Elemente der Mathematik (Online first)
DOI 10.4171/EM/487

Published by EMS Press

Unlikely leading digits of powers of 2

Sungkon Chang

Sungkon Chang is professor of mathematics at the Georgia Southern University. He received a PhD in mathematics at the University of Georgia, U.S.A, and his research areas are elementary number theory and the arithmetic of elliptic curves. He plays the classical guitar in his spare time.

The unit digits of the decimal expansion of 2^n for $n \in \mathbb{N}$ are uniformly distributed over $\{2, 4, 6, 8\}$, so for each $d \in \{2, 4, 6, 8\}$, the probability of the unit digit being d is $1/4 = 25\,\%$. However, the leading digits of 2^n are not uniformly distributed. It turns out that 1 is the most likely digit to occur in the leading digit, and 9 is the least likely digit to occur. In this note, we review the distribution of the leading digits of 2^n, and introduce a method of generating the least likely leading digits of 2^n using the continued fraction of $\log_{10}(2)$. Throughout the paper, the common log function $\log_{10}$ is denoted by $\log(x)$.

1 Benford's Law

In [1, 6], Benford and Newcomb observed that, for many extensive collections of real numbers written in decimal expansion, their leading digits have a certain logarithmic distribution, and it is often referred to as *Benford's Law*. Let us review in this section the logarithmic distribution and examples of sequences of positive integers whose leading digits have the logarithmic distribution.

Das bekannte Gesetz von Newcomb-Benford besagt, dass bei entsprechend verteilten echten Zufallszahlen, aber auch bei vielen Zahlenfolgen, die Ziffer n mit der Wahrscheinlichkeit $\log_{10}(1 + 1/n)$ als führende Ziffer auftaucht. Zahlen mit der Anfangsziffer 1 treten demnach rund sechsmal häufiger auf als Zahlen mit der Anfangsziffer 9. Analoges gilt für Blöcke von führenden Ziffern. Insbesondere ist es am unwahrscheinlichsten, dass eine Zweierpotenz mit einem Block der Form $99\ldots9$ beginnt. Zum Beispiel ist unter den ersten hundert Potenzen von 2 die Potenz 2^{93} die einzige, bei der die ersten beiden führenden Ziffern 99 sind. In der vorliegenden Arbeit wird gezeigt, wie man solche Potenzen mit Hilfe der Kettenbruchentwicklung von $\log_{10} 2$ erhält.

Given positive integers L and $m \geq 10^{L-1}$, let $\ell_L(m)$ denote the first L leading digits of m as an integer in $\{d \in \mathbb{Z} : 10^{L-1} \leq d \leq 10^L - 1\}$, and if m has less than L digits, i.e., $1 \leq m < 10^{L-1}$, define $\ell_L(m) = 0$. For example, $\ell_2(7^7) = \ell_2(823543) = 82 \in \mathbb{Z}$ and $\ell_7(7^7) = 0$. Let $\{A_n\}$ be a sequence of positive integers. Given positive integers L and $d \in \{10^{L-1}, \dots, 10^L - 1\}$, let $P(L,d)(n)$ denote the quotient $\frac{1}{n}\#\{k \leq n : \ell_L(A_k) = d\}$, and let $P(L,d)_\infty$ denote the limit of $P(L,d)(n)$ as $n \to \infty$ if it exists, which is a natural way of defining the probability of the first L leading digits of A_n being d. The sequence $\{A_n\}$ is called a *Benford sequence* if $P(1,d)_\infty = \log(1 + \frac{1}{d})$ for each $d \in \{1, \dots, 9\}$, and indeed, the nine values of $\log(1 + \frac{1}{d})$ do add up to 1. The sequence given by $A_n = n!$ for each $n \geq 1$ is a Benford sequence, but the sequences $A_n = n^a$ for any positive integer a are not Benford sequences; see [4, 5].

