# New Structural Properties of Strings Generated by Leading Digits of $2^N$*

A. K. Rajagopal†

*Department of Physics and Astronomy*
*Louisiana State University*
*Baton Rouge, Louisiana 70803*

V. R. R. Uppuluri

*Mathematics and Statistics Research Department*
*Computer Science Division*
*Union Carbide Corporation*
*Nuclear Division*
*Oak Ridge, Tennessee 37830*

David S. Scott

*Computer Science Department*
*University of Texas at Austin*
*Austin, Texas 78712*

S. Sitharama Iyengar

*Department of Computer Science*
*Louisiana State University*
*Baton Rouge, Louisiana 70803*

and

Mohan Yellayi

*Department of Electrical Engineering*
*Louisiana State University*
*Baton Rouge, Louisiana 70803*

Transmitted by Robert E. Kalaba

---

*Research sponsored by Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission under Interagency Agreement 40-550-75 with the U.S. Department of Energy under Contract W-7405-eng-26 with Union Carbide Corporation.




52 Vanderbilt Ave., New York, NY 10017 0096-3003/84/$03.00

ABSTRACT

Several new properties of the first decimal digits of the powers of 2 are presented. It is shown that the consecutive members of the sequence of the first digits of powers of 2 appear in one of five strings. Some statistical properties such as probabilities of occurrence of the strings, their transition probabilities, etc. are presented. The associated state transition graphs are also displayed. It is found that the process of generating strings follows a non-Markovian process but is ergodic. The relation of these properties to ergodic theory is mentioned, and possible applications to information theory, computer science, and statistical mechanics are briefly stated.

## I. INTRODUCTION

It has been known for over a century that if an extensive collection of numerical data expressed in decimal form is arranged according to the first digit, without consideration of the position of the decimal point, the resulting nine digits do not occur in equal proportion. In fact, they obey a logarithmic law. Newcomb [1] was the first to draw attention to this problem in 1881, and Frank Benford [2] supplied a lot of data in support of the law, collected from a wide variety of sources, such as telephone directories, lists of physical constants, electricity consumption records, etc. Since then a large number of investigators from a variety of different disciplines—mathematicians, economists, amateurs, etc.—have examined this phenomenon. A popular account by Ralph A. Raimi [3] appeared in Scientific American in 1969, and a technical review was written by him [4] just five years ago. The significance of this problem for information processing has recently been considered by Hamming [5], Tsao [6], and Knuth [7]. Hamming argued that the distribution of the fractional parts of floating-point numbers in arithmetic operations obey a reciprocal distribution and suggested that this could be applied to hardware, software, and general computing. Knuth drew attention to the fact that the most obvious techniques of average error estimation for floating point calculations are invalid in view of the logarithmic distribution of the first digits. Tsao has suggested its use in determining the roundoff errors in floating point considerations. The essence of the argument of Hamming and others lies in the earlier observation by Pinkham [8] that the reciprocal distribution mentioned above is the only scale invariant distribution. Also Pinkham showed that the logarithmic law of Newcomb and Benford is a consequence of the equidistribution property modulo unity of arithmetic sequences with irrational spacings in the interval $[0,1)$, and the scale invariance assumption on the underlying probability distribution function.

In mathematics, the theory of asymptotic distributions modulo one introduced by Hermann Weyl has been studied extensively; it has its roots in Kronecker's investigations of the behavior of fractional parts of linear forms with integer variables. It gave rise to one of the first ergodic theorems. This is related to Bohl's work on secular perturbations, Sierpinski's on irrational numbers, Borel and Bernstein's on probability, and Hardy and Littlewood's on diophantine approximations and Fourier series. A review of these may be found in Koksma [9]. These investigations were basic to the ergodic problems of classical mechanics, as has been recently expounded by Arnold and Avez [10, 11]. The application of these techniques to the first digit problem, in particular the derivation of the Newcomb-Benford law of the first digits of powers of 2, may be found in Reference [11]. The use of these ideas in statistical mechanics has recently been explained by Pomeau [12].

It should be clear from this brief account that the first digit problem is a significant one with applications to statistical mechanics, information theory, and computer science. In the present paper, we shall give an entirely new set of results concerning the first digits of powers of two. This is a prototype problem of a general class which will be mentioned at the end of the paper. It is observed that the first digits $1,2,\ldots,9$ occur in groups or strings (1248), (1249), (125), (136), and (137), and these strings appear in an irregular fashion. Calling the groups $a,b,c,d,e$ respectively, we find that if string $a$ occurred, what follows is either string $d$ or $e$ but not string $a$, $b$, $c$, etc., leading to an interesting tree structure. It will be shown explicitly that the process is non-Markovian. It was soon realized that the fractions generated by $x_n = n\log_{10}2 \pmod 1$ are uniformly distributed on an interval $[0,1)$, from which we can deduce not only the Newcomb-Benford law for the occurrence of integers in our case, but also the probabilities of occurrence of the strings $a$, $b$, $c$, $d$, and $e$. We also deduce, by a geometrical construction, the transition probabilities for these processes. Since we have here generated pseudo-"random" sequences, application to Monte Carlo methods, coding theory, and cryptography may perhaps be explored, in much the same way as the decimal sequences of rational numbers have recently been explored for such applications by S. C. Kak and A. Chatterjee [13].

In the next section we present the results concerning the probability distributions of the integers and the strings, of which the first (as already mentioned) is well known, but the second is new, to the best of our knowledge. In Section III, we give the transition probabilities for both the integers and the strings, both of which are new. In the Section IV we explicitly show that the process is non-Markovian. In the final section, a summary of our results and extensions of these ideas will be given.

## II. PROBABILITY DISTRIBUTIONS OF THE INTEGERS AND THE STRINGS

The powers of two may be generated by the linear recursion $x_{n+1} = 2x_n$ with $x_0 = 1$. (This form of generating the powers of 2 will prove useful in generalizing the ideas later on). A computer program in FORTRAN on an IBM 3033 system for listing the first digits of $2^n$ was written, and the results were compiled for $n = 0$ to 1999, and in fact to even larger numbers (we went up to 4000). The program was also designed to arrange the strings as they occurred, as well as to pick off the frequency of occurrences of the digits $1,\ldots,9$ in groups of size ranging from 99 to 103 (i.e., $n = 0$–99, 100–199, 200–298, etc.), as well as the strings $a,\ldots,e$ in groups of size varying between 29 and 31 from the *same* set of numbers. The sizes of the string groups varied a little because we wanted the groups of strings picked from the same sets as the integers, but it is not crucial for the discussion to follow. The computer programs were also designed to pick other groups of groups and their frequencies of occurrences, etc. A paper on this aspect of the work is already submitted for publication [14]. Table 1 gives the actual numbers generated by $2^n$ for $n = 0$ to 100 and their decimal representation, namely $p.xyz\ldots \times 10^k$; for $n$ up to 50, such a set of numbers is given in NBS tables. In Table 2 we give the string sequences generated from the first digits of $2^n$ for $n = 0$ to 4000. It is interesting to note that the first digits of $2^{-n}$, written in the form $q.abc\ldots \times 10^{-k}$, generate a similar pattern of string sequences.

