BENFORD'S LAW

James Tanton

www.jamestanton.com

The following represents a sample activity based on the January 2008 newsletter of the St. Mark's Institute of Mathematics (www.stmarksschool.org/math). Some content here also appears in *THINKING MATHEMATICS! Volume 2* available at www.jamestanton.com

This essay is a draft chapter from a forthcoming book.

The St. Mark's
Institute of Mathematics
Newsletter

JANUARY 2008

PUZZLER: Here is one of my favorite mathematical mysteries:

Consider the powers of two:

1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, …

Do any of these numbers begin with a 7?

If so, which is the first power of two that does? What is the second, and the third? Are there ten powers of two that begin with a seven? Are there infinitely many?

If, on the other hand, no power of two begins with a seven, why not?

ANOTHER PUZZLER:
The powers of three begin:

1, 3, 9, 27, 81, 243, 729, 2187, 6561, …

Several of these numbers end in 1. Do any powers of three end in 01? In 001? In 0001?

TODAY'S TIDBIT: BENFORD'S LAW
In 1881, whilst looking at books of numerical values in his local library – logarithm tables and books of scientific data, for instance - American astronomer Simon Newcomb observed that the first few pages of tables seemed to be more worn from public use than latter pages. He surmised that data values that begin with the digit 1 were looked at and used more often than values beginning with other digits. He thought this curious.

The same phenomenon was later independently noticed in 1938 by physicist Frank Benford, who went further and examined large collections of data tables from a wide variety of sources: population growth data, financial data, scientific observation, and the like. He observed that, with some consistency, about 30.1% of the entries in each data set began with a 1, about 17.6% with a 2, about 12.5% with a 3, all the way down to about 4.6% with a 9. This observation has since become known as Benford's Law.

FILING TAXES:
The IRS today believes and uses Benford's law to quickly scan for possibly falsified tax records: About 1/3 of the figures appearing on a tax form should begin with a "1," about 1/6 with a "2", and so forth.

EXPLAINING BENFORD'S LAW:
Benford's Law can indeed be justified, with relative "ease" in fact, as long as one is comfortable with logarithms and some sneaky thinking!

Let's work with the powers of two, 1, 2, 4, 8, 16, …and show that about 12.5% of them begin with a 3, for example. (We can show that about 30.1% of them begin with a 1 in the same way.)

Now … A power of two begins with a 3 if it lies between 30 and 40, or between 300 and 400, or between 3000 and 4000, and so on. That is, a power of two, 2^n, starts with a 3 if there is a value k for which:

$$3\times10^k < 2^n < 4\times10^k$$

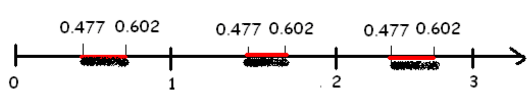
Taking logarithms we obtain:

$$\log 3 + k < n\log 2 < \log 4 + k$$

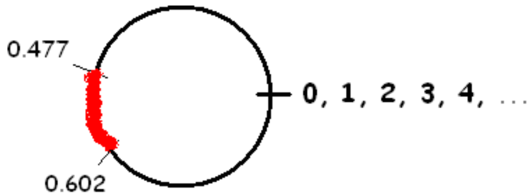
i.e.

$$k + 0.477 < n\log 2 < k + .602$$

That is, on a number line we are seeking multiples of $\log 2$ that lie between the positions 0.477 and 0.602 after some whole number k units along the line.



Let's take the number line and curl it into a circle so that all the whole number positions lie on top of one another and all the intervals of interest coincide:



We seek the proportion of the numbers $\log 2,\ 2\log 2,\ 3\log 2,\ 4\log 2, \ldots$ that land in the indicated part of the circle.

Note that the circle has circumference one unit and that the highlighted portion represents 0.602 – 0.477 = 0.125 = 12.5% of the circle. If we can show that the multiples of $\log 2$ evenly "fill up" the circle, then we can legitimately say that 12.5% of the powers of two begin with a 3.

Our goal is now to show that this is indeed the case.

