# Where Are All the Typical Numbers - 1

*B V Rao*

**The entire discussion below, except the first few paragraphs, takes place on the interval $(0, 1]$, so you can relax. In this two-part article, we show that decimal digits of numbers obey several patterns.**

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

## 1. Before the Beginning

You know that every number is either *rational* or *irrational*. Recall that a number is called rational if it is a ratio of two integers, that is, of the form $m/n$, where $m$ and $n$ are integers and $n \geq 1$. Since several ratios represent the same rational number – 3/6, 2/4, 18/36, all represent 1/2 – it is customary to consider such ratios where $m$ and $n$ have no common factors (except $\pm 1$). Of course, if you want the rational number to belong to the interval $(0, 1]$, as advertised at the beginning, the integer $m$ must satisfy $0 < m \leq n$. Recall that a set is countable if it is either an empty set or is in one-to-one correspondence with a subset of natural numbers. Otherwise the set is said to be uncountable. We can show that the set of rational numbers is a countable set by simply making the following correspondence with a subset of non-negative integers: the rational number $(m/n)$ where $m \geq 0$ corresponds to the integer $2^m 5^n$ (note that $n$ is a strictly positive integer); the rational number $(-m/n)$ where $m > 0$ corresponds to the integer $3^m 5^n$. But there are uncountably many numbers in the interval $(0, 1]$. Thus we are justified in feeling that a typical number is indeed irrational.

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

There is another classification of numbers: *algebraic* and

There are only countably many algebraic numbers.

*transcendental.* Recall that a real number $a$ is algebraic if there is a polynomial in one variable, say $P(x)$, with integer coefficients such that $P(a) = 0$. In other words, an algebraic number is a number that satisfies a polynomial equation with integer coefficients. The rational number $m/n$ satisfies $nx - m = 0$ and thus, every rational number is algebraic. There are many others. For example, the number $\sqrt{2}$ is algebraic because it satisfies the equation $x^2 - 2 = 0$. So is the number

$$(\sqrt[5]{3} + \sqrt{5}) \Big/ (\sqrt[31]{13} - \sqrt{11})$$

though it takes some time to exhibit a polynomial with integer coefficients satisfied by this number. Numbers which are not algebraic are called transcendental. For example, the number

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \cdots$$

is transcendental. So is the number $\pi$, namely, the smallest number $x > 0$ such that $\cos(x/2) = 0$. But to prove that these numbers are transcendental is not easy.

The set of algebraic numbers is again a countable set. This can be seen as follows. Polynomials with integer coefficients are in correspondence with finite, non-empty, sequences of integers

$$\langle n_0, n_1, \cdots, n_k \rangle \quad \longleftrightarrow \quad n_0 + n_1 x + n_2 x^2 + \cdots + n_k x^k.$$

So to prove that the set of polynomials with integer coefficients is a countable set, it suffices to show that the set $S$ of finite, non-empty, sequences of integers is a countable set. But $S = \cup_k S_k$, where $S_k$ is the set of sequences of integers of length exactly $k$. Suppose we have shown that each $S_k$ is a countable set and in fact $f_k : S_k \to \{1, 2, \cdots\}$ is a one-to-one map on $S_k$ onto a subset of natural numbers. We define $f$ on $S$ as follows. Take $s \in S$, then there is a unique $k$, namely length of $s$, such that $s \in S_k$. Let $m = f_k(s)$. Put $f(s) = 2^k 3^m$. This sets up a one-to-one map of $S$ with a subset of natural numbers. Incidentally, this argument shows that a countable union of sets, each of which is countable, is again a countable set. The only issue is that in the present case the sets $S_k$ are disjoint, which may not be true in general. When that happens, you only need to choose for $s$ in the union, the smallest $k$ such that $s \in S_k$ and use the same algorithm to define a function on $S$.

Here is how we show that $S_k$ is countable. Let $p_1, p_2, \cdots, p_{2k}$ be the first $2k$ primes starting with 2. Thus $p_1 = 2,\ p_2 = 3,\ p_3 = 5,\ p_4 = 7,\ p_5 = 11,\ \cdots$. If all the $n_i$ are non-negative integers, then make the corespondence

$$\langle n_1, n_2, \cdots, n_k \rangle \quad \longleftrightarrow \quad p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}.$$

If an integer, say $n_i$ in the sequence is negative, then use $(p_{k+i})^{|n_i|}$ instead of $p_i^{n_i}$ in the above correspondence. Then this is a one-one correspondence of $S_k$ with a subset of integers, showing that $S_k$ is a countable set.

