Primes and squares with preassigned digits

Cathy Swaenepoel

IMJ-PRG, Université Paris Cité.

32èmes Journées Arithmétiques,
Nancy, July 3, 2023.

Digits

Representation of an integer $k \in [0, g^n[$ in base $g \geqslant 2$:

$$k = \sum_{j=0}^{n-1} \varepsilon_j(k) g^j.$$

where $\varepsilon_j(k) \in \{0, \dots, g-1\}$ is the **digit** of k at the position j.

independence?

base g expansion $\longleftrightarrow$ multiplicative representation (as a product of prime factors)

Integers with preassigned digits

- $A \subset \{0, \ldots, n-1\}$: set of positions,
- $\boldsymbol{d} = (d_j)_{j \in A}$: preassigned digits at these positions.

	d_{n-2}					d_6		d_4			d_1	
$n-1$	$n-2$					6		4			1	0

$$\left| \underbrace{\{k < g^n : \forall j \in A,\ \varepsilon_j(k) = d_j\}}_{\substack{\text{sparse set} \\ \text{if } |A| \to +\infty \text{ as } n \to +\infty}} \right| = g^{n-|A|}$$

Prime numbers with preassigned digits

Historical background

Goal: estimate $|\{p < g^n : \forall j \in A, \varepsilon_j(p) = d_j\}|$ as $n \to +\infty$.

- **Kátai (1986)**.
- **Wolke (2005)**: asymptotic, $|A| \leqslant 2$
  ($|A| \leqslant (1-\varepsilon)\sqrt{n}$ under GRH).
- **Harman (2006)**: lower bound, $|A| \leqslant$ constant.
- **Harman-Kátai (2008)**: asymptotic, $|A| \ll \sqrt{n}(\log n)^{-1}$.
- **Bourgain (2013)**: asymptotic, $|A| \ll n^{4/7}(\log n)^{-4/7}$, in base 2.
- **Bourgain (2015)**: asymptotic, $|A| \leqslant cn$, in base 2 ($c > 0$ absolute constant).

Result

Theorem 1 (S. 2020)

For any $g \geqslant 2$, there exist an explicit $c = c(g) \in]0,1[$ and $\delta = \delta(g) > 0$ such that for any $n \geqslant 1$, for any $A \subset \{0, \ldots, n-1\}$ satisfying $\{0, n-1\} \subset A$ and

$$|A| \leqslant cn,$$

for any $(d_j)_{j \in A} \in \{0, \ldots, g-1\}^A$ such that $(d_0, g) = 1$ and $d_{n-1} \geqslant 1$, we have

$$|\{p < g^n : \forall j \in A,\ \varepsilon_j(p) = d_j\}| = \frac{g^{n-|A|}}{\log g^n} \frac{g}{\varphi(g)} \left(1 + O_g\left(n^{-\delta}\right)\right).$$

This generalizes Bourgain's result (2015) to any base.

Explicit values of c

Theorem 1 holds with $c(g)$ given by

g	2	3	4	5	10	10^3	2^{200}
$c(g) \cdot 10^2$	0.21	0.31	0.36	0.40	0.47	0.68	0.90

Squares with preassigned digits

Squares vs primes

Denote $\mathcal{S} = \{\ell^2,\ \ell \geqslant 0\}$ the set of squares.

Goal: estimate $|\mathcal{S} \cap \{k < g^n : \forall j \in A,\ \varepsilon_j(k) = d_j\}|$ as $n \to +\infty$.

- Squares are a priori easier to handle than primes (distribution in short intervals, in arithmetic progressions, ...).

But

- squares are sparser than primes,
- there are algebraic constraints on the digits of squares.

$\rightarrow$ New difficulties for squares.

Hypothesis $\mathcal{H}$ on the preassigned digits

$v_2(g) =$ 2-adic valuation of g.

- If g is odd or $v_2(g) \geqslant 3$,

$$\mathcal{H}(g): \quad \{0\} \subset A,\ (d_0, g) = 1,\ d_0 \text{ square mod } g.$$

- If $v_2(g) = 2$,

$$\mathcal{H}(g): \quad \{0,1\} \subset A,\ (d_0, g) = 1,\ d_1 g + d_0 \text{ square mod } g^2.$$

- If $v_2(g) = 1$ (e.g. $g = 2$ or $g = 10$),

$$\mathcal{H}(g): \quad \{0,1,2\} \subset A,\ (d_0, g) = 1,\ d_2 g^2 + d_1 g + d_0 \text{ square mod } g^3.$$

Result

Theorem 2 (S. 2023+)

