Benford's Law, A Growth Industry

Kenneth A. Ross

Abstract. Often data in the real world have the property that the first digit 1 appears about 30% of the time, the first digit 2 appears about 17% of the time, and so on with the first digit 9 appearing about 5% of the time. This phenomenon is known as Benford's law. This paper provides a simple explanation, suitable for nonmathematicians, of why Benford's law holds for data that have been growing (or shrinking) exponentially over time. Two theorems verify that Benford's law holds if the initial values and rates of growth of the data appear at random.

INTRODUCTION. To get started, consider the sequence $2, 4, 8, 16, 32, 64, 128, \ldots, 2^{1000}$. Each of these numbers has a first digit, starting out with 2, 4, 8, 1, 3, 6, 1. About how many of these numbers do you think have first digit 1? What about the other eight possible first digits? Consider the same questions for $9, 81, 729, \ldots, 9^{1000}$, where the first few first digits are 9, 8, 7, 6, 5, 5, 4, 4, 3, 3, 3, 2, 2, 2, 2. We will return to these questions after Table 3.

Way back in 1881, Simon Newcomb observed that certain sets of data from the real world have the property that the first digits of the numbers do not appear uniformly often. In fact, the first digit 1 appears about 30% of the time, while the first digit 9 appears about 5% of the time. As often happens in science, this observation was forgotten and rediscovered later. Frank Benford observed this phenomenon in 1938 [**2**], and it has become known as Benford's law or the "first-digit law." He gave numerous examples of data from many sources, including newspaper items, areas of rivers, street addresses, cost data, and populations. We give another example in Table 1, namely data based on the populations of 117 cities in Indiana, from Wikipedia based on the 2000 census.[1] For a good history up to 1975 with many references, see Raimi [**18**]. More references will be given at the end of the paper.

Table 1.

first digit →	1	2	3	4	5	6	7	8	9
number of cities	34	17	16	8	16	12	5	3	6
percent of cities	29.1	14.5	13.7	6.8	13.7	10.3	4.3	2.6	5.1

Recently, Benford's law has gone public. In fact, it was discussed recently on a public radio program, but listeners complained that it was not explained. That was because it isn't possible to explain it in a sound bite. Since Benford's law applies to many different sorts of data, it is natural to view it as a statistical phenomenon, so most serious attempts at understanding it use probability. The most successful seems to be Theodore Hill's very nice, but sophisticated, analysis in [**13**]. I found his analysis challenging, and I am reasonably acquainted with probability. Thus, neither Hill's result nor its proof is accessible to young students and others who have an interest in Benford's law. In my view, what is needed is a collection of special cases that are

http://dx.doi.org/10.4169/amer.math.monthly.118.07.571

[1]This is the first and only state that I examined for this purpose.

readily understandable. In fact, I think it might be easiest to start with nonstatistical, totally deterministic examples, which will be formalized in Theorem 1. Theorems 2 and 3 provide versions of Benford's law.

This article is intended for four audiences. The discussion up to Figure 1 could be used to explain Benford's law to people with essentially no mathematical background, while the discussion up to the proof of Theorem 1, just after Table 5, could be used to explain the law to any motivated person who understands logarithms. Upper-division undergraduate students, willing to assume the uniform distribution theorem, should be able to read the proof of Theorem 1. The remaining proofs involve a little measure theory and harmonic analysis.

MODELING BENFORD'S LAW. Here we will model populations of some old cities in some region, with the following assumptions. Though the populations of the cities probably started out as random, each of them grows (or shrinks) exponentially over time, so that a census is a snapshot of the population-growth curves of the various cities. We will see that this is why the observed populations years later will approximately satisfy Benford's law.

We will focus at first on one city. The population of a city satisfies the equation $P(n) = ar^n$. Here n represents units of time in integers (such as years or decades). We are only interested in the first digits of the populations. Since multiplying them by any power of 10 won't change the mathematics, we henceforth assume that the initial "population" a satisfies $1 \leq a < 10$ and the rate of growth r satisfies $1 < r < 10$.

More generally, suppose you have data collected regularly over time, and the data describe a quantity which is increasing or decreasing exponentially. The data might be populations as in Table 1, values of portfolios,[2] weights of vegetables at a county fair, etc. Suppose the length of time over which the data is collected is sufficiently large; this can be quantified and will depend on the rate of growth. Then Bedford's law applies. Roughly, that's because the time spent when the first digit is 1 is relatively flat and is considerably longer than the time spent when the curve is increasing more steeply. See Figure 1, showing how ar^t increases from 10^k to 10^{k+1}. (If $r < 1$, the picture is the mirror image of the one in Figure 1.) The first digit of ar^t is 1 for the time interval $[c_1, c_2)$, it is 2 for the time interval $[c_2, c_3)$, etc. So clearly the first digit is 1 for much more time than it is 9.

We want to quantify these observations, so look again at Figure 1. For $d = 1, 2, \ldots, 9$, the time that the first digit is d is the length of the interval $[c_d, c_{d+1})$. To determine these times, we observe that the graph of ar^t has height $d \cdot 10^k$ at c_d. Thus

$$ar^{c_d} = d \cdot 10^k.$$

Using the rules of logarithms, we obtain

$$\log_{10} a + c_d \log_{10} r = \log_{10} d + k$$

and similarly

$$\log_{10} a + c_{d+1} \log_{10} r = \log_{10}(d+1) + k.$$

Thus

$$(c_{d+1} - c_d) \log_{10} r = \log_{10}(d+1) - \log_{10} d,$$

[2] Several articles study applications of Benford's law to the detection of accounting fraud; see, e.g., [**7**] and [**15**].

