Fundamental Flaws in Feller's

Classical Derivation of Benford's Law

Arno Berger[*]

Mathematical and Statistical Sciences, University of Alberta

and

Theodore P. Hill[†]

School of Mathematics, Georgia Institute of Technology

22 January 2010

Abstract

Feller's classic text *An Introduction to Probability Theory and its Applications* contains a derivation of the well known significant-digit law called Benford's law. More specifically, Feller gives a sufficient condition ("large spread") for a random variable X to be approximately Benford distributed, that is, for $\log_{10} X$ to be approximately uniformly distributed modulo one. This note shows that the large-spread derivation, which continues to be widely cited and used, contains serious basic errors. Concrete examples and a new inequality clearly demonstrate that large spread (or large spread on a logarithmic scale) does not imply that a random variable is approximately Benford distributed, for any reasonable definition of "spread" or measure of dispersion.

Key words: Significant digit, Benford distribution, Kolmogorov–Smirnoff distance, uniform distribution modulo one.

In probability and statistics, a correct general explanation of a principle is often as valuable as a detailed formal argument. In his December 2009 column in the IMS Bulletin, UC Berkeley statistics professor T. Speed extols the virtues of derivations in statistics (Speed 2009):

> *I think in statistics we need derivations, not proofs. That is, lines of reasoning from some assumptions to a formula, or a procedure, which may or may not have certain properties in a given context, but which, all going well, might provide some insight.*

For illustration, Speed quotes two examples of the convolution property for the Gamma and Cauchy distributions from the classic 1966 text *An Introduction to Probability Theory and Its Applications* by W. Feller.

On page 63, Feller (1966) also gave a brief derivation, in Speed's sense, of the well known logarithmic distribution of *significant digits* called Benford's law (Benford 1938; Fewster 2009; Hill 1995a,b; Newcomb 1881; Raimi 1976). Recall that if a random variable is Benford (i.e. has a Benford distribution) then its first significant digit is "1" with probability $\log_{10} 2 \approx 0.3010$; similar expressions hold for the general joint Benford distributions of all the significant digits (Hill 1995a). For the purposes of this note, a simple and very useful characterization of a Benford distribution is

(1) A positive random variable X is Benford if and only if $\log_{10} X$ is uniformly distributed (mod 1).

Since Feller has inspired so many who teach probability and statistics today, and since many undergraduate courses now include a brief introduction to Benford's law, it is not surprising that Feller's derivation is still in frequent use to provide some insight about Benford's law. For example, a class project report for a 2009 upper-division course in statistics at UC Berkeley (Aldous and Phan 2009, p.3) said

> *...like the birthday paradox, an explanation [of Benford's law] occurs quickly to those with appropriate mathematical background ...To a mathematical statistician, Feller's paragraph says all there is to say ...Feller's derivation has been common knowledge in the academic community throughout the last 40 years.*

The online database (Berger and Hill 2009) lists about twenty published references since 2000

alone to Feller's argument (e.g. Aldous and Phan 2009; Fewster 2009) the crux of which is Feller's claim (trivially edited) that

(2) If the spread of a random variable X is very large, then $\log_{10} X \pmod 1$ will be approximately uniformly distributed on $[0, 1)$.

The implication of (1) and (2) is that all random variables with large spread will be approximately Benford distributed. That sounds quite plausible, but as C.S. Pierce observed (Gardner 1959, p.174), "in no other branch of mathematics is it so easy for experts to blunder as in probability theory". Indeed, even Feller blundered on Benford's law, and took many experts with him. Claim (2) is simply false under any reasonable definition of spread or measure of dispersion, including *range*, *interquantile range* (or distance between the $(1-\alpha)$- and the α-quantile), *standard deviation*, or *mean difference* (Gini coefficient), no matter how smooth or level a density the random variable X may have. To see this, one does not have to look far. Concretely, no positive uniformly distributed random variable even comes close to being Benford, regardless of how large (or small) its spread is. This statement can be quantified explicitly via the following new inequality; for its formulation, recall that the Kolmogorov-Smirnoff distance $d_{\mathrm{KS}}(X, Y)$ between two random variables X and Y with cumulative distribution functions F and G, respectively, is $d_{\mathrm{KS}}(X, Y) = \sup\{|F(x) - G(x)| : x \in \mathbb{R}\}$.

Proposition 1 (Berger 2010). *For every positive uniformly distributed random variable X,*

$$d_{\mathrm{KS}}\big(\log_{10} X \pmod 1, U(0,1)\big) \geq \frac{-9 + \ln 10 + 9 \ln 9 - 9 \ln \ln 10}{18 \ln 10} = 0.1334\dots,$$

and this bound is sharp.

There is nothing special about the usage of the Kolmogorov-Smirnoff distance or decimal base in this regard; similar universal bounds hold for the Wasserstein distance, for example, and other bases. Another way to see that (2) is false, in the discrete and significant-digit setting, is to observe that no matter how large n is, an integer-valued random variable uniformly distributed on the first $2 \cdot 10^n$ positive integers will have more than 50% of its values beginning with a "1", as opposed to the Benford probability of about 30%.

