# ELLIPTIC DIVISIBILITY SEQUENCES AND BENFORD'S LAW

## 1. Recurrence Sequences and Benford's Law

A *recurrence sequence* is a sequence of numbers $\{f(n)\}$ indexed by $n \in \mathbb{Z}$ and satisfying some given recurrence relation. Simple examples include the integers, which satisfy the relation

$$a(n) = a(n-1) + 1,$$

and the Fibonacci numbers, which satisfy the relation

$$f(n) = f(n-1) + f(n-2).$$

Generalizing the Fibonacci relation in a natural way, we have *linear recurrence sequences*, satisfying a general relation of the form

$$a(x+n) = s_1 a(x+n-1) + \cdots + s_{n-1} a(x+1) + s_n a(x),$$

with $x \in \mathbb{Z}$ and the constant coefficients $s_i$ in some coefficient ring.

Benford's Law, also known as the "leading digits phenomenon," was first observed by Newcomb and later by Benford, both of whom noticed that in books of logarithms, the earlier pages (those corresponding to smaller leading-digits) were more worn than later pages. Benford observed the same phenomenon across many different data sets, and he found it to be even stronger when he combined data from many different sources.

For any integer base $B$ and any $x \in \mathbb{R}^+$, we may write $x = M_B(x) \cdot B^k$ where $k \in \mathbb{Z}$ and $M_B(x) \in [1, B)$ is the *mantissa.* A sequence of positive numbers $\{a(n)\}$ is *Benford base* $B$ if

$$\lim_{N\to\infty} \frac{\#\{n \le N \colon 1 \le M_B\left(a(n)\right) \le s\}}{N} = \log_B s. \tag{1}$$

In other words, the probability that the mantissa of any element of the sequence is in the interval $[1, s]$ is precisely $\log_B s$.

**To see a modern development / explanation of Benford's law in terms of probability measures, see Section 2.3 of the file Adams_Benfords-Law.pdf.**

The leading digits phenomenon discovered by Newcomb and Benford amounts to saying that many data sets are Benford base 10.

**Theorem 1.1.** *A sequence $(u_n)$ is Benford base-$B$ if and only if the sequence $(\log_B u_n)$ is distributed uniformly mod* 1*.*

**Theorem 1.2** (Kronecker-Weyl Criterion)**.** *A sequence $(cn^2)$ is distributed uniformly mod* 1 *if and only if $c$ is irrational.*

Specifically, the Fibonacci sequence — along with any linear recurrence satisfying certain conditions — are Benford base 10.

**For a proof of the Kronecker-Weyl Criterion, see the first seven pages of the file Granville_UniDistn.pdf. For proofs about the Fibonacci sequence, see Fibonacci_Benford.pdf. For other linear recurrences, see Recursive_Benford.pdf.**

## 2. Elliptic Divisibility Sequences

Generalizing the work described in Section 1, we explore the distribution properties of more complicated recurrence sequences. We begin with some definitions.

An *integral divisibility sequence* is a sequence of integers $\{u(n)\}$ satisfying

$$u(n) \mid u(m) \quad \text{whenever } n \mid m.$$

An *elliptic divisibility sequences* is an integral divisibility sequence which satsifies the following recurrence relation for all $m \geq n \geq 1$:

$$u(m+n)u(m-n)u(1)^2 = u(m+1)u(m-1)u(n)^2 - u(n+1)u(n-1)u(m)^2. \tag{2}$$

We note that elliptic divisibility sequences are special cases of *Somos sequences*, which satisfy more general bilinear recurrences.

It is a simple matter to check that certain sequences trivially satisfy this recurrence. For example:

- The sequences of integers, where $u(n) = n$.
- The sequence $0, 1, -1, 0, 1, -1, \ldots$.
- The sequence $1, 3, 8, 21, 55, 144, 377, 987, 2584, 6765, \ldots$ (this is every-other Fibonacci number).

However, there are many more interesting examples of elliptic divisibility sequences, for example:

- The sequences which begins $0, 1, 1, -1, 1, 2, -1, -3, -5, 7, -4, -28, 29, 59, 129, -314, -65, 1529, -3689, -8209, -16264, 833313, 113689, -620297, 2382785, 7869898, 7001471, -126742987, -398035821, 168705471, -7911171597, \ldots$. (This is sequence A006769 in the *On-Line Encyclopedia of Integer Sequences* [5].)
- The sequence which begins $1, 1, -3, 11, 38, 249, -2357, 8767, 496036, -3769372, -299154043, -12064147359, \ldots$.

These sequences have been studied extensively, beginning with Morgan Ward's work [6], and continuing now because of their use in cryptography, in particular their application to the elliptic curve discrete log problem (see [3], for example). Ward showed that the first three examples above are degenerate examples of elliptic divisibility sequences, in that they are all either the integers or Lucas sequences of the form

$$u(n) = \frac{a^n - b^n}{a - b} \quad \text{where} \quad a + b \in \mathbb{Z} \text{ and } ab = 1,$$

(so they satisfy a simpler recurrence as well).

