The Law That Detects Fraud

Grab a newspaper. Write down all the numbers in it. Then tally the frequencies of the first digit of each number. You would expect each digit to appear equally frequently, so around 11.1% for each digit from 1 to 9. While this may seem logical, I can tell you that 1 will be the first digit more often than 2, which will be the first digit more often than 3, and so on; you might be even more surprised to learn that the frequencies follow the table to the right, with 1 appearing almost a third of the time. This is Benford's Law. Although seemingly counter-intuitive, this law holds true with frightening accuracy for many numerical data sets (particularly those covering multiple orders of magnitude)[1].

d	$P(d)$	Relative size of $P(d)$
1	30.1%	
2	17.6%	
3	12.5%	
4	9.7%	
5	7.9%	
6	6.7%	
7	5.8%	
8	5.1%	
9	4.6%	

Probability, P(d), or leading digit, d, appearing[1]

Although it may be difficult to prove why Benford's Law is true for numerous data sets, and even more difficult to see how it has any usefulness outside of a party trick, it has many applications in the real world, most notably in law, which I will explore.

What is Benford's Law exactly?

Benford's Law can apply to many different data sets. You've already seen that random numbers from a newspaper, where there is no consistency with units or measurements, follows Benford's Law very accurately. There are many sequences in maths that follow the law as well, such as the doubling sequence, the tripling sequence, and the Fibonacci sequence. In terms of specific criteria, a set of numbers will satisfy Benford's Law if

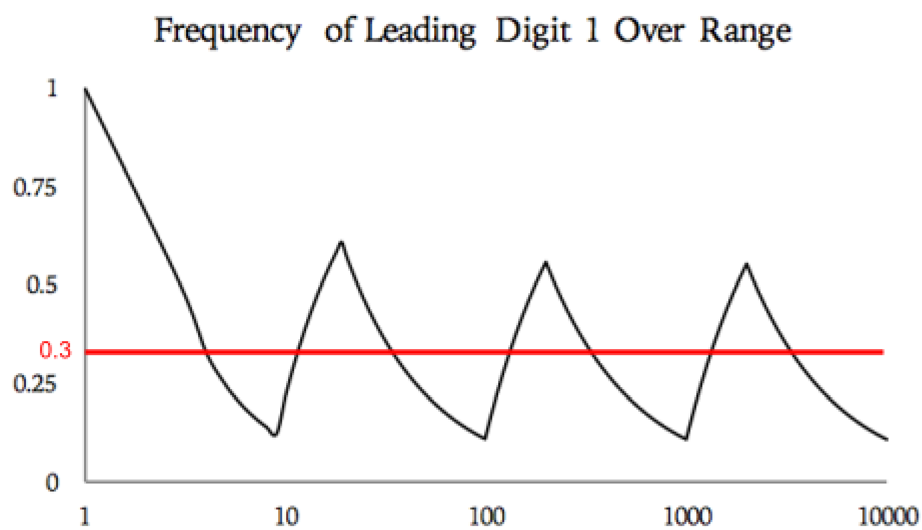
$P(d) = \log_{10}(d+1) - \log_{10}(d)$ where d is the leading digit

This can be simplified to give

$P(d) = (1+1/d)$

How does Benford's Law work?

It is difficult to prove Benford's Law, and there isn't one method of firmly confirming it. One way of understanding why it works is by thinking about a raffle of numbers[2]. If you put numbers into the raffle in numerical order so start with 1, the probability of a number starting with 1 winning is 100%. When you add 2, that becomes 50%. With 3, it becomes 33.3%, and it keeps on decreasing until you reach 9 where the probability is 11.1%. However, when you reach 10, the probability starts to increase again, up to 20%, then 27.3% with 11, and so on, reaching 57.9% at 19. But then it starts to decrease again as you keep adding more numbers, reaching the lowest probability of 11.1% again when you reach 99. And then as you go through the 100s, it rises, then decreases through the 200s, 300s, 400s, etc., and starts to increase again through the 1000s, and so on. If you plot this on a graph, it would look like this:



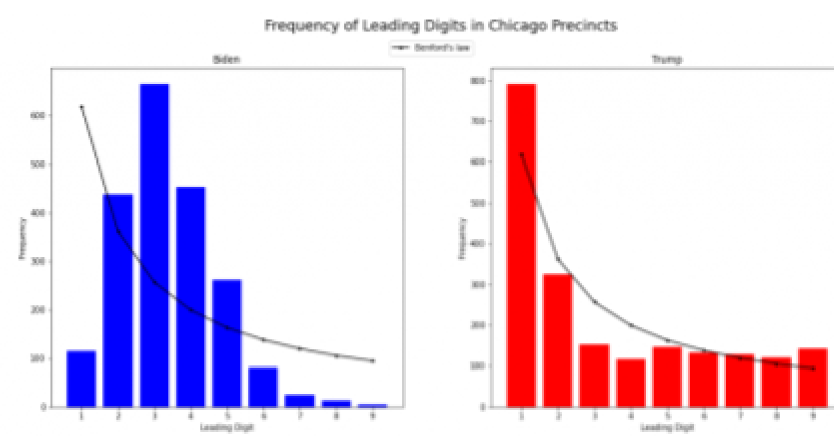
The x-axis is the last number added to the raffle, and the y-axis is the probability of a number picked from the raffle having a leading digit of 1[3]

The average of this graph is around 30%: this matches with Benford's Law.

