Quick Evaluation of Benford's Law

Paris Karahalios, TRIUS, Inc.

Benford's Law is experiencing a resurgence, in the last few decades, after having been forgotten since it was first evidenced that first digits of certain numbers behaved a certain way [(1)(2)].

I should point out here that in my opinion this should be called Newcombs' Law, since it was the American Astronomer Simon Newcomb in 1881 who observed that pages of logarithms starting with the digits 1 and 2 were grubbier (more used) that pages that contained logarithms starting with a larger digit. So, he suggested that the probability of the first or leading digit being ***d*** should be:

$$P(d) = log_{10}\left(\frac{d+1}{d}\right), \quad d = 1,2,3,4, \dots .9 \qquad \text{Eq. (1)}$$

Of course, it was not until 1938, when Frank Benford, a physicist with the General Electric Company, assembled over 20,000 numbers from diverse sources (Readers' Digest articles, street addresses of American Men of Science, atomic weights, population sizes, drainage rates of rivers, and physical constants) and showed that leading digits from a wide range of sources showed an uncanny adherence to this logarithmic rule

As Oded Kafri has noted, "Although it attracts considerable attention, there is no a priori probabilistic criterion when a data set should or should not obey the law."[(3)] Note that Oded Kafri, in the cited reference, in my opinion provides the best/simplest explanation why Benford's Law works (using an example of 3 balls placed in 3 boxes).

Furthermore, Benford's distribution of digits is at first counterintuitive. One would expect that a set of random integers starting with digits1-9 would result in uniformity of their first digits distribution,

$$P(9) = \frac{1}{9} \qquad \text{Eq. (2)}$$

as in the case of an unbiased lottery. This is precisely the reason why Benford's law is routinely used by income tax agencies of several nations to detect fraud of large companies and accounting businesses [(4)]. Usually, when fraud is takes place, the digits are invoked in equal probabilities and the distribution of digits does **not** follow Eq. (1).

Although no claim is made here that such a priori criterion exists, based on the simple analysis presented here, it appears evident that numbers that are themselves statistics of any sort, do not seem to follow the Law.

The first graph (Fig. 1) shown below plots the first digit of the areas of the countries of the world (204 data points, https://simple.wikipedia.org/wiki/List_of_countries_by_area) and the GDPs of 192 countries (https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)), together with the theoretical Benford's Law curve, as defined in the equation above.

It is clear from this graph that the two sets of numbers behave in a way predictable by Benford's Law.

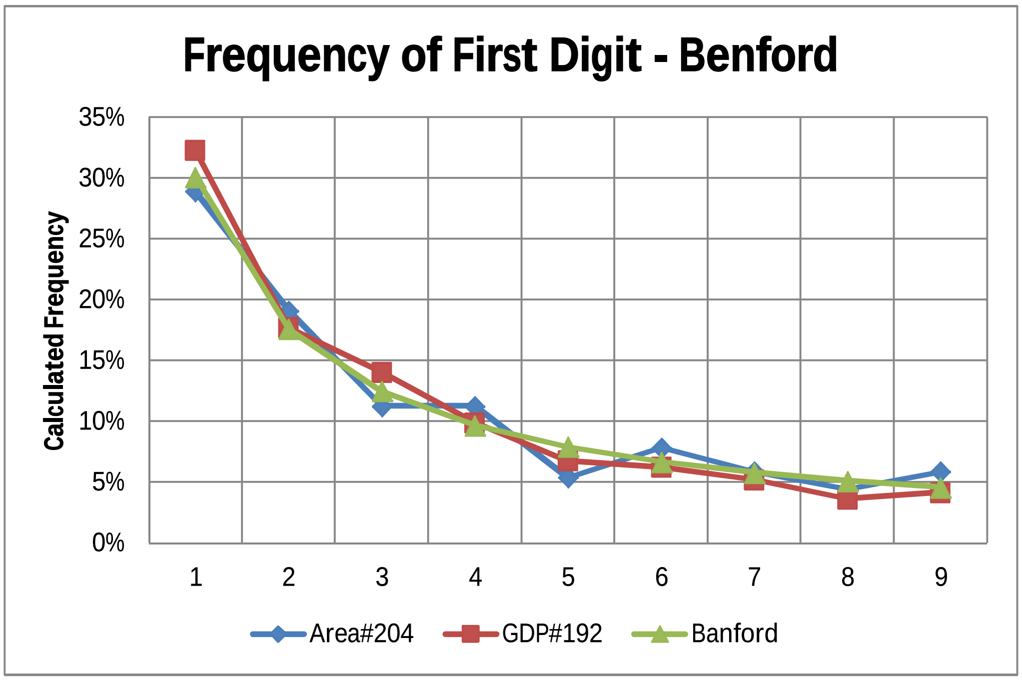

Fig.1

The second graph (Fig. 2) shown below plots the same three frequencies from Fig. 1 and in addition, it also plots two additional datasets, the Average temperature in each country (192 data points, https://en.wikipedia.org/wiki/List_of_countries_by_average_yearly_temperature) and the distribution of 300 one digit random number 1..9 generated in excel.

