Some distribution problems concerning arithmetic functions

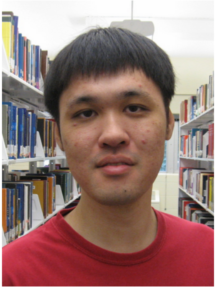
Paul Pollack
University of Georgia

LSU Number Theory Seminar

October 2022

This talk represents joint work with Fai Chandee & Xiannan Li (both at Kansas) and Akash Singha Roy (UGA).

Fai Chandee



Xiannan Li

Akash Singha Roy

Our theme will be'' the distribution of **integer-valued arithmetic functions**, i.e., functions from $\mathbb{Z}_{>0}$ to $\mathbb{Z}$.

"Function from $\mathbb{Z}_{>0}$ to $\mathbb{Z}$" is a fancy way of saying "integer sequence". So you should think of this as being about integer sequences, with special attention to sequences that arise naturally in number theory.

What "distribution" means can be interpreted in many ways. We will be interested in questions about digits, both leading and trailing.

PART I. Trailing digits

PART I. Trailing digits

If we work in base q, the rightmost digit of a positive integer n is simply the least nonnegative remainder when n is divided by q.

Thus, if f is an integer-valued arithmetic function, understanding the rightmost digit of $f(n)$ mod q amounts to understanding the distribution of $f(n)$ mod q.

Furthermore, we can access the last several digits by replacing the modulus q with a higher power of q.

Hence, the fundamental question is: How is f distributed in residue classes to a given modulus?

Definition

Let f be an integer-valued arithmetic function; that is, f is a function from $\mathbb{Z}_{>0}$ to $\mathbb{Z}$. Let q be a positive integer. We say f is **uniformly distributed modulo** q (or **equidistributed mod** q) if, for each integer a,

$$\frac{1}{x}\#\{n \le x : f(n) \equiv a \pmod{q}\} \to \frac{1}{q}, \quad \text{as } x \to \infty.$$

Example (trivial): $n \mapsto n$ is equidistributed mod q for every q.

Example (not so trivial): $n \mapsto F_n$ (nth Fibonacci number) is equidistributed mod q if and only if $q = 5^k$ for some k.
(Niederreiter, Kuipers–Shiue)

Let $A(n) = \sum_{p^k \| n} kp$ be the sum of the prime factors of n, counted with multiplicity; e.g.,

$$A(20) = 2 + 2 + 5 = 9.$$

Theorem (Alladi–Erdős)

$A(n)$ is equidistributed modulo 2. In fact,

$$\#\{n \le x : A(n) \equiv a \pmod 2\} = \frac{x}{2} + O(x \exp(-c\sqrt{\log x \log\log x})).$$

Theorem (Delange, Goldfeld)

$A(n)$ is equidistributed mod q for each fixed q. In fact,

$$\#\{n \le x : A(n) \equiv a \pmod q\} = \frac{x}{q} + O(x/\sqrt{\log x}).$$

Approach.

The philosophy is: If you want to show a sequence is equidistributed in a finite abelian group, show that it averages to zero when hit with any nontrivial character of the group.

Here, we project the integer sequence down to the corresponding sequence mod q. The q characters of the group $\mathbb{Z}/q\mathbb{Z}$ are the functions $e^{2\pi i h \cdot / q}$, for each h mod q. So to show $A(n)$ is equidistributed mod q, it is enough to show that each of the averages

$$\frac{1}{x}\sum_{n\leq x} e^{2\pi i h A(n)/q}$$

tends to 0 (for $h = 1, 2, \ldots, q-1$).

It is enough to show that each of the averages

$$\frac{1}{x}\sum_{n\leq x} e^{2\pi i h A(n)/q}$$

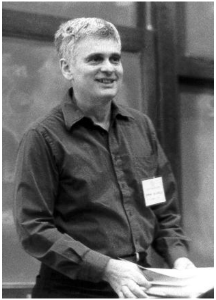
tends to 0 (for $h = 1, 2, \ldots, q-1$).