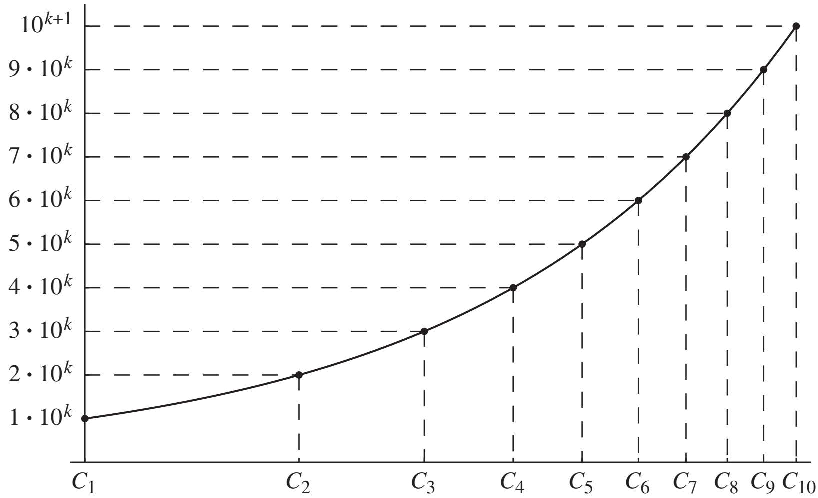

Figure 1.

so that

$$c_{d+1} - c_d = A[\log_{10}(d+1) - \log_{10} d],$$

where A is the constant $1/\log_{10} r$, which does not depend on k. Moreover, the time that the curve takes to rise from 10^k to 10^{k+1} is exactly $c_{10} - c_1 = A[\log_{10}(10) - \log_{10} 1] = A$. In other words, in this time interval, *the fraction of time that the first digit of ar^t is d is*

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

Moreover, this observation applies to all such intervals $[10^k, 10^{k+1})$, $k = 0, \pm 1, \pm 2, \ldots$. The numbers in equation (1) are the numbers that arise in Benford's law. Approximations of them are given in Table 2.

Table 2.

d	1	2	3	4	5	6	7	8	9
$\log_{10} \frac{d+1}{d}$	0.30103	0.17609	0.12494	0.09691	0.07918	0.06695	0.05799	0.05115	0.04576

Let's return to discrete time, i.e., return to $P(n) = ar^n$ for integers n, and visualize these values in Figure 1, i.e., visualize the points $(n, P(n))$ marching along the curve. From the statement involving (1), one would expect that the fraction of these points that would land in the interval $[c_d, c_{d+1})$ would be about $\log_{10}\left(\frac{d+1}{d}\right)$. This makes the following theorem very plausible, though its proof (given later) is somewhat more complicated.

For any $x \geq 1$, we write $D(x)$ for the first digit of x.

Theorem 1. *Suppose that $1 \leq a < 10$, $1 < r < 10$, and that r is rational. Then*

$$\lim_{n\to\infty} \frac{1}{n} \#\left\{k \leq n : D(ar^k) = d\right\} = \log_{10}(d+1) - \log_{10} d, \tag{2}$$

for $d = 1, 2, 3, 4, 5, 6, 7, 8, 9$.

The uninvited hypothesis, that r be rational, comes up in the proof. In fact, equation (2) holds for all but countably many real numbers r in (1, 10). Theorem 1 appears in Raimi [**18**] and [**19**], and it essentially appears in Diaconis [**6**].

Theorem 1 suggests that, for reasonable choices of r and a, the numbers $\#\{k \leq 100 : D(ar^k) = d\}$ will be close to $100[\log_{10}(d+1) - \log_{10} d]$. It also suggests that the numbers $\#\{11 \leq k \leq 110 : D(ar^k) = d\}$ will be close to the same values, and I prefer to look at this because the first few values of $D(ar^k)$ are likely to be predictable, especially if r is close to 1. Table 3 gives some examples of actual counts. For example, the third column indicates that the theorem predicts that about 30.1 out of 100 values of $a \cdot r^k$ will have first digit equal to 1. The fourth row indicates that exactly 31 of the values $\{9 \cdot 2^k : 11 \leq k \leq 110\}$ have first digit equal to 1.

Table 3.

expected # first digits		30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6
a	growth rate r	1	2	3	4	5	6	7	8	9
1	2	30	17	13	10	7	7	6	5	5
9	2	31	17	13	9	8	7	6	5	4
1	3	32	16	14	10	8	6	6	4	4
5	3	30	18	12	8	9	7	7	5	4
9	3	31	17	14	10	7	6	6	4	5
1	1.1	28	19	14	10	7	6	6	6	4
5	1.1	33	17	12	10	8	6	6	4	4
9	1.1	29	19	13	10	8	5	7	5	4
1	1.05	32	18	12	9	7	7	5	6	4
5	1.05	30	16	12	10	7	7	5	7	6
9	1.05	34	16	12	10	7	7	4	6	4
1	1.01	59	41	0	0	0	0	0	0	0
5	1.01	41	0	0	0	8	15	14	12	10
9	1.01	70	30	0	0	0	0	0	0	0

The answers to the questions at the beginning of the paper are similar. The nine answers to the first question about $2, 4, 8, 16, \ldots, 2^{1000}$ are 301, 176, 125, 97, 79, 69, 56, 52, 45. The answers to the question about $9, 81, 729, \ldots, 9^{1000}$ are 297, 177, 125, 98, 80, 67, 58, 52, 46.