Ward also showed that all examples of elliptic divisibility sequences which are not linear arise from elliptic division polynomials (see [4, Exercise 3.7]) in the following way. For an elliptic curve given by

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \qquad a_i \in \mathbb{Q} \tag{3}$$

we define the quantities

$$
\begin{aligned}
b_2 &= a_1^2 + 4a_2,\\
b_4 &= 2a_4 + a_1a_3,\\
b_6 &= a_3^2 + 4a_6,\\
b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a + 3^2 - a_4^2.
\end{aligned}
$$

Define a sequence of polynomials $(\psi_n)$ in $\mathbb{Q}[x, y]$ by:

$$
\begin{aligned}
\psi_0 &= 0,\\
\psi_1 &= 1,\\
\psi_2 &= 2y + a_1x + a_3,\\
\psi_3 &= 3x^4 + b_2x^3 + 3b_4x^2 + 3b_6x + b_8,\\
\psi_4 &= \psi_2\left(2x^6 + b_2x^5 + 5b_4x^4 + 10b_6x^3 + 10b_8x^2 + (b_2b_8 - b_4b_6)x + b_4b_8 - b_6^2\right).
\end{aligned}
$$

The recurrence, for $n \geq 2$ is given by

$$
\begin{aligned}
\psi_{2n+1} &= \psi_{n+2}\psi_n^3 - \psi_{n-1}\psi_{n+1}^3,\\
\psi_{2n}\psi_2 &= \psi_n\left(\psi_{n+2}\psi_{n-1}^2 - \psi_{n-2}\psi_{n+1}^2\right).
\end{aligned}
$$

The polynomial $\psi_n$ has

- zeroes at the $n$-torsion points of $E$ and
- a pole supported on $\mathbf{O} \in \mathbf{E}$.

Furthermore, we have the property that if $P \in E(\mathbb{Q})$ is given by

$$P = (x, y), \quad \text{then} \quad [n]P = \left(\frac{\phi_n(P)}{\psi_n(P)}, \frac{\omega_n(P)}{\psi_n(P)}\right).$$

**Theorem 2.1** (M. Ward, 1948)**.** *With the definitions above, let $P = (x, y)$ be a non-torsion rational point on the elliptic curve $E$. Then*

$$u(n) := \psi_n(P)$$

*forms an elliptic divisibility sequence. Furthermore, every non-linear elliptic divisibility sequence satisfying $u(1) = 1$ and $u(2)u(3) \neq 0$ arises in this way.*

In this construction, the intuition is as follows. Beginning with a point $P \in E(\mathbb{Q})$, we may form the sequence of multiples of $P$:

$$P, 2P, 3P, 4P, \ldots \tag{4}$$

These points are also in $E(\mathbb{Q})$, and if $P$ is not a torsion point, then this forms an infinite sequence. So instead of the sequence in (4), we may instead consider the sequence $(\psi_n(P))$.

Ward's theorem says that every nontrivial elliptic divisibility sequence is exactly this kind of sequence of denominators, where the convention that $u(1) = 1$ means that we start with a non-torsion *integral* point on some elliptic curve. The particular sequences is determined by both the curve and the initial point.

Note: Fix a rational point $P \in E$. The point $[n]P = (\phi_n(P)/\psi_n(P), \omega_n(P)/\psi_n(P))$ may not be written in lowest terms. Specifically, $\gcd(\phi_n(P), \psi_n(P))$ is supported on the primes dividing the discriminant of the curve $E$. So an elliptic divisibility sequence is almost (but not quite) the sequence of denominators of $[n]P$, where $P$ is a non-torsion rational point.

**For a more comprehensive introduction to elliptic divisibility sequences, including a proof of Ward's correspondence, see the file KateThesisCh2.**

## 3. Canonical Heights

Let $K$ be a number field with $[K\colon \mathbb{Q}] = d$ and $\alpha \in K$. Further, let $M_K$ be the set of valuations of $K$ with each valuation corresponding to the absolute value $|\cdot|_v$. The naive height $h(\alpha)$ is, in essence, a measure of the arithmetic complexity of $\alpha$. It is defined by

$$h(\alpha) = \frac{1}{d} \sum_{v \in M_K} \ln \max\{1, |\alpha|_v\}.$$

For a $K$-rational point $P$ on a elliptic curve, we set $h(P)$ to be the height of the $x$-coordinate of $P$, and $h(P) = 0$ for the point at infinity. For two points $P$ and $Q$ on an elliptic curve, the naive height satisfies an almost-parallelogram law:

$$h(P+Q) + h(P-Q) = 2h(P) + 2h(Q) + \mathcal{O}(1).$$

Tate defined a global canonical height for points on elliptic curves as follows:

$$\hat{h}(P) = \frac{1}{2} \lim_{n\to\infty} 4^{-n} h(2^n P).$$

The canonical height satisfies:

1. $\hat{h}(P) = 0$ if and only if $P$ is a torsion point of $E(K)$,
2. $\hat{h}(P+Q) + \hat{h}(P-Q) = 2\hat{h}(P) + 2\hat{h}(Q)$ (true parallelogram law), so in particular
3. $\hat{h}(nP) = n^2 \hat{h}(P)$, and finally
4. $\hat{h}(P) = h(P) + \mathcal{O}(1)$.