Several interesting observations are in order at this stage, which though evident, do not seem to have been noted before in the vast literature on the subject. In Figure 1 we display graphically how the first digits move around as we multiply by 2 successively. For example, the first digit 1 will go to 2 if the digit following it is less than 5 and will go to 3 if the digit following it is greater than or equal to 5, and so on. The structure of the graph shown in Fig. 1 is thus evident. In Fig. 2, we represent a similar movement of the strings. To understand this, we shall give here an intuitive geometric construction. Before we go into this, let us mention that in Fig. 3 we represent the nine digits $1,\ldots,9$ on a line segment on which points numbered 1 to 10 which are equally spaced and the segments are numbered sequentially from 1 to 9. Also shown in this figure is the mapping of the line segment $(1,10)$ to $(0,1)$ by a logarithmic transformation, to which we shall return presently. In Fig. 4, we construct a line segment from 1 to 2 with segments marked 1.125, 1.250, 1.5, and 1.750, and the names of strings are shown. To understand this construction, we make the following observation. We first pick an integer $n_0$ such that

$$2^{n_0} = 1.xyz\ldots \times 10^{k_0}, \tag{1}$$

TABLE 1

DECIMAL NUMBERS GENERATED BY $2^N$ FOR $n = 0$ TO 100

| | | |
|---|---|---|
| 0 | 0.1000E01 | 1 |
| 1 | 0.200E+01 | 2 |
| 2 | 0.400E+01 | 4 |
| 3 | 0.800E+01 | 8 |
| 4 | 0.160E+02 | 16 |
| 5 | 0.320E+02 | 32 |
| 6 | 0.640E+02 | 64 |
| 7 | 0.128E+03 | 128 |
| 8 | 0.256E+03 | 256 |
| 9 | 0.512E+03 | 512 |
| 10 | 0.102E+04 | 1024 |
| 11 | 0.205E+04 | 2048 |
| 12 | 0.410E+04 | 4096 |
| 13 | 0.819E+04 | 8192 |
| 14 | 0.164E+05 | 16384 |
| 15 | 0.328E+05 | 32768 |
| 16 | 0.655E+05 | 65536 |
| 17 | 0.131E+06 | 131072 |
| 18 | 0.262E+06 | 262144 |
| 19 | 0.524E+06 | 524288 |
| 20 | 0.105E+07 | 1048576 |
| 21 | 0.210E+07 | 2097152 |
| 22 | 0.419E+07 | 4194304 |
| 23 | 0.839E+07 | 8388608 |
| 24 | 0.168E+08 | 16777216 |
| 25 | 0.336E+08 | 33554432 |
| 26 | 0.671E+08 | 67108864 |
| 27 | 0.134E+09 | 134217728 |
| 28 | 0.268E+09 | 268435456 |
| 29 | 0.537E+09 | 536870912 |
| 30 | 0.107E+10 | 1073741824 |
| 31 | 0.215E+10 | 2147483648 |
| 32 | 0.429E+10 | 4294967296 |
| 33 | 0.859E+10 | 8589934592 |
| 34 | 0.172E+11 | 17179869184 |
| 35 | 0.344E+11 | 34359738368 |
| 36 | 0.687E+11 | 68719476736 |
| 37 | 0.137E+12 | 137438953472 |
| 38 | 0.275E+12 | 274877906944 |
| 39 | 0.550E+12 | 549755813888 |
| 40 | 0.110E+13 | 1099511627776 |
| 41 | 0.220E+13 | 2199 23255552 |
| 42 | 0.440E+13 | 4398 46511104 |

TABLE 1 (*Continued*)

| | | |
|---|---|---|
| 43 | 0.880E+13 | 8796 93022208 |
| 44 | 0.176E+14 | 17592186044416 |
| 45 | 0.352E+14 | 35184372088832 |
| 46 | 0.704E+14 | 70368744177664 |
| 47 | 0.141E+15 | 140737488355328 |
| 48 | 0.281E+15 | 281474976710656 |
| 49 | 0.563E+15 | 562949953421312 |
| 50 | 0.113E+16 | 1125899906842624 |
| 51 | 0.225E+16 | 2251799813685248 |
| 52 | 0.450E+16 | 4503599627370496 |
| 53 | 0.901E+16 | 9007199254740992 |
| 54 | 0.180E+17 | 18014398509481984 |
| 55 | 0.360E+17 | 36028797 18963968 |
| 56 | 0.721E+17 | 72057594 37927936 |
| 57 | 0.133E+18 | 144115188 75855872 |
| 58 | 0.288E+18 | 288230376151711744 |
| 59 | 0.576E+18 | 576460752303423488 |
| 60 | 0.115E+19 | 1152921504606846976 |
| 61 | 0.231E+19 | 2305843009213693952 |
| 62 | 0.461E+19 | 4611686018427387904 |
| 63 | 0.922E+19 | 9223372036854775808 |
| 64 | 0.184E+20 | 18446744073709551616 |
| 65 | 0.369E+20 | 36893488147419103232 |
| 66 | 0.738E+20 | 73786976294838206464 |
| 67 | 0.148E+21 | 147573952589676412928 |
| 68 | 0.295E+21 | 295147905179352825856 |
| 69 | 0.590E+21 | 590295810358705651712 |
| 70 | 0.118E+22 | 1180591620717411303424 |
| 71 | 0.236E+22 | 2361183241434822606848 |
| 72 | 0.472E+22 | 4722366482869645213696 |
| 73 | 0.944E+22 | 9444732965739290427392 |
| 74 | 0.189E+23 | 18889465931478580854784 |
| 75 | 0.378E+23 | 37778931862957161709568 |
| 76 | 0.756E+23 | 75557863725914323419136 |
| 77 | 0.151E+24 | 151115727451828646838272 |
| 78 | 0.302E+24 | 302231454903657293676544 |
| 79 | 0.604E+24 | 604462909807314587353088 |
| 80 | 0.121E+25 | 1208925819614629174706176 |
| 81 | 0.242E+25 | 2417851639229258349412352 |
| 82 | 0.484E+25 | 4835703278458516698824704 |
| 83 | 0.967E+25 | 9571406556917033397649408 |
| 84 | 0.193E+26 | 19342813113834066795298816 |

TABLE 1 (*Continued*)

| | | |
|---|---|---|
| 85 | 0.387E+26 | 38685626227668133590597632 |
| 86 | 0.774E+26 | 77371252455336267181195264 |
| 87 | 0.155E+27 | 154742504910672534362390528 |
| 88 | 0.309E+27 | 309485009821345068724781056 |
| 89 | 0.619E+27 | 618970019642690137449562112 |
| 90 | 0.124E+28 | 1237940039285380274899 1 |
| 91 | 0.248E+28 | 2475880078570760549788248448 |
| 92 | 0.495E+28 | 4951760157141521099596496896 |
| 93 | 0.990E+28 | 9903520314283042199192993792 |
| 94 | 0.198E+29 | 19807040628566084398385987584 |
| 95 | 0.396E+29 | 39614081257132168796771975168 |
| 96 | 0.792E+29 | 79228162514264337593543950336 |
| 97 | 0.158E+30 | 158456325028528675187 87900672 |
| 98 | 0.317E+30 | 316912650057 573503741758013 44 |
| 99 | 0.634E+30 | 633825300114114700748351602688 |
| 100 | 0.127E+31 | 1267650600228229401496703205376 |

where $k_0$ is an integer. Thus $1.xyz\ldots$ stands for the number $2^{n_0}$ expressed in decimal notation, and the first digit associated with $2^{n_0}$ is 1. Upon multiplying this by 2, we generate either 2 or 3 according as $x < 5$ or $x \geqslant 5$. If we begin with 1, then, on multiplying by 2 successively we obtain the string $(1, 2, 4, 8)$ as long as $1.xyz\ldots$ is in the interval $[1, 1.125)$. Thus these decimal members in the interval $[1, 1.125)$ generate the string $a$. Similarly the segment $[1.125, 1.25)$ generates the string $b$, and so on. To establish the pathways given in Fig. 2, let us take the case of the string $a$. The last digit in this string is of the form $8.pqr\ldots \times 10^k$, say. Upon multiplication by 2, this goes over to $1.p'q'r'\ldots \times 10^{k+1}$, and we need to check where $1.p'q'r'$ is in Fig. 4. Remembering that 8 maps to 1.6 in this representation, we see that it lands on the interval $d$ in Fig. 4. Now if $p < 5$ in $8.pqr\ldots$, $a$ lands on $d$ until, when $p$ becomes 9, $q$ becomes 9, etc. it goes to 1.8 at this outer limit. This lies in the interval marked $e$. Thus we observe that $a$ goes either to $d$ or $e$. By a similar analysis we can easily construct the graph associated with the movements of the strings $a, \ldots, e$ shown in Fig. 2.

