Smooth Integers with Restricted Digits

James Cumberbatch

April 15, 2025

Outline

What makes a number smooth?

- A number is called "smooth" or "friable" if it has no large prime factors.
- How large is generally a parameter. A number with no prime factor greater than y is called y-smooth.
- For example,

  $$4915308332089485243506261298391651 89 = 3^9 \cdot 7^2 \cdot 11^{13} \cdot 13 \cdot 17^6 \cdot 19^6$$

  is 19-smooth. It is also 20-smooth, and 21-smooth, and in general it is y-smooth for any $y \geq 19$.
- Meanwhile,

  $$943061746557897048715252259678330378$$

  $$= 2 * 47363 \times 9955680030381279149496994063703$$

  is not 19-smooth.
- Look at $\mathcal{S}(X, y)$, the smooth integers less than X with no prime factor larger than y.
- Let $\Psi(X, y)$ be the number of such integers.

How many smooth numbers are there?

- When $y = X^{1/u}$, a positive, fixed proportion of integers are smooth.

$$\lim_{X\to\infty} \frac{\Psi(X, X^{1/u})}{X} = \rho(u)$$

- The function ρ is the Dickman function, the solution to the delay differential equation

$$u\rho'(u) + \rho(u-1) = 0$$

  with initial conditions $\rho(u) = 1$ for $0 \leq u \leq 1$.

- Even when u increases with X, the value of $\Psi(X, y)$ still depends mainly on $u := \frac{\log X}{\log y}$.
- Provided $y \geq (\log X)^{1+\varepsilon}$, we have

$$\Psi(X, y) = Xu^{-u+o(u)}.$$

Why do we care about smooth numbers?

- Integers with no prime factor greater than $X^{1/u}$ for constant u are used in Waring's problem.
  - It is easier to show that n may be represented as the sum of kth powers of smooth numbers than it is to show that n may be represented as the sum of kth powers of any numbers.
- Integers with no prime factor greater than $y = \exp(c\sqrt{(\log X)(\log\log X)})$ are relevant in cryptography, where they are used to factor integers.
  - For these, we have $\frac{\Psi(X,y)}{X} \approx 1/y^{1/(2c)}$.
  - An algorithm which finds "random" smooth integers then does something with them which takes time polynomial in y will be optimized by this smoothness
- When $y = (\log X)^c$ for constant $c > 1$, the number of smooth integers is

$$\Psi(X, y) = X^{1-1/c+o(1)}$$

- When $y = (\log X)^{1+\varepsilon}$, whether or not $\Psi(X, y) \approx X\rho(u)$ is equivalent to the Riemann hypothesis.

Digitally Restricted Integers

- Given a base b, for any integer n we can represent n as

$$n = \sum_{i=0}^{k-1} a_i b^i$$

  with a_i taking values in $\{0, 1, ..., b-1\}$.
- $k := \left\lfloor \frac{\log n}{\log b} \right\rfloor + 1$ is the number of digits.
- A popular value of b is 10.
- If we limit a_i to only take values in $\mathcal{D} \subsetneq \{0, ..., b-1\}$ and can still represent n, we call n digitally restricted.
- Let $\mathcal{A}_{k,\mathcal{D}}$ be the set of all integers which obey the restriction $\mathcal{D}$ and have exactly k digits.

Examples of digitally restricted integers

- Let $\mathcal{A}_k$ be the set of digitally restricted integers which can be written with exactly k digits.
- For example, if we set $b := 10$ and $\mathcal{D} := \{0, ..., 9\} \setminus \{7\}$, then

$$491530833208948524350626129839165189$$

obeys this restriction and has 36 digits, and thus is in $\mathcal{A}_{36,\mathcal{D}}$. And $\mathcal{A}_{37,\mathcal{D}}$, since it can also be written as a 37-digit integer by adding a leading 0. Meanwhile,

$$9430617465578970487152522596783303 78$$

does not obey this restriction.