The function $A(n)$ is **additive**: $A(nm) = A(n) + A(m)$ whenever n and m are coprime. (In fact, coprimality is not needed here.) Hence, the function $e^{2\pi i h A(n)/q}$ is multiplicative, meaning

$$e^{2\pi i h A(nm)/q} = e^{2\pi i h A(n)/q} e^{2\pi i h A(m)/q}$$

whenever n, m are coprime. So our sum represents the average value of a multiplicative function taking values on the unit circle. Sums of this kind are well-studied and there are powerful tools available (Halász's theorem, the Landau–Selberg–Delange method).

Hubert Delange



Dorian Goldfeld

Equidistribution of $f(n)$ mod q is not always the right notion, even for "nice" functions f.

For example, suppose

$$f(n) = p_n, \quad n\text{th prime number.}$$

If a mod q is a residue class with $\gcd(a, q) > 1$, then $f(n)$ hits the residue class a mod q at most once. So equidistribution mod q fails for every $q > 1$.

On the other hand, it is well-known — and incredibly useful — that each **coprime** residue class mod q gets its fair share of the values p_n.

Equidistribution of $f(n)$ mod q is not always the right notion, even for "nice" functions f.

For example, suppose

$$f(n) = p_n, \quad n\text{th prime number.}$$

If a mod q is a residue class with $\gcd(a, q) > 1$, then $f(n)$ hits the residue class a mod q at most once. So equidistribution mod q fails for every $q > 1$.

On the other hand, it is well-known — and incredibly useful — that each **coprime** residue class mod q gets its fair share of the values p_n.

You might object that $f(n) = p_n$ is not so natural...

Let $\varphi(n)$ denote Euler's totient; that is, $\varphi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times$.

It is **certainly not** the case that $\varphi(n)$ is equidistributed modulo each fixed q. For example,

$$\varphi(n) \quad \text{is even once } n > 2.$$

In general, $\varphi(n)$ is divisible by q whenever $p \mid n$ for some prime $p \equiv 1 \pmod{q}$. For each fixed q, a positive proportion of primes p satisfy $p \equiv 1 \bmod q$. Moreover, most numbers n have many prime factors. So it should be rare for $\varphi(n)$ to **not** be 0 mod q.

Proposition (Landau? Erdős?)

Fix q. The limiting proportion of n with $\varphi(n) \not\equiv 0 \pmod{q}$ is 0.

Definition (Narkiewicz)

Let f be an integer-valued arithmetic function; that is, f is a function from $\mathbb{Z}_{>0}$ to $\mathbb{Z}$. Let q be a positive integer. We say f is **weakly uniformly distributed modulo** q if there are infinitely many n with $\gcd(f(n), q) = 1$ and if, for each a coprime to q,

$$\frac{\#\{n \leq x : f(n) \equiv a \pmod q\}}{\#\{n \leq x : \gcd(f(n), q) = 1\}} \to \frac{1}{\varphi(q)},$$

as $x \to \infty$.

Perhaps $\varphi(n)$ is usually weakly equidistributed mod q. We need q odd to satisfy $\gcd(\varphi(n), q) = 1$. But this is not enough. For example,

$$\#\{n \le x : \varphi(n) \equiv 1 \pmod 3\} \sim c_1 x/\sqrt{\log x},$$

while

$$\#\{n \le x : \varphi(n) \equiv -1 \pmod 3\} \sim c_{-1} x/\sqrt{\log x},$$

whereas

$$c_1 = 0.6109\ldots, \quad c_{-1} = 0.3284\ldots$$

(see Dence and Pomerance).

Thus, we can only hope for weak equidistribution modulo q when $\gcd(q, 6) = 1$.

Theorem (Narkiewicz)

Let q be any positive integer with $\gcd(q, 6) = 1$. Then $\varphi(n)$ is weakly equidistributed modulo q.

What goes wrong with $q = 3$? The numbers $p - 1$, for p prime and $p \neq 3$, either fail to be coprime to 3 or are "trapped" in the trivial subgroup of $(\mathbb{Z}/3\mathbb{Z})^{\times}$.

Approach. Similar in spirit to the $A(n)$ proof (and to the proof of Dirichlet's theorem on primes in progressions). One hits $\varphi(n)$ with the characters of the multiplicative group $(\mathbb{Z}/q\mathbb{Z})^{\times}$.