One of the simplest examples of Benford sequences is $A_n = 2^n$ for $n \geq 1$, and let us demonstrate it below. Given a positive integer $d \in \{1, 2, \dots, 9\}$, if $\ell_1(2^n) = d$ and N denotes the number of digits in the expansion of 2^n, then $d10^{N-1} \leq 2^n < (d+1)10^{N-1}$. Thus,

$$\begin{aligned} &\log(d) + (N-1) \leq n\log(2) < \log(d+1) + (N-1) \\ &\quad \Longrightarrow\ 0 \leq \log(d) \leq n\log(2) - (N-1) < \log(d+1) \leq 1. \end{aligned}$$

Thus, $n\log(2) - (N-1)$ is the fractional part of $n\log(2)$, and it is known as *the equidistribution theorem* [8, proof of Weyl's Theorem, pp. 105–113] that the irrationality of $\log(2)$ implies that the fractional part $\{n\log(2)\}$ of $n\log(2)$ is uniformly distributed in the interval $[0, 1)$. Therefore, the probability of $n\log(2) - (N-1)$ falling in the sub-interval $[\log(d), \log(d+1))$ is equal to the length of the interval, which is $\log(d+1) - \log(d) = \log(1 + \frac{1}{d})$. Nothing but the irrationality of $\log(2)$ was used for calculating the probability, and the argument easily extends to the proof of the fact that the sequence $A_n = b^n$ for $n \geq 1$ is a Benford sequence if b is not a power of 10 since $\log(b)$ is irrational for such values of b.

Using the above principle on the fractional parts of the integer multiples of an irrational number, we find that the probability $P(2,d)_\infty$ is equal to $\log(1 + \frac{1}{d})$, where $d \in \{10, 11, \dots, 99\}$, and in general, we have

$$P(L,d)_\infty = \log\left(1 + \frac{1}{d}\right), \tag{1}$$

where $d \in \{10^{L-1}, \dots, 10^L - 1\}$; see [4]. If a sequence $\{A_n\}$ has the distribution (1) for each L and d, it is called a *strong Benford sequence*, and hence, the sequence given by $A_n = b^n$ for $n \geq 1$ is a strong Benford sequence. For simplicity, we focus on the example $A_n = 2^n$ for $n \geq 1$ for the remainder of this work.

According to the probability (1), the digit of 9 is the most unlikely leading digit of 2^n with probability $\log(10/9) \approx 4.6\,\%$. For the first two leading digits of 2^n, the block 99 is the most unlikely one with probability $\log(100/99) \approx 0.4\,\%$, and in general, the block of L digits $99\dots9$ is the most unlikely one with probability $\log(10^L/(10^L - 1)) \approx 10^{-L}\log(e) \approx 0.43 \times 10^{-L}$ for larger values of L. We asked ourselves if there is a method of finding a positive integer n for which the first block of L leading digits of 2^n is the most unlikely one; of course, we want a method that is more effective than trying all consecutive powers of 2. In this note, we present an answer using the continued fraction of $\log(2)$.

2 Continued fractions

Diophantine approximation [3] is a subject in number theory that concerns the question of how to effectively approximate an irrational number β using a sequence of distinct rational numbers p_k/q_k in lowest terms, and the effectiveness is typically measured by a positive real number s independent of k such that

$$\left|\beta - \frac{p_k}{q_k}\right| < \frac{C}{q_k^s}$$

for all $k \geq 1$, where C is a positive constant independent of k. It is the celebrated result of Roth [7] that if β is an irrational *algebraic number*, i.e., an irrational zero of a polynomial with integer coefficients, then $s \leq 2$. Thus, $s = 2$ is the best effectiveness we can hope for an irrational algebraic number. For all irrational real numbers β, there is a standard method of constructing a sequence of distinct rational numbers with $s = 2$ and $C = 1$. It is called *the continued fraction of* β, and let us review the construction and the theory below, which are available in the standard textbooks of elementary number theory such as [2].