We now give the results concerning the probabilities of occurrence of the digits and the strings. In Table 3 we give the frequency distribution of the first digits in $2^n$ and in Table 4 their probabilities of occurrence. We state and prove the asymptotic results in the form of theorems, the first of which is known, as stated before.

TABLE 2

STRING SEQUENCES GENERATED FROM FIRST DIGITS OF $2^n$ FOR $n = 0$ TO 4000

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 |
| 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 |
| 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 |
| 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 |
| 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 |
| 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1248 | 137 | 125 | 1249 |
| 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 |
| 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 |
| 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 |
| 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 |
| 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 |
| 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 |
| 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 |
| 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 |
| 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 |
| 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 137 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 |
| 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 |
| 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 |
| 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1248 |
| 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 |
| 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 |
| 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 |
| 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 |
| 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1248 | 137 | 125 | 1249 | 137 |
| 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 |
| 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 |

TABLE 2 (*Continued*)

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 |
| 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 |
| 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 |
| 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 |
| 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 |
| 136 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 137 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 |
| 137 | 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 |
| 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 |
| 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 |
| 136 | 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 136 | 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 |
| 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 |
| 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 |
| 1249 | 137 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 |
| 125 | 1248 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 |
| 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 |
| 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 125 |
| 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 136 | 125 | 1248 | 137 | 125 |
| 1249 | 137 | 125 | 1249 | 137 | 125 | 1249 | 137 | 136 | 1249 | 137 | 136 | 1249 | 137 | 136 |
| 125 | 1248 | 136 | 125 | | | | | | | | | | | |

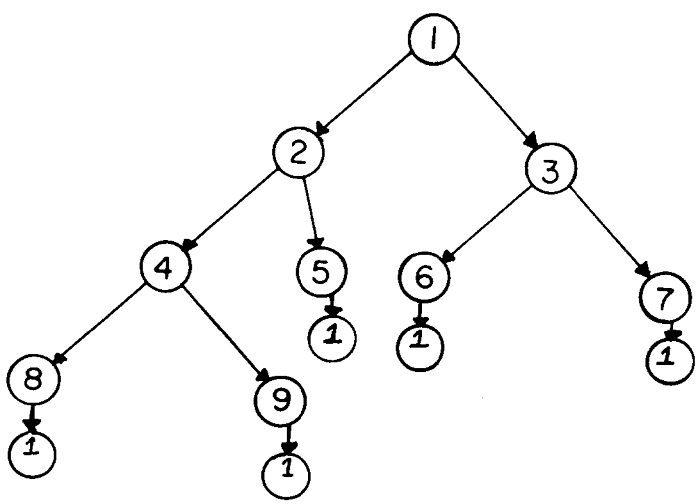



FIG. 1. Formation of binary tree structure generated by leading digits of $2^N$.

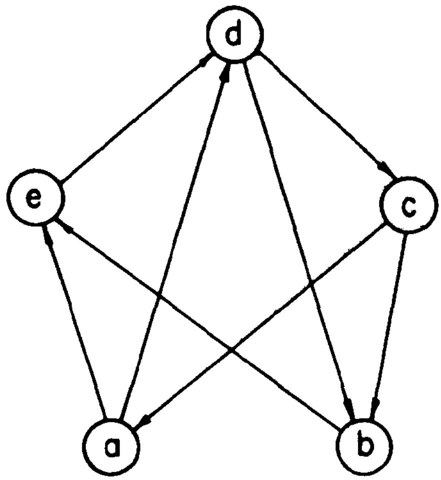



FIG. 2. State transition graph of string sequences.

THEOREM I. *The asymptotic probabilities of the integers $k$, denoted by $P(k)$, $k=1,\ldots,9$, are given by*

$$P(k)=\log_{10}\left(1+\frac{1}{k}\right). \tag{2}$$

For a discussion of this, see Reference [1], [2], [3], [4], [8], and [11]. Our derivation is similar to the one given in Reference [11] and depends on the fact that the fractions defined by

$$x_n = n\log_{10}2 \ (\mathrm{mod}\ 1) \tag{3}$$

are uniformly densely distributed on a line segment between 0 and 1. Our proof shows a method of deducing other asymptotic properties such as transition probabilities, etc. From Fig. 3, we see that the line segments marked 1–9 map into the line segments shown at the bottom, whose union is the segment $[0,1)$. Since the digits $1,\ldots,9$ are generated by the first digits of $2^n$ for a very large collection of $n$, one collects these integers into bins as in Table 3. To obtain the probability, one finds the fraction of times a given integer occurred relative to the total number of integers sampled. In other

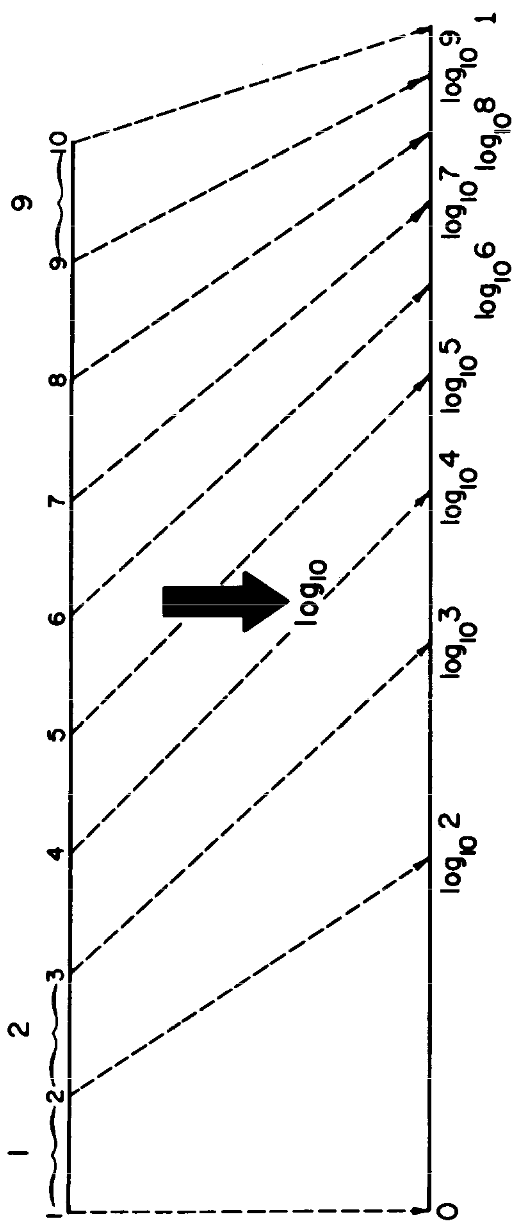



FIG. 3. Mapping of line segment (1, 10) to (0, 1) by logaritonic transformation.

