Complexity of Leading Digit Sequences

Xinwei He, Yuchen Li, Yunyi Zhang, A.J. Hildebrand (Faculty Mentor)

University of Illinois at Urbana-Champaign

Joint Mathematics Meetings, San Diego, CA, January 10-13, 2018

Measuring Complexity of Sequences

Block Complexity of Sequence S, $p_S(n)$

- **Definition (Allouche, 2012):** Given a sequence S, the block complexity $p_S(n)$ of S is defined by

  $\mathbf{p_S(n)} = \#$ of blocks of length n appearing in the sequence S
- **Example:** S = sequence of leading digits of 3^n in base 5.
  - **Sequence 3^n: 1 3 14 102 311 1433 10404 32222 202221...**
  - **Leading digit sequence: 13111311321321421421421131113213...**
  - 4 blocks of length 1: **1 2 3 4**. Thus $p_S(1) = 4$.
  - 7 blocks of length 2: **13 31 11 32 21 14 42**. Thus $p_S(2) = 7$.

Cyclomatic Complexity of Graph G, c_G

- **Definition (Berge, 1973; McCabe, 1976):** Given a directed graph G, the cyclomatic complexity c_G of G is defined by

  $$\mathbf{c_G} = e - n + p,$$

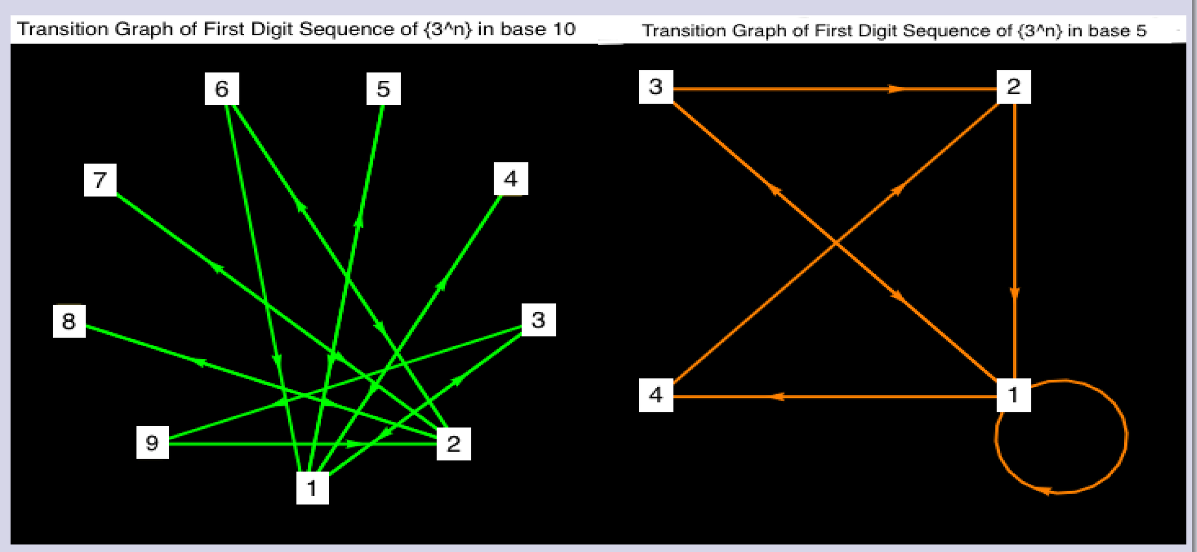
  where e is the # of edges, n is the # of vertices and p is the # of connected components in the graph G.
- **Remark:** Cyclomatic complexity was proposed by McCabe as a complexity measure of computer programs.

Cyclomatic Complexity of Sequence S, c_S

- **Definition (Iyengar et al., 1983):** Given a sequence S, the cyclomatic complexity c_S of S is defined by

  $$\mathbf{c_S} = \mathbf{c_G},$$

  where G is the directed transition graph associated with the sequence S (i.e. the graph $G = (V, E)$, where V = set of distinct digits of S, E = set of tuples representing adjacent digits of S).



- **Example**: Shown above are the transition graphs of the sequences of leading digits of 3^n in base 10 and base 5. By the formula above, the cyclomatic complexities for these sequences are $c_S = 15 - 9 + 1 = 7$ and $c_S = 7 - 4 + 1 = 4$, respectively.
- **Remark**: The use of cyclomatic complexity in the context of sequences was first proposed by Iyengar et al. (1983), who showed that the sequence S of leading digits of 2^n in base 10 has cyclomatic complexity $c_S = 5$.

