USA TODAY CyberSpeak

10/18/00- Updated 11:59 AM ET

Search
the site the Web
GO
POWERED BY LYCOS

Inside Tech
Talk Tech
FAQ/Tips
Web Column
Hot Sites
Tech News
Tech Investor
Tech Reviews
Answer Desk
Game Zone
Daily Digest
Shareware Shelf
Web Potholes
Web Resources
Consumer Sites
Tech Front

Marketplace
Hardware
Accessories
Software

Print Edition
Today
Yesterday

Technology

Baffled by math?

Wait 'til I tell you about Benford's Law

By Kevin Maney, USA TODAY

Certain realities of our world defy logic. There's tongue piercing. And the notion that caviar is tasty. Or how a TV network ever funded a cartoon show called *Sponge Bob Square Pants.*

One of the more illogical, bizarre things getting some attention right now is a mathematical oddity called Benford's Law. And what's amazing is that the bizarreness of the law, when paired with the number-crunching capabilities of computers, is actually the reason it's so useful. The law has turned into a way to uncover fraudulent data in medical tests, identify families of plankton and design better hard drives.

Kevin Maney

For other columns by Kevin click here.

More Tech columnists:

- Edward C. Baig
- Tamara Holmes
- Larry Johnson
- Kim Komando
- John Makulowich
- Sam Meddis
- Robin Raskin
- Bruce Schwartz
- Joel Smith
- Elizabeth Weise
- John Yaukey

I learned about the law from Ted Hill, a ruddy-faced former Army Ranger captain who's now a math professor at Georgia Tech and enjoying a sabbatical at the University of Goettingen in Goettingen, Germany. So what is Benford's Law and why is it weird?

First, Hill says, think of a large and random set of numbers that is somehow derived from other numbers. Closing Nasdaq stock prices would be a good example, because stock prices are essentially derived from a host of other numbers, such as growth, cost of labor, prevailing interest rates and so on. Another

of numbers in an Al Gore debate answer.

Now, you'd think that those numbers, which are basically assembled randomly, would be spread out randomly. For instance, there'd be just as many numbers beginning with nine or four or one. But that's where you'd be wrong. Some unseen and unknown universal force — possibly similar to the cosmic force that impels all preteen girls to like the same pop star at the same moment — bunches these kinds of random numbers into very predictable patterns.

Benford's Law maintains that certain digits show up more than others. A one will appear as the first non-zero digit 30% of the time; two will be the leading digit 18% of the time; nine will lead off just 4.6% of the time. Zero is most likely to be the second digit, popping in there 12% of the time. It's all very predictable. Benford's Law never fails to work.

"It even surprises mathematicians," Hill says. "It's counterintuitive."

The first inkling of this was discovered in 1881 by astronomer Simon Newcomb. He'd been looking up numbers in an old book of logarithms and noticed that the pages that began with one and two were far more tattered than the pages for eight and nine. He published an article, but because he couldn't prove or explain his observation, it was considered a mathematical fluke. In 1963, Frank Benford, a physicist at General Electric, ran across the same phenomenon, tried it out on 20,229 different sets of data (baseball statistics, numbers in newspaper stories and so on) and found it always worked.

Still, Benford's Law wasn't accepted until Hill, in 1996, worked out a rock-solid proof. Then anyone who wanted to apply the law to the real world could be confident the results were correct.

At this point, you're probably thinking: What the heck good does it do to predict the non-randomness of random numbers?

For one, you can tell if someone fakes data that are derived from other data. As Hill points out, people who fake data — say, on tax returns or results of medical research — usually think the numbers should be evenly spread out to seem real. So they begin as many numbers with nine as one. If you run those faked numbers through a computer program that's testing for Benford's Law, it will sound an alert if the numbers are too random. It can't tell what might be faked, but it can tell that something is wrong. "If you're testing for fraud, it doesn't prove anything, but it certainly raises the level of suspicion," Hill says.

The IRS, the U.S. Institute of Internal Auditors and the International Institute for Drug Development in Brussels are all working on ways to use the law. Mark Nigrini, a professor at Southern Methodist University, has been the pioneer pushing Benford's Law into

accounting.

The law can be used to test predictions. The numbers showing the populations of all U.S. counties conform to Benford's Law. So if someone creates a computer model that's supposed to predict county populations 50 years from now, the outcome can be tested against Benford's Law. If the numbers don't conform, then there's something wrong with the model.

Recently, Hill helped out on applying the law to the study of ocean plankton. Data on one set of plankton — a collection of microscopic organisms — conformed to the law, while another set did not. That suggests that the first set is probably made up of several species, because the law only works when the numbers are derived from underlying numbers. The other set of plankton is probably more pure.

Some mathematicians are working on ways to use the law to make computer disk drives more efficient. If the drive knows that it needs to allocate a certain amount of space for data that start with one, a certain amount for two and so on, it apparently can pack these together and find the data much faster later.

I might like to suggest a little program for laptops that would let you run Benford's Law against scores on a golf card, just to see if your friends are cheating. It seems to meet the criteria — a golf score is derived from other numbers, such as number of swings and number of mulligans allowed. But Hill tells me the set of data is too small.

Why does any of this work? Mathematicians have an answer, I think. It goes something like this: "It's natural that statistical data for a phenomenon that obeys one of the power laws is biased toward the lower part of the range, whereas that for a phenomenon with saturation tends to be biased toward the upper part of the range."

Could be. But as Mrs. Dodd of my 11th-grade calculus class would tell you, don't expect me to vouch for that.

For most of us, Benford's Law is another mystery. Much like the mystery of how the entire planet agreed that red means stop. Boggles the mind.

Kevin Maney writes about technology for USA TODAY.