Applications of Benford's Law

While Benford's Law may not seem to have any purpose, it has some interesting applications in real world scenarios. After learning about Benford's Law, the financial investigator, Darrell D. Dorrell, applied this knowledge to his work of detecting fraud: he would check the leading digits of bank accounts (financial data is spread over multiple orders of magnitude so should comply fairly closely with Benford's Law) and if they don't conform, he would analyse the data further to see if there is any clear evidence of fraud[4]. Of course, there may be several reasons why the data doesn't fit the law, such as a recurring purchase of a good priced at £90 skewing the data, but it does flag any indication of malpractice which could prompt further investigation. In fact, Benford's Law is admissible in federal, state and local courts of law. Dorrell successfully managed to convict a local financial advisor, Wesley Rhodes, after he stole millions of dollars from investors; Rhodes' financial statements failed to pass the first-digit test, prompting Dorrell to investigate further and discover that the data was fake.

Benford's Law most notably came to light when conspiracists claimed that it proved Biden's election victory was rigged. On the right, we see the results for the Chicago area, with Biden's in blue and Trump's in red, and a line representing what Benford's Law would look like with the data set. Clearly, Trump's data conforms with Benford's Law, and Biden's doesn't. This was clear evidence that the voting was rigged, so Biden was sent to jail and Trump served his second term. Not quite. The problem with applying Benford's Law to district election data is that it doesn't span multiple orders or magnitude:



2020 US presidential election data for the 2054 precincts in Chicago[5] precincts usually only register a 3 digit number of votes. If a precinct registers between 400 and 900 votes, and the votes were split in some proportion between Biden and Trump, you would expect the most frequent leading digits to be 2, 3, 4 and 5, as shown in Biden's diagram, instead of following Benford's Law[6]. So, the Trump supporters were wrong in using the law to claim rigging in this case. However, the 2009 Iranian elections also showed signs of electoral fraud when political scientist Walter Mebane analysed the ballot-by-ballot results using a version of Benford's Law that looks at the frequency of the second digit of numbers, and discovered that the winner, Mahmoud Ahmadinejad, had results that showed high discrepancies with the law. These accusations of rigging had a much greater impact than those against Biden, with millions of Iranians protesting around the world as part of the Iranian Green Movement[7]. It was never proven that the results were rigged and Ahmadinejad served his second term as president, but it was still interesting that Benford's Law was able to spark such outrage.

Benford's Law is certainly one of the stranger mathematical theorems that exists, and the fact that there is little to no proof to explain it does put its reliability into question. But I hope that I have shown you how it has the incredible capability of flagging fraud, on as big a scale as presidential elections.

[1] (Author unknown) En.wikipedia.org. 2022. Benford's law - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Benford%27s_law> [Accessed 30 January 2022].

[2] Mould, S., 2013. Number 1 and Benford's Law - Numberphile. [online] Youtube.com. Available at: <https://www.youtube.com/watch?v=XXjlR2OK1kM> [Accessed 30 January 2022].

[3] (Author unknown) Reddit.com. 2016. Why does Benford's Law work?. [online] Available at: <https://www.reddit.com/r/math/comments/4smdbf/why_does_benfords_law_work/> [Accessed 30 January 2022].

[4] Bellos, A., 2015. Alex Through the Looking Glass. London: Bloomsbury Publishing PLC, pp.27-38.

[5] Howgate, K., 2021. Fraud in the 2020 US Election?!?!. [online] Lancaster.ac.uk. Available at: <https://www.lancaster.ac.uk/stor-i-student-sites/katie-howgate/2021/03/12/fraud-in-the-2020-us-election/> [Accessed 30 January 2022].

[6] Dacey, J., 2020. Benford's law and the 2020 US presidential election: nothing out of the ordinary. [online] Physicsworld. Available at: <https://physicsworld.com/a/benfords-law-and-the-2020-us-presidential-election-nothing-out-of-the-ordinary/> [Accessed 30 January 2022].

[7] (Author unknown) En.wikipedia.org. 2021. 2009 Iranian presidential election. [online] Available at: <https://en.wikipedia.org/wiki/2009_Iranian_presidential_election> [Accessed 30 January 2022].