Motivation: "Simple" vs "Complex" Sequences

- 3 3 3 4 1 4 3 1 1 4 2 4 4 3 4 1 1 2 2 1 4 3 2 2 4 2 3 2 4 1 1 4 1 1 2 1 4 4 4 4 3 4 3 2 1 4 3 1 3 ... (**Random Digits in base 5**)
- 1 3 1 1 3 1 1 3 2 1 3 2 1 4 2 1 4 2 1 4 2 1 1 3 1 1 3 1 1 3 2 1 3 2 1 4 2 1 4 2 1 4 2 1 1 3 1 1 3 ... (**First Digits of 3^n in base 5**)
- 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 3 4 2 1 ... (**Last Digits of 3^n in base 5**)

Main Problem: Complexity of Leading Digit Sequences

Random sequences (such as the first sequence above) are in some sense the most complex, whereas periodic sequences (such as the third sequence above) are in some sense the least complex. Between the two extremes, how complex are leading (i.e., first) digit sequences such as the second sequence above?

Results and Conjectures

Definition: Leading Digit Sequences $S_{a,b}$

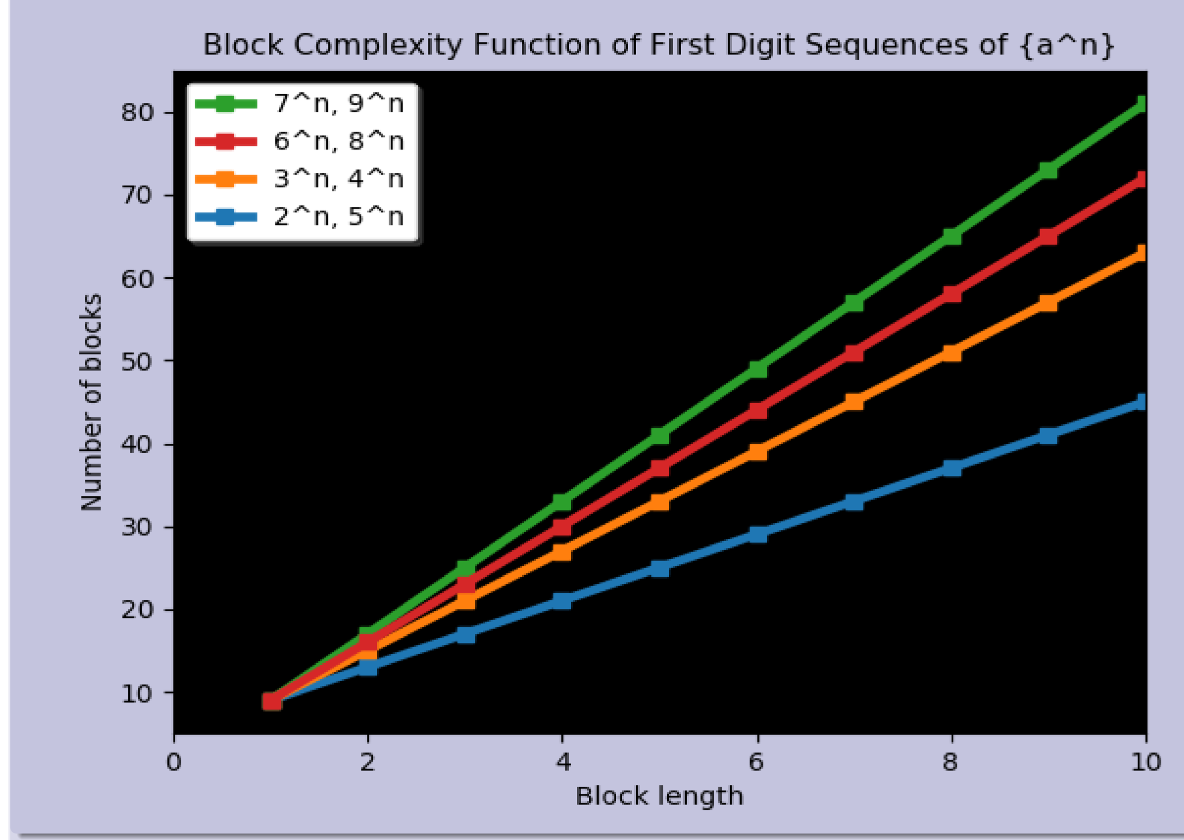
Let $b \in \mathbb{N}$, $b \geq 3$, $a \in \mathbb{R}$, $a > 1$. The sequence $S_{a,b}$ is defined as

$$S_{a,b} = \{D_b(a^n) : n = 1, 2, 3, \dots\},$$

where $D_b(x)$ denotes the leading digit of x in base b. i.e.

$$D_b(x) = d \iff d \cdot b^k \leq x < (d+1)b^k \text{ for some integer } k.$$

Block Complexity of Sequences $S_{a,b}$: Numerical Data



Main Theorem (Block and Cyclomatic Complexity of $S_{a,b}$)

Let $a \in (1, b)$, such that a is not a rational power of b, and let $S = S_{a,b}$. Then:

- If $a \in \mathbb{Q}$, $a = p/q$, where $p, q \in \mathbb{N}$ and $(p, q) = 1$, then

  $$\boxed{p_S(n) = c \cdot n + d, \quad c_S = c + 1}$$

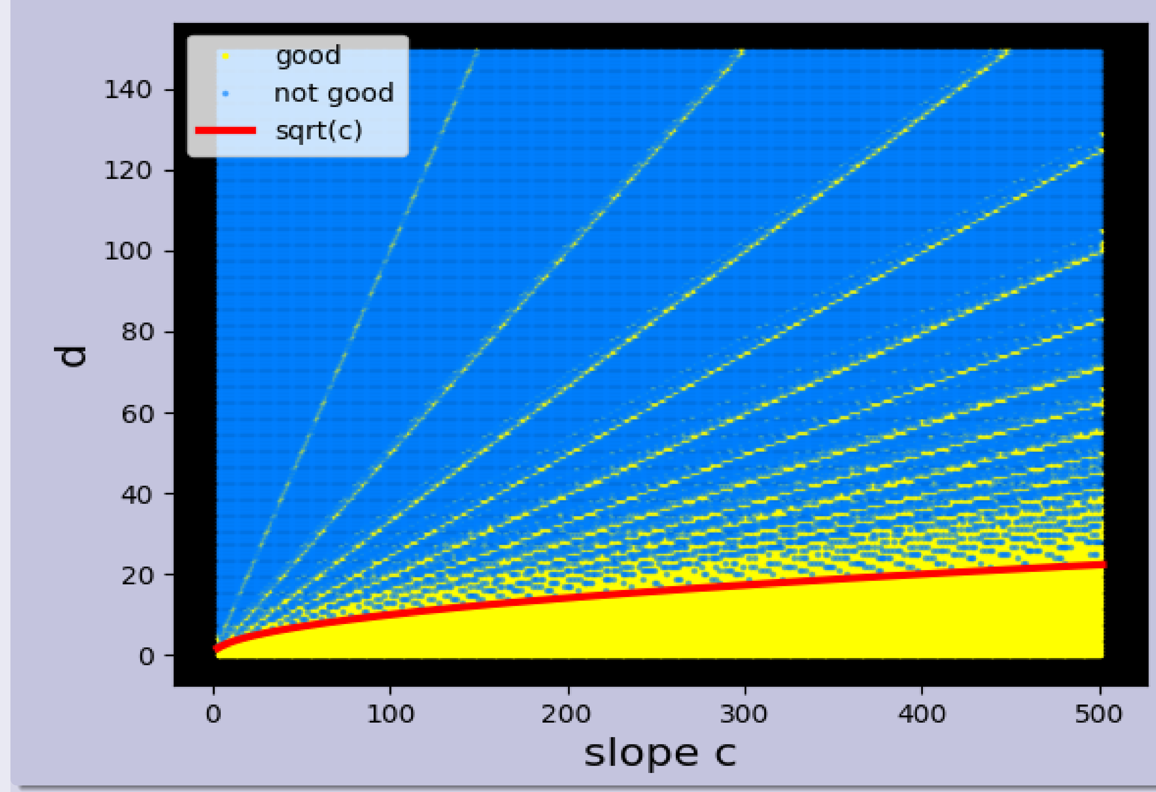
  where $d = \left\lceil \frac{b-1}{p} \right\rceil + \left\lceil \frac{(b,p)-1}{q} \right\rceil$ and $c = b - 1 - d$.
- If $a \notin \mathbb{Q}$, then

  $$\boxed{p_S(n) = (b-1) \cdot n, \quad c_S = b}$$

Definition: Leading Digit Complexity Function

- A linear function $cn + d$ is called a **leading digit complexity function** if there exist integers $a \geq 2$ and $b \geq 3$ such that $cn + d$ is the complexity function of the sequence $S_{a,b}$.
- A pair (c, d) of integers is called **good** if it is the pair of coefficients of a leading digit complexity function.

Good Pairs (c, d): Numerical Data



Conjecture (Good Pairs (c, d))