Specifically, if χ is a nontrivial Dirichlet character mod q, one needs to show that

$$\sum_{n \le x} \chi(\varphi(n))$$

exhibits cancelation relative to

$$\sum_{n \le x} \chi_0(\varphi(n)),$$

(with χ_0 the trivial character) as $x \to \infty$.

Again, the sums here are sums of multiplicative functions of modulus ≤ 1. With some elbow grease, this can be deduced from a special case of Halász's theorem due to Wirsing (or by following Landau–Selberg–Delange).

For general integer-valued additive arithmetic functions, Delange has a practical necessary and sufficient condition for uniform distribution. For "polynomial-like" multiplicative functions, Narkiewicz has a practical condition for weak uniform distribution.

These generalizations can all be thought of as working in the f-aspect.

For general integer-valued additive arithmetic functions, Delange has a practical necessary and sufficient condition for uniform distribution. For "polynomial-like" multiplicative functions, Narkiewicz has a practical condition for weak uniform distribution.

These generalizations can all be thought of as working in the f-aspect.

Question. What about the q-aspect? Can we prove (weak) equidistribution theorems when q is allowed to vary with our stopping point x?

Model. Prime numbers, again!

Theorem (Siegel–Walfisz)

As $x \to \infty$,

$$\frac{\#\{p \le x : p \equiv a \pmod q\}}{\frac{1}{\varphi(q)}\#\{p \le x\}} \to 1,$$

uniformly in coprime residue classes a mod q, as long as $q \le (\log x)^A$ (any fixed A).

Probably more is true. Under GRH, $(\log x)^A$ can be replaced with $x^{\frac{1}{2}-\epsilon}$. It is plausible q can even be taken as large as $x^{1-\epsilon}$, but this is far beyond reach.

Theorem (Singha-Roy, P., 2022+)

Fix $K > 0$. As $x \to \infty$,

$$\frac{\#\{n \le x : A(n) \equiv a \pmod q\}}{x/q} \to 1,$$

uniformly for residue classes $a \bmod q$ with $q \le (\log x)^K$.

Theorem (Singha-Roy, P., 2022+)

Fix $K > 0$. As $x \to \infty$,

$$\frac{\#\{n \le x : \varphi(n) \equiv a \pmod q\}}{\frac{1}{\varphi(q)}\#\{n \le x : \gcd(\varphi(n), q) = 1\}} \to 1,$$

uniformly for coprime residue classes $a \bmod q$ with $\gcd(q, 6) = 1$ and $q \le (\log x)^K$.

(w/ Lebowitz-Lockard: special case $q = p$, prime)

A word on the proofs

What we don't do: We don't use characters! Reducing the problem to one about mean values of multiplicative functions is the right thing to do for fixed q. But the standard methods for estimating these sums (such as Landau–Selberg–Delange, or quantitative versions of Halász's theorem) seem to yield the desired asymptotics only in much more limited ranges of q.

Instead, we develop a quasi-elementary method suggested by work of Banks–Harman–Shparlinski, who proved a theorem of the same kind on the weak-equidistribution of $P^{+}(n)$, the largest prime factor of n.

I restrict attention to the case of $A(n)$, which is simpler here than $\varphi(n)$. I also make things easier on myself by assuming q is odd.

We split off the K largest prime factors of n, say $n = mP_K \cdots P_1$, where

$$P_K \leq P_{K-1} \leq \cdots \leq P_2 \leq P_1.$$

Observe that

$$A(n) = A(m) + P_K + \cdots + P_1.$$

The idea is to fix m and obtain the equidistribtuion from the 'mixing' in $\mathbb{Z}/q\mathbb{Z}$ coming from $P_K + \cdots + P_1$.

Using the Siegel–Walfisz theorem, one can model the P_i as random coprime residue classes mod q. One then calculates that $P_1 + \ldots P_K$ is close to uniformly distributed mod $\mathbb{Z}/q\mathbb{Z}$. Here we mean that one approaches uniform distribution whenever $K \to \infty$, uniformly in q. (This step uses that q is odd.)

