Prime numbers with preassigned digits

Cathy Swaenepoel

Institut de Mathématiques de Marseille,
Université d'Aix-Marseille, France.

Inter'Actions 2019, Bordeaux.

Digits

Let $g \geq 2$. Any integer $k \geq 0$ can be written in base g as

$$k = \sum_{j \geq 0} \varepsilon_j(k) g^j$$

where, for any $j \geq 0$, $\varepsilon_j(k) \in \{0, \ldots, g-1\}$ is the j-th digit of k.

$$\text{base } g \text{ expansion} \xleftrightarrow{\text{independence?}} \text{multiplicative representation}$$

(as a product of prime factors)

$$\Big\downarrow$$

many important and interesting problems

Motivation I

Are the digits of primes “random”?

- Gelfond’s conjecture: the sum of digits of primes is well distributed.
- Mauduit-Rivat, Drmota-Mauduit-Rivat
- Maynard: primes with missing digits

Prime numbers in sparse sets

- Primes of the form $2^n - 1$ (Mersenne primes) ?
- Primes of the form $n^2 + 1$?
- Primes of the form $m^2 + n^4$ (Friedlander-Iwaniec)
- Primes of the form $m^3 + 2n^3$ (Heath-Brown)
- Primes with missing digits (Maynard)

Integers with preassigned digits

Consider the set of integers $k < g^n$.

For some positions (between 0 and $n-1$),
we preassign (i.e. prescribe) the value of the digits at these positions.

The larger the number of preassigned digits is,
the smaller the number of integers $< g^n$ with these digits is.

If the number of preassigned digits tends to $+\infty$ as $n \to +\infty$ then those integers form a "thin" or "sparse" subset.

An interesting problem

Let n large, $A \subset \{0, \ldots, n-1\}$ and $(d_j)_{j \in A} \in \{0, \ldots, g-1\}^A$.

Question: Estimate $|\{p < g^n : \forall j \in A, \varepsilon_j(p) = d_j\}|$ with $|A|$ as large as possible and (almost) no restriction on the set A itself and on the digits d_j.

Very natural conditions: $0 \in A$ and $(d_0, g) = 1$.

Expected formula?

An interesting problem

Let n large, $A \subset \{0, \dots, n-1\}$ and $(d_j)_{j \in A} \in \{0, \dots, g-1\}^A$.

Question: Estimate $|\{p < g^n : \forall j \in A,\ \varepsilon_j(p) = d_j\}|$ with $|A|$ as large as possible and (almost) no restriction on the set A itself and on the digits d_j.

Very natural conditions: $0 \in A$ and $(d_0, g) = 1$.

Expected formula?

$$\sum_{\substack{p < g^n \\ \forall j \in A,\ \varepsilon_j(p) = d_j}} 1 \approx \frac{g^{n-|A|}}{\log g^n} \frac{g}{\varphi(g)} \tag{1}$$

Previous results on this problem

- **Kátai (1986)**.
- **Wolke (2005)**: asymptotic, $|A| \leq 2$.
- **Harman (2006)**: lower bound, $|A| \leq$ 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)**: There exists a constant $c > 0$ such that, for any $A \subset \{0, \ldots, n-1\}$ satisfying $0 \in A$ and $|A| \leq cn$, for any $(d_j)_{j \in A} \in \{0,1\}^A$ such that $d_0 = 1$,

$$\sum_{\substack{0 \leq k < 2^n \\ \forall j \in A,\, \varepsilon_j(k) = d_j}} \Lambda(k) = 2^{n-|A|+1}\,(1 + o(1))$$

as $n \to \infty$, where Λ is the von Mangoldt function.

My aims

Some new questions arised from Bourgain's paper (2015):

- give an explicit admissible value for the proportion c (as large as possible) in base $g = 2$,
- generalize this result in any base $g \geq 2$,
- for each $g \geq 2$, give an explicit admissible value for the proportion c,
- provide and clarify some arguments which are not developped in Bourgain's paper.

New result

Theorem 1 (S.)