All this shows that the set of polynomials with integer coefficients is countable. Since each polynomial has finitely many roots, we conclude that the set of algebraic numbers is a countable union of sets, each of which is finite; and so is itself a countable set. Thus a 'typical' number is indeed transcendental.

We discuss another typical behaviour of numbers. This concerns patterns among their decimal digits. We show that typical numbers exhibit interesting patterns. Of course now the word 'typical' no longer means 'complement of a countable set' as happened in the earlier two situations.

Typical numbers exhibit interesting patterns.

Every number has a decimal expansion.

## 2. Decimal Expansion

First observe that given any number $0 < x \leq 1$, we can get integers $x_1, x_2, \cdots$ such that each $x_i$ is one of the digits $\{0, 1, 2 \cdots, 9\}$ and

$$x = \frac{x_1}{10} + \frac{x_2}{10^2} + \frac{x_3}{10^3} + \cdots . \quad (1)$$

This is also expressed as

$$x = 0 \cdot x_1\, x_2\, x_3\, \cdots$$

and is called the *decimal expansion* of the number $x$. The integer $x_n$ is called the $n$-th decimal digit of $x$. Since later we plan to replace 10 by a positive integer $r > 1$ and expand numbers in powers of $1/r$, it is well worth recalling how the above decimal expansion is obtained.

Given a number $x$ with $0 < x \leq 1$, here is an algorithm for obtaining its decimal expansion. Denote by $I_k$ the interval

$$I_k = \left(\frac{k}{10}, \frac{k+1}{10}\right]; \quad k = 0, 1, \cdots, 9.$$

These intervals are disjoint, each having length 1/10. They make up all of (0,1]. Thus $x$ must be in exactly one of these intervals. Let that interval be $I_k$ and put $x_1 = k$. Since

$$\frac{x_1}{10} < x \leq \frac{x_1 + 1}{10},$$

we see that

$$0 < x - \frac{x_1}{10} \leq \frac{1}{10}.$$

Now divide this interval $I_k$ into ten parts by conisdering

$$I_{kl} = \left(\frac{k}{10} + \frac{l}{10^2}, \frac{k}{10} + \frac{l+1}{10^2}\right]; \quad l = 0, 1, \cdots, 9.$$

Since $x \in I_k$ and the intervals $\{I_{kl} : 0 \le l \le 9\}$ are disjoint making up all of $I_k$, there is exactly one $l$ such that $x \in I_{kl}$. Put $x_2 = l$ and immediately observe, as earlier, that

$$0 < x - \frac{x_1}{10} - \frac{x_2}{10^2} \le \frac{1}{10^2}.$$

By induction one can obtain the required digits and show the stated properties too by using the inequalities deduced at each stage.

Thus we have exhibited a decimal expansion for every number in $(0, 1]$. But remember that the only thing needed for a decimal expansion is that the equation (1) be satisfied. Sometimes there are two such expansions. For example,

$$\frac{3}{10} + \frac{0}{10^2} + \frac{0}{10^3} + \cdots = \frac{3}{10} = \frac{2}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \cdots.$$

The algorithm outlined above will give us the second expansion. This is called the non-terminating expansion and the first is called the terminating expansion. More precisely, an expansion is called terminating if all the digits are zero after some stage, while an expansion is called non-terminating if infinitely many digits are different from zero. It is easy to see that a number can have at most two decimal expansions. Moreover if it has two expansions, one of them must be non-terminating and the other terminating. One can be more precise. Suppose there are two such expansions

$$a; = \cdot x_1\, x_2\, x_3 \, \cdots = \cdot y_1\, y_2\, y_3 \, \cdots ,$$

where the $x$-expansion is terminating and the $y$-expansion is non-terminating. Then there must be an integer $n \geq 1$, such that the following holds.

$$x_i = y_i \text{ for } i < n; \quad x_n = y_n + 1; \quad x_j = 0, \quad y_j = 9 \text{ for} j > n.$$

A number can have at most two decimal expansions.