Given a positive real number β_0, let a_0 be the integer part of β_0, and let β_1 be the reciprocal of the fractional part of β_0, i.e., $\beta_1 = 1/(\beta_0 - a_0)$ if $\beta_0 - a_0 \neq 0$. If $\beta_0 - a_0 = 0$, define $\beta_1 = 0$. Recursively define a_k to be the integer part of β_k, and β_{k+1} to be the reciprocal of the fractional part of β_k, i.e., $\beta_{k+1} = 1/(\beta_k - a_k)$ for all $k \geq 0$ if $\beta_k - a_k \neq 0$; if $\beta_k - a_k = 0$, define $\beta_{k+1} = 0$. For example, if $\beta_0 = 55/89$, then $(a_0, a_1, \ldots, a_9) = (0, 1, 1, \ldots, 1, 2)$, and $a_k = 0$ for all $k \geq 10$. If we unfold the reciprocals of the fractional parts, we can write $55/89$ as in (2), and the integers a_k are visible in that expansion. The integers a_k for $k \geq 1$ are called the *partial denominators* of this fraction:

$$\frac{55}{89} = \cfrac{1}{1 + \cfrac{1}{\cdots + \cfrac{1}{1 + \cfrac{1}{2}}}}, \quad \frac{1+\sqrt{5}}{2} = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}},$$
$$\log(2) = \cfrac{1}{3 + \cfrac{1}{3 + \cfrac{1}{9 + \cfrac{1}{2 + \cdots}}}}. \tag{2}$$

In general, β_0 is rational if and only if the sequence $\{a_k\}$ terminates with 0, and $\{a_k\}$ is periodic with non-zero period if and only if β_0 is a zero of an irreducible quadratic polynomial with integer coefficients; see the example of the golden ratio in (2). Thus, for transcendental numbers such as $\log(2)$, the sequence $\{a_k\}$ is not periodic; see the expansion of $\log(2)$ in (2).

Let β_0 be a real number, let a_0 be the integer part of β_0, and let a_k for $k \geq 1$ be the partial denominators of β_0. Given a positive integer n, the rational number p_n/q_n in lowest terms is called the *nth convergent* of β_0 if the integer part b_0 and the partial denominators b_k of p_n/q_n coincide with a_k for $k = 0, \dots, n$, and $b_k = 0$ for all $k > n$. For example, the second convergent of log(2) is the fraction $\frac{1}{3+1/3} = \frac{3}{10}$, and the third convergent is $\frac{28}{93}$. If k is odd, the kth convergent is called *an odd convergent* of β_0.

The following is well known in the theory of continued fractions [2, Chapter 13.4], and according to Roth's result [7], it shows that the continued fraction achieves the best approximation in general.

Theorem 1. *If p_n/q_n is the nth convergent to an irrational number β_0, then*

$$\left|\beta_0 - \frac{p_n}{q_n}\right| < \frac{1}{q_{n+1}q_n} < \frac{1}{q_n^2}.$$

Moreover, if n is odd, then $0 < \frac{p_n}{q_n} - \beta_0 < \frac{1}{q_{n+1}q_n} < \frac{1}{q_n^2}$.

For example, if $\beta_0 = \log(2)$, then the third convergent is $\frac{28}{93}$, and $\frac{28}{93} - \log(2) \approx 0.000045$, which is less than $1/93^2 \approx 0.00012$. Just using a two-digit denominator, we could obtain the accuracy up to the 4th decimal place!

3 The unlikely leading digits

Our task is to find an effective method of finding a positive integer n for which the first block of L leading digits of 2^n is $99\dots9$, and the following theorem is one answer.

Theorem 2. *Let $n \ge 3$ be the denominator of an odd convergent of the continued fraction of* log(2)*, and let L be the smallest non-negative integer* $> \log(n/3) - 1$. *Then the first L leading digits of 2^n are all* 9.

For example, the rational number $\frac{1838395}{6107016}$ is the 13th convergent of log(2), and

$$\log\frac{6107016}{3} - 1 \approx 5.3 \implies L = 6.$$

The computer calculation shows that $2^{6107016} = 9999996\dots36$, and it verifies the theorem. The first convergent of log(2) is $1/3$, so $n = 3$ is the smallest value that applies for the theorem. If $n = 3$, then $L = 0$, and the theorem is vacuously true. The third convergent is $28/93$. So $L = 1 > \log(93/3) - 1 \approx 0.49$, and $2^{93} = 9903\dots92$, which verifies the theorem.

Proof of Theorem 2. Given a positive integer n, the real number $n\log(2)$ is not an integer, and hence, there is a unique integer M such that $M - 1 < n\log(2) < M$, i.e., $10^{M-1} < 2^n < 10^M$. Since 10^M is the smallest positive integer with $M + 1$ digits, the inequality implies that 2^n has M digits, and hence, $M = \lceil n\log(2)\rceil$ is the number of digits of 2^n. Let $N_n := \lceil n\log(2)\rceil$ and $\beta_n := N_n - n\log(2)$.