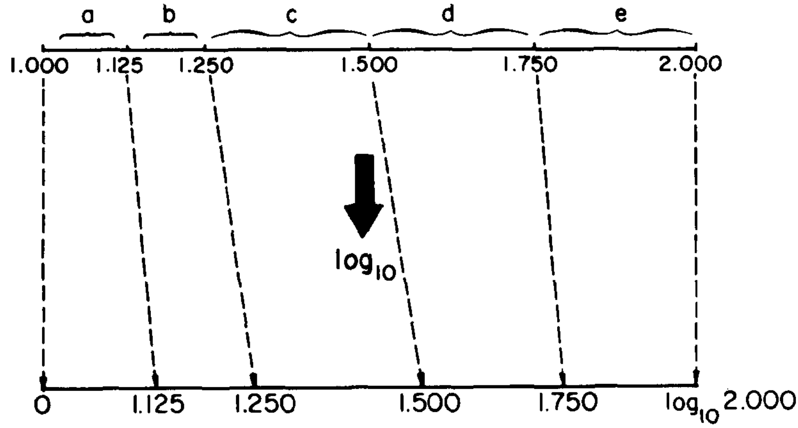



FIG. 4. Construction of a line segment from 1 to 2 with segments marked 1.125, 1.250, 1.5, and 1.750 (the names of the corresponding strings are shown).

words, the sizes of the segments in the interval $[0,1)$ give us the frequency of occurrence of the corresponding integer. For example, the segment 1 maps to $[0, \log_{10} 2)$, and the asymptotic probability of occurrence of 1 is thus $\log_{10} 2$ (total length of segment is 1). Similarly, segment 2 maps to $[\log_{10} 2, \log_{10} 3)$, and so $P(2) = \log_{10} 3 - \log_{10} 2$, whence

$$P(k) = \log_{10}(k+1) - \log_{10} k = \log_{10}\left(1 + \frac{1}{k}\right),$$

as stated in the theorem.

In Table 4 we have listed these asymptotic probabilities also. In Fig. 5, we have displayed the numerical results as points and the asymptotic answers in the form of a curve.

We noted earlier that the strings $a, b, \ldots, e$ also occur in a random fashion. These strings have the following unique markers. $a$ is associated with 8, $b$ with 9, $c$ with 5, $d$ with 6, and $e$ with 7. We have given in Table 5 the frequency of occurrence of these string sequences, and in Table 6 the corresponding probabilities of occurrence, for the same set of 2000 numbers generated in Table 3. As explained earlier, these strings too can be generated by points distributed uniformly on a line segment, as in Fig. 4. In the limit of large $n$ then we can obtain the asymptotic probabilities $\pi(\alpha)$ of occurrence of the

TABLE 3
FREQUENCY DISTRIBUTION OF FIRST DIGITS 1 THROUGH 9 IN $2^n$ FOR $n$ UP TO 1000

| Period of length $P_k$ | Occurrence of the first digit 1 through 9 in $2^n$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total |
| 0–99 | 30 | 17 | 13 | 10 | 7 | 7 | 6 | 5 | 5 | 100 |
| 100–199 | 30 | 19 | 11 | 10 | 9 | 6 | 5 | 6 | 4 | 100 |
| 200–298 | 30 | 17 | 13 | 9 | 8 | 7 | 6 | 4 | 5 | 99 |
| 299–398 | 30 | 17 | 13 | 10 | 7 | 8 | 5 | 6 | 4 | 100 |
| 399–498 | 30 | 18 | 12 | 10 | 8 | 6 | 6 | 5 | 5 | 100 |
| 499–597 | 30 | 17 | 13 | 9 | 8 | 7 | 6 | 5 | 4 | 99 |
| 598–700 | 31 | 19 | 12 | 10 | 9 | 7 | 5 | 6 | 4 | 103 |
| 701–800 | 30 | 17 | 13 | 10 | 7 | 7 | 6 | 5 | 5 | 100 |
| 801–900 | 30 | 18 | 12 | 10 | 8 | 7 | 5 | 6 | 4 | 100 |
| 901–999 | 30 | 17 | 13 | 9 | 8 | 7 | 6 | 4 | 5 | 99 |
| 1000–1099 | 30 | 17 | 13 | 10 | 7 | 7 | 6 | 6 | 4 | 100 |
| 1100–1199 | 30 | 19 | 11 | 10 | 9 | 6 | 5 | 6 | 4 | 100 |
| 1200–1298 | 30 | 17 | 13 | 9 | 8 | 7 | 6 | 4 | 5 | 99 |
| 1299–1398 | 30 | 18 | 12 | 10 | 8 | 7 | 5 | 6 | 4 | 100 |
| 1399–1498 | 30 | 18 | 12 | 10 | 8 | 6 | 6 | 5 | 5 | 100 |
| 1499–1597 | 30 | 17 | 13 | 9 | 8 | 7 | 6 | 5 | 4 | 99 |
| 1598–1700 | 31 | 19 | 12 | 10 | 9 | 6 | 6 | 6 | 4 | 103 |
| 1701–1800 | 30 | 17 | 13 | 10 | 8 | 7 | 6 | 5 | 5 | 100 |
| 1801–1900 | 30 | 19 | 11 | 10 | 9 | 6 | 6 | 5 | 5 | 100 |
| 1901–1999 | 30 | 17 | 13 | 9 | 8 | 6 | 7 | 5 | 5 | 99 |
| Totals | | | | | | | | | | |
| 0–999 | 301 | 176 | 125 | 97 | 79 | 69 | 56 | 52 | 45 | 1000 |
| 1000–1999 | 301 | 178 | 123 | 97 | 82 | 65 | 59 | 53 | 45 | 1000 |
| 0–1999 | 602 | 354 | 248 | 194 | 161 | 134 | 115 | 105 | 90 | 2000 |

strings $\alpha$ standing for $a, b, \ldots, e$. They are given by

THEOREM II. *The asymptotic probabilities of occurrence of the strings are given by*

$$\pi(a) = \log_2(1+\tfrac{1}{8}), \qquad \pi(b) = \log_2(1+\tfrac{1}{9}),$$

$$\pi(c) = \log_2(1+\tfrac{1}{5}), \qquad \pi(d) = \log_2(1+\tfrac{1}{6}), \tag{4}$$

