Where Are All the Typical Numbers – 2

B V Rao

In this two-part article, we are trying to discover patterns in the decimal digits of numbers between zero and one. In Part I we started showing that each digit occurs with frequency 1/10. We reduced the problem with the help of Borel–Cantelli Lemma and Markov inequality to show certain sums of integrals are finite. In this concluding part we bring in an essentially new ingredient and conclude the proof.

B V Rao was a Professor at the Indian Statistical Institute, Kolkata. Currently he is with the Chennai Mathematical Institute.

1. Randomness of Decimal Digits

The decimal digits X_i and hence the functions Z_i satisfy a property which is essentially probabilistic in nature. We shall explain this, though we do not use any probabilistic results. Suppose you have a bag containing three red balls and two black balls. If you pick a ball at random, what are the chances that it is red? There are five possible outcomes – any one of the five balls is an outcome. Of these, three outcomes correspond to red ball. So the chances are $3/5$.

Part 1. *Resonance,* Vol.16, No.10, pp.933–949, 2011.

We now ask a similar question: if you pick a number at random from $(0, 1]$ what are the chances that its third decimal digit is 5? If we blindly try to follow the same rule as in the case of balls, we run into trouble. This is because there are an infinite number of outcomes and of them infinitely many outcomes satisfy the criterion and infinitely many do not. But remember, we have decided to quantify sets by their length and not by their cardinality. Thus probability of an event (that is, a nice set of outcomes) is length of the event divided by the length of the set of all outcomes. But the set of all outcomes is the interval $(0, 1]$ which has length one. Thus to calculate probability of an event, we only need to calculate

Keywords
Frequency of decimal digits, Borel–Cantelli Lemma, Markov inequality, independence.

X_n takes each of the values 0,1...,9 with equal probability.

length of the event. In other words, length has an alternative interpretation as probability. Thus instead of 'length', we could use 'probability'. It is clear, if you understood the decimal expansion, that there are 10^2 disjoint intervals each of length 10^3 which comprise the event in question and hence the length or probability of the event is $1/10$.

2. Independence of Decimal Digits

We already saw that the event 'third decimal digit is 5' consists of 10^2 intervals each of length 10^{-3}, so its length equals $1/10$. There is nothing special about the third digit nor of the digit 5. For any n, the n-th digit X_n takes each of the values $0, 1, \cdots, 9$ with equal probability, namely, $1/10$. This only means that the length of the set $\{0 < \omega \leq 1 : X_n(\omega) = j\}$ is $1/10$, whatever be the integer j, $0 \leq j \leq 9$.

Something more is true. Suppose we fix integers $1 \leq n_1 < n_2 < \cdots < n_k$ and digits $j_1, j_2, \cdots, j_k$ and ask: what are the chances that the n_i-th digit is j_i for $1 \leq i \leq k$? The answer is 10^{-k} – the product of individual probabilities. More precisely, let E_i be the event $(\omega : X_{n_i}(\omega) = j_i)$, then $\text{length}(\cap E_i) = \prod \text{length}(E_i)$. This is routine pen and paper calculation, nothing deep is involved. First do it for $k = 1, 2, 3$ and then you will get the general idea. Of course, for a rigorous proof of the general case, you would use induction on k. You first consider the simpler case when $n_i = i$, that is, consecutive decimal places and then deduce the general case. In terms of Z_i, these observations amount to the following:

$$\text{length}\{Z_{n_1} = \epsilon_1, \cdots, Z_{n_k} = \epsilon_k\} = \prod_{i=1}^{k} \text{length}\{Z_{n_i} = \epsilon_i\}, \tag{1}$$

$Z_i(x) = 1$ if i-th decimal digit of x equals 5; zero otherwise.

for any sequence $(\epsilon_1, \cdots, \epsilon_k)$ of zeros and ones and for any $(n_1, \cdots, n_k)$. If you replace 'length' by 'probability' – since both are same as mentioned in the earlier para – the equation (1) above has a special name in probability theory. One says that the functions (Z_i) are independent.

Concept of independence, simple as it is, is extremely powerful.

3. A Consequence of Independence

Equation (1), simple as it is, is extremely powerful. One useful thing is that, it implies that integrals also multiply. We give a more precise statement below. Since we shall apply this consequence not only for the variables (Z_i) but also for some simple related variables, it is better to get out of the shackles of the notation above.