Let $g \geq 2$. There is an explicit $c_0 = c_0(g) \in]0, 1/2[$ such that: for any $0 < c < c_0$, there exist $n_0 = n_0(g,c) \geq 1$ and $\delta = \delta(g,c) > 0$ such that for any $n \geq n_0$, $A \subset \{0, \dots, n-1\}$ satisfying $0 \in A$ and

$$|A| \leq cn,$$

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

$$\sum_{\substack{0 \leq k < g^n \\ \forall j \in A, \varepsilon_j(k) = d_j}} \Lambda(k) = g^{n-|A|} \frac{g}{\varphi(g)} \left(1 + O_{g,c}\left(n^{-\delta}\right)\right).$$

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

$\rightarrow$ In any base, the number of primes with a given proportion $< c_0$ of preassigned digits is asymptotically as expected (under some technical conditions on A and the digits d_j).

Theorem 1 holds with $c_0(g)$ given by

g	2	3	4	5	10	10^3	$2 \cdot 3^{100}$	2^{200}
$c_0(g) \cdot 10^3$	2.1	3.1	3.6	4.0	4.7	6.8	0.7	9.0

Notations for the proof of Theorem 1

Let $n \in \mathbb{N}$ be large, $A \subset \{0, \dots, n-1\}$ and $(d_j)_{j \in A} \in \{0, \dots, g-1\}^A$ such that $0 \in A$ and $(d_0, g) = 1$. Denote $N = g^n$ and

$$f(k) = \begin{cases} 1 & \text{if } 0 \leq k < g^n \text{ and for any } j \in A,\ \varepsilon_j(k) = d_j, \\ 0 & \text{otherwise.} \end{cases}$$

We want to establish that if $|A| \leq cn$ with c small enough then

$$\sum_{k \leq N} \Lambda(k) f(k) = g^{n-|A|} \frac{g}{\varphi(g)} + o_{g,c}(g^{n-|A|})$$

as $n \to \infty$.

Actually, we will obtain a quantitative version.

Method for the proof of Theorem 1

Use the **circle method**:

$$\sum_{k\leq N} \Lambda(k)f(k) = \int_0^1 S(\alpha)\overline{R(\alpha)}d\alpha$$

where

$$\underbrace{S(\alpha) = \sum_{k\leq N} \Lambda(k)\,\mathrm{e}(k\alpha)}_{\substack{\text{can be large only when } \alpha \text{ is close to} \\ \text{a rational with small denominator} \\ \text{i.e. } \alpha \text{ is in a major arc}}} \quad \text{and} \quad \underbrace{R(\alpha) = \sum_{k\leq N} f(k)\,\mathrm{e}(k\alpha).}_{\text{depends on the digital conditions}}$$

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

$$F(\lambda) = g^{-n} \sum_{0 \le k < g^n} f(k) \, \mathrm{e}(-k\lambda) = \; N^{-1} \, \overline{R(\lambda)}$$

By denoting $\Phi_g(t) = \left| \sum_{v=0}^{g-1} \mathrm{e}(vt) \right| = \left| \frac{\sin \pi g t}{\sin \pi t} \right|$ and writing k in base g,

$$|F(\lambda)| = g^{-|A|} \prod_{\substack{0 \le j \le n-1 \\ j \notin A}} \frac{\Phi_g\left(\lambda g^j\right)}{g}.$$

For $g = 2$, $$|F(\lambda)| = 2^{-|A|} \prod_{\substack{0 \le j \le n-1 \\ j \notin A}} \left| \cos \pi \lambda 2^j \right|.$$

We need very sharp upper bounds of $\|F\|_1$ and $|F(a/q)|$.

Major and minor arcs

$B_1 \leq B$ "small" powers of N with $B_1 = o(B)$.

- Major arcs:
$$\mathfrak{M} = \bigcup_{1 \leq q \leq B_1} \bigcup_{\substack{1 \leq a \leq q \\ (a,q)=1}} \mathfrak{M}(q,a)$$
where $\mathfrak{M}(q,a)$ is the interval $\left|\alpha - \frac{a}{q}\right| \leq \frac{B}{qN}$ modulo 1.
- Minor arcs:
$$\mathfrak{m} = [0,1[\backslash \mathfrak{M}.$$

Note that Bourgain does not introduce B_1 (i.e. $B_1 = B$) but later a localization argument in the study of the major arcs does not seem to work in all cases.