For any $g \geqslant 2$, there exist an explicit $c = c(g) \in]0, 1/2[$ and $\delta = \delta(g) > 0$ such that for any $n \geqslant 3$, for any $A \subset \{0, \dots, n-1\}$ and $\boldsymbol{d} = (d_j)_{j \in A} \in \{0, \dots, g-1\}^A$ satisfying $\mathcal{H}(g)$, $n - 1 \in A$, $d_{n-1} \geqslant 1$ and

$$|A| \leqslant cn,$$

we have

$$|\mathcal{S} \cap \{k < g^n : \forall j \in A,\, \varepsilon_j(k) = d_j\}| = \mathfrak{S}(g, n, A, \boldsymbol{d}) \left(1 + O_g\left(n^{-\delta}\right)\right)$$

where

$$\mathfrak{S}(g, n, A, \boldsymbol{d}) = \sum_{\substack{k < g^n \\ \forall j \in A,\, \varepsilon_j(k) = d_j}} \frac{\eta(g)}{2\sqrt{k}}, \quad \eta(g) = \begin{cases} 2^{\omega(g)}, & g \text{ odd}, \\ 2^{\omega(g)+1}, & g \text{ even}. \end{cases}$$

In particular, the order of magnitude of $|\mathcal{S} \cap \{k < g^n : \forall j \in A,\, \varepsilon_j(k) = d_j\}|$ is $g^{\frac{n}{2} - |A|}$.

Explicit values of c

Theorem 2 holds with $c(g)$ given by

g	2	3	4	5	10	16	2^{32}	2^{64}
$c(g) \cdot 10^2$	0.5	0.9	1.1	1.3	1.6	1.8	3.6	4

Lemma (S.)

Let m such that $\frac{n}{4} - m \to +\infty$ as $n \to +\infty$. Choose

$$A = \{0, 2, 4, \dots, 2(m-1), n-1\}.$$

Let s such that $s \equiv 1 \bmod 8$ and $s \equiv 0 \bmod 5^{2m-1}$ and let $d \in \{0, \dots, 9\}$. Choose

$$d_{2i} = \varepsilon_{2i}(s) \text{ for } i = 0, \dots, m-1, \quad d_{n-1} = d.$$

Then we have

$$|\mathcal{S} \cap \{k < 10^n : \forall j \in A,\, \varepsilon_j(k) = d_j\}| = \frac{C(d)}{2^{|A|}}\, 10^{\frac{n}{2} - |A|}\,(1 + o(1))$$

where $C(d) > 0$ depends only on d.

So **the order of magnitude may be smaller** than $10^{\frac{n}{2} - |A|}$.

Idea: at the positions $1, 3, \dots, 2m-3$, the digits of k have to be the digits of s.

Notations for the proof of Theorem 2

- $\mathrm{e}(x) = \exp(2i\pi x)$, $x \in \mathbb{R}$.
- $\mathcal{S} = \{\ell^2,\ \ell \geqslant 0\}$ the set of squares.
- $\mathcal{D}(n, A, \boldsymbol{d}) = \{k < g^n : \forall j \in A,\ \varepsilon_j(k) = d_j\}$.
- $N = g^n$.

We want to estimate

$$\sum_{N_0 \leqslant k < N_1} \mathbf{1}_{\mathcal{S}}(k) \mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}(k)$$

where $N_0 = d_{n-1} g^{n-1}$ and $N_1 = (d_{n-1} + 1) g^{n-1}$.

Method

Use the **circle method**:

$$\sum_{N_0 \leqslant k < N_1} \mathbf{1}_{\mathcal{S}}(k)\mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}(k) = \int_0^1 S(\alpha)\overline{R(\alpha)}d\alpha$$

where

$$\underbrace{S(\alpha) = \sum_{N_0 \leqslant k < N_1} \mathbf{1}_{\mathcal{S}}(k)\,\mathrm{e}(k\alpha)}_{\substack{\text{can be large only when } \alpha \text{ is close to} \\ \text{a rational with small denominator} \\ \text{i.e. } \alpha \text{ is in a major arc}}} \quad \text{and} \quad \underbrace{R(\alpha) = \sum_{N_0 \leqslant k < N_1} \mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}(k)\,\mathrm{e}(k\alpha)\,.}_{\text{depends on the digital conditions}}$$

- integral over major arcs $\rightarrow$ main term ($+$ error term)
- integral over minor arcs $\rightarrow$ error term

Fourier transform of $\mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}$

$$F_n(\alpha) = \frac{1}{g^{n-|A|}} \sum_{k<g^n} \mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}(k)\, \mathrm{e}(k\alpha) = \frac{1}{g^{n-|A|}} R(\alpha).$$