- Every pair (c, d) of integers such that $c \geq 3$ and

  $$0 \leq d \leq (1/2)\sqrt{c}$$

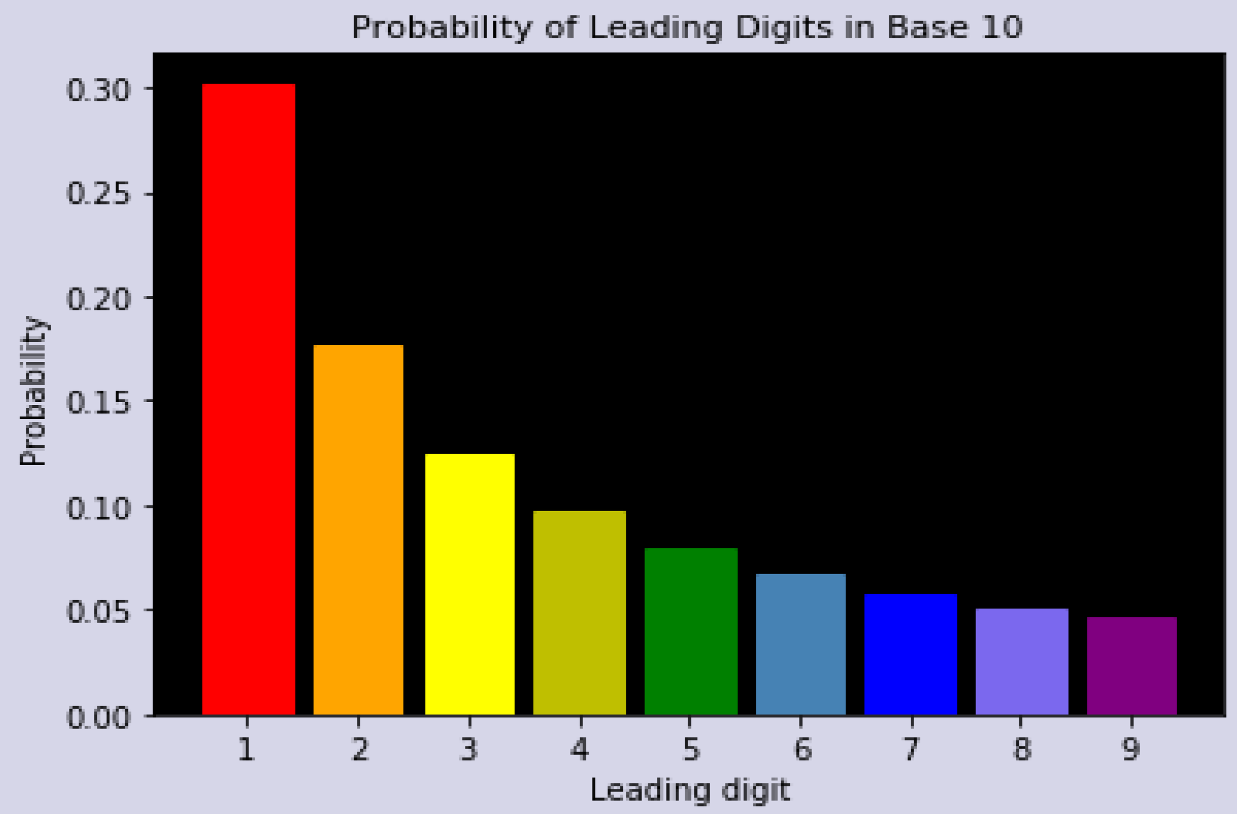
  is good, i.e., $cn + d$ is a leading digit complexity function for every such pair (c, d).
- For any fixed integer $c \geq 3$ there are at most $4\sqrt{c}$ integers d such that (c, d) is good.
- The total number of good pairs (c, d) with $c \leq N$ is of order $\Theta(N^{3/2})$ as $N \to \infty$.

Related Results

Benford's Law

- **Benfords's Law (1938)**: A sequence S of real numbers satisfies Benford's law for base b, if for each $d = 1, \dots, b-1$,

  $$P(\text{leading digit in base } b \text{ is } d) = \log_b(1 + \frac{1}{d}).$$
- **Diaconis (1997)**: The sequence $\{a^n\}$ satisfies Benford's law in base b if $\log_b(c) \notin \mathbb{Q}$.



References

- **Allouche, J.-P.,** *Surveying some notions of complexity for finite and infinite sequences, RIMS Kokyuroku Bessatsu, B34, pp. 27-37, 2012*
- **Benford, F.,** *The Law of Anomalous Numbers, Proc. Amer. Philos. Soc., 78(4) (1938), pp. 551-572.*
- **Berge, C.,** *Graphs and Hypergraphs. New York: American Elsevier Publishing Company, Inc., 1973.*
- **Diaconis, P.,** *The Distribution of Leading Digits and Uniform Distribution Mod 1, Annals of Probability, Vol.5, No. 1 (1997), pp. 72-81*
- **Iyengar, S., Rajagopal, A. K., Uppuluri, V. R. R.,** *String Patterns of Leading Digits. Applied Mathematics and Computation 12, (1983), pp. 321-337.*
- **Kak, S.,** *Strings of First Digits of Powers of a Number. Indian Journal of Pure and Applied Mathematics 14, no. 7 (1983), pp. 896-907.*
- **McCabe, T. J.,** *A Complexity Measure. IEEE Transactions on Software Engineering SE-2, no. 4 (1976), pp. 308-320.*