REPORT ON BENFORD'S LAW ANALYSIS OF 2020 PRESIDENTIAL ELECTION DATA

Alex E. Kossovsky and Steven J. Miller

We have decades of experience in the theory and application of Benford's law; this is a mathematical result that describes many data sets and is often used to detect if a data set has been modified. As there has been a lot of mixed discussion on the internet as to whether or not Benford's law is applicable to detect possible fraud in the 2020 election, we analyzed some of the data. **The standard base 10 test, as well as a new base 3 test of Benford's Law, do not show fraud in the 2020 election results for the states, counties, and precincts that we examined.**

Benford's law (originally conjectured by Newcomb) states there is a tendency in many data sets to have more numbers with low rather than high leading digit (the leading digit of 2020 is 2, of .0341 is 3). Specifically

$$\text{Prob(First digit is d)} = \log\left(1+\frac{1}{d}\right), \quad and \quad \text{Prob(Second digit is d)} = \sum_{k=1}^{9} \log\left(1+\frac{1}{10k+d}\right),$$

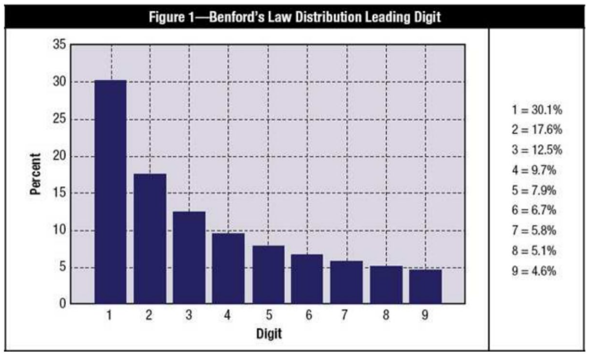
The table below gives the probability of a first or second digit being d base 10; it is possible to adjust these formulas for other bases. For example,

$$\text{Prob(First digit is d base B)} = \log\left(1+\frac{1}{d}\right) \;/\; \log(B).$$

d	Probability first digit d	Probability second digit d
0		0.1197
1	0.3010	0.1139
2	0.1761	0.1088
3	0.1249	0.1043
4	0.0969	0.1003
5	0.0792	0.0967
6	0.0669	0.0934
7	0.0580	0.0904
8	0.0512	0.0876
9	0.0458	0.0850

Table 1.1 Newcomb's conjecture for the probabilities of observing a first digit of d or a second digit of d; all probabilities are reported to four decimal digits.

Figure 1—Benford's Law Distribution Leading Digit



Not all data sets should follow Benford's law; for example, if most of the precincts have approximately the same population and each candidate's support is the same in each precinct, there will be a clustering of leading digits. One solution is to look at second digits. Our ***new approach*** is to write the number of votes in base 3 instead of base 10; the advantage of this is that our numbers are now spread out over several more magnitudes (81 is a two digit number in base 10, but in base 3 it is 10000_3, five digits, and this spreads out clumped data).

Our goal was to look at a variety of statistics and the results of the two major candidate (Biden and Trump) in several settings. This was done to see if it is reasonable to expect Benford's law to hold, and if so if the data in Pennsylvania follows Benford's law. In the table below the number in parentheses by the candidate indicates the base used for the comparison. Below the data by county (or whatever the grouping is called). The higher the chi-square value, the further the observed distribution is from Benford's law.

(--------------------------QUICK DIGRESSION ON CHI-SQUARE VALUES--------------------------)

When we test to see if data follows Benford's law, it is important that we have something better than 'it looks like a good fit' and 'it looks like a bad fit'. A common test is to compute the chi-square statistic, and compare it to the 95% and 99% values. If E(d) is the expected number that have a first digit of d base B, and O(d) is the observed number, then the chi-square statistic is

$$\sum_{d=1}^{B} \frac{(O(i) - E(i))^2}{E(i)}.$$

If we work base B=10 then as we have 9 digits and the probabilities sum to 1, we only have 8 degrees of freedom (the probability of the last digit is forced from the other 8, and zero is never a first digit). If the digit distribution is drawn from Benford's law, then 95% of the time we should have a chisquare value of 15.5 or less, and 99% of the time it should be 20.1 or less. To put it another way, if your data follows Benford's law and you observe a value greater than 20.1, there is only a 1% chance of that happening *given that your data is Benford;* if you observe values larger than 20.1 the probability falls rapidly.

(--------------------------QUICK DIGRESSION ON CHI-SQUARE VALUES--------------------------)

Below are the results for several states; similar results were found at the county and precinct level.