Minor arcs contribution

$$\int_{\mathfrak{m}} \left| S(\alpha)\overline{R(\alpha)} \right| d\alpha \leq N \|F\|_1 \sup_{\alpha \in \mathfrak{m}} |S(\alpha)|$$

Use a very sharp upper bound of $\|F\|_1$.

Use Vinogradov Lemma to bound $|S(\alpha)|$ over minor arcs (classical).

Major arcs contribution

$$\int_{\mathfrak{M}} S(\alpha)\overline{R(\alpha)}d\alpha = \sum_{1\le q\le B_1} \sum_{\substack{1\le a\le q\\(a,q)=1}} \int_{\left|\alpha-\frac{a}{q}\right|\le\frac{B}{qN}} S(\alpha)\overline{R(\alpha)}d\alpha$$

First step: replace the indicator function of the interval $\left|\alpha - \frac{a}{q}\right| \le \frac{B}{qN}$ by the smooth function

$$\alpha \mapsto w\left(\frac{qN}{B}\left(\alpha - \frac{a}{q}\right)\right)$$

where w satisfies: $0 \le w \le 1$, $w = 1$ on $[-1, 1]$, $\operatorname{supp} w \subset [-2, 2]$, $w \in \mathcal{C}^\infty(\mathbb{R})$, $\widehat{w}(y) = O\left(e^{-|y|^{1/2}}\right)$ for any $y \in \mathbb{R}$
(this follows from a construction of Ingham).

$\rightarrow$ This creates an error term which is bounded by $\int_{\mathfrak{m}} \left|S(\alpha)\overline{R(\alpha)}\right| d\alpha$.

$\rightarrow$ The decreasing speed of $\widehat{w}$ will be essential.

Switch to multiplicative characters

$$S(\alpha)\overline{R(\alpha)} = \text{double sum}$$

Up to an admissible error, replace $S(\alpha)$ by

$$\frac{1}{\varphi(q)} \sum_{\chi \bmod q} \tau(\overline{\chi})\chi(a) \sum_{k \le N} \chi(k)\Lambda(k)\, \mathrm{e}(k\beta) \tag{2}$$

where $\beta = \alpha - \frac{a}{q}$.

- principal characters $\rightarrow$ main term ($+$ error term)
- nonprincipal characters $\rightarrow$ error term

Contribution of the principal characters

- Use the decreasing speed of $\widehat{w}$ and $B_1 = o(B)$ to restrict the summation over $k \leq N$ in (2) to k in a "short" interval.
- Use an estimate for primes in short intervals with a good enough error term (e.g. Huxley, Karatsuba).

$$\text{principal characters} \rightarrow \sum_{\substack{q \leq B_1 \\ q \text{ sf}}} \sum_{1 \leq k_2 \leq N} f(k_2) \frac{\mu((q, k_2))}{\varphi\left(\frac{q}{(q,k_2)}\right)} + \text{ error}$$

- Use sharp upper bounds of $|F(a/q)|$.

$$\text{double sum } = g^{n-|A|} \frac{g}{\varphi(g)} + \text{ error}$$

Rem: the q's which have a non-zero contribution in the main term are the squarefree divisors of the base g.

Contribution of the nonprincipal characters

Use sharp results on zeros of Dirichlet L-functions:

- zero density estimates,
- an improved zero-free region for special moduli (Iwaniec).

Use sharp upper bounds of $|F(a/q)|$.

If g has several prime factors then new difficulties occur.

Conclusion

Choosing appropriately the parameters B_1 and B and taking c sufficiently small, we finally obtain

$$\sum_{k \leq N} \Lambda(k) f(k) = g^{n-|A|} \frac{g}{\varphi(g)} \left(1 + O_{g,c}(n^{-\delta})\right)$$

for some $\delta > 0$ depending only on g and c.

Conclusion

Choosing appropriately the parameters B_1 and B and taking c sufficiently small, we finally obtain

$$\sum_{k \leq N} \Lambda(k) f(k) = g^{n-|A|} \frac{g}{\varphi(g)} \left(1 + O_{g,c}(n^{-\delta})\right)$$

for some $\delta > 0$ depending only on g and c.

Thank you for your attention!