Number One for a Reason

If I gave you a big, messy list of numbers such as river lengths, country populations, electricity bills, stock prices, or just whatever you can think of and asked you which digit appears the most often at the start, you would probably say they should all show up around equally.

Seems fair enough.

There are nine different first digits from 1 through 9. So, you would expect each one to appear about one-ninth of the time so it should be roughly 11%. Nice and balanced. Everyone gets a turn.

That answer feels sensible.

It is also completely wrong.

In lots of different real-world datasets, numbers starting with 1 appear about 30% of the time whereas the number 9 shows up less than 5% of the time. So, the digit 1 is basically the teacher's favourite, while 9 is sat at the back wondering what it did to deserve this.

This strange pattern is called Benford's Law, and it is one of those bits of maths that sounds made up until you realise it keeps turning up everywhere such as in science or economics or geography and especially in finance. It is even used in fraud investigations.

A law about digits turning into a detective tool is not something you see every day.

So, what is Benford's Law?

$$P(D_1 = d) = \log_{10}\left(1 + \frac{1}{d}\right), \quad d = 1, 2, \ldots, 9$$

This formula helps us determine the probability that first digit of a number is d, where d can be any digit from 1-9.

This formula looks slightly intimidating, but the results are surprisingly simple.

If we substitute d=1, we get:

P(1) = Log10(2) ≈ 0.301

So about 30% of numbers begin with 1.

If we substitute d=9, we get:

$P(9) = \log10(1.11…) \approx 0.046$

So only about 4.6% begin with 9.

That means the chances that the starting digit is 1 would be roughly six times more common than numbers starting with 9. Which feels wildly unfair, but maths is not really about fairness but rather patterns.

What this formula is measuring is not how big numbers are, but how much space they take up on a logarithmic scale. The logarithm tells us the proportion of numbers that fall between one value and the next when we measure growth multiplicatively rather than additively. Basically, what this means is it is calculating how much of the number line belongs to each leading digit when numbers grow by factors instead of fixed amounts. That is why the digit 1 ends up dominating because it just occupies a larger share of the logarithmic scale.

Why does this happen?

This is the part where everything clicks and makes sense.

Let's say for example on a normal number line, the distance from 1 to 2 is the same as the distance from 8 to 9. Both are 1 unit. Very equal. Very boring.

But Benford's Law is not about equal distances. There is something called multiplicative change, and that means logarithms start running the show. A number starting with 1 stays that way, from 1 up to just before 2. So, to stop starting with 1, the number must double. A number starting with 9 however only stays that way from 9 up to just before 10. To stop starting with 9, it only needs to increase by about 11%. So, numbers spend much longer "living" in the 1-zone than the 9-zone.

That is the key idea. First digits are not equally likely because the intervals they occupy are not equally large on a logarithmic scale. The interval for numbers starting with 1 is much wider, so numbers naturally spend more time beginning with 1.

Another way we can look at this is through ratios rather than differences.

The interval for numbers starting with 1 runs from 1 to 2 which is a multiplication by 2.

The interval for numbers starting with 9 runs from 9 to 10 which is only a multiplication by approximately 1.11.

Because logarithms measure multiplicative change, the interval from 1 to 2 is mathematically larger than the interval from 9 to 10, even though both appear similar on a normal number line, so numbers naturally spend more time beginning with smaller digits.

You can think of it like this:
1 gets a whole mansion.
9 gets a studio flat.

Where it shows up in real life

Benford's Law tends to appear in datasets that cover a wide range of values especially when those values grow naturally.

Country populations is an obvious example. Some countries have populations in the tens of thousands, while others have populations over a billion. This would be the type of data that can be applied to Benford's. What makes population data especially useful here is that it spans several orders of magnitude, from very small countries to extremely large ones. That widespread is exactly the sort of condition under which Benford's Law is most likely to appear.

Financial data is another big one. Companies grow, shrink, invest, and spend in percentages rather than fixed amounts. That kind of multiplicative behaviour is the perfect territory for Benford's Law.

Scientists have even found patterns in:

- river lengths
- earthquake magnitudes
- electricity usage
- stock market data
- physical constants

Real datasets are not expected to match Benford's Law perfectly. What matters is whether the overall shape is broadly like the predicted distribution.

Figure 1: Comparison between observed first-digit frequencies in population data and the theoretical distribution predicted by Benford's Law.

I have used real world data from the World Bank Group's total population data to help represent the first digit frequencies showing that the larger numbers where least likely to show up as the first digit. Although the counts were not perfectly decreasing, the overall trend matched the theoretical model. This suggests that population data, which spans several orders of magnitude, naturally follows the logarithmic pattern described by Benford’s Law.

