Proof of A236402 Natural Density

Xiaotong Chen

April 2018

Lemma 1. *Suppose A is a set of positive integers, and the natural density of A is 1. Then the natural density of any superset of A is 1.*

Proof. Let A be a set of positive integers. Suppose A has natural density 1. Let $x \in \mathbf{Z}_{>0}$. Let $A(x)$ be the number of elements of A that are less than or equal to x. Hence,

$$\lim_{x\to\infty} \frac{A(x)}{x} = 1$$

Suppose $B \subseteq \mathbf{Z}_{>0}$ such that $A \subseteq B$. Let $B(x)$ be the number of elements of B that are less than or equal to x. Then,

$$A(x) \leq B(x) \leq x$$

Therefore,

$$\frac{A(x)}{x} \leq \frac{B(x)}{x} \leq 1$$

Since $\lim\limits_{x\to\infty} \frac{A(x)}{x} = 1$, according to the Squeeze Theorem,

$$\lim_{x\to\infty} \frac{B(x)}{x} = 1.$$

Therefore, the natural density of B is 1. □

Lemma 2. *The set of all integers containing a given substring has natural density 1.*

Proof. Let S be a substring of length l. S takes the form of $a_1 a_2 \cdots a_l$ where a_1, a_2, ..., a_l are digits from 0 to 9.

To show the natural density of sequence with integers containing S is equivalent to show the probability of a randomly selected integer containing S. Hence, consider the process of randomly selecting a integer digit by digit from the most significant to the least significant. That is, each time we select a digit, we would append it to the right of previously selected ones. For each selection, every digit has a 0.1 chance of appearing.

The process of forming S contains $l+1$ states. Starting from the zero state where the most recently selected digit is not a_1, state 1 where the most recently selected digit is a_1, state 2 where two most recently selected digits are a_1 and a_2 in chronological order, ... state $k(k \leq l)$ where k most recently selected digits are a_1 to a_k in chronological order, and the process ends at state l when we get substring S.

Above mentioned process can be represented by a Markov chain. And the transition matrix is a $l+1$ by $l+1$ transition matrix that takes the form below:

$$T = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & \cdots & 0 \\ 0.8 & 0.1 & 0.1 & 0 & \cdots & 0 \\ 0.8 & 0.1 & 0 & 0.1 & \cdots & 0 \\ \vdots & \ddots & & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

where

$$\begin{aligned} &T_{1,1} = 0.9 \\ &T_{i,1} = 0.8 (i \neq 1) \\ &T_{i,2} = 0.1 (i \neq l+1) \\ &T_{i,i+1} = 0.1 \\ &T_{l+1,i} = 0 (i \neq l+1) \\ &T_{l+1,l+1} = 1 \end{aligned}$$

And $T_{i,j}$ represents the probability of transitioning from state $i-1$ to state $j-1$ in one step.

Hence, the entries of T^n, the nth power of matrix T, represents the probability of transitioning from state i to state j in k steps. Specifically, $T^n_{1,l+1}$, the last entry of the first column, represents the probability of transitioning from the zero state to the absorbing state, where the substring S appears, in n steps. This is equivalent to the probability of a n-digit integer(including the leading zeros if any) containing substring S. In other words,

$$T^n_{1,l+1} = \frac{N}{10^{k-1}}$$

where N is the number of n-digit integer(including the leading zeros if any) containing substring S.

Notice that the natural density of sequence with integers containing S is represented as,

$$\lim_{x \to \infty} \frac{N'}{x} = \lim_{k \to \infty} \frac{N}{10^{k-1}} = \lim_{x \to \infty} T^n_{1,l+1}$$

where x is the range of the sequence, and N' is the number of terms in the sequence.

From the expression above, it is clear that the natural density of sequence containing substring S is 1 iff $T^n_{1,l+1} = 1$.