How could Feller's error have persisted in the academic community, among students and experts alike, for over 40 years? Part of the reason, as one colleague put it, is simply that Feller, after all, is Feller, and Feller's word on probability has just been taken as gospel. Another reason

for the long-lived propagation of the error has apparently been the confusion of (2) with the similar claim

(3) If the spread of a random variable X is very large, then X (mod 1) will be approximately uniformly distributed on $[0, 1)$.

For example, (Aldous and Phan 2009, p.3) cites Feller's claim (2), but (Aldous and Phan 2009, p.8) cites Feller's claim as (3). A third possible explanation for the persistence of the error is the common assumption that (3) implies (2). For example, Gauvrit and Delahaye (2009, p.1) state:

> *An elementary new explanation has recently been published, based on the fact that any X whose distribution is "smooth" and "scattered" enough is Benford. The scattering and smoothness of usual data ensures that $\log(X)$ is itself smooth and scattered, which in turn implies the Benford characteristic of X.*

Now (3) is also intuitive and plausible, but unlike (2), it is often accurate if the distribution is fairly uniform. And if the distribution is not fairly uniform, then without further information, no interesting conclusions at all can be made about the significant digits — most of the values could for instance start with a "7". Since X has very, very large spread if and only if $\log X$ has very large spread, on the surface (2) and (3) appear to be equivalent. After all, what difference can one tiny extra "very" mean? On the other hand, as Proposition 1 clearly implies, they are not the same, and (2) is false.

Although (3) is perhaps more accurate than (2), unfortunately it does not explain Benford's law at all, since the criterion in (1) says that X is Benford if and only if the logarithm of X — and not X itself — is uniformly distributed (mod 1). Some authors partially explain the ubiquity of Benford distributions based on an assumption of a "large spread on a logarithmic scale" (e.g. Aldous and Phan 2009, 2010; Fewster 2009; Wagon 2010). Others (e.g. Aldous and Phan 2010, p.17) claim that "what Feller obviously *meant*" [italics in original] by spread was log spread, i.e. that when Feller wrote (2) he really meant to say that

(3') If $\log_{10} X$ has very large spread, then $\log_{10} X$ (mod 1) will be approximately uniformly distributed on $[0, 1)$,

which is but an unnecessarily convoluted version of (3). They then apply (3) or (3') to conclude that if $\log_{10} X$ has large spread, then X is approximately Benford. This avoids Feller's error (2), but still leaves open the question of why it is reasonable to assume that the *logarithm* of the spread, as opposed to the spread itself or, say, the log log spread should be large. As seen above, those assumptions contain a subtle difference, and lead to very different conclusions about the distributions of the significant digits. Using the same logic, for instance, an assumption of large spread on the log log scale would imply that $\log X$ is Benford, whereas none of the usual Benford random variables such as X_k with densities $1/(x \ln 10)$ on $(10^k, 10^{k+1})$ are also Benford on the log scale. Moreover, via (1) and (3), assuming large spread on a logarithmic scale is equivalent to assuming an approximate Benford distribution. Quite likely, Feller realized this, and in (2) specifically did *not* hypothesize that the log of the range was large.

A related and apparently widespread misconception is that claim (2), notwithstanding its incorrectness, or claim (3) implies that a larger spread or log spread automatically means closer conformance to Benford's law. For example, Wagon (2010) concludes that "datasets with large logarithmic spread will naturally follow the law, while datasets with small spread will not", and the Conclusion of the study (Aldous and Phan 2010, p.12) states

> On a small stage (18 data-sets) we have checked a theoretical prediction. Not just the literal assertion of Benford's law - that in a data-set with large spread on a logarithmic scale, the relative frequencies of leading digits will approximately follow the Benford distribution - but the rather more specific prediction that distance from Benford should decrease as that spread increases. In one sense it's not surprising this works out.

But distance from the Benford distribution does not generally decrease as the spread increases, regardless of whether the spread is measured on the original scale or on the logarithmic scale. A simple way to see this is as follows: Let Y be a random variable uniformly distributed on $(0,1)$, and let $X = 10^Y$ and $Z = 10^{3Y/2}$. Then by (1), X is exactly Benford, since $\log_{10} X = Y$, and Z is not close to Benford since $3Y/2 \pmod 1$ is not close to uniform on $(0,1)$. Yet for any reasonable definition of spread, including all those mentioned earlier, the spread of Z is larger than the spread of X, and the spread of $\log_{10} Z = 3Y/2$ is larger than the spread of $\log_{10} X = Y$. Another way to see that the distance from the Benford distribution does not

decrease as the spread increases is contained in the proof of Proposition 1: For X_T a random variable uniformly distributed on $(0, T)$, it is shown that the Kolmogorov-Smirnoff distance between $\log_{10} X_T$ and $U(0, 1)$ is a continuous 1-periodic function of $\log_{10} T$. Moreover, when employing a logarithmic scale it is important to keep in mind that what is considered large generally depends on the base of the logarithm. For example, as noted earlier, if Y is uniformly distributed on $(0, 1)$ then $X = 10^Y$ is exactly Benford base 10, yet it is not Benford base 2 even though its spread on the $\log_2$-scale is $\log_2 10 \approx 3.3219$ times as large.