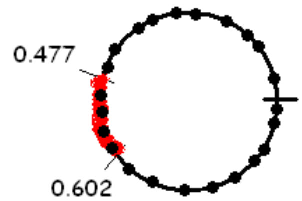
Now log2 is quite a complicated number: It is irrational. [Reason: If $\log 2 = \frac{a}{b}$, then $10^a = 2^b$, which says that a power of two is divisible by five – not true.]

This means that no two multiples of $\log 2$ will ever land on the same point of the circle. [Reason: If they do, then we'd have an equation of the form $a + m\log 2 = b + n\log 2$ giving $\log 2 = (b-a)/(m-n)$ a fraction - not true.]

So all the multiples of $\log 2$ - infinitely many of them - mapped on this circle march around the circle and never land in the same position twice.

We're almost done. We just(!) need to show that they "spread out" around the circle and fill it up evenly without gaps.

St. Mark's Institute of Mathematics.
mathinstitute@stmarksschool.org



To see this … Plot, say, the first 101 multiples of $\log 2$ on the circle. They all can't be 0.01 units or more apart (there are 101 points), so two multiples of $\log 2$ are within a distance of 0.01 of each other. In fact, according to my calculator, $5\log 2$ and $98\log 2$ are less than 0.01 units apart. This means that if I were to add another 93 multiples of $\log 2$ to consider $191\log 2$, and then another 93 to obtain $284\log 2$, and so on, I'd obtain a string of multiples of $\log 2$ all less than 0.01 units apart from each other, spreading themselves evenly around the entire circle. In fact, the choice 0.01is arbitrary, and one can find a sequence of multiples of $\log 2$ that spread themselves across the entire circle in steps less than 0.001 units, or 0.00000001 units, and so forth.

So there we have it! The multiples of $\log 2$ do densely fill up the circle and $\log 4 - \log 3 = 12.5\%$ of them correspond to powers of two that begin with three.

By the same method … $\log 8 - \log 7 = 5.8\%$ of the infinitely many powers of two begin with 7 (answering the opening puzzler) , $\log 78 - \log 77 \approx 0.56\%$ of the powers of two begin with 77 and $\log 778 - \log 777 \approx 0.056\%$ with 777. There are also infinitely many powers of two that begin with your birth date and infinitely many that begin with the first billion digits of pi!

Any physical phenomenon that involves powers of a quantity (population growth – powers of *e*, banking and interest – powers of *e*, the Fibonacci numbers and their appearance in nature – essentially powers of phi) will follow Benford's law.

RESEARCH CORNER: How many powers of two have second digit 7? Third digit 7?

BENFORD'S LAW
January 2008

COMMENTARY, SOLUTIONS and THOUGHTS

One must hunt for quite some time to find a power of two that begins with a seven. The first such power is 2^{46}, and then it is a surprise to see that 2^{56}, 2^{66}, 2^{76}, 2^{86}, 2^{96} all begin with seven but 2^{106} does not! Next come $2^{149}, 2^{159}, 2^{169}, 2^{179}, 2^{189}$ but not 2^{199}. Why these runs of ten in the exponents?

They come from the fact that $2^{10} = 1024$ is approximately equal to 1000, and multiplication by 1000 does not change the first digit of a number. Thus if 2^n begins with a 7, then there is a chance that 2^{n+10} will as well. But of course, the error of "$+24$" in this analysis will eventually build to break this pattern and put an end to any run of tens.

One can see cycles of ten if one lists the first digits of the powers of two in a table. Here are the first digits of 2^0 up to 2^{99} listed in rows of ten:

1	2	4	8	1	3	6	1	2	5
1	2	4	8	1	3	6	1	2	5
1	2	4	8	1	3	6	1	2	5
1	2	4	8	1	3	6	1	2	5
1	2	4	8	1	3	7	1	2	5
1	2	4	9	1	3	7	1	2	5
1	2	4	9	1	3	7	1	3	5
1	2	4	9	1	3	7	1	3	6
1	2	4	9	1	3	7	1	3	6
1	2	5	9	2	4	7	1	3	6

QUESTION: What structure lies in a table like this? When a change in a column occurs, is the change sure to be an increase of just one? What about tables of the first digits of powers of three, or powers of 173? (Notice: The proof of Benford's Law in the newsletter is not unique to the powers of two.)