Note that $r = 1.01$ is too close to 1 to illustrate the theorem only checking $D(ar^k) = d$ for k up to 110. On the other hand, with $a = 1$ and $r = 1.01$, the counts for $\#\{101 \leq k \leq 1100 : D(ar^k) = d\}$, which should be close to $1000[\log_{10}(d+1) - \log_{10} d]$, turn out to be 279, 173, 144, 112, 87, 61, 54, 48, and 42.

Theorem 1 is interesting, but it doesn't explain the model involving old cities. It explains the model if we focused on *one* city and its various populations over a long period of time. With several cities in mind, it seems likely that their initial populations and growth rates are random. Therefore, the following version of Benford's law seems quite plausible.

Theorem 2. $\lim_{n\to\infty} \Pr[D(ar^n) = d] = \log_{10}(d+1) - \log_{10} d$.

Probability? We assume that the initial population a is selected at random from $[1, 10)$ by any random process, but the reader won't lose anything by assuming that the choices are uniformly distributed, i.e., for each interval I in $[1, 10)$, the values a fall into the interval I with probability length$(I)/9$. We assume that the growth rate is selected uniformly from a fixed interval $[s, u)$ in $[1, 10)$ so that, for each interval I in $[s, u)$, the values are selected with probability length$(I)/(u - s)$.

Theorem 2 implies that if n is sufficiently large, and if we take a sufficiently large sample (say of size N) of values (a, r) uniformly selected from the rectangle $[1, 10) \times [s, u)$, then

$$\#\{(a, r) \text{ in the sample satisfying } D(ar^n) = d\} \approx N[\log_{10}(d+1) - \log_{10} d], \quad (3)$$

for $d = 1, 2, 3, 4, 5, 6, 7, 8, 9$. We illustrate counts, by taking random samples from $[1, 10) \times [1, 1.5)$ of size $N = 1000$. See Table 4.

Table 4.

expected # first digits $\longrightarrow$	301	176	125	97	79	67	58	51	46
value of n $\downarrow$	1	2	3	4	5	6	7	8	9
10	273	190	140	129	66	60	47	54	41
10	312	175	135	92	70	66	61	46	43
20	305	176	127	89	74	73	58	58	40
20	304	154	143	85	75	69	69	66	35
50	295	174	127	103	78	66	53	49	55
50	289	158	130	97	93	64	69	43	57
100	303	194	122	109	72	65	50	36	49
100	306	179	132	91	78	58	61	45	50

The table seems consistent with Theorem 2, though the counts based on random samples vary more than I had expected.

Theodore P. Hill has suggested that Theorem 2 would be a more realistic model if the rates of growth varied over time. Theorem 3 below was motivated by this observation.

Theorem 3. $\lim_{n\to\infty} \Pr[D(ar_1 r_2 \cdots r_n) = d] = \log_{10}(d+1) - \log_{10} d$.

As with Theorem 2, a is selected at random from $[1, 10)$ by any random process, but now each r_i is selected uniformly from some fixed interval $[s, u)$ in $[1, 10)$ and the selections of r_i are assumed to be independent. Table 5 gives some results, each based on samples of size 1000 from $[1, 10) \times [1, 1.5)$.

The requirements of uniform selection in Theorems 2 and 3 can be substantially relaxed, as will be explained later.

Table 5.

expected # first digits $\longrightarrow$	301	176	125	97	79	67	58	51	46
value of n $\downarrow$	1	2	3	4	5	6	7	8	9
10	321	172	126	99	71	43	61	48	59
10	332	178	116	71	82	63	57	46	55
20	306	181	104	90	94	69	65	38	53
20	315	162	141	98	80	50	55	45	54
50	306	176	120	95	76	77	56	44	50
50	311	163	132	99	74	69	53	52	47
100	316	190	116	89	78	65	47	51	48
100	308	168	114	108	73	69	54	49	57

PROOF OF THEOREM 1. We will use the function $g(x) := \log_{10} x \bmod 1$, for x in $[1, \infty)$. Here mod 1 means that we are taking the decimal or fractional portion of $\log_{10} x$. Thus g is a function from $[1, \infty)$ onto $[0, 1)$, though we will sometimes restrict it to $[1, 10)$ where it is one-to-one. We actually prefer to identify the image set $[0, 1)$ with the circle of circumference 1 by using the wrapping function:

$$W(x) := \frac{1}{2\pi}(\cos 2\pi x, \sin 2\pi x) \text{ or, if you prefer, } W(x) := \frac{1}{2\pi}e^{2\pi i x}.$$