$$\pi(e) = \log_2(1+\tfrac{1}{7}).$$

TABLE 4
PROBABILITIES OF OCCURRENCE OF FIRST DIGITS
1 THROUGH 9 IN $2^n$ FOR $n$ UP TO 2000

| Period of length $P_k$ | Probability of occurrences of the first digit 1 through 9 in $2^n$: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 0–99 | 0.300 | 0.17 | 0.13 | 0.10 | 0.07 | 0.07 | 0.06 | 0.05 | 0.05 |
| 100–199 | 0.300 | 0.19 | 0.11 | 0.1 | 0.09 | 0.06 | 0.05 | 0.06 | 0.04 |
| 200–298 | 0.303 | 0.172 | 0.131 | 0.091 | 0.081 | 0.071 | 0.061 | 0.04 | 0.05 |
| 299–398 | 0.30 | 0.17 | 0.13 | 0.10 | 0.07 | 0.08 | 0.05 | 0.06 | 0.04 |
| 399–498 | 0.30 | 0.18 | 0.12 | 0.10 | 0.08 | 0.06 | 0.06 | 0.05 | 0.05 |
| 499–597 | 0.303 | 0.172 | 0.13 | 0.09 | 0.08 | 0.07 | 0.06 | 0.051 | 0.04 |
| 598–700 | 0.301 | 0.184 | 0.117 | 0.097 | 0.0874 | 0.068 | 0.041 | 0.058 | 0.039 |
| 701–800 | 0.30 | 0.17 | 0.13 | 0.10 | 0.07 | 0.07 | 0.06 | 0.05 | 0.05 |
| 801–900 | 0.30 | 0.18 | 0.12 | 0.10 | 0.08 | 0.07 | 0.05 | 0.06 | 0.04 |
| 901–999 | 0.30 | 0.17 | 0.13 | 0.09 | 0.08 | 0.07 | 0.06 | 0.04 | 0.05 |
| 1000–1099 | 0.30 | 0.17 | 0.13 | 0.10 | 0.07 | 0.07 | 0.06 | 0.06 | 0.04 |
| 1100–1199 | 0.30 | 0.19 | 0.11 | 0.10 | 0.09 | 0.06 | 0.05 | 0.06 | 0.04 |
| 1200–1298 | 0.303 | 0.172 | 0.131 | 0.091 | 0.081 | 0.071 | 0.061 | 0.041 | 0.051 |
| 1299–1398 | 0.30 | 0.18 | 0.12 | 0.10 | 0.08 | 0.07 | 0.05 | 0.06 | 0.04 |
| 1399–1498 | 0.30 | 0.18 | 0.12 | 0.10 | 0.08 | 0.06 | 0.06 | 0.05 | 0.05 |
| 1499–1597 | 0.303 | 0.172 | 0.131 | 0.09 | 0.08 | 0.07 | 0.06 | 0.051 | 0.04 |
| 1598–1700 | 0.301 | 0.185 | 0.117 | 0.0971 | 0.0871 | 0.0583 | 0.0583 | 0.0583 | 0.0388 |
| 1701–1800 | 0.30 | 0.17 | 0.13 | 0.10 | 0.08 | 0.07 | 0.06 | 0.05 | 0.05 |
| 1801–1900 | 0.30 | 0.19 | 0.11 | 0.10 | 0.09 | 0.06 | 0.06 | 0.05 | 0.05 |
| 1901–1999 | 0.303 | 0.172 | 0.131 | 0.091 | 0.081 | 0.061 | 0.071 | 0.051 | 0.051 |
| Totals | | | | | | | | | |
| 9–999 | 0.301 | 0.176 | 0.125 | 0.097 | 0.079 | 0.069 | 0.056 | 0.052 | 0.045 |
| 1000–1999 | 0.301 | 0.178 | 0.123 | 0.097 | 0.082 | 0.065 | 0.059 | 0.053 | 0.045 |
| 0–1999 | 0.301 | 0.177 | 0.124 | 0.097 | 0.0805 | 0.067 | 0.0575 | 0.0525 | 0.045 |
| Asymptotic probability | 0.301 | 0.176 | 0.125 | 0.097 | 0.079 | 0.0669 | 0.0580 | 0.0511 | 0.0458 |

Note that the logarithms are now taken to base 2.

We first deduce these probabilities by a straightforward probabilistic argument and then rederive them by an elegant geometrical construction patterned after the proof of Theorem I. As observed earlier, each string is uniquely determined by the last member in its group, and hence the asymptotic probabilities of the strings are proportional to the probabilities of

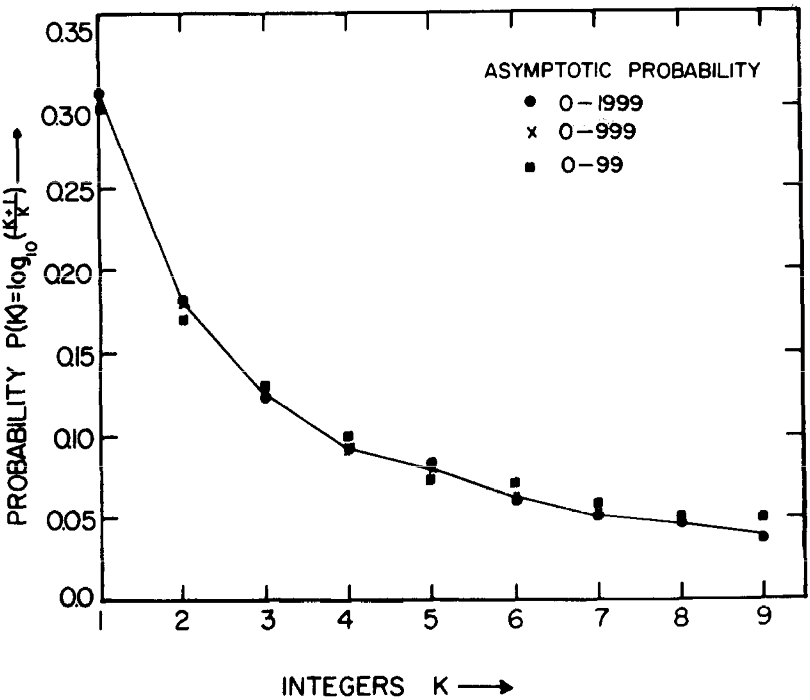



FIG. 5. The asymptotic probability of integers for $n = 0$ to 1999.

occurrence of these last members, according to Theorem I. Hence

$$\pi(a) = \lambda \log_{10}(1+\tfrac{1}{8}), \qquad \pi(b) = \lambda \log_{10}(1+\tfrac{1}{9}),$$
$$\pi(c) = \lambda \log_{10}(1+\tfrac{1}{5}), \qquad \pi(d) = \lambda \log_{10}(1+\tfrac{1}{6}), \tag{5}$$
$$\pi(e) = \lambda \log_{10}(1+\tfrac{1}{7}).$$

The constant of proportionality, $\lambda$, is determined by the condition that the total probability $\sum_{\alpha=a}^{e} \pi(\alpha) = 1$, and so $\lambda = (\log_{10} 2)^{-1}$ and thus we arrive at (4).

The geometric proof uses the uniform distribution of $x_n = n \log_{10} 2 \pmod 1$ as follows. This proof, as with that of Theorem I, will be most useful in deducing the transition probabilities to be discussed in the next section. We now consider Fig. 4, and recall how the line segment $[1,2)$ gives rise to the strings $a, \ldots, e$. Since the points on the line segment become uniformly

TABLE 5

FREQUENCY DISTRIBUTION OF STRING SEQUENCES OF DIGITS

| String sequence: | $a \leftarrow 1248$ | $b \leftarrow 1249$ | $c \leftarrow 125$ | $d \leftarrow 136$ | $e \leftarrow 137$ |
|---|---|---|---|---|---|
| 0–99 | 5 | 5 | 7 | 7 | 6 |
| 100–199 | 6 | 4 | 9 | 6 | 5 |
| 200–298 | 4 | 5 | 8 | 7 | 6 |
| 299–398 | 6 | 4 | 7 | 8 | 5 |
| 399–498 | 5 | 5 | 8 | 6 | 6 |
| 499–597 | 5 | 4 | 8 | 7 | 6 |
| 598–700 | 6 | 4 | 9 | 7 | 5 |
| 701–800 | 5 | 5 | 7 | 7 | 6 |
| 801–900 | 6 | 4 | 8 | 7 | 5 |
| 901–999 | 4 | 5 | 8 | 7 | 6 |
| 1000–1099 | 6 | 4 | 7 | 7 | 6 |
| 1100–1199 | 6 | 4 | 9 | 6 | 5 |
| 1200–1298 | 4 | 5 | 8 | 7 | 6 |
| 1299–1398 | 6 | 4 | 8 | 7 | 5 |
| 1399–1498 | 5 | 5 | 8 | 6 | 6 |
| 1499–1597 | 5 | 4 | 8 | 7 | 6 |
| 1598–1700 | 6 | 4 | 9 | 6 | 6 |
| 1701–1800 | 5 | 5 | 7 | 7 | 6 |
| 1801–1900 | 5 | 5 | 9 | 6 | 5 |
| 1901–1999 | 5 | 4 | 8 | 6 | 7 |
| 0–999 | 52 | 45 | 79 | 69 | 56 |
| 1000–1999 | 53 | 44 | 81 | 65 | 58 |
| 0–1999 | 105 | 99 | 161 | 134 | 114 |