Small deviations from the theoretical distribution are expected because real-world datasets are finite and influenced by many different factors. Even so, the overall downward pattern still supports Benford’s Law quite well.

In short, whenever numbers come from natural processes and span over a very wide range, Benford’s Law starts lurking in the background.

Why changing units does not break it

Here is one of the coolest and more interesting mathematical properties of Benford’s Law. It is scale-invariant.

What I mean by this is that the pattern stays the same even if you change units.

Here’s a practical example:

If you measure a river in kilometres instead of miles or record money in pounds instead of dollars, the overall distribution of first digits usually stays very similar. This happens because multiplying every value in a dataset by a constant would just shift the numbers along a logarithmic scale without changing the shape of the distribution. That’s a huge clue that Benford’s Law is not a coincidence or luck. It’s built into the structure of how numbers grow.

Where it doesn’t work

This part matters just as much as where it does work.

Benford’s Law isn’t some magical “lie detector” that works on every dataset.

Take life expectancy. Across countries, life expectancy usually falls between about 50 and 85 years meaning it is a pretty narrow range, so the first digits are forced to be 5, 6, 7 or 8 most of the time so the leading digits become almost identical. The range is just not wide enough for the logarithmic pattern to appear.

The same goes for:

- exam scores
- human heights
- shoe sizes
- temperatures
- phone numbers

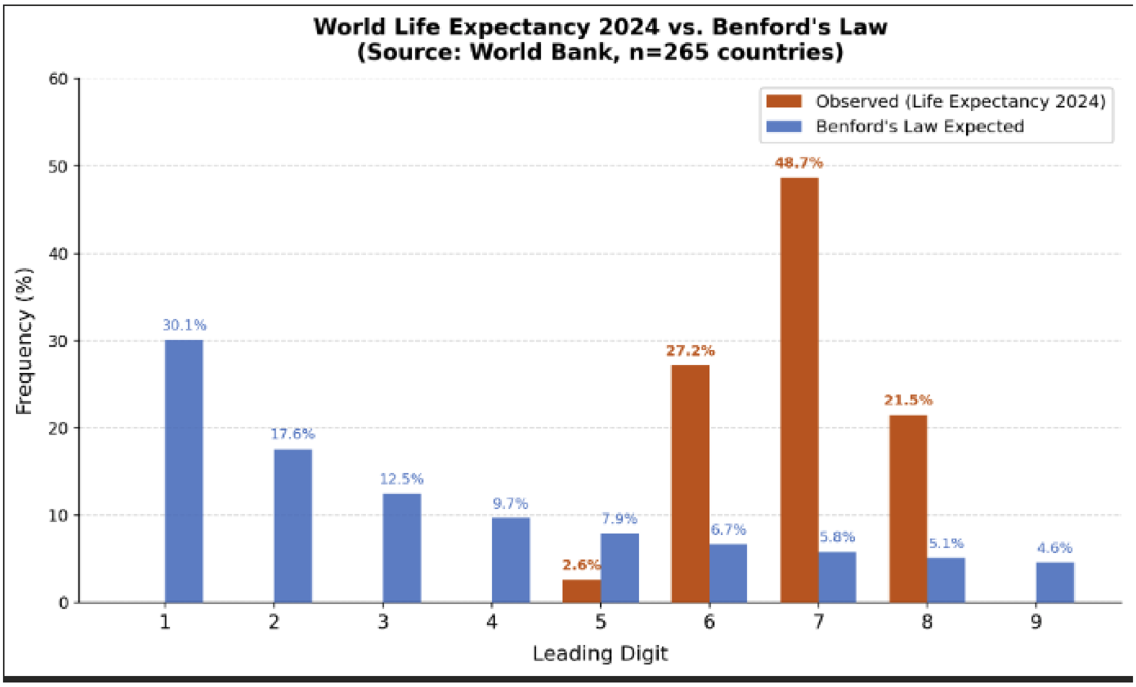

Figure 2 shows that the digit 7 makes up close to 49 per cent of the leading digits in its life expectancy data – over eight times the 5.8% that Benford's Law would predict. In contrast, digits 1, 2, 3, 4 and 9 occur exactly zero times across all 265 countries. Not rarely. Zero. That's not a very small deviation from Benford's Law. That's the opposite of it.

These values are either restricted, designed by humans, or clustered tightly together. Benford's Law was never meant to work there. So if someone applies it to the wrong dataset and acts shocked when it fails, that is not a maths discovery, that is just the user's error with extra confidence.

Fraud, fake numbers, and why accountants care

This is where Benford's Law is unexpectedly used.