By laws of logarithms,

$$g(ar^k) = (\log_{10} a + k \log_{10} r) \bmod 1 \text{ for all } k. \tag{4}$$

First we assume $a = 1$, so that $g(r^k) = k \log_{10} r \bmod 1$ for all k. Observe that $D(x) = d$ if and only if $d \cdot 10^j \leq x < (d + 1) \cdot 10^j$ for some integer j, so that $D(x) = d$ if and only if $\log_{10} d \leq \log_{10} x \bmod 1 < \log_{10}(d + 1)$. Hence

$$D(x) = d \text{ if and only if } \log_{10} d \leq g(x) < \log_{10}(d + 1), \tag{5}$$

and

$$D(r^k) = d \text{ if and only if } \log_{10} d \leq k \log_{10} r \bmod 1 < \log_{10}(d + 1). \tag{6}$$

Figure 2 shows a picture of the circle of circumference 1 and a picture of the regions corresponding to where $D = d$ for $d = 1, 2, \ldots, 9$. Old-timers may be interested that Raimi [**18**] included a picture similar to our second picture in Figure 2 and that he discussed its relationship to a circular slide rule.

We need the following well-known fact.

Uniform Distribution Theorem. For any irrational x in $[0, 1)$, the sequence $(kx \bmod 1)$ is uniformly distributed. That is, for any interval I in $[0, 1)$, we have

$$\lim_{n\to\infty} \frac{1}{n}\#\{k \leq n : kx \bmod 1 \text{ is in } I\} = \text{length}(I).$$

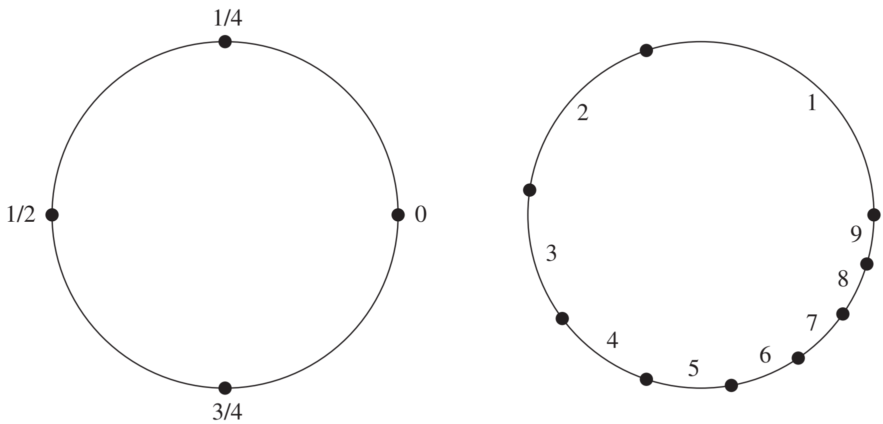

Figure 2.

This result is false if x is rational, because then the sequence (kx mod 1) takes on only finitely many values. Before we discuss and prove this theorem, we apply it to $x = \log_{10} r$, so we need $\log_{10} r$ to be irrational; but this is true by [**17**, Theorem 2.11] since we're assuming r is rational and $1 < r < 10$. We conclude that for any integer d, $1 \le d \le 9$,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\#\{k \le n : \log_{10} d \le k \log_{10} r \bmod 1 < \log_{10}(d+1)\}\\ = \log_{10}(d+1) - \log_{10} d.\end{aligned}$$

Equation (2) for $a = 1$ follows from this and the observation in (6). The proof works even if $a > 1$, because the uniform distribution theorem remains true if kx mod 1 is replaced by $(c + kx)$ mod 1 for some constant c, since the sequence $((c + kx)$ mod 1$)$ is just the sequence (kx mod 1) rotated a constant angle around the circle. This completes the proof of Theorem 1.

As noted in Raimi [**19**], the conclusion

$$\lim_{n\to\infty} \frac{1}{n}\#\{k \le n : D(FIB(k)) = d\} = \log_{10}(d+1) - \log_{10} d,$$

for the Fibonacci sequence ($FIB(k)$), has been noted by D. Singmaster and others since the 1960s. This is true because this sequence is asymptotic with the golden-mean geometric sequence $(((1+\sqrt{5})/2)^k)$.

The proof of the uniform distribution theorem is somewhat complicated, but the result is reasonable and is what one would hope would be true. The theorem was proved, independently, by P. Bohl, W. Sierpinski, and H. Weyl in 1909–1910. For two proofs, see [**17**, Theorem 6.3]. One proof uses continued fractions; the other uses Fourier series. An intricate proof, using only limits and epsilon-delta arguments, is given in [**16**, Chapter 2, Theorem 8.7]. In [**16**, Section 5.9], the uniform distribution theorem is discussed as an application of Birkhoff's individual ergodic theorem. We will give another, and more modern, proof.

First, we collect together some standard facts in the setting of the compact group $G = [0, 1)$; here 0 and 1 are identified. For a measure μ on G, its *Fourier-Stieltjes transform* $\widehat{\mu}$ is defined by $\widehat{\mu}(m) := \int_G \overline{\chi_m}\, d\mu$ for integers m, where χ_m is defined by $\chi_m(x) := \exp(2\pi i m x)$ for $x \in G$. A good reference is [**14**, Chapter I, §7.4, §7.12].

I thank my friend Robert B. Burckel for pointing out how the use of this transform simplifies our proofs.