The second puzzler in the newsletter asks about final digits of the powers of three:

Do any of the numbers 1, 3, 9, 27, 81, 243, 729, 2187, 6561, … end in 001?

Absolutely!

Divide each of the numbers 1, 3, 9, 27, 81, 243, … by 1000 and look at their remainders. As there are infinitely many powers of three and only a finite number of possible remainders, there must be two powers, say 3^n and 3^m with $n > m$, that have the same remainder upon division by 1000. This means that there difference is a multiple of this number:

$$3^n - 3^m = 1000k$$

for some k. Thus:

$$3^m\left(3^{n-m} - 1\right) = 1000k$$

Since 1000 has no factors in common with 3^m, this means that $3^{n-m} - 1$ must itself be a multiple of 1000:

$$3^{n-m} - 1 = 1000a$$

for some number a Thus $3^{n-m} = 1000a + 1$ is a power of three that ends in 001.

One can repeat this argument for any of the strings 1, 01, 0001, 00001, and so on.

QUESTION: No power of three ends with 002. (Why?) Is there a power of three that ends with 007?

The Exponential Nature of the Fibonacci Numbers

The newsletter claimed that the sequence of Fibonacci numbers, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, …, are "exponential" in nature and also follow Benford's Law. This claim requires some explanation.

Let F_n denote the nth Fibonacci number with $F_0 = 1$, $F_1 = 1$, $F_2 = 2$, and so on. We have:

$$F_{n+2} = F_{n+1} + F_n$$

Our goal is find a general formula for the nth Fibonacci number and so reveal its "exponential character."

A standard technique in analyzing a recursive sequence is to ask if there is a geometric sequence $1, x, x^2, x^3, x^4, \ldots$ that satisfies the same recursion relation. In this case, we seek a value x that satisfies $x^{n+2} = x^{n+1} + x^n$ for all n. Dividing through by x^n we see that this is equivalent to solving:

$$x^2 = x + 1$$

which has two solutions:

$$\varphi = \frac{1+\sqrt{5}}{2} \text{ and } \tau = \frac{1-\sqrt{5}}{2}$$

(The first value is the famous "golden ratio" ([LIVIO]).) Thus the two sequences:

$$1, \varphi, \varphi^2, \varphi^3, \ldots$$
$$1, \tau, \tau^2, \tau^3, \ldots$$

each follow the same Fibonacci relation: any term (beyond the second) is the sum of the previous two terms. It is not hard to check that the same is true for any linear combination of these two sequences. That is, if a and b are any two constants and we set:

$$G_n = a\varphi^n + b\tau^n$$

then we also have:

$$G_{n+2} = G_{n+1} + G_n$$

for all $n \geq 2$.

Let's now choose constants a and b to create a sequence that has the same initial start as the Fibonacci numbers. That is, let's choose constants that make $G_0 = 1$ and $G_1 = 1$.

$$G_0 = a + b = 1$$
$$G_1 = a\varphi + b\tau = 1$$

A little algebra gives $a = \frac{\varphi}{\sqrt{5}}$ and $b = -\frac{\tau}{\sqrt{5}}$.

We now have a sequence G_n that starts with the same initial values of as the Fibonacci sequence and follows the same recursive relation as the Fibonacci sequence. It must be the Fibonacci sequence!

$$F_n = \frac{\varphi}{\sqrt{5}} \cdot \varphi^n - \frac{\tau}{\sqrt{5}} \cdot \tau^n = \frac{\varphi^{n+1} - \tau^{n+1}}{\sqrt{5}} = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right)$$

This is Binet's Formula for the Fibonacci numbers. (This formula is today attributed to the French mathematician Jacques Binet (1786-1856) though it was known by other scholars decades earlier ([LIVIO]).)