$\mathcal{A}_k$

- There aren't many digitally restricted integers.
- $\mathcal{A}_{k,\mathcal{D}}$ is the set of all restricted integers less than b^k, and it has $|\mathcal{D}|^k$ elements.
- When $|\mathcal{D}| = 9$, $b = 10$, the number of digitally restricted less than X is approximately $X^{\frac{\log 9}{\log 10}} \approx X^{.954}$.
- These behave similar to real fractals, but in the integers.
- Can also look at missing digit sets of real numbers, for example, the Cantor Set is the set of real numbers in $[0, 1]$ with no 1s in their base 3 representation.

Historical Results

- Dartyge and Mauduit proved in 2000 that for any restriction in any base, there were infinitely many almost-primes with restricted digits.
- Maynard used the circle method in 2019 to prove that there were infinitely many primes with no 7s (Or with no copies of any other given digit) in base 10.
- Green proved that for any restriction where $\mathcal{D}$ contains at least two coprime digits, any sufficiently large integer can be expressed as the sum of at most b^{160k^2} many kth powers of digitally restricted integers.

Why do we care about digital restrictions?

- Digitally restricted integers have lots of additive structure, therefore we expect them to have no multiplicative structure.
- They come up in additive combinatorics, where they provide several important counterexamples, usually with $\mathcal{D} \subset [0, b/2)$.

Other studied restrictions

- Banning certain digits are not the only digital restriction which have been studied.
- Can prescribe some digits, this creates something similar to an arithmetic progression of short intervals.
  - For example, consider all numbers of the form X1XXXX357X24XX1
  - Bourgain in 2015 proved that there are infinitely many primes with a positive proportion of the binary digits prescribed
  - Swaenepoel proved that there are infinitely many primes with a positive proportion of digits in any base prescribed, also that there are infinitely many such squares.
- Can look at the sum of the digits.
  - Mauduit and Rivat in 2010 proved that the sum of digits of primes is evenly distributed mod n

Question: How many smooth numbers are there with restricted digits?

- When $0 \in \mathcal{D}$, we can see that 10^n (Or $2 \cdot 10^n = 200...0$ if 1 is banned) is trivially both digitally restricted and 5-smooth.
- Although there are infinitely many values of 10^n, there are only $O(\log X)$ many values less than X.
- When $0 \notin \mathcal{D}$, one can use polynomial identities to obtain some degree of smoothness.
- For example, $10^n - 1 = 999...9$, or $(10^n - 1) * d/9 = ddd...d$ for any repeated digit d, is approximately $10^{\varphi(n)}$-smooth.
- This is not particularly smooth (At best we get $y = \exp(c(\log X)/\log\log\log X))$, and there are even fewer of these values.

How many digitally restricted smooth integers would we expect?

- Naively, we might guess that the two sets are independent, that $P(\text{Restricted and Smooth}) \approx P(\text{Restricted})P(\text{Smooth})$.
  - This would lead to $|\mathcal{A}_{k,\mathcal{D}} \cap \mathcal{S}(X, y)| \approx |\mathcal{A}_{k,\mathcal{D}}|\Psi(X, y)X^{-1}$

Results

Theorem (C., 2025+)

For $b = 10$ or b large, for any $\mathcal{D}$ such that $|\mathcal{D}| = b - 1$, there is some $\delta > 0$ such that for for any large $X = b^k$, and any y such that $X^\delta > y > \exp((\log\log X)^7)$, the number of y-smooth numbers in $\mathcal{A}_{k,\mathcal{D}}$ is

$$\frac{\Psi(X, y)|\mathcal{A}_{k,\mathcal{D}}|}{X}(1 + o(1))$$

- This should be true even when $y > X^\delta$, but that would require completely different methods, it's more like counting primes or almost-primes.

How many digitally restricted smooth integers would we expect?