Lemma 1. *Given a sequence (μ_n) of probability measures on $G = [0, 1)$ and a probability measure μ on G, statements (a)–(d) below are equivalent.*

(a) $\mu_n \to \mu$ *in the weak-$*$ (or vague) topology, i.e.,* $\lim_n \int_G f \, d\mu_n = \int_G f \, d\mu$ *for all continuous functions on G.*[3]

(b) *If $F \subseteq G$ is closed, then* $\limsup_n \mu_n(F) \leq \mu(F)$.

(c) *If $V \subseteq G$ is open, then* $\liminf_n \mu_n(V) \geq \mu(V)$.

(d) $\lim_n \widehat{\mu_n}(m) = \widehat{\mu}(m)$ *for all integers m.*

Also, if (a)–(d) *are true, we have:*

(e) *If $\mu(\{x\}) = 0$ for all x in* $[0, 1)$*, then*

$$\lim_n \mu_n(I) = \mu(I) \text{ for each interval } I \text{ in } G. \tag{7}$$

Proof. The equivalence of (a)–(c) is an elementary fact that holds in all metric spaces; see, for example, [**25**, Theorem 3.1.5].

We have (a) $\Longrightarrow$ (d) because each function χ_m is a continuous function on G. We have (d) $\Longrightarrow$ (a) because the space of linear combinations of the functions χ_m is uniformly dense in the space of continuous functions on G, by an application of the Stone-Weierstrass theorem.

For assertion (e), note that there exist a and b so that $(a, b) \subseteq I \subseteq [a, b]$. Then (7) follows from (b) and (c) via the inequalities:

$$\begin{aligned}\limsup_n \mu_n(I) &\leq \limsup_n \mu_n([a, b]) \leq \mu([a, b]) \\ &= \mu((a, b)) \leq \liminf_n \mu_n((a, b)) \leq \liminf_n \mu_n(I). \qquad \blacksquare\end{aligned}$$

PROOF OF THE UNIFORM DISTRIBUTION THEOREM. Let P_n be the probability $\frac{1}{n}\sum_{k=1}^n \delta_{kx}$ on $G = [0, 1)$, where δ_{kx} is the point mass at kx mod 1. The conclusion of the theorem is equivalent to

$$\lim_n P_n(I) = \lambda(I) \text{ for each interval } I \text{ in } G,$$

where λ is Lebesgue measure on $[0, 1)$. Since $\widehat{\lambda}(0) = \widehat{P_n}(0) = 1$ and $\widehat{\lambda}(m) = 0$ for $m \neq 0$, Lemma 1 shows that it suffices to confirm

$$\lim_n \widehat{P_n}(m) = 0 \text{ for } m \neq 0.$$

Since $\widehat{\delta_{kx}}(m) = \exp(-2\pi i m k x)$, we need to prove that

$$\lim_n \frac{1}{n}\sum_{k=1}^n \alpha^k = 0 \text{ for } \alpha := \exp(-2\pi i m x). \tag{8}$$

But, since mx is not an integer, we have $\alpha \neq 1$ and so this geometric sum is equal to

$$\alpha \frac{\alpha^n - 1}{\alpha - 1},$$

which is clearly bounded in n. Hence (8) holds, completing the proof.

[3]Since 0 and 1 are identified, continuity of f includes the requirement $\lim_{x \to 1^-} f(x) = f(0)$.

INDUCED MEASURES. If ϕ is a measurable function from a measurable space $(X, \mathcal{S})$ to a measurable space $(Y, \mathcal{T})$, then every measure ν on X induces a measure $\nu \circ \phi^{-1}$ on Y defined by

$$\nu \circ \phi^{-1}(E) = \nu(\phi^{-1}(E)) \text{ for measurable sets } E \subset Y. \tag{9}$$

Observe that

$$\int_Y f(y) d(\nu \circ \phi^{-1})(y) = \int_X f(\phi(x)) d\nu(x) \tag{10}$$

for bounded measurable functions f on Y.

PROOF OF THEOREM 2. In this theorem, the r's are selected at random from a fixed interval $[s, u)$ in $[1, 10)$. Let U be the uniform probability on $[s, u)$, so that

$$U(I) := \frac{\text{length}(I)}{u - s} \text{ for each interval } I \subseteq [s, u).$$

Also, the a's are selected, independently of the r's, using any probability measure μ on $[1, 10)$. We again use the function $g : [1, 10) \to [0, 1)$ given by $g(r) = \log_{10} r$. We have $D(ar^n) = d$ if and only if $d \cdot 10^k \leq ar^n < (d + 1)10^k$ for some k and hence if and only if

$$\log_{10} d \leq g(a) + ng(r) < \log_{10}(d + 1);$$

here addition is in the group $[0, 1)$. Thus we have

$$\Pr[D(ar^n) = d] = \int_1^{10} \int_1^{10} \chi_{I_d}(g(a) + ng(r)) d\mu(a) dU(r), \tag{11}$$

where χ_{I_d} is the indicator function of the interval $I_d = [\log_{10} d, \log_{10}(d + 1))$. Our goal is to show that the quantity in (11) converges to $\log_{10}(d + 1) - \log_{10} d = \lambda(I_d)$.