Which set of numbers is large: having digit 5 or not having digit 5?

It follows, in particular, that all but countably many numbers have exactly one decimal expansion. Indeed, if a number $x$ has two expansions, then one of them must be terminating, hence ends with zeros. Put $f(x)$ to be the finite sequence of decimal digits upto and including the last non-zero digit in this terminating expansion. Then $f(x)$ is a finite sequence of integers, each belonging to $\{0, 1 \cdots, 9\}$ and the sequence ends with a non-zero digit. This is a one-to-one map from the set of numbers with two decimal expansions to a subset of finite sequences of integers. But the latter set, as seen above, is countable.

## 3. Occurrence of Digit 5

We are interested in the pattern of the decimal digits of numbers. For instance, let $A$ be the set of numbers in whose decimal expansion, the digit 5 occurs. Which is large: $A$ or $A^c$? But how do you measure largeness? We shall shortly see that both the sets have the same number of elements, that is, they have the same cardinality. Before we show this, let us clarify the terms involved.

Recall that two sets have the same number of elements – equivalently, have the same cardinality – if there is a one-one map defined on one of the sets having range all of the other set. A set $B$ is at least as large as $A$ if we can find a one-one map $f : A \to B$. Here $f$ is *not necessarily onto* $B$. The idea is that if you look at the elements $\{f(a) : a \in A\} \subset B$, you do feel that $B$ has at least as many elements as the set $A$. In such a case, let us say that $A \leq B$. This is only a suggestive notation for the phrase we used and nothing more than that. However, it is a non-trivial and useful theorem that if $A \leq B$ and $B \leq A$, then indeed $A$ and $B$ have the same number of elements. One interesting consequence of this fact is the following. If you want to show that two sets have the same number of elements, it is not necessary to define a one-one function on one of the sets with range all of the other set. It suffices to show that each set is at least as large as the other. It is easy to show that two non-empty open intervals $(a, b)$ and $(c, d)$ have the same number of elements. However, the criterion mentioned above will be helpful in showing that the intervals $(a, b)$ and $[c, d]$, with $a < b$ and $c < d$, also have the same number of elements. They are said to have the cardinality of the continuum.

Returning to our problem of the previous paragraph, all numbers in the interval $(\frac{5}{10}, \frac{6}{10})$ must have five as their first decimal digit. Thus the set of numbers having 5 in their decimal expansion has the cardinality of the continuum. On the other hand consider all the decimal expansions with only the digits 0, 1 and having infnitely many ones. All these numbers do not have 5 in their decimal expansion. However this set is also of the cardinality of the continuum, simply because the map

$$x = \cdot x_1 \, x_2 \, \cdots, \quad \longleftrightarrow \quad f(x) = \sum \frac{x_i}{2^i},$$

sets up a correspondence with the interval $(0, 1]$. Thus counting the number of elements is of no indication to decide which of the sets we are considering is larger. We should use a different scale of measurement.

## 4. Length

Recall the concept of length. An interval $(a, b]$ has length $b - a$. Length of a single point set is zero. In a sense the interval $(0, \frac{1}{4})$ is smaller than the interval $(\frac{1}{2}, 1)$ just because the former set has a smaller length. However, note that both sets have the same number of elements, in fact, $f(x) = 1 - 2x$ sets up a one-one correspondence

Length provides a way of measuring which set is large.

Can you define length for all subsets?

from the interval $(0, \frac{1}{4})$ onto the interval $(\frac{1}{2}, 1)$. Thus even if two sets have the same number of elements, the concept of length provides us a way of measuring which is large. So we shall use the notion of length to compare sets.

We will be dealing with slightly more complicated sets than just intervals. The only thing you should accept about the notion of length is the following: (i) length of an interval $(a, b]$ is $b-a$; (ii) length is countably additive, that is, for a sequence of disjoint sets $(A_n)$ with $\cup_n A_n = A$, we have $\text{length}(A) = \sum \text{length}(A_n)$. You should have no problem accepting property (i). However you may not be too willing to accept property (ii). But be assured that the following is a mathematical fact: If you have a sequence of disjoint intervals $(I_n)$ whose union is again an interval $I$ then the lengths of $I_n$ do add up to the length of $I$. At this stage several questions come to mind. In particular, can you define such a notion of length for *all* subsets of $(0, 1]$? We shall not enter into this discussion because it will take us too far away from our goal. It is enough to know that the sets we come across in our discussion are so simple that they will all have length.