Suppose that $(V_i : 1 \le i \le k)$ are functions defined on $(0, 1]$ each taking finitely many values on finitely many intervals. Suppose for any sequence of values $\langle v_1, \cdots, v_k \rangle$ of the variables $\langle V_1, \cdots, V_k \rangle$ we have

$$\text{length}\{V_i = v_i, \cdots, V_k = v_k\} = \prod_{i=1}^{k} \text{length}\{V_i = v_i\}. \tag{2}$$

That is, the functions (V_i) are independent. Then we claim that

$$\int \left(\prod_i V_i \right) = \prod_i \int V_i. \tag{3}$$

Just let us remind ourselves that the product $\prod_i V_i$ is the function whose value at any point x is obtained as follows: calculate each $V_i(x)$ and multiply these numbers. Thus, this function $\prod_i V_i$ also takes finitely many values on finitely many intervals, just because each V_i does so. To see the truth of equation (3), put for each sequence $s = \langle v_1, \cdots, v_k \rangle$ of values of the variables, the set $A_s = \{x : V_1(x) = v_1, \cdots, V_k(x) = v_k\}$. Remember that on this set, which is made up of a finite union of intervals, the product function takes the value $v_1 v_2 \cdots v_k$. The left side of (3), by definition of integral, equals

$$\sum_s v_1 v_2 \cdots v_k \ \text{length}\{x : V_1(x) = v_1, \cdots, V_k(x) = v_k\}.$$

If V_1, V_2 are independent, so are $f_1(V_1)$ and $f_2(V_2)$.

Now use hypothesis (2) to see that this equals

$$\sum_s v_1 v_2 \cdots v_k \prod_i \text{length}\{x : V_i(x) = v_i\}$$
$$= \sum_s \prod_i (v_i \cdot \text{length}\{x : V_i = v_i\}).$$

We have distributed each v_i to one term in the product. Now use the fact $\sum_{i,j} a_i b_j = (\sum_i a_i)(\sum_j b_j)$ – actually, a similar equation for more than two indices – to see that the above sum is

$$= \prod_i \left(\sum_{v_i} v_i \cdot \text{length}\{V_i = v_i\} \right) \;\; = \;\; \prod_i \int V_i.$$

Here the last equality is again from the definition of integral.

4. More on Independence

Let us also note the following useful fact. Suppose the sequence $(V_i : 1 \leq i \leq k)$ is independent. Remember, our functions take finitely many values. Suppose that for each i, U_i is a function of V_i, say $U_i = f_i(V_i)$, then the sequence (U_i) is also independent. Indeed, for $1 \leq i \leq k$, let u_i be a value of U_i. We need to show

$$\text{length}\{U_i = u_i;\ 1 \leq i \leq k\} = \prod_i \text{length}\{U_i = u_i\}.$$

To show this, let F_i be the finite set of values x of V_i such that $f_i(x) = u_i$. Clearly, the set $\{x : U_i(x) = u_i;\ 1 \leq i \leq k\}$ is same as the set $\{x : V_i(x) \in F_i : 1 \leq i \leq k\}$, which, in turn, is disjoint union of the sets $\{x : V_i(x) = v_i; 1 \leq i \leq k\}$ as v_i ranges over F_i for $1 \leq i \leq k$. Thus its length, by independence of the functions (V_i), equals

$$\sum_{v_i \in F_i : 1 \;\;\; i \,\leq\, k \leq i} \prod \text{length}\{x : V_i(x) = v_i\}$$
$$= \prod_i \;\; \sum_{v_i \in F_i} \text{length}\{V_i = v_i\}.$$

Since for each i,

$$\sum_{v_i \in F_i} \text{length}\{V_i = v_i\}$$
$$= \text{length}\{V_i \in F_i\} \;=\; \text{length}\{U_i = u_i\},$$

we have proved

$$\text{length}\{U_i = u_i;\ 1 \le i \le k\} = \prod_i \text{length}\{U_i = u_i\},$$

showing the independence of the functions (U_i).

Recall $W_j(x) = Z_j(x) - 1/10$.

5. Completion of Proof

Returning to our functions, observe that W_i is a function of Z_i and (Z_i) are independent. From the discussion above we conclude that (W_i) are also independent. Thus

$$\text{length}\{W_{n_1} = \epsilon_1, \cdots, W_{n_k} = \epsilon_k\} = \prod_{i=1}^{k} \text{length}\{W_{n_i} = \epsilon_i\} \tag{4}$$

for any sequence $(\epsilon_1, \cdots, \epsilon_k)$ of values of the functions W_i and for any $(n_1, \cdots, n_k)$.