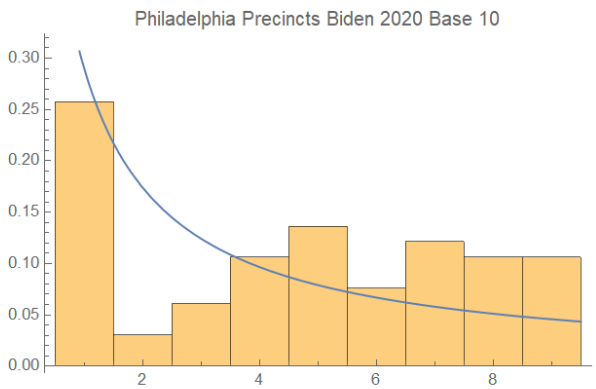
Locale	# Points	Biden (10) chi-square	Biden (3) chi-square	Trump (10) chi-square	Trump (3) chi-square
PA in person	67	9.3	2.1	12.0	0.3
PA mail	67	15.7	3.8	12.9	1.2
Texas	254	4.5	0.4	4.6	0.2
Arkansas	75	9.5	0.1	15.9	2.3
California	58	12.5	0.0	10.6	3.2
Georgia	159	13.1	1.2	10.4	0
Illinois	102	4.4	1.1	13	10.9
Indiana	92	16.8	0.7	15	0.4
Kentucky	120	7.5	1.4	21.8	0.3
Michigan	83	10.5	0.3	9.1	1.1
Minnesota	87	6.2	0	10.8	0.2
North Carolina	87	10.8	0.2	9.8	1.5
South Carolina	65	10.9	5.1	18.2	0.1
Wisconsin	72	11.4	0.4	4.4	4.2
95%		15.5	3.8	15.5	3.8
99%		20.1	6.6	20.1	6.6

Our data was drawn from not just PA but also several battleground states, as well as a few uncontested states. The data is, for the most part, consistent with Benford's law. The first digits base 10 typically had values below the 95% threshold. As expected, the fit to Benford was better when we shifted to base 3, as there the data covers more orders of magnitude.

The analysis supports election data is consistent with Benford at this scale; however, if a party were to modify only a few precincts that would not be detectable by such analysis. Additionally, if a small fixed number of votes were added across precincts that would not be detected (as it would almost surely not change the leading digit).

There was only one place above the 95% threshold for both the base 10 and base 3 test: the Biden PA mail in vote, but the two chi-square values are below the 99% cutoff threshold. The largest chi-square base 10 was Trump in Kentucky, the largest base 3 was Trump in Illinois. We have similar data at the precinct level for many of these states, indicating similar behavior (though the clustering effects are a bit stronger base in 0).

For example, below are data for Philadelphia. Note that before one uses Benford's law one must prove that Benford's law is applicable. As remarked by many, there is good reason to believe data at the precinct level should not follow Benford's law. If most precincts are between 1000 and 2000 people and between 70% and 80% of the people vote and one candidate gets between 75% and 85% of the vote, then their vote totals range from 525 to 1360, never having a first digit of 2, 3, or 4 (among other issues). To deal with such issues, people often look at the second digits, or our new trick is to look at the totals base three (base 10 there is barely a factor of 2 between the low and the high, while base 3 it is almost a full order of magnitude).



Data for Philadelphia Precincts Biden 2020 Base 10

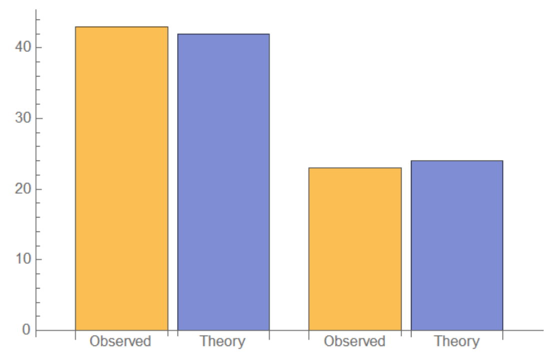
Chisquare first digit: 27.1094. Chisquare second digit: 6.81388.

Chisquare thresholds (8 degrees of freedom): 15.5 (95%), 20.1 (99%)

Chisquare thresholds (9 degrees of freedom): 16.9 (95%), 21.7 (99%)

First digit test fails Benford at the 99% level.

Second digit test passes at the 95% confidence level.



Data for Philadelphia Precincts Biden 2020 Base 3

Chisquare first digit: 0.120108. Chisquare second digit: 1.57438.

Chisquare thresholds (1 degrees of freedom): 3.84 (95%), 6.63 (99%)

Chisquare thresholds (2 degrees of freedom): 5.99 (95%), 9.21 (99%)

First digit test passes at the 95% confidence level.

Second digit test passes at the 95% confidence level.

Send email queries to akossovsky@gmail.com and stevemiller1701071@gmail.com.

REFERENCES

1. A. Berger and T. P. Hill: Benford Online Bibliography: http://www.benfordonline.net .
2. Chi-square calculator: https://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html
3. Chi-square test statistic: https://www.statisticshowto.com/probability-and-statistics/chi-square/
4. A. E. Kossovsky: BENFORDS LAW: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications. World Scientific Publishing Company, 2014. https://www.worldscientific.com/worldscibooks/10.1142/9089
5. J. C. McGinty, Can an Accounting Tool Detect Election Fraud? WSJ December 4, 2020: https://www.wsj.com/articles/can-accounting-tool-detect-election-fraud-11607077800.
6. S. J. Miller (editor): Theory and Applications of Benford's Law: Princeton University Press, 2015: http://press.princeton.edu/titles/10527.html
7. Wikipedia: https://en.wikipedia.org/wiki/Benford%27s_law