- Naively, we might guess that the two sets are independent, that $P(\text{Restricted and Smooth}) \approx P(\text{Restricted})P(\text{Smooth})$.
  - This would lead to $|\mathcal{A}_{k,\mathcal{D}} \cap \mathcal{S}(X,y)| \approx |\mathcal{A}_{k,\mathcal{D}}|\Psi(X,y)X^{-1}$
  - True when $\exp((\log\log X)^7) \leq y \leq X^{\delta}$.
- Being slightly less naive, we can look at large-scale distribution and distribution in residue classes.

Distribution of smooth integers at a large scale

- Our goal here is to approximate
$$\frac{\Psi(tX, y)}{\Psi(X, y)}$$
as a function of t.
- When y is much larger than any fixed power of $\log X$, it is true that
$$\frac{\Psi(tX, y)}{\Psi(X, y)} \approx t,$$
so smoothness is independent of size.
- When $y = (\log X)^c$, the above is not true. Instead, we have
$$\frac{\Psi(tX, y)}{\Psi(X, y)} \approx \frac{(tX)^{1-1/c}}{X^{1-1/c}} \approx t^{1-1/c}.$$

Distribution of smooth integers in residue classes

- Define

$$\Psi(X, y; q, a) := \#\{n < X : y \in \mathcal{S}(X, y) \text{ and } n \equiv a \mod q\}$$

- When y is much larger than any fixed power of $\log X$, we have

$$\Psi(X, y; q, a) \approx \frac{\Psi(X, y)}{q}$$

- Observe that if qn is y-smooth then so is n. If n and q are both y-smooth then so is qn. Hence,

$$\Psi(X, y; q, 0) = \begin{cases} \Psi(X/q, y) & q \text{ is } y\text{-smooth} \\ 0 & q \text{ is not } y\text{-smooth} \end{cases}$$

- Hence we have

$$\Psi(X, (\log X)^c; q, 0) \approx q^{-(1-1/c)}\Psi(X, (\log X)^c)$$

Distribution of digitally restricted integers on a large scale

- Let $N(u)$ be the number of elements in $\mathcal{A}_{k,\mathcal{D}}$ which are at most u. A way to count these elements is to look at the base-10 representation of u. For example, let's look at

$$N(59834721568193789547)$$

- If u has any banned digits, replace everything after the first one with 0s.

$$59834721568193789547 \to 59834700000000000000$$

- Next, shift everything above a banned digit down by 1, and interpret the result in base 9.

$$59834700000000000000 \to 58734700000000000000_9$$

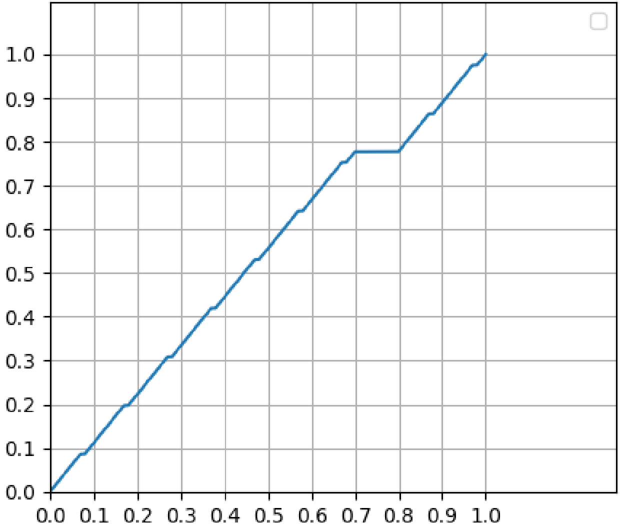
$$= 8078298705280738242$$

- This procedure resembles the Cantor function. Let $\mathscr{C}_{b,\mathcal{D}}(t)$ be the above procedure when applied to real numbers in $[0, 1]$.

Distribution of digitally restricted integers on a large scale

- If we want to count how many elements of $\mathcal{A}_{k,\mathcal{D}}$ are at most tX for any real number t, doing the above procedure to t the real number yields the correct answer up to an error of at most 1.
- Thus $\frac{\#\{n\in\mathcal{A}_{k,\mathcal{D}}:n<tX\}}{|\mathcal{A}_{k,\mathcal{D}}|} = \mathscr{C}_{\boldsymbol{b},\mathcal{D}}(t)$.

