Benford's Law

Jeremy Rouse

Wake Forest University

30 Sep 2010
Math 165: Freshman-only Math Seminar

Outline (1/2)

- Each person should get two post-its.
- I will ask each of you to write down a random number on one of your post-its. All these will be collected in a stack.
- Once those are collected, I will ask you for a non-random number (some sort of real-world data) and ask you to write it down on the other post-it. (This non-random number probably has between 2 and 5 digits).
- All these post-its will be collected and put in a stack.

Outline (2/2)

- I will leave the room, and Dr. Parsley will arrange the stacks on the desk and tell you which stack is which.
- I won't know which is which.
- I will come back in, look through the numbers, and tell you which stack is which.

Pick a random number

Pick a random whole number and write it on one of your index cards. [Don't pick some "special number" like 1729, 2187, 16384, etc.]

Non-random number

Write down the house number of the house you lived in before coming to college. [If you lived in several houses, pick one. If you don't remember, write down your room number in the dorm].

Benford's law

Theorem

In most sources of real-world data, about 30.1% of the numbers begin with a 1. More precisely, the probability that the first digit of a number is d is

$$\log_{10}\left(1+\frac{1}{d}\right).$$

First digit	Probability
1	.301
2	.176
3	.125
4	.097
5	.079
6	.067
7	.058
8	.051
9	.046

When does it work?

• Benford's law works in situations where the data varies by several orders of magnitude.

• Some examples: Lengths of rivers (independent of units), areas of countries or states, numbers appearing on the front page of a newspaper, household income.

• Exceptions: Values of $1/n$ or $\sqrt{n}$, physical constants, quantities with little variation.

History

• In 1881, Simon Newcomb noticed that the pages of tables of logarithms were much more worn at the beginning than at the end. He stated Benford's law.

• In 1938, Frank Benford noticed the same and tested many different sets of data and found excellent agreement with the prediction.

• The first rigorous proof was given in 1995 by Ted Hill.

Example 1

Populations of cities, towns, villages, etc. in North Carolina

There are 656 cities, towns, villages, ... recorded in North Carolina from the 2000 census.

First digit of pop	Number	Prediction
1	183	197
2	128	115
3	82	81
4	71	63
5	44	52
6	43	44
7	39	38
8	33	33
9	33	30

Example 2

Charges on my credit card in the last 2 months.

- There were 75 charges in the last 2 months.

First digit of amount	Number	Prediction
1	27	22
2	13	13
3	14	9
4	5	7
5	3	6
6	4	5
7	5	4
8	2	4
9	2	3

Example 3

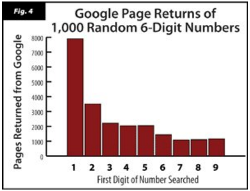
The numbers 2^k, $k \leq 100000$.

First digit	Number	Prediction
1	30102	30103
2	17611	17609
3	12492	12493
4	9692	9691
5	7919	7918
6	6695	6695
7	5797	5799
8	5116	5115
9	4576	4576

Example 4

Q: Given a number, what is the best way to determine how many times it appears in the "real-world"?

A: Google it!



Source: `http://www.thecleverest.com/benfords_law.html`

Reformulation of Benford's law

- If we take a number $n = m \cdot 10^k$, where $1 \leq m < 10$, then

$$\log_{10}(n) = k + \log_{10}(m).$$

- Benford's law is equivalent to the statement that $\log_{10}(m)$ is distributed uniformly in the interval $[0, 1]$.

- This means that the probability a number starts with the same digits as the number d is $\log_{10}\left(1 + \frac{1}{d}\right)$.

How is this useful?

- The magic trick shows that Benford's law can easily distinguish real-world data from randomly generated numbers.

- If a sizeable set of numbers fails to show the first-digit bias, this is an indication that the numbers were "cooked."

Example

• In 1993, Wayne Nelson was found guilty of defrauding Arizona of almost 2 million dollars.

• Nelson directed 23 checks to bogus vendors. The amounts of most of the checks were close to 100,000 dollars.

1st digit	Number
1	1
2	1
7	3
8	9
9	9

Use of Benford's law

- Evidence based on Benford's law is admissible in criminal trials at all levels.

- Steven J. Miller (a number theorist at Williams College) has made presentations to the IRS about the theory and applications of Benford's law.

Homework problems

1. Show that the probability that a number starts with either 10, 11, 12, 13, or 14 is *the same* as the probability that it starts with a 2.

2. (a) Draw the interval $[0, 3]$ and mark the points $\log_{10}(2^k)$ for $1 \leq k \leq 9$.
(b) If x is a real number, let $\lfloor x \rfloor$ denote the result of rounding x down to a whole number. For example, $\lfloor \pi \rfloor = 3$, and $\lfloor 0.99999 \rfloor = 0$. Show that if n is any integer, there are exactly

$$\lfloor n \log_{10}(2) \rfloor$$

values of k, $1 \leq k \leq n$, so that 2^k begins with a 1.