|b_n| \leq 1$, we have

$$\max_{\frac{x}{qM}<t\leq\frac{xq}{M}} \sum_{\frac{M}{q}<m\leq M} \left| \sum_{\frac{x}{qm}<n\leq t} g(mn) \right| \leq U \quad \text{for} \quad M \leq x^{\beta_1} \quad \text{(type I).,}$$

$$\left| \sum_{\frac{M}{q}<m\leq M} \sum_{\frac{x}{qm}<n\leq\frac{x}{m}} a_m b_n g(mn) \right| \leq U \quad \text{for} \quad x^{\beta_1} \leq M \leq x^{\beta_2} \quad \text{(type II)}$$

Then

$$\left| \sum_{x/q<p\leq x,\ p \text{ prime}} g(p) \right| \ll U\,(\log x)^2.$$

Idea of the proof for primes

Proof idea of Lemma 1: Type I - sums

For $q \ge 2$, $x \ge 2$, and for every α such that $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$ we have

$$\max_{\frac{x}{qM} < t \le \frac{xq}{M}} \sum_{M/q < m \le M} \left| \sum_{\frac{x}{qm} < n \le t} e(\alpha\, s_q(mn)) \right| \ll_q x^{1-\kappa_q(\alpha)} \log x$$

for $1 \le M \le x^{1/3}$ and

$$0 < \kappa_q(\alpha) := \min\left(\frac{1}{6}, \frac{1}{3}(1-\gamma_q(\alpha))\right)$$

where $0 \le \gamma_q(\alpha) < 1$ is defined by

$$q^{\gamma_q(\alpha)} = \max\left(1, \max_{t \in \mathbb{R}} \sqrt{\varphi_q(\alpha+t)\,\varphi_q(\alpha+qt)}\right)$$

with

$$\varphi_q(t) = \begin{cases} |\sin \pi q t| / |\sin \pi t| & \text{if } t \in \mathbb{R} \setminus \mathbb{Z}, \\ q & \text{if } t \in \mathbb{Z}. \end{cases}$$

Idea of the proof for primes

Proof idea of Lemma 1: Type II - sums

For $q \geq 2$, there exist β_1, β_2 and δ verifying $0 < \delta < \beta_1 < 1/3$ and $1/2 < \beta_2 < 1$ such that for all α with $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$, there exist $\xi_q(\alpha) > 0$ for which, uniformly for all complex numbers b_n with $|b_n| \leq 1$, we have

$$\sum_{q^{\mu-1} < m \leq q^{\mu}} \left| \sum_{q^{\nu-1} < n \leq q^{\nu}} b_n \, e(\alpha s_q(mn)) \right| \ll_q (\mu + \nu) q^{(1 - \frac{1}{2}\xi_q(\alpha))(\mu+\nu)},$$

whenever

$$\beta_1 - \delta \leq \frac{\mu}{\mu + \nu} \leq \beta_2 + \delta.$$

Idea of the proof for primes

Proof idea of Lemma 1: Methods

- Van-der-Corput Inequality

- Neglecting "large" digits $\longrightarrow$ periodic structure

- **Discrete Fourier analysis** with Fourier terms

$$F_\lambda(h,\alpha) = q^{-\lambda} \sum_{0 \le u < q^\lambda} e\left(\alpha s_q(u) - huq^{-\lambda}\right).$$

Idea of the proof for primes

Proof idea of Lemma 2:

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* $|\alpha| \le (\log N)^{\eta - \frac{1}{2}}$.

Idea of the proof for primes

Proof idea of Lemma 2: Interpretation as sum of random variables

$$S_N = S_N(p) = s_q(p) = \sum_{j \le \log_q N} \varepsilon_j(p).$$

Lemma 2 is equivalent to

$$\boxed{\varphi_1(t) := \mathbb{E}\, e^{it(S_N - L\mu_q)/(L\sigma_q^2)^{1/2}} = e^{-t^2/2}\left(1 + O\left(\frac{|t|}{\sqrt{\log N}}\right)\right) + O\left(\frac{|t|}{(\log N)^{\frac{1}{2}-\nu}}\right)}$$

that is uniform for $|t| \le (\log N)^{\eta}$.