Obviously $dU(r) = h(r)\, dr$ for a bounded measurable function h on $[1, 10)$, i.e.,

$$\int_1^{10} f(r)\, dU(r) = \int_1^{10} f(r) h(r)\, dr \tag{12}$$

for bounded measurable functions f on $[1, 10)$. We mention this here because we will use only (12) in the proof, so that it will be easy to see how to generalize this result. Let P and Q be the induced measures $U \circ g^{-1}$ and $\mu \circ g^{-1}$ on $[0, 1)$. Then from (11) and (10), with $X = [1, 10)$ and $Y = [0, 1)$, we obtain

$$\Pr[D(ar^n) = d] = \int_0^1 \int_0^1 \chi_{I_d}(b + nx)\, dQ(b)\, dP(x).$$

For each n, let P_n be the induced measure $P \circ \phi_n^{-1}$ on $[0, 1)$, where $\phi_n(x) = nx \bmod 1$ for x in $[0, 1)$. Then, using (10) again, with $X = Y = [0, 1)$, we obtain

$$\begin{aligned} \Pr[D(ar^n) = d] &= \int_0^1 \int_0^1 \chi_{I_d}(b + \phi_n(x))\, dP(x)\, dQ(b) \qquad (13) \\ &= \int_0^1 \int_0^1 \chi_{I_d}(b + y)\, dQ(b)\, dP_n(y). \end{aligned}$$

This is exactly the integral with respect to the convolution $Q * P_n$; see [**14**, Chapter I, §7.10] or [**10**, (19.10)]. Hence

$$\Pr[D(ar^n) = d] = \int_0^1 \chi_{I_d}\, d(Q * P_n) = (Q * P_n)(I_d),$$

and we want to show $\lim_n (Q * P_n)(I_d) = \lambda(I_d)$. As in the proof of the uniform distribution theorem, it suffices to show

$$\lim_n \widehat{Q * P_n}(m) = 0 \text{ for } m \neq 0.$$

Since $\widehat{Q * P_n}(m) = \widehat{Q}(m)\widehat{P_n}(m)$ for all m, it suffices to show

$$\lim_n \widehat{P_n}(m) = 0 \text{ for } m \neq 0.$$

Since $\chi_m(\phi_n(x)) = \exp(2\pi i m \phi_n(x)) = \exp(2\pi i m n x) = \chi_{mn}(x)$, we have

$$\widehat{P_n}(m) = \int_0^1 \overline{\chi_m(x)}\, dP_n(x) = \int_0^1 \overline{\chi_m(\phi_n(x))}\, dP(x) = \int_0^1 \overline{\chi_{mn}(x)}\, dP(x) = \widehat{P}(mn).$$

Using (12) and (10), it can be shown that $dP = H\, d\lambda$ for a bounded measurable function H on [0, 1). In other words,

$$\int_0^1 f(x)\, dP(x) = \int_0^1 f(x) H(x)\, dx$$

for bounded measurable functions f on [0, 1). In fact, $H(x) = \log_e(10)\, h(10^x) 10^x$. Therefore, $\widehat{P} = \widehat{H}$ and, since H is integrable, $\lim_{n\to\infty} \widehat{H}(n) = 0$ by the Riemann-Lebesgue lemma; see, e.g., [**14**, Chapter I, Theorem 2.8]. We conclude that for $m \neq 0$:

$$\lim_n \widehat{P_n}(m) = \lim_n \widehat{P}(mn) = \lim_n \widehat{H}(mn) = 0,$$

as desired.

PROOF OF THEOREM 3. We use the following notation from the proof of Theorem 2: g, U, μ, I_d, P, and Q. This time, each r_i is selected at random from [1, 10) using U, and the a's and r_i's are selected independently. Now, all we need about U is

$$\text{the probability } U \text{ is not supported on a finite subset of } [1, 10). \tag{14}$$

We have $D(ar_1 r_2 \cdots r_n) = d$ if and only if

$$\log_{10} d \leq g(a) + g(r_1) + g(r_2) + \cdots + g(r_n) < \log_{10}(d+1).$$

Thus we have

$$\begin{aligned} &\Pr[D(ar_1 r_2 \cdots r_n) = d] \\ &\quad = \int_1^{10}\int_1^{10}\int_1^{10}\cdots\int_1^{10} \chi_{I_d}(g(a) + g(r_1) + g(r_2) + \cdots + g(r_n)) \\ &\qquad\qquad d\mu(a)\, dU(r_1)\, dU(r_2) \cdots dU(r_n), \quad (15) \end{aligned}$$

and we need to show that the quantity in (15) converges to $\log_{10}(d+1) - \log_{10} d$. Our tool will be $Q * P^n$, where P^n is the convolution power $P^n = P * P * \cdots * P$. In this setting, (10) tells us that

$$\int_1^{10} f(g(r))\, d\nu(r) = \int_0^1 f(x)\, d(\nu \circ g^{-1})(x)$$

for measures ν on $[1, 10)$ and for bounded measurable functions f on $[0, 1)$. Using this $n+1$ times, the iterated integral in (15) becomes

$$\int_0^1 \int_0^1 \int_0^1 \cdots \int_0^1 \chi_{I_d}(x + x_1 + x_2 + \cdots + x_n)\, dQ(x)\, dP(x_1)\, dP(x_2) \cdots dP(x_n).$$

This is the integral with respect to the convolution $Q * P^n$, so that

$$\Pr[D(ar_1 r_2 \cdots r_n) = d] = \int_0^1 \chi_{I_d} d(Q * P^n) = (Q * P^n)(I_d),$$

and our goal is reduced to showing that

$$\lim_{n\to\infty} (Q * P^n)(I_d) = \log_{10}(d+1) - \log_{10} d = \lambda(I_d).$$

Assumption (14) implies that $P = U \circ g^{-1}$ is not supported on any finite subset of $[0, 1)$. As in the proof of Theorem 2, it suffices to prove

$$\lim_n \widehat{P^n}(m) = 0 \text{ for all integers } m \neq 0. \tag{16}$$

This will be verified in the proof of the following lemma, which is an easy consequence of Lemma 1.