Because real datasets often follow this pattern, fake datasets sometimes stand out. Imagine someone inventing financial figures because they want to avoid their shocking tax rate. They might assume digits are random and spread them too evenly, or they might subconsciously favour certain numbers and maybe repeat the same number too many times. Either way, the distribution of first digits can end up looking suspicious.

Auditors and investigators sometimes compare financial records against the expected Benford distribution. If the numbers deviate significantly, it can be a sign that something deserves a closer look and something fishy is going on.

Important detail to keep in mind though: **Benford's Law does not prove fraud.**

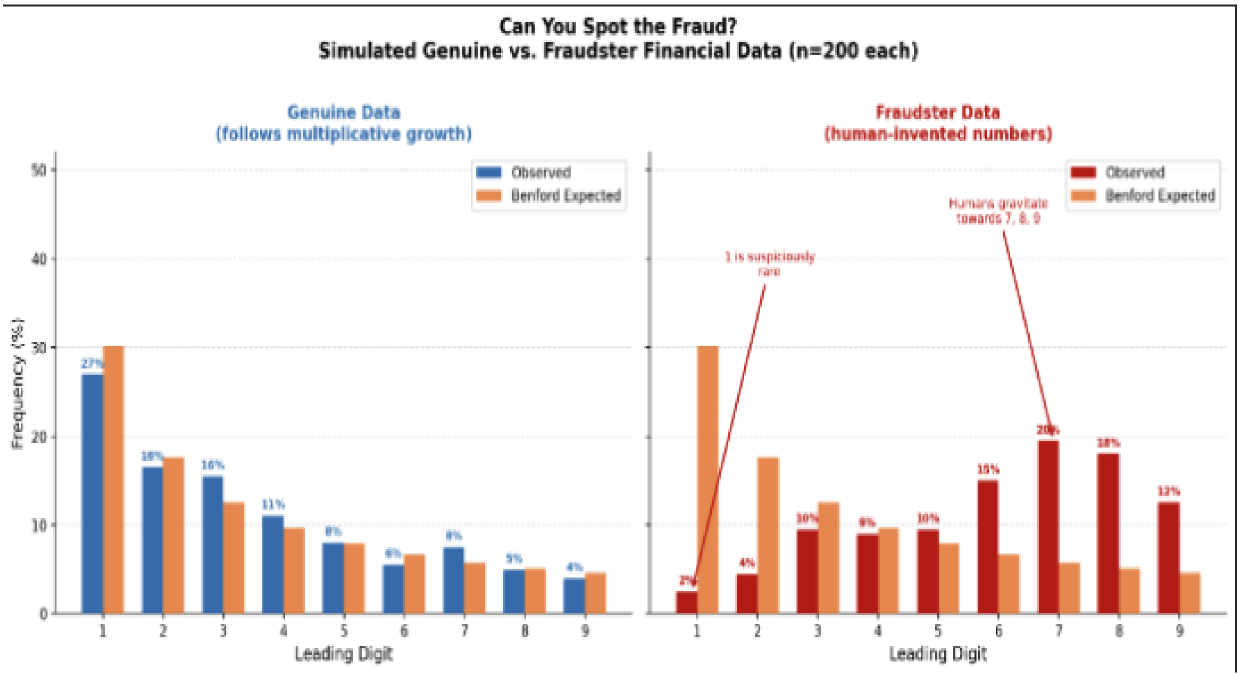

Figure 3 simulates exactly this. The left panel shows 200 numbers generated from a genuine multiplicative process — the distribution closely tracks Benford's expected curve. The right panel shows 200 numbers invented by a simulated fraudster. The spike in digits 7, 8 and 9 is immediate, and digit 1 is almost absent. No auditor needs to run a single calculation to see something is wrong. The graph gives the game away instantly. It just raises suspicion.

Think of it as a smoke alarm, not a police officer.

What makes Benford's Law genuinely fascinating is not just that it is useful but also that it challenges our intuition. We expect numbers to behave evenly. We expect fairness and symmetry however the real world is full of processes that grow by percentages rather than fixed amounts.

Populations grow proportionally. Investments rise and fall proportionally. Measurements span huge scales.

Once you realise that, Benford's Law stops looking strange and starts making more sense.

It is not a weird trick. It reflects how the world works.

Benford's Law is a perfect example of how maths can reveal hidden structure in an ordinary data. At first the idea seems simple, just count the first digits of the number. But the result would lead into logarithms, growth, probability, and real-world applications like fraud detection. It also shows that numbers are not always as random as they look and that even the smallest detail such as a single digit can carry surprising information.