Returning to our problem, let us see lengths of $A$ and $A^c$. Recall that $A$ is the set of numbers in whose decimal expansion the digit 5 appears. The sets $A$ and $A^c$ are not intervals. Luckily these sets are not complicated either and a little careful thought allows us to evaluate their lengths.

The set $A$ is a union of countably many intervals. In fact, $A = \cup A_n$, where $A_n$ is the set of numbers in whose decimal expansion the digit 5 does not appear in the first $(n-1)$ places but appears at the $n$-th place. These sets are disjoint. Since lengths add up, let us see if we can calculate lengths of these sets $A_n$. The set $A_1 = I_5$ which has length 1/10. The set $A_2$ is not an interval, but is union of finitely many intervals – more precisely, 9 intervals each of length $1/10^2$, namely $I_{k5}$ for $k \neq 5$. So length of $A_2$ equals $9/10^2$. Similarly $A_3$ is made up of the 81 intervals $I_{kl5}$, where $k \neq 5$ and $l \neq 5$ and hence has length $9^2/10^3$. In general, it is not difficult to see that $A_n$ has length $9^{n-1}/10^n$. Indeed, $A_n$ is the union of intervals $I_{s5}$, where $s$ varies over sequences of length $(n-1)$ consisting of decimal digits other than 5. Thus if we add all these lengths we get

For a typical number, the digit 5 occurs in its decimal expansion.

$$\frac{1}{10} + \frac{9}{10^2} + \frac{9^2}{10^3} + \cdots = 1.$$

Thus length of $A$ equals one. Since length is additive and $I$ has length one, we conclude that $A^c$ has length zero. We have now found that for 'almost all' numbers the digit 5 occurs in their decimal expansion, that is, the set of such numbers has length one. And of course consequently, the complementary set has length zero. This is expressed by saying that typical numbers have the digit 5 in their decimal expansion. Of course, there is nothing special about the digit 5.

## 5. Infinitely Many Occurrences of Digit 5

Let us get a little more ambitious. How many numbers have the digit 5 at infinitely many decimal places? It is difficult to handle this set, but its complement is easy to handle, namely the set $B$ of all those numbers in whose expansion 5 appears only finitely many times. This is because, $B = \cup B_n$ where $B_n$ is the set of numbers in whose expansion the digit 5 appears at exactly $n$ decimal places. If each $B_n$ has length zero then $B$ too would have length zero. This is again because, lengths add up even

For a typical number, the digit 5 occurs infinitely many times in its decimal expansion.

for an infinite sequence of sets, provided the sets are disjoint.

In turn, we can write $B_n$ as a countable union of simpler sets as follows. Let $s = \langle j_1, j_2, \cdots, j_n \rangle$ be a sequence of length $n$ consisting of positive integers in strictly increasing order. For such a finite sequence $s$, let $B_s$ be the set consisting of those numbers for which the digit 5 occurs in the places $j_1, j_2, \cdots, j_n$ and at no other decimal places. Then, $B_n = \cup B_s$ where the sum extends over all sequences as described above. Indeed, if the digit occurs exactly $n$ times then there must be $n$ places – which can be arranged in increasing order – where the digit appears and nowhere else. If we show that each of these sets $B_s$ has length zero, then it follows that $B_n$ would also have length zero. You only need to note that the set of such sequences is countable and these sets $B_s$ are disjoint and lengths add up.

As an illustration, consider the set $S$ of points for which the first $n$ places are 5 and the digit 5 occurs nowhere else. If you take a number $z$ in whose expansion the digit 5 does not appear then $\sum_1^n 5/10^j + z/10^n$ is in $S$. Conversely, if you take a number $x$ in $S$, since its first $n$ decimal places are 5 it makes sense to subtract $\sum_1^n 5/10^j$ from $x$ to get a number $w$ in whose decimal expansion the first $n$ digits are zero. Thus $10^n w = z \in I$ and the given number $x$ equals $\sum_1^n 5/10^j + z/10^n$. This argument shows the following. Let $S'$ be the set of numbers in whose decimal expansion the digit 5 does not appear. Then $S$ is obtained by translating $S'$. But, as seen above, this last set $S'$, has length zero and hence so is its translate. This argument can be carefully formulated to work for any of the $B_s$ by using induction on the length of $s$.