Based on the second Figure, it appears evident that numbers that are themselves statistics, e.g. average temperature, or random numbers generated based on some/any distribution built-in in excel, do not follow Benford's Law.

As stated at the beginning of this paper, it is not supposed to be some absolute proof of Benford's Law, but rather a quick look at why some data sets appear to follow the Law closely, while others do not follow it at all.

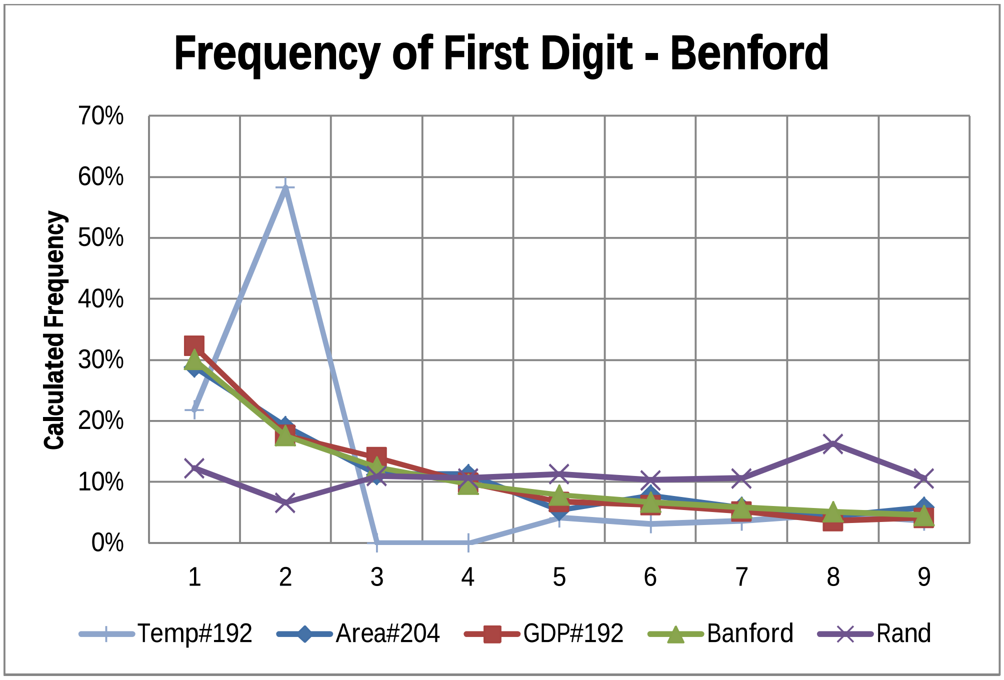

Fig. 2

If time permits, at a later date, another more detailed analysis may be performed and the results shared with the rest of those interested in this intriguing subject.

References

1. "***A Simple Explanation of Benford's Law***", R. M. FEWSTER, The American Statistician, February 2009, Vol. 63, No. 1
2. "***The First Digit Phenomenon***", T.P. Hill, American Scientist, 1998, No. 4
3. "**Entropy Principle in Direct Derivation of Benford's Law**", Oded Kafri Varicom Communications, Tel Aviv 68165 Israel
4. "**Numbers follow a surprising law of digits, and scientists can't explain why**", L. Zyga, www.PhysOrg.com (2007).
5. "***Benford's Law of First Digits: From Mathematical Curiosity to Change Detector***", Malcolm Sambridge, Hrvoje Tkalˇci´c and Pierre Arroucau, Asia Matematics Newsletter, October 2011, Volume 1 No 4
6. "**Benford's law**" Haack Christina , Kersch Julie , Van der Weken Lynn June 1, 2015
7. "**Entropy Principle in Direct Derivation of Benford's Law**", Oded Kafri Varicom Communications, Tel Aviv 68165 Israel