A graph of $\mathscr{C}_{10,\mathcal{D}}(t)$ for $\mathcal{D} = \{0, 1, ..., 9\} \setminus \{7\}$



Distribution of digitally restricted integers in residue classes

- Let $N(u; q, a)$ be the number of elements of $\mathcal{A}_{k,\mathcal{D}}$ which are at most u and which are a mod q.
- When $(q, 10) = 1$ and $q < \exp(c(\log X)/(\log\log X))$, $N(X; q, a) \approx N(X)/q$.
- The same does NOT hold when $q > \exp(c(\log X)/(\log\log X))$
- When we look at the distribution of $\mathcal{A}_{k,\mathcal{D}}$ modulo 10^j, we see that the intersection of $\mathcal{A}_{k,\mathcal{D}}$ with any given interval $(n10^j, (n+1)10^j]$ is either empty or a copy of $\mathcal{A}_{j,\mathcal{D}}$.
- Therefore we have that

$$N(u; 10^j, a) = 1_{\mathcal{A}_{j,\mathcal{D}}}(a)\left(\frac{N(u)}{|\mathcal{A}_{j,\mathcal{D}}|} + O(1)\right)$$

How many smooth integers do we expect? Continued

- When trying to estimate the size of $\mathcal{A}_{k,\mathcal{D}} \cap \mathcal{S}(X, (\log X)^c)$, we find...
- The "relative probability" an integer tX of unknown residue class is smooth is $\frac{\mathrm{d}t^{1-1/c}}{\mathrm{d}t} = (1-1/c)t^{-1/c}$.
- The "relative probability" an integer tX of unknown residue class is digitally restricted is $\frac{\mathrm{d}\mathscr{C}(t)}{\mathrm{d}t}$
- Hence the real density provides a corrective factor of

$$(1-1/c)\int_0^1 t^{-1/c}\,\mathrm{d}\mathscr{C}_{10,\mathcal{D}}(t)$$

- When we account for the distribution modulo powers of 10, we obtain a corrective factor of

$$\left(\prod_{p|10} p^{1/c}\frac{p-p^{1/c}}{p-1}\right)\lim_{j\to\infty}|\mathcal{A}_{b^j,\mathcal{D}}|^{-1}\sum_{n\in\mathcal{A}_{10^j,\mathcal{D}}}(n,10^j)^{1/c}$$

How many smooth integers do we expect? Continued

- When trying to estimate the size of $\mathcal{A}_{k,\mathcal{D}} \cap \mathcal{S}(X, (\log X)^c)$, we find...
- The "relative probability" an integer of unknown size which is a mod 10^j is smooth is

$$\left(\prod_{p|10} p^{1/c} \frac{p - p^{1/c}}{p-1}\right) (a, 10^j)^{1/c}$$

.

- The "relative probability" an integer of unknown size which is a mod 10^j is digitally restricted is

$$\begin{cases} |\mathcal{A}_{b^j,\mathcal{D}}|^{-1} & a \in \mathcal{A}_{b^j,\mathcal{D}} \\ 0 & o.w. \end{cases}$$

- Hence the local density provides a corrective factor of

$$\left(\prod_{p|10} p^{1/c} \frac{p - p^{1/c}}{p-1}\right) \lim_{j\to\infty} |\mathcal{A}_{b^j,\mathcal{D}}|^{-1} \sum_{n \in \mathcal{A}_{10^j,\mathcal{D}}} (n, 10^j)^{1/c}$$

Result 2

Theorem (C., 2025+)