Lemma 2. *For a probability measure P on $G = [0, 1)$, we have*

$$P^n \to \lambda \text{ in the weak-}* \text{ (or vague) topology}$$

provided that the support of P is not a subset of any coset of a finite subgroup of G.

Proof. As before, it suffices to verify (16). Since $\widehat{P^n}(m) = \widehat{P}(m)^n$ for all positive integers n, it suffices to show that $|\widehat{P}(m)| < 1$ for $m \neq 0$. If $|\widehat{P}(m)| = 1$, then $\int_G \chi_m\, dP = e^{i\theta}$ for some real number θ. This implies that the support of P is a subset of $A = \{x \in G : \chi_m(x) = e^{i\theta}\}$. Given a fixed x_0 in A, it follows that $A - x_0 = \chi_m^{-1}(1)$, which is a closed proper subgroup of G since $\chi_m \neq 1$. This violates the hypothesis regarding the support of P, since all closed proper subgroups of G are finite. ∎

NOTES FOR THE SPECIALIST. Since (14) is the only property of U used in the proof of Theorem 3, Theorem 3 holds so long as the r_i's are selected independently at random using a probability on $[1, 10)$ that is not supported on any finite subset of $[1, 10)$. In particular, Theorem 3 holds if the r_i's are selected from $[1, 10)$ using Benford's probability measure β where $\beta([a, b)) = \log_{10} b - \log_{10} a$ for all $[a, b) \subseteq [1, 10)$. However, in this case, $P := \beta \circ g^{-1}$ and all its convolution powers P^n are equal to λ, so Lemma 2 is not needed.

The equivalence (a) $\Longleftrightarrow$ (d) in Lemma 1 holds for any compact abelian group G, where assertion (d) becomes: $\lim_n \widehat{\mu_n}(\chi) = \widehat{\mu}(\chi)$ for all characters of G.

Using essentially the same proof, Lemma 2 holds for any compact abelian group G, provided the support of P is not a subset of any coset of a *proper closed* subgroup of G. For general compact groups, Lemma 2 holds provided the support of P is not a subset of any coset of a *proper closed normal* subgroup of G, but the proof is more complicated than that of Lemma 2. The result goes back to Yukiyosi Kawada and Kiyosi Itô (1940) and Karl Stromberg (1960) [**24**]. This generality is interesting, even for finite groups, because this theorem implies that repeated selections based on a fixed random process, such as shuffling cards, lead to a well-mixed selection, i.e., have the effect of a uniform selection. In the case of shuffling cards, the underlying group is a large finite nonabelian group of permutations. For much more about card shuffling, see [**1**].

Under very general assumptions, Peter Schatte [**23**, §5] shows that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Pr\left[D(ar_1 r_2 \cdots r_n) = d\right] = \log_{10}(d+1) - \log_{10} d.$$

In particular, for each i, the r_i's can be selected using different probabilities U_i. His conclusion is weaker than ours, but his hypotheses are also weaker.

MORE REFERENCES. At this time, the website Benford Online Bibliography [**3**] has a little over 600 entries regarding Benford's law, mostly articles but some website and video items as well. Here are further comments on some papers published after Raimi's fine summary [**19**]. In [**11**], Theodore P. Hill gives a formulation of the significant-digit problem based on the natural assumption of base invariance, and he proves a theorem about the probabilities of digits in different positions. He also notes, in [**12**], that these probabilities are dependent. For example, the (unconditional) probability that the second digit is 2 is ≈ 0.109, whereas the conditional probability that the second digit is 2, *given* that the first digit is 1, is ≈ 0.115. Other interesting mathematical treatments of Benford's law appear in Raimi [**20**], Cohen and Katz [**5**], and the new book by Rodney Nillsen [**16**]. For a survey of applications of Benford's law to the natural sciences, including observations from the fields of physics, astronomy, geophysics, chemistry, and engineering, see [**22**].

SUGGESTED STUDY. For what sequences does Theorem 1 hold? That is, for what sequences (a_k) do the limits

$$\lim_{n\to\infty} \frac{1}{n} \#\{k \le n : D(a_k) = d\} \tag{17}$$

exist and, if they do, equal $\log_{10}(d+1) - \log_{10} d$ for $d = 1, 2, \ldots, 9$? It is shown in [**21**] that the limit in (17) does not exist for the sequences $(k^N)_{k=1}^{\infty}$ for integers $N \ge 2$. However, an interesting subsequence converges to $(\sqrt[N]{d+1} - \sqrt[N]{d})/(\sqrt[N]{10} - 1)$. Perhaps the problem with $(k^N)_{k=1}^{\infty}$ is that it diverges too slowly to infinity. Consider any $r > 1$, that is not a power of 10, and any integer $N \ge 2$. Experimental evidence suggests that the limits in (17) exist for the sequence (a_k) defined by $a_0 = r$ and $a_{k+1} = a_k^N$ for $k \ge 1$, and equal $\log_{10}(d+1) - \log_{10} d$, for $d = 1, 2, \ldots, 9$. The case $r = N = 2$ is interesting. But, we have no proofs.