By writing k in base g, we obtain:

$$|F_n(\alpha)| = \prod_{\substack{0\leqslant j\leqslant n-1 \\ j\notin A}} \frac{\Phi_g\left(g^j\alpha\right)}{g} \qquad \text{where } \Phi_g(t) = \left|\sum_{v=0}^{g-1} \mathrm{e}(vt)\right| = \left|\frac{\sin \pi g t}{\sin \pi t}\right|.$$

For $g = 2$,

$$|F_n(\alpha)| = \prod_{\substack{0\leqslant j\leqslant n-1 \\ j\notin A}} \left|\cos \pi 2^j \alpha\right|.$$

We need very strong upper bounds for $\|F_n\|_1$ and some (weighted) averages of $|F_n(a/q)|$.

Minor arcs contribution

$$\int_{\mathfrak{m}} \left| S(\alpha)\overline{R(\alpha)} \right| d\alpha = g^{n-|A|} \int_{\mathfrak{m}} \left| S(\alpha)\overline{F_n(\alpha)} \right| d\alpha \leqslant g^{n-|A|} \, \|F_n\|_1 \sup_{\alpha \in \mathfrak{m}} |S(\alpha)|$$

- Use the strong upper bound:

$$\|F_n\|_1 \ll N^{\xi - 1} \log N \qquad \text{(trivial: } 1)$$

where ξ is explicit and $\xi \to 0$ as $|A|/n \to 0$.

- Use a classical estimate on Weyl sums to bound $|S(\alpha)|$ over the minor arcs:

$$\sup_{\alpha \in \mathfrak{m}} |S(\alpha)| = \sup_{\alpha \in \mathfrak{m}} \left| \sum_{\sqrt{N_0} \leqslant \ell < \sqrt{N_1}} \mathrm{e}(\ell^2 \alpha) \right| \ll \frac{\sqrt{N}}{\sqrt{B_1}} \qquad \text{(trivial: } \sqrt{N})$$

where B_1 is a small power of N.

Contribution of all major arcs around a/q, q fixed

Up to an admissible error, the contribution of all major arcs around a/q (q fixed) is

$$\mathcal{C}(q) := \sum_{N_0 \leqslant k < N_1} \frac{\mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}(k)}{2\sqrt{k}} H(q,k)$$

where

$$H(q,k) = \frac{1}{q} \sum_{\substack{1 \leqslant a \leqslant q \\ (a,q)=1}} G(q,a)\, \mathrm{e}\left(\frac{-ka}{q}\right) \quad \text{and} \quad G(q,a) = \sum_{u=1}^{q} \mathrm{e}\left(\frac{au^2}{q}\right).$$

- $q \mapsto H(q,k)$ is multiplicative, simple formulas for $H(p^\nu, k)$.
- Write $q = sq'$ where $(p \mid s \Rightarrow p \mid g)$ and $(q', g) = 1$.
- Distinguish two cases depending on s and q'.
  - In one case, $\mathcal{C}(q)$ is large $\to$ main term.
  - In the other case, $\mathcal{C}(q)$ is small $\to$ error term (use a strong bound for the Fourier transform of $\mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}$).

Conclusion of the proof

Taking c sufficiently small, we get

$$\sum_{k<g^n} \mathbf{1}_{\mathcal{S}}(k)\mathbf{1}_{\mathcal{D}(n,A,\boldsymbol{d})}(k) = \mathfrak{S}(g,n,A,\boldsymbol{d})\left(1 + O_g\left(n^{-\delta}\right)\right)$$

for some $\delta > 0$, where

$$\mathfrak{S}(g,n,A,\boldsymbol{d}) = \sum_{\substack{k<g^n \\ \forall j\in A,\, \varepsilon_j(k)=d_j}} \frac{\eta(g)}{2\sqrt{k}}, \quad \eta(g) = \begin{cases} 2^{\omega(g)}, & g \text{ odd}, \\ 2^{\omega(g)+1}, & g \text{ even}. \end{cases}$$

The main term comes from the major arcs around a/q with

- $q \in \{1, p\}$ if g is a prime $p \geqslant 3$,
- $q \in \{1, 4, 8\}$ if $g = 2$,
- $q \in \{1, 4, 5, 8, 20, 40\}$ if $g = 10$.

Conclusion

- In any base $g \geqslant 2$, we obtain an asymptotic formula for the number of primes and squares with a positive proportion of preassigned digits.
- We give explicit values for the proportion of digits this method allows us to preassign.

Conclusion

- In any base $g \geqslant 2$, we obtain an asymptotic formula for the number of primes and squares with a positive proportion of preassigned digits.
- We give explicit values for the proportion of digits this method allows us to preassign.

Thank you for your attention!