Notice that $|\tau| = \left|\frac{1-\sqrt{5}}{2}\right| \approx 0.618 < 1$ and so $\tau^{n+1} \to 0$ as $n \to \infty$. Thus for n large we have:

$$F_n \approx \frac{\varphi^{n+1}}{\sqrt{5}}$$

CHALLENGE: Prove that for each n, F_n is the closest integer value to $\frac{\varphi^{n+1}}{\sqrt{5}}$.

Thus in this sense the Fibonacci numbers are indeed essentially exponential in structure.

COMMENT: For more on Benford's law see [WEISSTEIN] and all the references therein.

FINAL THOUGHT

I once gave the following challenge as an e-mail puzzle for the St. Mark's Institute followers:

LIKING MY THREES:
I like numbers that contain the digit 3. The number $N = 751153$ contains the digit three, as does its double $2N = 1502306$ and $3N = 2253459$ and $4N = 3004612$ and $5N = 3755765$. (But alas $6N$ does not contain a three!)

a) Find me a number N such that N, $2N$, $3N$, $4N$, $5N$ and $6N$ each contain the digit three.

b) Find me a number N such that N, $2N$, ..., $10N$ each contain the digit three.

c) For each number k is there guaranteed to be some number N with the property that all of the multiples N, $2N$, $3N$, $\ldots$, kN contain the digit three?

d) Is there a number N such that each and every multiple of N contains the digit 3?

One can check that for $N = 19507893$ each of N, $2N$, $3N$, …, $25N$ contain a three! This certainly answers parts a) and b).

CHALLENGE: What is the smallest number that answers a) and the smallest number that answers b)?

And it is fairly easy to settle part d) to the negative:

Lemma 1: *There is no positive integer* N *so that each and every one of its multiples,* N*,* $2N$*,* $3N$*, …, contains the digit three.*

Proof: Suppose we are given a positive integer N. We'll construct a multiple of N that does not contain the digit three.

Consider the infinite list of integers:

1, 11, 111, 1111, 11111, ….

Divide each of these numbers by N and look at their remainders. As there are only a finite number of possible remainders, two of the numbers in this infinite list must have the same remainder. This means that their difference is a multiple of N.

But their difference must be a number of the form $11\ldots100\ldots0$ and so is a multiple of N that contains no three. □

To answer part c) we need an intermediate result.

Lemma 2: *For each positive integer* k *there is an integer* N *so that* kN *contains the digit three.*

Proof: Look at the infinite list of numbers 3, 33, 333, 3333, 33333, …. Divide each of these by k and look at their remainders. As there are only a finite number of possible remainders, two numbers in this infinite list leave the same remainder. Their difference, a number of the form $33\ldots300\ldots0$, is a multiple of k and contains a three. □

For example:

For $k=1$ the number $N=3$ has $1\cdot N$ containing a three.
For $k=2$, the number $N=15$ has $2N$ containing a three.
For $k=3$, the number $N=10$ has $3N$ containing a three.
For $k=4$, the number $N=75$ has $4N$ containing a three.

We can construct a number that "works simultaneously" for each of these four values of k by stringing together the values of N we have so far, separating them by a large number of zeros. For example, with

$$N = 75\ \ 00000\ \ 10\ \ 00000\ \ 15\ \ 00000\ \ 3$$

we have that each of N, $2N$, $3N$ and $4N$ contain the digit three. In fact, with the aid of lemma 2 we can construct a single value N whose multiples up to any high value we decree each contain the digit three. (One needs to place sufficiently many zeros between the individual terms in the construction of this number N so as to absorb any "carries" that might occur. One can construct a pattern for the number of zeros that will suffice.)

This establishes:

Theorem: *For each positive integer k there is an integer N such that each of the multiples N, $2N$, $3N$, ..., kN contain the digit three.*

REFERENCES:

[LIVIO]
Livio, M., *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number,* Broadway Books, New York, NY, 2002.

[WEISSTEIN]
Weisstein, E. W. "Benford's Law" from *MathWorld*-A Wolfram Web Resource. URL: http://mathworld.wolfram.com/BenfordsLaw.html