The reader can try his hands on this. However you need not be alarmed because we do not use this fact in what follows.

Nature did not show any bias towards a particular digit.

Thus for a typical number infinitely many decimal digits equal 5. Of course there is nothing special about the digit 5. Thus one can say that for a typical number, each of the digits appears infinitely many times in the decimal expansion. This simply means that the set of such numbers has length one and hence its complement (in the interval $(0, 1]$) has length zero. Can we say anything more? Yes. Nature did not show any bias towards a particular digit, the relative proportion is one-tenth for each of the ten digits. To make this precise, we should first understand what it means to say 'proportion' because there are infinitely many decimal places.

## 6. Frequency of Digit 5

For a number $x$, let us denote by $N_n(x)$ the number of occurrences of the digit 5 in its first $n$ decimal places. Then $N_n(x)/n$ denotes frequency or proportion of the digit 5 among the first $n$ decimal places. Let us say that a number is 5-regular if this ratio has a limit. It is easy to construct numbers which are not 5-regular. In fact make a number by prescribing its decimal digits as follows. Take the first digit to be 1. Then put 5 until the frequency of 5 exceeds $1/2$. Thus you put 5 in the second and third decimal places. Then put ones till the proportion of 5 falls down to $1/4$. Then put 5 till the proportion of 5 exceeds $1/2$. Continue like this forever. This gives us decimal expansion of a number, say, $x$. The number $x$ so obtained is not 5-regular simply because $N_n(x)/n$ does not have a limit – these proportions oscillate between $1/4$ and $1/2$.

We shall now show that a typical number is 5-regular

For a typical number, frequency of each digit is one tenths.

and indeed the frequency is 1/10. That is, for a typical number $x$, $N_n(x)/n \to 1/10$, which means the set of numbers for which this holds has length one. Or equivalently, the set of numbers which fail this property has length zero. Remember that failure of this property can happen in two ways: either the frequencies $N_n(x)/n$ do not have a limit; or have a limit different from 1/10. In the earlier para, we saw that there are numbers $x$, for which the frequencies $N_n(x)/n$ do not have a limit. It is not difficult to show that such numbers have again the cardinality of continuum. It is easier to construct numbers for which these proportions do indeed converge, *but* to a number different from 1/10. For instance, put every seventh decimal digit as 5 and others as zero.

## 7. Limsup of Sequence of Sets

How do we recognize when a set has length zero. We describe one criterion. First, we need some notation. Suppose we have a sequence of sets $(A_n)$. We define the limit superior of these sets, denoted by $A^*$, to be the set of all those points which belong to infinitely many of the $A_n$. In other words, those points $x$ which satisy the following: for every positive integer $n$ there is an $m > n$ such that $x \in A_m$. In fact, if $x$ belonged to infinitely many of the sets then this clearly holds. Conversely, if this holds, take $n = 1$ and get $p > 1$ such that $x \in A_p$; take this $p$ as $n$ and get $q > p$ such that $x \in A_q$; take this $q$ as $n$ and get $r > q$ such that $x \in A_r$. Proceeding this way you can show that $x$ is in infinitely many of the sets $A_n$. In other words, to use symbols, $x \in A^*$ iff for any $n$, $x \in \cup_{m>n} A_m$. Or simply put $A^* = \cap_n \cup_{m>n} A_m$.

## 8. Recognizing Zero Length

Here is then a way to recognize when a set has length zero. Suppose we have a sequence of sets $(A_n)$, all sets contained in $(0,1]$.

$$\text{If } \sum \text{length}(A_n) < \infty \quad \text{then } \text{length}(A^*) = 0. \tag{2}$$

If lengths add up to a finite number, then points cannot belong to infinitely many sets.