Idea of the proof for primes

Proof idea of Lemma 2: Truncation of digits

$L = \log_q N$, $L' = \#\{j \in \mathbb{Z} : L^\nu \le j \le L - L^\nu\} = L - 2L^\nu + O(1)$ $(0 < \nu < \frac{1}{2})$,

$$\boxed{T_N = T_N(p) = \sum_{L^\nu \le j \le L - L^\nu} \varepsilon_j(p) = \sum_{L^\nu \le j \le L - L^\nu} D_j},$$

$$\varphi_2(t) := \mathbb{E}\, e^{it(T_N - L'\mu_q)/(L'\sigma_q^2)^{1/2}}$$

We have, uniformly for all real t

$$\boxed{|\varphi_1(t) - \varphi_2(t)| = O\left(\frac{|t|}{(\log N)^{\frac{1}{2}-\nu}}\right)}$$

Idea of the proof for primes

Proof idea of Lemma 2: Approximation by sum of iid random variables

Z_j $(j \ge 0)$ iid random variables with range $\{0, 1, \ldots, q-1\}$ and

$$\mathbf{P}\{Z_j = \ell\} = \frac{1}{q}.$$

$$\boxed{\overline{T}_N := \sum_{L^\nu \le j \le L - L^\nu} Z_j}.$$

We have

$$\mathbb{E}\,\overline{T}_N = L'\mu_q \quad \text{and} \quad \mathbb{V}\,\overline{T}_N = L'\sigma_q^2$$

and

$$\boxed{\varphi_3(t) := \mathbb{E}\, e^{it(\overline{T}_N - L'\mu_q)/(L'\sigma_q^2)^{1/2}} = e^{-t^2/2}\left(1 + O\left(\frac{|t|}{\sqrt{\log N}}\right)\right)}$$

uniformly for $|t| \le (\log N)^{\frac{1}{2}}$. (This is the classical central limit theorem.)

Idea of the proof for primes

Proof idea of Lemma 2: quantification

We have uniformly for real t with $|t| \leq L^{\eta}$

$$\boxed{|\varphi_2(t) - \varphi_3(t)| = O\left(|t| e^{-c_1 L^{\kappa}}\right)},$$

where $0 < 2\eta < \kappa < \nu$. and $c_1 > 0$ is a constant depending on η and κ.

This estimate directly **proves Lemma 2**.

Remark: Taylor' theorem gives for random variables X, Y:

$$\begin{aligned} \mathbb{E}e^{itX} - \mathbb{E}e^{itY} &= \sum_{d<D} \sum_{0 \leq d < D} \frac{(it)^d}{d!} \left(\mathbb{E}\, X^d - \mathbb{E}\, Y^d\right) \\ &+ O\left(\frac{|t|^D}{D!} \left|\mathbb{E}\,|X|^D - \mathbb{E}\,|Y|^D\right| + \frac{|t|^D}{D!} \mathbb{E}\,|Y|^D\right). \end{aligned}$$

Idea of the proof for primes

Proof idea of Lemma 2: comparision of moments

We have uniformly for $1 \le d \le L'$

$$\mathbb{E}\left(\frac{T_N - L'\mu_q}{\sqrt{L'\sigma_q^2}}\right)^d = \mathbb{E}\left(\frac{\overline{T}_N - L'\mu_q}{\sqrt{L'\sigma_q^2}}\right)^d + O\left(\left(\frac{4q}{\sigma_q}\right) L^{(\frac{1}{2}+\nu)d} e^{-c_4 L^\nu}\right),$$

which can be reduced to **compare frequencies**

$$\begin{aligned}\mathbf{Pr}\{D_{j_1,M} = \ell_1, \ldots, D_{j_d,N} = \ell_d\} \\ = \mathbf{Pr}\{Z_{j_1} = \ell_1, \ldots, Z_{j_d} = \ell_d\} + O\left((4L^\nu)^d e^{-c_4 L^\nu}\right)\end{aligned}$$

Recall:

$$T_N := \sum_{L^\nu \le j \le L - L^\nu} D_j, \qquad \overline{T}_N := \sum_{L^\nu \le j \le L - L^\nu} Z_j$$

Idea of the proof for primes

Proof idea of Lemma 2: comparision of moments

Proof of this property uses

- exponential sum estimats over primes

$$\sum_{p \le x} e\left(\frac{A}{Q}\, p\right) \ll (\log x)^5\, x\, q^{-K/2}$$

- Erdős-Turán inequality

$$\text{Discrepancy of } (x_n) \ll \frac{1}{H} + \sum_{h=1}^{H} \frac{1}{h} \left| \frac{1}{N} e(h x_n) \right|,$$

- trivial observation:

$$\epsilon_j(n) = d \quad \Longleftrightarrow \quad \left\{ \frac{n}{q^{j+1}} \right\} \in \left[\frac{d}{q}, \frac{d+1}{q} \right).$$

Open problem

Cubes:

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.

Thank You!