ACKNOWLEDGMENTS. It is a pleasure to thank Stan Wagon, Richard Koch, Theodore P. Hill, Robert B. Burckel, and John Clay for their help and interest in this project.

1. D. Bayer and P. Diaconis, Trailing the dovetail shuffle to its lair, *Ann. Appl. Probab.* **2** (1992) 294–313. http://dx.doi.org/10.1214/aoap/1177005705
2. F. Benford, The law of anomalous numbers, *Proc. Amer. Philos. Soc.* **78** (1938) 551–572.
3. A. Berger and T. P. Hill, Benford Online Bibliography, available at http://www.benfordonline.net.
4. ———, Benford's law strikes back: No simple explanation in sight for mathematical gem, *Math. Intelligencer* **33**, no. 1 (2011) 85–91. http://dx.doi.org/10.1007/s00283-010-9182-3
5. D. I. A. Cohen and T. M. Katz, Prime numbers and the first digit phenomenon, *J. Number Theory* **18** (1984) 261–268. http://dx.doi.org/10.1016/0022-314X(84)90061-1
6. P. Diaconis, The distribution of leading digits and uniform distribution mod 1, *Ann. Probab.* **5** (1977) 72–81. http://dx.doi.org/10.1214/aop/1176995891
7. C. Durtschi, W. Hillison, and C. Pacini, The effective use of Benford's law to assist in detecting fraud in accounting data, *J. Forensic Accounting* **1524–5586** (2004) 17–34.
8. W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley, New York, 1966.
9. R. M. Fewster, A simple explanation of Benford's Law, *Amer. Statist.* **63** (2009) 26–32. http://dx.doi.org/10.1198/tast.2009.0005
10. E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis*, vol. I, 2nd ed., Springer-Verlag, Berlin, 1979.
11. T. P. Hill, Base-invariance implies Benford's law, *Proc. Amer. Math. Soc.* **123** (1995) 887–895.
12. ———, The significant-digit phenomenon, *Amer. Math. Monthly* **102** (1995) 322–327. http://dx.doi.org/10.2307/2974952
13. ———, A statistical derivation of the significant-digit law, *Statist. Sci.* **10** (1995) 354–363.
14. Y. Katznelson, *An Introduction to Harmonic Analysis*, 3rd ed., Cambridge University Press, Cambridge, 2004.
15. M. Nigrini, A taxpayer compliance application of Benford's law, *J. Amer. Taxation Assoc.* **18** (1996) 72–91.
16. R. Nillsen, *Randomness and Recurrence in Dynamical Systems: A Real Analysis Approach*, Carus Monograph #31, Mathematical Association of America, Washington, DC, 2010.
17. I. Niven, *Irrational Numbers*, Carus Monograph #11, Mathematical Association of America, Rahway, NJ, 1956.
18. R. A. Raimi, The peculiar distribution of first digits, *Scientific American* **221** (Dec. 1969) 109–120. http://dx.doi.org/10.1038/scientificamerican1269-109
19. ———, The first digit problem, *Amer. Math. Monthly* **83** (1976) 521–538. http://dx.doi.org/10.2307/2319349
20. ———, The first digit phenomenon again, *Proc. Amer. Philos. Soc.* **129** (1985) 211–219.
21. K. A. Ross, First digits of squares and cubes, *Math. Mag.* (to appear).
22. M. Sambridge, H. Tkalčić, and A. Jackson, Benford's law in the natural sciences, *Geophysical Research Letters* **37** (2010) L22301, available at http://www.agu.org/journals/gl/gl1022/2010GL044830/.
23. P. Schatte, On measures of uniformly distributed sequences and Benford's law, *Monatsh. Math.* **107** (1989) 245–256. http://dx.doi.org/10.1007/BF01300347
24. K. Stromberg, Probabilities on a compact group, *Trans. Amer. Math. Soc.* **94** (1960) 295–309.
25. D. W. Stroock, *Probability Theory: An Analytic View*, Cambridge University Press, Cambridge, 1993.

KENNETH A. ROSS holds a B.S., University of Utah (1956), and also a Ph.D., University of Washington (1960). He taught at the University of Rochester, 1961–1964, and the University of Oregon, 1965–2000. In the 1980s and 1990s, he served as MAA Secretary, Associate Secretary, and President. His mathematical interests include abstract commutative harmonic analysis, elementary probability, and expository writing. His books include *Elementary Analysis: The Theory of Calculus*, originally published in 1980 by Springer. His most recent publication was Repeating decimals: A period piece, *Math. Mag.* **83** (2010) 33–45. Ken has been a member of both the AMS and MAA for over fifty years. His sedentary hobbies in retirement include baseball, mathematics, and mentoring his 12-year-old grandson in mathematics.
Department of Mathematics, University of Oregon, Eugene, OR 97403
rossmath@pacinfo.com