This is known as *Borel–Cantelli Lemma*, a simple but powerful criterion. To see its truth, fix $\epsilon > 0$, we show length$(A^*) < \epsilon$. Since the series $\sum$ length$(A_m) < \infty$, tail sums of this series converge to zero. Here we used the fact that if a series $\sum_n a_n$ of numbers converges then the tail sums $\sum_{m>n} a_m$ converge to zero as $n$ becomes large. So fix $n$ such that $\sum_{m>n}$ length$(A_m) < \epsilon$. Since $A^* \subset \cup_{m>n} A_m$, we see

$$\text{length}(A^*) \leq \text{length}(\cup_{m>n} A_m) \leq \sum_{m>n} \text{length}(A_m) \leq \epsilon.$$

Here the middle inequality needs justification. But this is easy. Suppose we have sets $(B_i : i \geq 1)$, not necessarily disjoint, and $B = \cup B_i$. For each $i \geq 1$, define $C_i$ to be the set of those points which are in $B_i$ but not in any of the $B_j$ for $j < i$. Clearly these sets $C_i$ are disjoint and their lengths add up. Further, $\cup C_i = \cup B_i$. Since $C_i \subset B_i$, we also have length$(C_i) \leq$ length$(B_i)$. Thus,

$$\begin{aligned} \text{length}(\cup B_i) = \text{length}(\cup C_i) &= \sum \text{length}(C_i) \\ &\leq \sum \text{length}(B_i). \end{aligned}$$

## 9. Reduction of the Problem

Returning to our problem, suppose we could show the following: for every fixed $\epsilon > 0$.

$$\sum_n \text{length}\left(x : \left|\frac{N_n(x)}{n} - \frac{1}{10}\right| > \epsilon\right) < \infty. \tag{3}$$

Denote the set in braces above by $A_n(\epsilon)$. Denote by $A^*(\epsilon)$ the limsup of the sequence of sets $\{A_n(\epsilon) : n \geq 1\}$.

Then (2) and (3) tell us that $A^*(\epsilon)$ has length zero. Using the values $\epsilon = 1, 1/2, 1/3, \cdots$, let us now take the countable union $A = \cup_k A^*(1/k)$. Being countable union of sets of length zero, the set $A$ has length zero. Thus length of $A^c$ is one. If $x \in A^c$, then whatever be the integer $k \geq 1$, we have $x \notin A^*(1/k)$. The definition of limsup of sequence of sets now tells us that for only finitely many values of $n$ we can have $|\frac{N_n(x)}{n} - \frac{1}{10}| > 1/k$. Thus for all large values of $n$, we have $|\frac{N_n(x)}{n} - \frac{1}{10}| \leq 1/k$. This being true for every integer $k \geq 1$, we see that the ratios $N_n(x)/n$ converge to $1/10$. This is so for every $x \in A^c$ and this latter set has length one. This shows that a typical number is indeed 5-regular and in fact the frequency of the digit 5 is $1/10$. Of course, we still need to show (3).

## 10. Estimating Length

Towards this end, we realize that we need an estimate for lengths of sets that appear in (3). This is provided by the following inequality. Suppose $f$ is a nice function defined on $(0, 1]$. Then

$$\text{length}(x : |f(x)| > \epsilon) \;\leq\; \frac{1}{\epsilon}\int_0^1 |f|, \quad \text{for any } \epsilon > 0. \quad (4)$$

This is known as *Markov's inequality*, a simple yet powerful inequality. It saves the day in many a calculation. Do not worry about the word 'nice function'. We did not mean in the sense of real analysis, like continuous or differentiable. We are going to apply it only to functions which take finitely many values on finitely many intervals. For such functions the integral is easy to calculate.

For a non-negative function, if integral is small, function cannot be large.

You must have noticed that in writing the integral we did not show explicitly the variable of integration. This is what we do, at times, later too. If you are familiar with Lebesgue integral, you would even ask if we are using Lebesgue integral or Riemann integral. As mentioned already, our functions are so simple that they take finitely many values on finitely many sub-intervals of $(0,1]$ and as a consequence, you need not use any complicated theories.

To see the truth of the inequality (4), let $A$ denote the set in braces and $I_A$ denote its indicator function, that is, the function which takes the value one for points in the set $A$ and zero for points outside the set $A$. Then we have $\epsilon I_A \leq |f|$. Such an inequality for functions means that if you take a point and evaluate the functions at this point on both sides, then the inequality holds. A smaller function has a smaller integral, right? Integrating both sides of this inequality, we get $\epsilon$ length$(A) \leq \int |f|$ which is the stated inequaltiy.