The global canonical height is fundamental in the study of the arithmetic of elliptic curves, and is a part of major conjectures such as the Birch-Swinnerton-Dyer Conjecture and the Elliptic Lehmer Problem. It is interesting to note that it is unknown, for even a single non-torsion point on a single elliptic curve, whether canonical heights of rational points are rational or irrational. See [4] for more background on heights.

## 4. Math Project

It should be the case that an elliptic divisibility sequence is Benford base $b$ for almost all $b$. Here's a heuristic argument:

- It's well-known that elliptic divisibility sequences satisfy a growth condition like $\psi_n(P) \approx c^{n^2}$ where the constant $c$ depends on the arithmetic height of the point $P$ and on the curve $E$.
- Weyl's theorem tells us that $\{n^k \alpha\}$ is uniform distributed mod 1 iff $\alpha \notin \mathbb{Q}$.
- So we should at least be able to conclude that a given EDS is Benford base $b$ for almost every $b$.
- **But:** There is some subtlety in the error term.

Let $B_n = c^{n^2}$ for some constant $c$. Then the argument above applies to this sequence. But we have $\psi_n(P) = c^{n^2} + \mathcal{O}(1)$. Though the error term is bounded, it could still mess things up. It could be just enough to make your original sequence periodic mod 1, for example.

I recently had the following correspondence with Noam Elkies about the question, which gives me hope that there's something we can prove here:

> "Anyway, of course $n^2(\hat{h}(P)/\log b) + \mathcal{O}(1)$ need not be equidistributed mod 1 in general, but here we know more than that formula, because there's a closed form for the "elliptic divisibility sequence" in terms of theta functions that gives that $\mathcal{O}(1)$ correction as a periodic function of $n$ plus an almost-periodic function $s(n)$, and $s(n)$ depends smoothly on $[n]P$ modulo the real period Omega of the curve, except for a logarithmic singularity at multiples of Omega. Now the Weyl equidistribution criterion can be used to show that the joint distribution of $(cn^2, rn) \mod \mathbb{Z} \times \mathbb{Z}$ is asymptotically uniform as long as $c$ and $r$ are both irrational (which we conjecture is true for $c = \hat{h}(P)/\log(b)$, and know is true for $r$ because $P$ is non-torsion). The desired result soon follows; to deal with the logarithmic singularity, ignore $n$ for which $rn$ is within $\epsilon$ of the nearest integer (which in the limit account for only $2\epsilon$ of all $n$ values), proof equidistribution for the rest, and then note that $\epsilon$ can be taken arbitrarily small."

I don't totally understand this argument, but my hope is that we can unwind it and fill in the details and prove the result about elliptic divisibility sequences.

In the case where we take the base $b = e$ (so we are taking natural logarithms and asking if the resulting sequence is uniform mod 1), we actually have

$$\ln \psi_n(P) \sim n^2 \hat{h}(P).$$

So the elliptic divisibility sequence given by $P$ should be Benford base-$e$ iff the canonical height $\hat{h}(P)$ is irrational. This isn't known for any single rational point on any elliptic curve. (Though it's certainly likely given the way the canonical height is defined.) It would be interesting to show particularly good (or bad?) fits to a Benford distribution for lots of points on lots of elliptic curves to provide more solid evidence of the irrationality of canonical heights of rational points on elliptic curves.

> **I suspect the problems on division polynomials in Silverman's *Arithmetic of Elliptic Curves* will be helpful. (See problems 3.7, 3.35, and especially 6.15 and 6.16.) The stuff about Weierstrass $\sigma$ in Silverman's *Advanced Topics...* book might also be of use.**

## References

[1] W.G. Brady. More on benford's law. *Fibonacci Quart.*, 16(1):51–52, 1978.

[2] G. Everest and T. Ward. The canonical height of an algebraic point on an elliptic curve. *New York J. Math.*, 6:331–342 (electronic), 2000.

[3] Rachel Shipsey. *Elliptic divisibility sequences.* PhD thesis, Goldsmith's College (University of London), 2000.

[4] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1992.

[5] N. Sloane and T. On. Line encyclopedia of integer sequences.

[6] Morgan Ward. Memoir on elliptic divisibility sequences. *Amer. J. Math.*, 70:31–74, 1948.

[7] William Webb. Distribution of the first digits of Fibonacci numbers. *Fibonacci Quart.*, 13(4):334–336, 1975.

[8] J. Wlodarski. Fibonacci and lucas numbers tend to obey benford's law. *Fibonacci Quart.*, 9(1):87–88, 1971.