distributed upon taking the logarithm to base 10, we arrive at the line segment displayed at the bottom of Fig. 4, which now goes from 0 to $\log_{10} 2$ and the segments $a,\dots,e$ correspondingly get mapped as shown there. Hence the asymptotic probabilities of occurrence of the strings are given by the ratio of the corresponding lengths on the logarithmically mapped segment to the total length:

$$\pi(a) = \frac{\log_{10} 1.125}{\log_{10} 2} = \log_2 1.125 = \log_2\left(1 + \tfrac{1}{8}\right),$$

$$\pi(b) = \frac{\log_{10} 1.25 - \log_{10} 1.125}{\log_{10} 2} = \log_2 \frac{1.25}{1.125} = \log_2\left(1 + \tfrac{1}{9}\right)$$

and so on, as given by Equation (4).

TABLE 6

PROBABILITY DISTRIBUTION OF STRING SEQUENCES OF DIGITS

| String sequence: | $a \leftarrow 1248$ | $b \leftarrow 1249$ | $c \leftarrow 125$ | $d \leftarrow 136$ | $e \leftarrow 137$ |
|---|---|---|---|---|---|
| 0–99 | .1667 | .1667 | .2333 | .2333 | .2000 |
| 100–199 | .2000 | .1333 | .3000 | .2000 | .16667 |
| 200–298 | .1333 | .16667 | .26667 | .2333 | .2000 |
| 299–398 | .2000 | .1333 | .2333 | .26667 | .2000 |
| 399–498 | .16667 | .16667 | .26667 | .2000 | .2000 |
| 499–597 | .16667 | .13333 | .26667 | .2333 | .2000 |
| 598–700 | .1936 | .1290 | .2903 | .2258 | .1613 |
| 701–800 | .16667 | .1667 | .2333 | .2333 | .2000 |
| 801–900 | .16667 | .1337 | .26667 | .2333 | .16667 |
| 901–999 | .1333 | .1667 | .26667 | .2333 | .2000 |
| 1000–1099 | .2 | .1333 | .2333 | .2333 | .2 |
| 1100–1199 | .2 | .1333 | .3 | .2 | .1667 |
| 1200–1298 | .1333 | .1667 | .2667 | .2333 | .2 |
| 1299–1398 | .2 | .1333 | .2667 | .2333 | .1667 |
| 1399–1498 | .1667 | .1667 | .2667 | .2 | .2 |
| 1499–1597 | .1667 | .1333 | .2667 | .2333 | .2 |
| 1598–1700 | .1935 | .1290 | .2903 | .1935 | .1935 |
| 1701–1800 | .1667 | .1667 | .2333 | .2333 | .2 |
| 1801–1900 | .1667 | .1667 | .3 | .2 | .1667 |
| 1901–1999 | .1667 | .1333 | .2667 | .2 | .2333 |
| 0–999 | .1728 | .1495 | .2625 | .206 | .1861 |
| 1000–1999 | .1761 | .1462 | .2691 | .2160 | .1927 |
| 0–1999 | .1713 | .1615 | .2626 | .2186 | .186 |
| Asymptotic probability | .169925 | .152003 | .2630344 | .222392 | .19264 |

In Fig. 6, we have plotted these asymptotic probabilities $\pi(\alpha)$ versus $\alpha$ as lines, and the numerically obtained values from our samples are shown for comparison.

In the next section, we shall discuss the transition probabilities of the integers and the strings, as they seem to exhibit interesting features and they clearly show the non-Markov character of the underlying stochastic process. This has not been studied before.

## III. TRANSITION PROBABILITIES OF THE INTEGERS AND THE STRINGS

The idea of transition probabilities is a useful one in discussing the statistical properties, as these give a feel for the "movement" of the events

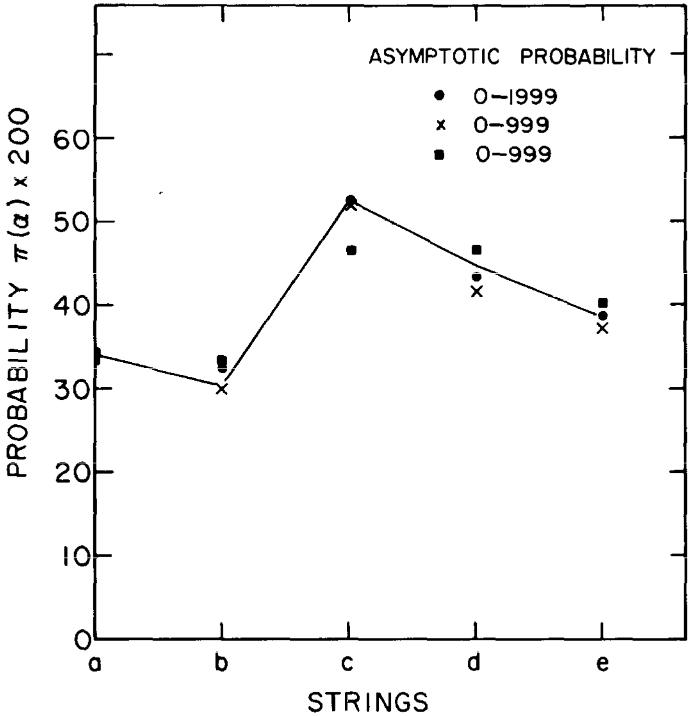



FIG. 6. The asymptotic probability of strings for $n = 0$ to 1999.

studied. The transition probability $T(p \to q)$ is the probability that an event $p$ will go to an event $q$ in the next trial. It is usually written in the form of a matrix.

From Figures 1 and 2, it is clear what happens in our system. For integers, the derivation of the asymptotic transition matrix is easy to deduce from Figure 1. Clearly 1 can go only to 2 or 3, and so $T(1 \to 2)$ and $T(1 \to 3)$ are nonzero while all other $T(1 \to k)$ are zero. Since the paths are exhausted, the sum $T(1 \to 2) + T(1 \to 2)$ must equal unity. In this way, in conjunction with Fig. 1, we obtain the matrix $T(p \to q)$ after a little computation. We will now give the algebraic derivation and then a geometric method based on the line segment theorem and these results are stated in the form of a theorem.

THEOREM III. *The asymptotic transition probability matrix $T(p \to q)$ for going from $p$ to $q$ $(p, q = 1, \ldots, 9)$ is as given in Table* 7.

TABLE 7

| $p \backslash q$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | $\log_{4/2}(\frac{3}{2})$ | $\log_{4/2}(\frac{4}{3})$ | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | $\log_{6/4}(\frac{5}{4})$ | $\log_{6/4}(\frac{6}{5})$ | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | $\log_{8/6}(\frac{7}{6})$ | $\log_{8/6}(\frac{8}{7})$ | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\log_{10/8}(\frac{9}{8})$ | $\log_{10/8}(\frac{10}{9})$ |
| 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The direct algebraic method is to observe that, for example, $T(1 \to 2)$ is proportional to the probability of occurrence of 2, viz. $P(2) = \log_{10}(\frac{3}{2})$, and $T(1 \to 3)$ is proportional to $P(3) = \log_{10}(\frac{4}{3})$. But since their sum must be unity, the proportionality constant is just $[\log_{10}(\frac{4}{2})]^{-1}$, and hence we find the entries in the transition matrix given in Table 7.