Lemma 3. *Let n be the positive integer defined in Theorem 2. Then $\beta_n < 1/n$.*

Proof. Let m/n be the odd convergent defined in Theorem 2. Then, by Theorem 1,

$$0 < \frac{m}{n} - \log(2) < \frac{1}{n^2} \implies 0 < m - n\log(2) < \frac{1}{n} < 1.$$

If $\gamma := m - n\log(2)$, then $m = n\log(2) + \gamma$, where $0 < \gamma < 1$, and hence, we find $m = \lceil n\log(2) \rceil$. Thus, $\beta_n = m - n\log(2) < 1/n$. ∎

Let n be an integer defined in Theorem 2, and let r be the length of the maximal block of repeated digits of 9 in the leading digits of 2^n, e.g., $r = 2$ if $n = 93$ since $2^{93} = 990\ldots792$. Suppose that $r \le \log(n/3) - 1$; since $r \ge 0$, this can happen only when $n \ge 30$. Notice that $N_n > n\log(2) > r$ since $N_n = r$ implies $2^n = 99\ldots99$, whose RHS is an odd integer. Then

$$\begin{aligned} 2^n &< 9\sum_{k=N_n-r}^{N_n-1} 10^k + 9\cdot 10^{N_n-r-1} = 10^{N_n}\left(1 - \frac{1}{10^{r+1}}\right) \\ &< 10^{n\log(2)+\beta_n}\left(1 - \frac{1}{10^{r+1}}\right) \le 2^n 10^{\beta_n}\left(1 - \frac{1}{10^{\log(n/3)}}\right) \\ &\implies 1 < 10^{\beta_n}\left(1 - \frac{1}{10^{\log(n/3)}}\right) = 10^{\beta_n}\left(1 - \frac{1}{n/3}\right) < 10^{1/n}\left(1 - \frac{1}{n/3}\right). \end{aligned} \tag{3}$$

The function $f(x) = 10^{1/x}(1 - 3/x)$ is increasing for $x > 0$, and

$$\begin{aligned} f(x) &= \left(10\left(1 - \frac{1}{x/3}\right)^x\right)^{1/x} \\ &= \left(10\left(1 - \frac{1}{x/3}\right)^{-x/3\cdot(-3)}\right)^{1/x} \to (10e^{-3})^0 = 1 \quad \text{as } x \to \infty. \end{aligned}$$

Thus, $10^{1/n}(1 - \frac{1}{n/3}) < 1$, but this contradicts the above inequality (3). Therefore, we have $r > \log(n/3) - 1$. Since $r \ge L$, we prove Theorem 2. ∎

References

[1] F. Benford, The law of anomalous numbers. *Proc. Am. Philos. Soc.* **78** (1938), 551–572

[2] D. M. Burton, *Elementary number theory*. Allyn and Bacon, Inc., Boston, Mass.-London, 1980

[3] J. W. S. Cassels, *An introduction to Diophantine approximation*. Cambridge Tracts in Math. Math. Phys. 45, Cambridge University Press, New York, 1957

[4] P. Diaconis, The distribution of leading digits and uniform distribution mod 1. *Ann. Probability* **5** (1977), no. 1, 72–81

[5] W. Hürlimann, Integer powers and Benford's law. *Int. J. Pure Appl. Math.* **11** (2004), no. 1, 39–46

[6] S. Newcomb, Note on the frequency of use of the different digits in natural numbers. *Amer. J. Math.* **4** (1881), no. 1–4, 39–40

[7] K. F. Roth, Rational approximations to algebraic numbers. *Mathematika* **2** (1955), 1–20; corrigendum, 168

[8] E. M. Stein and R. Shakarchi, *Fourier analysis*. Princeton Lectures in Anal. 1, Princeton University Press, Princeton, 2003

Sungkon Chang
Georgia Southern University
11935 Abercorn St
Savannah GA, 31419, U.S.A.
schang@georgiasouthern.edu