It is interesting to note that when Feller credited Pinkham in his derivation in 1966, it was not widely known that Pinkham's argument (Pinkham 1961) for the scale-invariant characterization of Benford's law also contains an irreparable and fundamental flaw. Raimi (1976, sec. 7) explains Pinkham's error in detail, and credits Knuth (1997) for the discovery that the error was in Pinkham's unwritten implicit assumption that there exists a scale-invariant probability distribution on the positive real numbers — when clearly there does not, since the largest median of every positive random variable changes under changes of scale. The first correct proof that the Benford distribution is the unique scale-invariant probability distribution on the significant digits (and the unique continuous base-invariant distribution) is in (Hill 1995b).

In conclusion, classroom experiments based on Feller's derivation or on an assumption of large range on a logarithmic scale (e.g. Aldous and Phan 2009, 2010; Fewster 2009; Wagon 2010) should be used with caution. As an alternative or supplement, teachers might also ask students to compare the significant digits in the first 20-30 articles in tomorrow's *New York Times* against Benford's law, thereby testing real-life data against the explanation given in the main theorem in (Hill 1995b) which, without any assumptions on magnitude of spread, shows that mixing data from different distributions in an unbiased manner leads to a Benford distribution.

Although some experts may still feel that "like the birthday paradox, there is a simple and standard explanation" for Benford's law (Aldous and Phan 2010, p.6) and that this explanation occurs quickly to those with appropriate mathematical background, there does not appear to be a simple derivation of Benford's law that both offers a "correct explanation" (Aldous and Phan 2010, p.7) and satisfies Speed's goal to provide insight. In that sense, although Benford's law now rests on solid mathematical ground, most experts seem to agree with (Fewster 2009) that its ubiquity in real-life data remains mysterious.

References

Aldous, D., and Phan, T. (2009), "When Can One Test an Explanation? Compare and Contrast Benford's Law and the Fuzzy CLT," Class project report, Statistics Department, UC Berkeley; retrieved Jan. 9, 2010 from `http://www.stat.berkeley.edu/~aldous/Papers/me-benford` and from (Berger and Hill 2009).

Aldous, D., and Phan, T. (2010), "When Can One Test an Explanation? Compare and Contrast Benford's Law and the Fuzzy CLT," Preprint dated Jan. 3, 2010, Statistics Department, UC Berkeley; retrieved Jan. 16, 2010 from (Berger and Hill 2009).

Benford, F. (1938), "The law of anomalous numbers," *Proceedings of the American Philosophical Society* **78**, 551–572.

Berger, A. (2010), "Large spread does not imply Benford's law," Preprint; retrieved Jan. 16, 2010 from `http://www.math.ualberta.ca/~aberger/papers`.

Berger, A., and Hill, T.P. (2009), *Benford Online Bibliography*, `http://www.benfordonline.net`; accessed Jan. 16, 2010.

Feller, W. (1966), *An Introduction to Probability Theory and Its Applications* vol 2, 2nd ed., J. Wiley, New York.

Fewster, R. (2009), "A simple Explanation of Benford's Law," *American Statistician* **63**(1), 20–25.

Gardner, M. (1959), "Mathematical Games: Problems involving questions of probability and ambiguity," *Scientific American* **201**, 174–182.

Gauvrit, N., and Delahaye, J.P. (2009), "Loi de Benford générale," *Mathématiques et sciences humaines* **186**, 5–15; retrieved Jan. 9, 2010 from `http://msh.revues.org/document11034.html`.

Hill, T.P. (1995a), "Base-Invariance Implies Benford's Law," *Proceedings of the American Mathematical Society* **123**(3), 887–895.

————— (1995b), "A Statistical Derivation of the Significant-Digit Law," *Statistical Science* **10**(4), 354–363.

Knuth, D. (1997), *The Art of Computer Programming*, pp 253-264, vol. 2, 3rd ed, Addison-Wesley, Reading, MA.

Newcomb, S. (1881), "Note on the frequency of use of the different digits in natural numbers," *American Journal of Mathematics* **4**(1), 39–40.

Pinkham, R. (1961), "On the Distribution of First Significant Digits," *Annals of Mathematical Statistics* **32**(4), 1223–1230.

Raimi, R. (1976), "The First Digit Problem," *American Mathematical Monthly* **83**(7), 521–538.

Speed, T. (2009), "You want proof?," *Bulletin of the Institute of Mathematical Statistics* **38**, p 11.

Wagon, S. (2010), "Benford's Law and Data Spread;" retrieved Jan. 20, 2010 from Wolfram Online Demonstrations Projects `http://demonstrations.wolfram.com/BenfordsLawAndDataSpread`.