For example, taking $V_1 = W_1^2$, $V_2 = W_2$ and $V_3 = W_3$, we see that these V_1, V_2, V_3 are also independent and in particular (3) tells us

$$\int W_1^2 W_2 W_3 = \int W_1^2 \int W_2 \int W_3 = 0,$$

simply because $\int W_3 = 0$. Similarly,

$$\int W_1 W_2 W_3 W_4 = 0, \quad \int W_1^3 W_2 = 0.$$

This shows finally that each summand of each of the second, fourth and fifth sums on the right side of (8) (in Part 1) do integrate to zero. This completes the proof of our statement that for a typical number the frequency of the digit 5 in its decimal expansion is 1/10.

What about proportion of the pattern: digit 5 followed by digit 7.

There is nothing special about the digit 5. Suppose i is any integer $0 \leq i \leq 9$ and $N_n(x)/n$ is the proportion of the digit i in the first n decimal places of x. Of course, N_n now depends not only on x but also on i. Then this proportion $N_n(x)/n$ converges to $1/10$ for almost all points x. That is, if A_i is the set of those points x for which this does not converge to $1/10$, then length$(A_i) = 0$. If $A = \cup_i A_i$, we see that length$(A) = 0$. If $x \notin A$ then whatever be the digit i, $0 \leq i \leq 9$ the proportion of the digit i is $1/10$ in the decimal expansion of x.

Thus for a typical number, each digit occurs with the right frequency in its decimal expansion.

6. More Patterns

We can now get a little more ambitious. What about the proportion of the pair $\langle 5, 7\rangle$? As expected it is $1/100$. Of course, we have to again make the question precise. Let us consider any finite sequence of decimal digits,

$$s = \langle a_1, a_2, \cdots, a_k\rangle; \quad a_i \in \{0, 1, 2 \cdots, 9\} \text{ for } 1 \leq i \leq k.$$

Such a finite sequence of decimal digits is also called a *pattern*. We define length of the sequence s as k and also denote it by $|s|$. Clearly, there are 10^k such sequences of length k. Let us fix now one such sequence s. If $x \in (0, 1]$ with decimal expansion $x = \cdot x_1\ x_2\ \cdots$, we put

$$\begin{aligned} N_n(x) \quad = \quad & \#\{1 \leq i \leq n : x_i = a_1,\ x_{i+1} = a_2, \\ & \qquad \cdots,\ x_{i+k-1} = a_k\}. \end{aligned}$$

Thus $N_n(x)$ is the number of times the pattern s appears in the decimal expansion of x starting somewhere in the first n decimal places. Equivalently, it is the number of times the pattern s occurs in the first $n + k$ decimal places. We used the same notation N_n as earlier and you should not get confused.

We now claim that for almost all numbers x, the proportion $N_n(x)/n$ converges to $1/10^k$. The proof, plan and execution, goes along similar lines as in the case of one digit carried out above. The definition of Z_i is similar, namely,

$$Z_i(x) = 1 \ \text{ if } \ \langle x_i, x_{i+1}, \cdots, x_{i+k} \rangle = s;$$
$$Z_i(x) = 0 \ \text{ otherwise.}$$

Every pattern occurs with right frequency.

The difference in calculations is the following. Now Z_i takes the value one with probability $1/10^k$ instead of $1/10$. Accordingly, we define $W_i = Z_i - (1/10^k)$. Thus, as earlier, we need to estimate

$$\frac{1}{n^4} \int \left(\sum_{i=1}^{n} W_i \right)^4 ,$$

and then sum over n in order to arrive at (6) (in Part 1).

The variables Z_i are not independent, but for any i, Z_i and Z_{i+k+1} are independent and in particular most of the terms on the right side of (8) (in Part 1) do integrate to zero. The ones that do not integrate to zero are bounded by one (in modulus) and there are at most $n^2 P(k)$ many such terms, where P is an explicitly computable polynomial of degree at most 4. As mentioned at the end of Section 12 (in Part 1), this is enough to deduce the required result. You can try your hands on $k = 2$, that is, pattern of length two.

Thus if you fix a finite sequence s of decimal digits, there is a set A_s of length zero such that for $x \notin A_s$, the pattern s occurs with frequency $1/10^{|s|}$ in the decimal expansion of x. Since length is additive, the countable union $A = \cup_s A_s$ also has length zero. For $x \notin A$ any given pattern s of decimal digits occurs with frequency $1/10^{|s|}$. The difference between this sentence and the earlier sentence is this. Earlier we said that if we take a pattern s, there is a set of length one (which depends on the pattern s) such that for all x in this set the pattern

All this holds for expansion to any base.

occurs with the right frequency. Now we are saying that there is a set of length one (without reference to any pattern) such that whatever pattern s you now take, it occurs with the correct frequency for all points in this set.