The above inequality can be used in several variants, whichever is useful in a given instance. For example, one could apply this inequality to $f^2$ and $\epsilon^2$ leading to

$$\begin{aligned} \text{length}(x : |f(x)| > \epsilon) &= \text{length}(x : |f(x)|^2 > \epsilon^2) \\ &\leq \frac{1}{\epsilon^2} \int f^2. \end{aligned}$$

In a similar way,

$$\text{length}(x : |f(x)| > \epsilon) \leq \frac{1}{\epsilon^4} \int f^4. \tag{5}$$

## 11. Further Reduction of the Problem

In view of (5), the inequality (3) follows if we show

$$\sum_n \int \left[\frac{N_n(x)}{n} - \frac{1}{10}\right]^4 \mathrm{d}x < \infty. \tag{6}$$

Here is a convenient way of expressing the quantity of our interest, namely, $\frac{N_n(x)}{n} - \frac{1}{10}$. We define a function $Z_j$

on the interval $(0,1]$. Put $Z_j(x) = 1$ if the $j$-th decimal digit of $x$ is 5; otherwise $Z_j(x) = 0$. Remember, the set $\{x : Z_j(x) = 1\}$ is made up of finitely many intervals. In terms of these new functions, $N_n(x)$ is nothing but $\sum_1^n Z_j(x)$, and

$$\frac{N_n(x)}{n} - \frac{1}{10} = \frac{1}{n}\sum_{j=1}^{n}\left(Z_j(x) - \frac{1}{10}\right) = \frac{1}{n}\sum_{j=1}^{n} W_j(x), \quad \text{say.} \tag{7}$$

Thus, returning to the claimed inequality (6), we need to estimate

$$\frac{1}{n^4}\int\left(\sum_{k=1}^{n} W_k(x)\right)^4 \mathrm{d}x,$$

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

## 12. Calculations

Towards this end note that, if we have numbers $a_1, a_2, \cdots, a_n$; then

$$\left(\sum_i a_i\right)^4 = \sum_{i,j,k,l} a_i a_j a_k a_l.$$

The summands here can be organized depending on how many distinct indices appear in the term. If all the suffixes are same, we get $a_i^4$. If there are only two distinct suffixes then we get terms $a_i^3 a_j$ and $a_i^2 a_j^2$. If there are three distinct suffixes, we get terms $a_i^2 a_j a_k$. Finally if all the suffixes are distinct, we get product of four distinct $a$'s. Thus,

$$\left(\sum_1^n W_k\right)^4 = \sum_1^n W_i^4 + 4\sum_{i\neq j} W_i^3 W_j + 6\sum_{i<j} W_i^2 W_j^2 + 6\sum_{i\neq j\neq k} W_i^2 W_j W_k + \sum_{i\neq j\neq k\neq l} W_i W_j W_k W_l. \tag{8}$$

We now need to integrate the quantities on the right side of (8). Since $Z_j$ takes only the values zero and one, $W_j = Z_j - \frac{1}{10}$ assumes values: $-1/10$ and $9/10$. In any case $|W_j| \leq 1$ and hence

$$0 \leq \int W_i^4 \leq 1.$$

Thus the first term in (8), after integration, is at most $n$. For the same reason the third term is at most $6n^2$. We shall now show that each of the other three terms contribute zero. If this is done then it follows that

$$\frac{1}{n^4} \int \left(\sum W_j\right)^4 \leq (n + 6n^2)/n^4 \leq 7/n^2. \qquad (9)$$

If we now sum over $n$, (7) shows that the inequality (6) holds. Here we have used the well-known fact that the series $\sum \frac{1}{n^2}$ converges. The proof of our statement is complete.

We now proceed to show that the second, fourth and fifth terms on the right side of (8) contribute zero. We show this by arguing that in each of the summations, each summand integrates to zero. Even though we show this, it is useful to note the following. Suppose we could only show that all except possibly $10^{99}n^2$ summands integrate to zero. This would also suffice for the above argument. This is because, as observed above, each of those summands that do not integrate to zero contribute at most one in modulus. You obtain $(7 + 10^{99})/n^2$ in place of $7/n^2$ in (9); in any case finiteness of the series involved holds.

We now need a truly new ingredient which helps understand the behaviour of the decimal digits.

*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