The Suspicious Case of the Number One

Benford's Law and the Hidden Pattern in Real-World Data

Imagine you are a fraud investigator. You have just been handed a spreadsheet of 10,000 expense claims filed by an accountant at a large company. A lunch for £47, a taxi for £82, a hotel for £312. Nothing looks obviously wrong. But you know something that most people do not. You know about Benford's Law, and you are about to use it to figure out whether any of these numbers were made up.

Benford's Law says this: in a naturally occurring set of numbers, about 30% of them will start with the digit 1. Not 11%, which is what you would guess if all nine digits had an equal chance. Not some complicated formula that depends on the type of data. Just 1, plain old 1, turning up nearly a third of the time. And 2 shows up about 17.6% of the time. And 9 shows up less than 5% of the time.

I know. It sounds made up. But it is one of the most well-tested and genuinely strange patterns in mathematics, and once you understand why it happens, you start seeing it everywhere.

How it was discovered

The story starts in 1881 with an astronomer named Simon Newcomb who was flipping through a book of logarithm tables. These were the calculators of the time, thick books full of numbers that scientists used to do multiplication and division by hand. Newcomb noticed something odd. The first pages of the book, the ones with numbers beginning with 1 and 2, were much more worn and dirty than the later pages. People were clearly looking up numbers starting with 1 far more often than numbers starting with 8 or 9.

He wrote a short paper suggesting that the probability of a number having first digit *d* follows the formula $\log_{10}(1 + 1/d)$. For d = 1 that gives you $\log_{10}(2)$, which is roughly 0.301, so about 30%. For d = 9 you get about 4.6%. The digits are not equally likely at all.

Nobody really paid attention. The paper sat in a journal for over fifty years until a physicist at General Electric called Frank Benford independently found the same thing in 1938. He was much more thorough about it. He collected 20,229 numbers from completely different places: the surface areas of rivers, the atomic weights of elements, street addresses from random directories, population figures, death rates, numbers pulled out of random magazine articles. The same pattern appeared in all of them. The first digit was 1 about 30% of the time, across the board. Benford published and got his name attached to the law, which is a little unfair on Newcomb, but that is how these things tend to go.

So why does it happen?

This is the part I find genuinely beautiful, because the explanation is not really about the numbers themselves. It is about how things grow.

Think about what it takes to go from 1 to 2. That is a 100% increase. To go from 2 to 3 is only a 50% increase. From 8 to 9 is just a 12.5% increase. So on a logarithmic scale, which is the natural scale for anything that grows by multiplication, the gap between 1 and 2 is huge and the gap between 8 and 9 is tiny.

Real-world data tends to grow multiplicatively. Populations double. Investments compound. The size of a river's drainage area scales with geography in a multiplicative way. For any quantity that evolves through multiplication, its first digits will follow Benford's Law almost inevitably, because the time it spends with each leading digit is proportional to the logarithmic size of that digit's range.

There is a more formal way to say this. A distribution follows Benford's Law if it is scale-invariant, meaning that if you measure everything in different units, say switching from miles to kilometres, the pattern of first digits does not change. It turns out the only distribution with this property is the one Newcomb described. So Benford's Law is essentially the fingerprint that scale-invariant, multiplicatively growing data leaves behind.

The fraud detection part

Here is the uncomfortable thing that Benford's Law reveals about human psychology: when people invent numbers, they are terrible at it.

If you ask someone to fabricate a list of expenses, they will try to make the numbers look random and spread out. What they will not do is front-load the list with numbers starting in 1, because their intuition tells them that 1 is just one of nine digits and should appear about one ninth of the time. That intuition is wrong, and Benford's Law will catch it.

Forensic accountants now use this as a standard check. If the first-digit distribution of a set of financial records looks too flat, with too many 6s and 7s and not enough 1s, that is a signal worth investigating. It does not prove anything on its own, but it suggests someone may have been inventing rather than recording.

It has been applied in some notable cases. Analyses of Greek economic data submitted to the EU before the Eurozone crisis found suspicious deviations from Benford's Law, suggesting the figures had been altered. Similar analyses have been done on election results, corporate earnings, and scientific datasets where fabrication was suspected. The law does not read minds, but it reads numbers with a kind of quiet, relentless patience.

Where it does not work

It is worth being honest about the limits here, because a good tool deserves to be used properly.

Benford's Law does not apply to data with a narrow range. Shoe sizes cluster between about 4 and 12. There is no spanning of multiple orders of magnitude, so the logarithmic pattern never gets a chance to emerge. Phone numbers, postcodes, and lottery results are designed to be random and will not follow the law either.

Human heights in centimetres are another example. Everyone is somewhere between about 140cm and 210cm, so all the first digits are either 1 or 2 anyway. That is not Benford's Law at work, it is just the range of the data.

The law works best on datasets that span several orders of magnitude and come from natural, compounding processes: financial transactions across a large economy, city populations across a country, the lengths of rivers worldwide. These are the places where 1 quietly dominates, and where any attempt to fake the data tends to get caught.

A strange kind of comfort

There is something almost philosophical about Benford's Law once you sit with it for a while.

We tend to think of the digits 1 through 9 as neutral and interchangeable, just arbitrary labels for quantity. But Benford's Law suggests that in a world governed by growth and compounding, these symbols are anything but equal. The digit 1 is privileged in a very real mathematical sense. It sits at the low end of every exponential leap. It is where things start before they grow, and it turns out that beginnings are far more common than endings.

Every time a population increases by an order of magnitude it spends proportionally more time with a 1 at the front than with any other digit. Every stock price, every city population, every river length that has ever grown has lingered longest near its starting point. And starting points begin with 1.

So the next time you open a spreadsheet or look at a table of statistics, try counting the leading digits. If 1 does not come up roughly 30% of the time, if the distribution looks suspiciously uniform or if the higher digits are showing up far more than they should, it is worth asking the question that Benford's Law has been quietly asking for over a hundred years.

Did these numbers grow? Or were they invented?

AI tools were used to help with editing and refining this essay.
- Hasvi Muriki