The argument for φ is similar. In this case, one looks at products

$$(P_K - 1) \cdots (P_1 - 1)$$

and shows that these approach equidistribution in $(\mathbb{Z}/q\mathbb{Z})^\times$.

Here each P_i is modeled as a random residue class a mod q subject to the condition that $a(a-1)$ relatively prime to q.

PART II: Leading digits

Our interest in this half of the talk will be on arithmetic functions whose values follow (or fail to follow) **Benford's law**.

To a first approximation, Benford's law asserts that for many natural data sets, the initial digits are not uniformly distributed. Rather, smaller initial digits appear more frequently than larger ones.

This phenomenon was observed by Simon Newcomb in 1881 and rediscovered by Frank Benford in 1938.

TABLE I

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average.......		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error		±0.8	±0.4	±0.4	±0.3	±0.2	±0.2	±0.2	±0.2	±0.3	—

table taken from *The Law of Anomalous Numbers* (1938)

Benford claims: The digit $d \in \{0, \dots, 9\}$ should appear with frequency $\log_{10} \frac{d+1}{d}$. For example, 1 should appear as the leading digit

$$\log_{10} 2 = 0.301029995\dots,$$

and 2 should appear with frequency

$$\log_{10} \frac{3}{2} = 0.176091259\dots.$$

Benford claims: The digit $d \in \{0, \ldots, 9\}$ should appear with frequency $\log_{10} \frac{d+1}{d}$. For example, 1 should appear as the leading digit

$$\log_{10} 2 = 0.301029995\ldots,$$

and 2 should appear with frequency

$$\log_{10} \frac{3}{2} = 0.176091259\ldots.$$

The same should hold (in large data sets) for longer strings of digits. For example, 2022 should be the leading 4 digits with frequency

$$\log_{10} \frac{2023}{2022} = 0.000214731515\ldots.$$

For finite data sets, one can expect only approximate conformance to Benford's law.

Definition

Let f be a positive integer-valued arithmetic function. We say f **obeys Benford's law** (or f **is Benford**) if, for any positive integer D, the limiting proportion n for which $f(n)$ begins with the digits of D is

$$\log_{10} \frac{D+1}{D}.$$

Reminder: The "limiting proportion" of the set of integers with property P is

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \leq x : n \text{ has } P\}.$$

(natural density/asymptotic density)

Where does Benford's law come from?

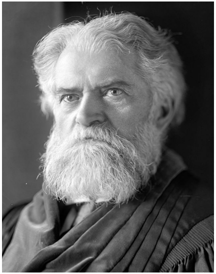
Observe that the leading digits do not depend on where one places the decimal point! That is, the leading digit of α is a function of the fractional part of its base 10 logarithm.

For example, a positive real number α has leading digit 3 precisely when

$$\log_{10} \alpha \in [\log_{10} 3, \log_{10} 4) + \mathbb{Z}.$$

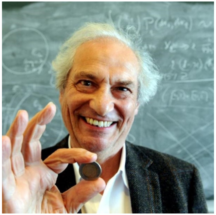
Proposition (Diaconis)

A positive integer-valued arithmetic function f is Benford $\iff$ $\log_{10} f(n)$ is uniformly distributed modulo 1.



Simon Newcomb

Frank Benford



Persi Diaconis

An unlikely criminal

Which arithmetic functions obey Benford's law?

Obeying Benford's law is a more delicate condition that is apparent on first glance.

Proposition

The identity function $f(n) = n$ is not Benford.

Proof.

Up to $x = 10^k$, there are $10^{k-1} + 10^{k-2} + \cdots + 10^1 + 1$ integers beginning with 1, which is $\approx x/9$. Up to $x' = 2 \cdot 10^k$, there are $10^k + 10^{k-1} + 10^{k-2} + \cdots + 10^1 + 1$ such integers, and this is $\approx 5x'/9$. These oscillations mean that there is **no** limiting frequency of n beginning with the digit 1.

Proposition

The identity function $f(n) = n$ is not Benford.

In a similar way, one can show ....

Proposition

The function $n \to p_n$ (nth prime) is not Benford.

Proposition

The identity function $f(n) = n$ is not Benford.

In a similar way, one can show ....