Thus for a typical number any pattern of decimal digits occurs with the right frequency, namely, $1/10^k$ where k is the length of the pattern.

7. Expansions with Other Base

Fix any integer $r \geq 2$. Given any number $0 < x \leq 1$ we can get $x_1, x_2, \cdots$ such that each x_i is one of the digits $\{0, 1, 2 \cdots, r-1\}$ and

$$x = \frac{x_1}{r} + \frac{x_2}{r^2} + \frac{x_3}{r^3} + \cdots . \tag{5}$$

This is called expansion of x to the base r, or simply r-expansion. When $r = 2$, it is called *binary* expansion; when $r = 3$, it is called *ternary* expansion; when $r = 10$, it is decimal expansion. The proof of existence of such an expansion is imitation of decimal case of Section 2 in Part 1. Further there can be at most two expansions. If there are two expansions, then one of them must be terminating and the other non-terminating. As earlier, terminating means that after some stage all the digits equal zero and non-terminating means that infinitely many digits are different from zero. If there are two expansions, one can show that in the non-terminating expansion, all digits after some stage equal $r-1$. Thus there are only countably many numbers having two r-expansions. By an r-pattern we mean a finite sequence s consisting of digits from among $\{0, 1, 2 \cdots, r-1\}$. The proof, word for word, carries over to conclude that there is a set A_r of length zero with the following property. Given any r-pattern s, it occurs with frequency $1/r^{|s|}$ in the r-expansion of x for any $x \notin A_r$.

Finally, by considering the countable union $A = \cup_{r \geq 2} A_r$, we get a set of length zero such that the following happens. If we take any $x \notin A$, take any integer $r \geq 2$, and consider any r-pattern s; then the r-pattern s occurs with frequency $1/r^{|s|}$ in the r-expansion of x. Thus we have the following.

Champernowne number and Copeland–Erdös number are normal.

For a typical number x, whatever be the base r, whatever be the r-pattern s, in the r-expansion of x the pattern s occurs with the right frequency, namely $1/r^k$, where k is the length of the pattern.

Such numbers x for which this happens are called 'absolutely normal' numbers. If for some integer $r \geq 2$, each r-pattern appears with the right frequency in the r-expansion of x, then x is said to be 'normal to the base' r. When $r = 10$, the r-normal numbers are usually referred to as normal numbers.

8. What Next?

The considerations above are essentially due to the French mathematician Emile Borel at the beginning of the last century. Even though almost all numbers are absolutely normal, one has to work hard to construct such numbers! Here are some well-known numbers which are normal. Of course, there are others.

Champernowne number: This is the number whose decimal expansion consists of all positive integers written in the usual order.

$$0 \cdot 1234567891011121314151617181920212223 \cdots\cdots$$

Copeland–Erdös number: This is the number whose decimal expansion consists of all primes among natural numbers written in increasing order.

$$0 \cdot 235711131719232931374143475359616 7 \cdots\cdots$$

These are normal to the base 10. However it is not known if they are absolutely normal, that is, if they are normal to every base. There is a number that arises in

Chaitin constant is absolutely normal.

the study of algorthmic complexity in computer science, known as Chaitin constant (Gregory John Chaitin). It is the chance that a randomly chosen prefix-free program halts. It was shown by Cristian Calude [2] that this number is absolutely normal. See the book of Calude [2] for relevant definitions.

Of course, no rational number is normal. More than hundred years ago Borel conjectured that *non-rational algebraic numbers* are normal. We have no clue, either way, regarding the conjecture. The numbers that you and I pretend to know well – like $\sqrt{2}$ or π or e – are they normal? No one knows! Worse still, some number theorists I talked to tell me that we do not know if, for example, the digit zero appears infinitely many times in the decimal expansion of π.

So where are the typical numbers hiding? Are the numbers we are familiar with not typical?

Address for Correspondence
B V Rao
Chennai Mathematical Institute
Plot H1, SIPCOT IT Park
Siruseri, Padur Post
Chennai 603 103, TN, India.
Email: bvrao@cmi.ac.in

Suggested Reading

[1] P Billingsley, *Probability and Measure*, Wiley India, 2008.

[2] C Calude, *Information and Randomness – An Algorithmic Perspective*, Springer, 2002.

[3] G H Hardy and E M Wright, *An Introduction to the Theory of Numbers*, English Language Book Society/Oxford University Press, 1975.