In the geometric derivation, we appeal to Figure 3. We observe that if we start with 1 we land on 2 or 4. In other words, the segment [2, 4] in the top of Figure 3 is reached by all decimal numbers of the form $1.xyz\ldots$ upon multiplication by 2. If $x$ in $1.xyz$ is less than 5, we get points in the segment [2,3), and if $x \geqslant 5$, we get points in [3,4]. Using the logarithmic mapping underneath, the asymptotic transition probabilities are deduced. It is important to observe that the first numbers beginning with 1 are all of the form $1.xyz\ldots \times 10^k$, which move to the points in the segment lying in [2,4) upon multiplication by 2. Similar arguments hold for $p.xyz\ldots \times 10^m$, for finding the transitions from $p$ to the corresponding $q$'s. It is quite clear that the integers 5, 6, 7, 8, 9 all go to 1, and hence that part of the transition matrix is obtained trivially.

The calculation of the transition probabilities for the strings is a little more subtle. To establish the pathways the strings take, we find the geometric construction given in Fig. 4 is most transparent. Let us examine the string $a$. The last digit in this string is of the form $8.xyz\ldots \times 10^k$, say. Upon multiplication by 2, this will go over to $1.x'y'z'\ldots \times 10^{k+1}$, and we need to check where this fraction is on the top segment of Fig. 4. Remembering that the string $a$ corresponds to points in the segment [1, 1.125) it is clear that 8 goes to 1.6 in this representation and hence lands in the segment marked $d$ in Fig. 4. Since 8 occurs in the form $8.xyz\ldots$, it is clear that all numbers with $x < 5$ will land to the right of 1.6 in segment $d$, and this goes on till $x = 9$, $y = 9$, etc., which implies that such numbers $8.999\ldots$ *go up to* 1.8 *at the outer limit.* In other words, $a$ goes to $d$ or $e$.

It is now important to realize from this discussion that the segments $d$ and $e$ are not fully covered by the points so generated. Hence the transition probabilities are proportional to the respective lengths of the segments after a $\log_{10}$ mapping. The string $b$, which has 9 in it, lands on $e$ only because upon multiplication by 2, $9.xyz\ldots$ is at least 1.8 and at most 2. Since $b \to e$ only, this transition probability is unity. We thus obtain

THEOREM IV. *The asymptotic transition probability matrix* $\mathfrak{T}(\alpha \to \beta)$ *of the strings* $(\alpha, \beta = a, b, \ldots, e)$ *is as given in Table* 8.

In the actual evaluation of this matrix, we employed the above description of the process and the fact that

$$\mathfrak{T}(a \to d) = \lambda(\log_{10} 1.75 - \log_{10} 1.6)$$

$$\mathfrak{T}(a \to e) = \lambda(\log_{10} 1.8 - \log_{10} 1.75)$$

TABLE 8

| $\alpha \backslash \beta$ | $a$ | $b$ | $c$ | $d$ | $e$ |
|---|---|---|---|---|---|
| $a$ | 0 | 0 | 0 | $\log_{9/8}\left(\frac{1.75}{1.6}\right)$ | $\log_{9/8}\left(\frac{1.8}{1.75}\right)$ |
| $b$ | 0 | 0 | 0 | 0 | 1 |
| $c$ | $\log_{6/5}(1.125)$ | $\log_{6/5}\left(\frac{1.2}{1.125}\right)$ | 0 | 0 | 0 |
| $d$ | 0 | $\log_{7/6}\left(\frac{1.125}{1.2}\right)$ | $\log_{7/6}\left(\frac{1.4}{1.125}\right)$ | 0 | 0 |
| $e$ | 0 | 0 | $\log_{8/7}\left(\frac{1.5}{1.4}\right)$ | $\log_{8/7}\left(\frac{1.6}{1.5}\right)$ | 0 |

TABLE 9
THE ASYMPTOTIC TRANSITION PROBABILITY MATRIX $\tau_0(\alpha \to \beta)$ OF THE STRINGS $(\alpha, \beta = a, b, c, d, e)$

| | $1248 \leftarrow a$ | $1249 \leftarrow b$ | $125 \leftarrow c$ | $136 \leftarrow d$ | $137 \leftarrow e$ |
|---|---|---|---|---|---|
| $1248 \leftarrow a$ | 0 | 0 | 0 | 0.7767 | 0.2233 |
| $1249 \leftarrow b$ | 0 | 0 | 0 | 0 | 1 |
| $125 \leftarrow c$ | 0.6519 | 0.34810 | 0 | 0 | 0 |
| $136 \leftarrow d$ | 0 | 0.2654 | 0.7346 | 0 | 0 |
| $137 \leftarrow e$ | 0 | 0 | 0.5085 | 0.4915 | 0 |

and others are zero. Since the sum should be unity, $\lambda = [\log(1.8/1.6)]^{-1}$ and the results are as given in the matrix.

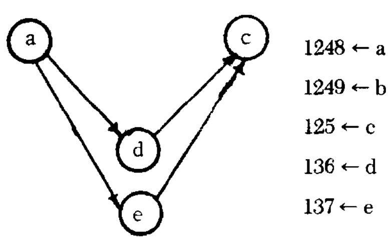
We have obtained these quantities numerically also for 613 strings ($2^n$, $n = 0$–1999), and they are given in Table 9.

We have developed computer codes for calculating probabilities of occurrence of pairs, triples, and quadruples of strings as well as path probabilities for the occurrence of strings in successive trials such as $\pi(\alpha \to \beta\gamma)$ etc., which have also been represented as tree structures. These results are given in Ref. [14]. In the next section we shall give an explicit proof that the string sequences are non-Markovian.

## IV. NON-MARKOVIAN NATURE OF THE PROCESS

The process of branching from a string $I$ to a string $J$ can take place in $n$ intermediate steps, where $n$ can be $0, 1, 2, 3, \ldots$. An example of a one step process is treated below.

The process of string $a$ going to string $c$ is a one step process. This can take place in two ways—string '$a$' can go to string '$b$' through string '$d$' or string '$c$'. The following diagram describes the process:




Mathematically we can define the above branching process as follows. Let $\mathfrak{T}_0(p, q)$ be the transition probability matrix of a zero step process. Let $\mathfrak{T}_1$ be the transition probability matrix of a one step process. For a non-Markovian process, $\mathfrak{T}_1 \neq \mathfrak{T}_0^2$, i.e. the transition probability from one state to another is not independent of the previous states visited. A computation of this path process of strings indicates that the underlying process is non-Markovian (see Tables 9, 10, and 11). This is also intuitively evident, as stated in the Introduction.