Proposition

The function $n \to p_n$ (nth prime) is not Benford.

On the other hand...

Proposition (Diaconis)

The function $n!$ is Benford.

Proposition (Massé-Schneider)

The primorial function $\prod_{k=1}^{n} p_k$ is Benford.

Chandee, Li, Singha-Roy, and I investigate ‘Benfordness’ for positive integer-valued multiplicative functions.

Here is our strategy:

$$f \text{ is Benford} \iff \log_{10} f \text{ is uniformly distributed mod } 1.$$

By Weyl's criterion for uniform distribution,

$$\begin{aligned}&\log_{10} f \text{ is uniformly distributed mod } 1\\ &\qquad \iff \frac{1}{x}\sum_{n\le x} e^{2\pi i h \log_{10} f(n)} \to 0, \quad \text{for all nonzero integers } h.\end{aligned}$$

For each fixed h, the function $n \mapsto e^{2\pi i h \log_{10} f(n)}$ is a multiplicative function taking values on the complex unit circle. Halász has a criterion for when such functions have mean value 0.

Let F, G be multiplicative functions taking values in the complex unit disc. The **pretentious distance** between F, G, denoted $\mathbf{D}(F, G)$, is defined by

$$\mathbf{D}(F, G)^2 = \sum_p \frac{1 - \Re(F(p)\overline{G(p)})}{p} \quad (\in [0, \infty]).$$

Theorem (Halász)

Let F be a multiplicative arithmetic function with $|F(n)| \leq 1$ for all n. If $\mathbf{D}(F, n^{i\alpha}) = \infty$ for every real α, then F has mean value 0. Otherwise, there is a unique $\alpha \in \mathbb{R}$ with $\mathbf{D}(F, n^{i\alpha}) < \infty$. In that case, F has mean value 0 if and only if $F(2^k) = -2^{ik\alpha}$ for every positive integer k.

We apply this to our functions $F_h(n) = e^{2\pi i h \log_{10} f(n)}$. This gives a criterion for f to be Benford (solely) in terms of the values of f at primes p.

While it may seem unwieldy at first, the pretentious distance is actually a well-behaved quantity. For example, it obeys a triangle inequality:

$$\mathbf{D}(FF', GG') \leq \mathbf{D}(F, G) + \mathbf{D}(F', G').$$

This makes it feasible to assess the Benfordity of f whenever we understand the distribution of f at prime numbers p.

Theorem (C-L-P-SR)

Euler's φ function and the classical sum of divisors function σ are not Benford.

Theorem (C-L-P-SR)

The k-fold divisor function

$$d_k(n) = \#\{(d_1, \ldots, d_k) \in (\mathbb{Z}_{>0})^k : d_1 \cdots d_k = n\}$$

is Benford if and only if k is not a power of 10.

These functions are "trivial" to access on the primes: $\varphi(p) = p - 1$, $\sigma(p) = p + 1$, and $d_k(p) = k$.

Define Ramanujan's τ-function, $\tau(n)$, by the power series identity

$$\sum_{n\geq 1} \tau(n) q^n = q \prod_{n=1}^{\infty} (1 - q^n)^{24}.$$

It was known to Ramanujan that $\tau(n)$ is multiplicative. These are the coefficients of a holomorphic cusp form of weight 12 and level 1.

In 1974, Deligne proved that $|\tau(p)| \leq 2p^{11/2}$ for all primes p.

As of about 10 years ago, it is known that the numbers

$$\frac{\tau(p)}{2p^{11/2}}$$

are distributed in $[-1, 1]$ according to the Sato-Tate measure

$$\frac{2}{\pi}\sqrt{1-x^2}\,\mathrm{d}x.$$

This is a special case of the **Sato–Tate conjecture**, proved by Barnet-Lamb, Geraghty, Harris, and Taylor.

This information on the distribution of $\tau(p)$ is enough to apply our criterion, and we find that τ is Benford... *relative to the set of n with* $\tau(n) \neq 0$. (It is still an open problem whether there are any such n.)

Here τ is for purposes of illustration; we can handle other modular form coefficient sequences for which Sato–Tate applies.

Thank you for your attention!