There exists c_1 such that the following is true. If $b = 10$ or b is sufficiently large, and if $\mathcal{D} \subsetneq \{0, 1, \ldots, b-1\}$ with $|\mathcal{D}| = b - 1$, then for all large $X = b^k$ and y with $(\log X)^{c_1} < y < \exp((\log X)^{1-\varepsilon})$, we have the asymptotic

$$|\mathcal{A}_{k,\mathcal{D}} \cap \mathcal{S}(X, y)| \sim \mathfrak{S}_b(\alpha, b, \mathcal{D})\mathfrak{S}_\infty(\alpha, b, \mathcal{D}) \frac{|\mathcal{A}_{k,\mathcal{D}}||\mathcal{S}(X, y)|}{X}$$

where $\alpha := 1 - \frac{\log\log X}{\log y}$,

$$\mathfrak{S}_b(\alpha, b, \mathcal{D}) := \left(\prod_{p|b} p^{-(1-\alpha)} \frac{p - p^{1-\alpha}}{p - 1}\right) \lim_{j\to\infty} |\mathcal{A}_{j,\mathcal{D}}|^{-1} \sum_{n\in\mathcal{A}_{b^j,\mathcal{D}}} (n, b^j)^{1-\alpha},$$

and

$$\mathfrak{S}_\infty(\alpha, b, \mathcal{D}) := \alpha \int_0^1 t^{-(1-\alpha)} \, \mathrm{d}\mathscr{C}_{b,\mathcal{D}}(t)$$

A sketch of the proof

- We use the Hardy-Littlewood circle method.
  - Usually this method requires lots of variables, so it is surprising that we can apply it to the two-variable equation $n = m$.
  - A digitally restricted integer is secretly the sum of many variables.

$$n = n_0 + 10n_1 + 10^2 n_2 + \ldots + 10^{k-1} n_{k-1}$$

- Take a Fourier transform of both sets and use orthogonality.

$$f(\theta) := \sum_{n \in \mathcal{A}_{k,\mathcal{D}}} e^{2\pi i n\theta}$$

$$g(\theta) := \sum_{n \in \mathcal{S}(X,y)} e^{2\pi i n\theta}$$

The intersection is counted by

$$\frac{1}{X} \sum_{v=0}^{X-1} f(v/X) g(-v/X)$$

Proof Sketch: The Major Arcs

The intersection is counted by

$$\frac{1}{X}\sum_{v=0}^{X-1} f(v/X)g(-v/X)$$

- In sums like these, points near rationals with small denominator contribute information on the large-scale distribution and distribution in residue classes. These are called the major arcs, and are used to find the main term.
- In sums like these, points which are not near a rational with small denominator hopefully do not contribute any information. These are called the minor arcs, and are used to find the error term.
- When y is large, the major arcs other than $v = 0$ are negligible, so the main term is $f(0)g(0)/X$.
- When y is smaller, major arcs around denominators with denominator a dividing X contribute the densities in the theorem.

Proof Sketch: The Minor Arcs

$$\frac{1}{X}\sum_{v=0}^{X-1} f(v/X)g(-v/X)$$

- When v/X is not close to a rational with small denominator, $|g(-v/X)|$ is small.
- When b is large we have that the ℓ^1 norm of f is small
- Hence the sum of $|fg|$ over the minor arcs is small.
- When $b = 10$, a more complicated analysis is required and we need to view g as a bilinear sum.

Results summary

- When $\exp((\log\log X)^7) < y < X^\delta$, being smooth and being digitally restricted are independent.

$$|\mathcal{S}(X,y) \cap \mathcal{A}_{k,\mathcal{D}}| = \frac{\Psi(X,y)|\mathcal{A}_{k,\mathcal{D}}|}{X}(1+o(1))$$

- When $\exp((\log X)^{1-\varepsilon}) > y > (\log X)^{c_1}$, we need to account for the real and local densities.

$$|\mathcal{A}_{k,\mathcal{D}} \cap \mathcal{S}(X,y)| = \mathfrak{S}_b \mathfrak{S}_\infty \frac{|\mathcal{A}_{k,\mathcal{D}}||\mathcal{S}(X,y)|}{X}(1+o(1))$$

Thank you