TABLE 10

COMPUTATION OF PATH PROCESS OF STRINGS FOR A ONE-STEP TRANSITION MATRIX $T_1$

| | $1248 \leftarrow a$ | $1249 \leftarrow b$ | $125 \leftarrow c$ | $136 \leftarrow d$ | $137 \leftarrow e$ |
|---|---|---|---|---|---|
| $1248 \leftarrow a$ | 0 | 0 | 1 | 0 | 0 |
| $1249 \leftarrow b$ | 0 | 0 | 0.3933 | 0.6067 | 0 |
| $125 \leftarrow c$ | 0 | 0 | 0 | 0.4969 | 0.5031 |
| $136 \leftarrow d$ | 0.7463 | 0 | 0 | 0 | 0.2537 |
| $137 \leftarrow e$ | 0.0430 | 0.9750 | 0 | 0 | 0 |

TABLE 11

COMPUTATION OF $\mathcal{T}_0^2$, THE SQUARE OF THE ZERO STEP TRANSITION MATRIX FOR STRINGS

| | $1248 \leftarrow a$ | $1249 \leftarrow b$ | $125 \leftarrow c$ | $136 \leftarrow d$ | $137 \leftarrow e$ |
|---|---|---|---|---|---|
| $1248 \leftarrow a$ | 0 | 0.2061 | 0.6841 | 0.1098 | 0 |
| $1249 \leftarrow b$ | 0 | 0 | 0.5085 | 0.4915 | 0 |
| $125 \leftarrow c$ | 0 | 0 | 0 | 0.5063 | 0.4937 |
| $136 \leftarrow d$ | 0.4789 | 0.2557 | 0 | 0 | 0.2654 |
| $137 \leftarrow e$ | 0.3315 | 0.3075 | 0.3611 | 0 | 0 |

In the next section we shall give a summary of the results reported here as well as a class of other extensions and observations and possible applications of our investigations.

## V. SUMMARY AND CONCLUDING REMARKS

We have examined the statistical properties of the first digits and the strings generated by powers of 2. It is clear from the asymptotic transition probability matrices for both digits and strings that the underlying stochastic process is non-Markovian. From a mathematical point of view, however, the process is ergodic in the sense of the asymptotic distribution modulo one [10, 11]. This case study is an elegant example of an ergodic but non-Markovian process. We have now shown [15], using the theory of tower structures in ergodic theory, that the strings found here can be given an elegant mathematical derivation. In fact, it turns out that the number of strings associated with the $2^n$ problem is five, but the choice of the entities in the strings themselves and hence their probabilities are not unique. This is intuitively expected. Another important point that emerges out of this work is that unlike the digits $1,\ldots,9$, the strings $a,\ldots,e$ may be an interesting set as candidates for generating quasi random events. Observing that $\pi(a) = 0.169925$, $\pi(b) = 0.152003$, $\pi(c) = 0.2630344$, $\pi(d) = 0.2223923$, and $\pi(e) = 0.192645$ are all, crudely speaking, close to $\frac{1}{5} = 0.2$, and thus the strings occur almost randomly —unlike the integers, whose probabilities ranged from 0.3010 to 0.0458, quite far from $\frac{1}{9}$ ($\approx 0.111$). From the structure of the transition probabilities $\pi(\alpha \to \beta)$, it seems we may have possibilities for application to coding and cryptography. It may not be out of place here to point out that Uppuluri [16] has examined a certain probability aspect of the prime numbers within a Markov theoretic framework. It is intriguing to find that the first digit problem turns out to be non-Markovian.

The Newcomb-Benford law (Theorem I above) is in fact quite general and does not depend on the first digits of powers of 2. In fact, the same law obtains for the first digits of $a^n$. Even more generally, in base $b$ (instead of 10 used in our work), the probabilities of occurrence of the integers $p = 1, \dots, b - 1$ are $\log_b(1 + 1/p)$. But it is expected that the string structures and their transition probabilities will not be independent of the generator $a$ in $a^n$. Recently Kak [17] has examined $3^n$, $7^n$, $11^n$, etc. in some detail and has found interesting string structures: graphs associated with the movements of the integers $1, \dots, 9$. J. Robertson et al. [15] show how these string structures depend on the base (e.g. 3, 7) and are derivable from towers generated by the underlying ergodic process; they also show that while the number of strings associated with a given base $b$ is unique, what constitutes these strings is not unique. Kak [17] surmises that there may be a relationship between the statistics of the first digits at the input and the output of an exponentiator modulo $n$, a transformation which has been suggested for public key encryption.

The problem of determining the structure of the sequence $x_n = f(x_{n-1})$ for different classes of $f$ is of considerable current importance in physics and computer science. Feigenbaum [18, 19] has recently suggested that a behavioral universality is associated with a large class of systems obeying the recursion relation above. The analysis used by us may be of some relevance to these questions also.

*We thank Dr. Mark Robinson of ORNL for many discussions on the topic of this paper with one of us (A. K. R.), who also thanks the Division of Solid State, ORNL, for their warm hospitality and support. Dr. S. C. Kak of the Electrical Engineering Department, LSU, has been very kind in spending his time to read, criticize, and offer suggestions concerning the applications of the results of this paper. We thank him profusely.*

*We are grateful to Professor Raimey of the University of Rochester for his critical comments on our paper.*

## REFERENCES


1 Simon Newcomb, Note on the frequency of use of the different digits in natural numbers, *Amer. J. Math.* 4:39–40 (1881).
2 Frank Benford, The law of anomalous numbers, *Proc. Amer. Phil. Soc.* 78:551–572 (1938).
3 Ralph A. Raimi, The peculiar distribution of first digits, *Scientific American* 221:109–120 (1969).
4 Ralph A. Raimi, The first digit problem, *Amer. Math. Monthly* 83:521–538 (1976).
5 R. W. Hamming, On the distribution of numbers, *Bell System Tech. J.* 40:1609–1625 (1970).
6 N. K. Tsao, The distribution of significant digits and round off error, *Comm. ACM* 17:269–271 (1974).
7 D. Knuth, *The Art of Computer Programming*, Vol. 2, Addison-Wesley, Reading, Mass., 1969 pp. 219–229.
8 Roger S. Pinkham, On the distribution of first significant digits, *Ann. Math. Statist.* 32:1223–1230 (1961).
9 J. F. Koksma, The theory of asymptotic distribution modulo one, *Compositio Math.* 16:1–22 (1964). This volume of the journal has a collection of articles on a variety of aspects of this theory.
10 V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer, New York, 1978, Chapter 3.
11 V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, Benjamin, New York, 1968; see especially Appendix 12, pp. 134–137.
12 Y. Pomeau, Stochastic behaviors of simple dynamical systems, in *Stochastic Processes in Nonequilibrium Systems*, (L. Garrido, P. Seglar, and P. J. Shepherd, Eds.), Springer Berlin, 1978, pp. 235–277.
13 S. C. Kak and A. Chatterjee, *IEEE Trans. Inform. Theory*, Sept. 1981.
14 A. K. Rajagopal, S. Sitharama Iyengar, and Mohan Yellayi, Statistical properties of strings generated by first digits of powers of two, submitted for publication, *BIT*, Sept. 1981.
15 J. Robertson, V. R. R. Uppuluri, and A. K. Rajagopal, First digit phenomenon and ergodic theory, to be submitted to *J. Combin. Theory Ser. B.*
16 V. R. R. Uppuluri, in *Essays in Probability and Statistics*, a volume in honor of Professor Junjiro Ogawa, Japan, 1976, Chapter 39, pp. 645–650.
17 S. C. Kak, New results on the first digit problem, LSU Technical Report EE#607, Aug. 1981.
18 M. J. Feigenbaum, Quantitative universality for a class of non-linear transformations, *J. Statist. Phys.* 19:25–52 (1978).
19 M. J. Feigenbaum, Universal behavior of non-linear systems, *Los Alamos Sci.*, 1980, pp. 4–27.
20 S. Sitharama Iyengar, A. K. Rajagopal, and V. R. R. Uppuluri, String patterns of leading digits, *J. Appl. Math. and Comp.* (in press).
21 S. Sitharama Iyengar, A. K. Rajagopal, and Frank Ramos, On the distribution of string sequences of leading digits, *J. Comb. Syst. Inf. Theory* (in press).