Notice that for $T^n_{i,l+1}$, any entry of the last column T^n,

$$T^n_{i,l+1} \geq 0.8(0.1)^k$$

which means that we have a positive probability to transition from any state to state $l+1$, the absorbing state, in n steps

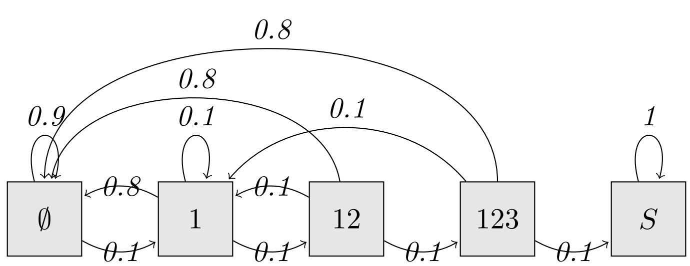
Therefore, this Markov chain is absorbing since it has one absorbing state, and from every state, there is a positive probability to go to the absorbing state. Based on the properties of absorbing Marbov chains, as $n \to \infty$, the probability of reaching the absorbing state is 1. Therefore, the natural density of sequence containing substring S is 1.

□

Example 1. *Now, let's look at an example of this process. For $S = 1234$, the process of forming S has 5 states, they are:*

1. *The most recently selected digit is not 1.*
2. *The most recently selected digit is 1.*
3. *The second most recently selected digit is 1, and the most recently selected digit is 2.*
4. *The third most recently selected digit is 1, the second most recently selected digit is 2, and the third most recently selected digit is 3.*
5. *The substring* 1234 *appears.*

We can represent this process using a Markov Chain, as shown below:



Hence, we obtain a 5 by 5 transition matrix for this Markov Chain:

$$T' = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 \\ 0.8 & 0.1 & 0.1 & 0 & 0 \\ 0.8 & 0.1 & 0 & 0.1 & 0 \\ 0.8 & 0.1 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Take an arbitrary integer power of this matrix to demonstrate the relationship between the last entry of the first row and the number of integers containing substring 1234*, below is the* 5*th power of matrix* T'*:*

$$T'^5 = \begin{bmatrix} \frac{17741}{20000} & \frac{9869}{100000} & \frac{247}{25000} & \frac{99}{100000} & \frac{1}{5000} \\ \frac{2739}{3125} & \frac{9753}{100000} & \frac{977}{100000} & \frac{97}{100000} & \frac{119}{100000} \\ \frac{977}{1250} & \frac{1087}{12500} & \frac{109}{12500} & \frac{11}{12500} & \frac{63}{6250} \\ \frac{8881}{12500} & \frac{247}{3125} & \frac{99}{12500} & \frac{1}{1250} & \frac{1251}{12500} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where $T'^5_{1,5} = \frac{1}{5000}$ *represents the probability to randomly select an at most 5-digit integer, and the selected number contains the substring* 1234*. Therefore, there are* $10^5 \times \frac{1}{5000} = 20$ *integers that contains the substring* 1234*. It is easy to check that there are indeed 20 at most 5-digit integers that contains the substring. Also, as the power of matrix* $n \to \infty$, $T'^5_{1,5} \to 1$, *Hence, the natural density of sequence contains substring* 1234 *is 1, which confirms Lemma 1.*

Theorem 1. *The sequence of positive integers A236402 is defined as numbers with the property where the sum of any pairs of adjacent digits is a substring of the number. The natural density of A236402 is 1.*

Proof. Let P be the set of every possible sum of two digits. Then,

$$P = \{0, 1, 2, \cdots, 18\}.$$

Let S be the concatenation of all the element of P in descending order. That is,

$$S = 181716 \cdots 3210.$$

Let A be the set of numbers where S is a substring of its digits. Let B be the set containing all terms of sequence A236402.

According to Lemma 2, the natural density of A is 1. Since B is a superset of A, the natural density of B is 1. Hence, the natural density of A236402 is 1. □

Theorem 2. *The sequence of positive integers A203565 is defined as numbers with the property where the product of any pairs of adjacent digits is a substring of the number. The natural density of A203565 is 1.*

Proof. Let P be the set of all integers in the range of product of two digits. Then,

$$P = \{0, 1, 2, \cdots, 81\}.$$

Let S be the concatenation of all the element of P in descending order. That is,

$$S = 818079 \cdots 3210.$$

Let A be the set of numbers where S is a substring of its digits. Let B be the set containing all terms of sequence A203565.

According to Lemma 2, the natural density of A is 1. Since B is a superset of A, the natural density of B is 1. Hence, the natural density of A203565 is 1. $\square$