university of groningen

faculty of mathematics
and natural sciences

The Constant of Champernowne

Bachelor Project Mathematics

July 2018

Student: G. Pomstra

First supervisor: Prof.dr. J. Top

Second supervisor: Dr. A.E. Sterk

Abstract

The constant of Champernowne first saw light when David G. Champernowne introduced it in 1933. He proved that the constant given by $0.123456789101112\ldots$ is normal in base ten. Four years later, Kurt Mahler proved that this number also is transcendental by showing the transcendence of constants with a similar structure to that of the constant of Champernowne. In addition to these two properties, the constant also has a peculiar continued fraction expansion. It namely contains exceptionally large terms throughout the expansion.

The contribution of this thesis will be to give a better understanding of the constant of Champernowne, its properties, the proofs and the techniques used for these proofs.

The notion of normality will be explained and proven in the way Champernowne did and with the use of the notion of (ϵ, k)-normality. Thereafter we will introduce the notion of transcendence and this will be proven for the constant following an argument of Kurt Mahler. Lastly, we will consider the procedure of continued fraction expansions and apply it to the constant. Some general notions for continued fraction expansions will be shown. With these notions, we show the connection between the fractions used in the argument of Kurt Mahler and the large terms of the continued fraction in the case of the constant of Champernowne.

Contents

1 Introduction

In this thesis, we will be considering the constant of Champernowne. The constant is named after the economist David G. Champernowne, who proved the normality of the constant in base ten in a paper he published in 1933. Champernowne lived from 1912 to 2000 and thus he published the paper as an undergraduate. Since the notion of normality is not trivial to prove for a constant, it is quite remarkable he proved it at that age. His knowledge of mathematics was probably not very extensive and indeed the proof is very intuitive and uses basic mathematical reasoning.

The constant for which Champernowne has proved normality in base ten and thus the constant we will be focusing on is given as follows.

$$C_{10} = 0.12345678910111213\ldots$$

Formally, the constant is defined as a concatenation of the natural numbers. The subscript indicates the base of the natural numbers we are considering for constructing the constant. If we for example take the constant with natural number in base two, we get the following constant.

$$C_2 = 0.11011100101110111\ldots$$

This constant is equal to $0.86224\ldots$ if we consider its value in base ten. For this thesis, we only will be considering the constant of Champernowne constructed with natural numbers in base ten.

The paper of Champernowne inspired mathematicians to further investigate the constant and constants with the same kind of structure. This, for example, led to the combinatorial method of normality proofs, which proofs the normality of constants with certain characteristics with the use of the notion of (ϵ, k)-normality. Other properties were also proven for the constant of Champernowne. Kurt Mahler proved that the constant is transcendental and it turned out that the constant has a remarkable continued fraction expansion which contains very large terms.

2 Normality

The paper of Champernowne, published in 1933, is concentrated on proving normality of some constants in base ten. Hence, we will begin with the normality property. Let us first consider an intuitive explanation of normality.

A constant is normal in base ten if every string constructed out of the digits $0, 1, \ldots, 9$ appears equally often as each other string with the same number of digits. In practice this means that a constant with infinite number of decimals is normal in base ten if, for example, the digit 1 appears equally often as the digit 2 and the string 025 appears equally often as the string 999. To be able to work with the notion of normality, we will consider the following formal definition of normality.

Definition 2.1. *A constant x is said to be normal in base g if, for every string s consisting of k digits of base g, we have*

$$\lim_{N\to\infty} \frac{v(x,N,s)}{N} = \frac{1}{g^k}, \tag{1}$$

where $v(x, N, s)$ denotes the number of times string s occurs in the first N digits of constant x considered in base g. [2, p.1]

To get a more intuitive feel to this definition we try it out on strings s of length 1, thus on digits, with base ten. We get

$$\lim_{N\to\infty} \frac{v(x,N,s)}{N} = \frac{1}{10^1}, \qquad s \in \{0, 1, \ldots, 9\}.$$

Thus each single digit string s will appear one tenth of the time when N tends to infinity. Since it doesn't depend on which single digit string we take, all possibilities will occur equally often. Analogously, this holds for every k. There will be 10^k different possibilities to create a strings with k digits. When we divide the occurrences of each of those strings in the first N digits of x by N, the result will tend to $1/10^k$ as N tends to infinity for any string of k digits.

Now we have defined normality, let us view how this is proven. Because we focus on the constant of Champernowne in base ten, we will work with that base in these proofs. We also need to set some notation for the following proofs.

c_i A sequence of digits consisting of all 10^i possible strings of i digits, ordered in the following way: $0\ldots00, 0\ldots01, \ldots, 9\ldots98, 9\ldots99$. These specific strings are called *members* of c_i. (Note that commas are used to emphasize the different strings, this is used throughout the paper.)

C^j A chain of sequences c_i where every sequence c_i is repeated j times in the following way: $C^j = c_1\ldots c_1 c_2\ldots c_2 c_3\ldots c_3\ldots$. For example
$C^1 = c_1c_2c_3\ldots = 0, 1, \ldots, 9, 00, 01, \ldots, 99, 000, 001, \ldots$.

C An ordered sequence of the natural numbers.
This sequence is given by: $C = 1, 2, \ldots, 9, 10, 11, \ldots, 99, 100, 101, \ldots$.

$L(\cdot)$ The function that denotes the length of a string, that is, the number of digits that make up the string.

$O(\cdot)$ Known as *the big-O notation.* When a function $f(n)$ has that $f(n) \in O(g(n))$, it means that for some constant K we have $|f(n)/g(n)| \leq K$ for every n.

$o(\cdot)$ Known as *the little-O notation.* When a function $f(n)$ has that $f(n) \in o(g(n))$, it means that $f(n)/g(n) \to 0$ as $n \to \infty$. If we have a term $x + f(n)$ in this case, it also can be denoted as $x + o\big(g(n)\big)$. The same holds for the big-O notation.

2.1 Normality proof by Champernowne

As it was done in the paper of Champernowne [1], let us first state the following three theorems. The first two theorems are needed to eventually prove that the constant of Champernowne is indeed normal.

Theorem 2.1. *The decimal number* $0.C^1$ *is normal.*

Theorem 2.2. *The decimal number* $0.C^j$ *is normal,* $\forall j \in \mathbb{N}$*.*

Theorem 2.3. *The constant of Champernowne, given by* $0.C$*, is normal.*

Now that we have formulated the theorems, let us start with proving the first theorem, Theorem 2.1.

2.1.1 Normality proof of $0.C^1$

We will be proving the normality of the constant $0.C^1$ with the use of the following steps.

1. First we will count the number of arbitrary strings s in sequences c_i;
2. Using the previous result, we will find the number of arbitrary string s occurring in the chain C_r^1;
3. Then we need to know which member of the sequence c_i contains the Nth digit of c_i;
4. Next, we determine the positions of strings s in members of the sequence c_i;
5. We will determine the possibilities of digits before and after the string s in every member;
6. Using steps 4 and 5, we will find out the number of undivided occurrences of strings s in c_i;
7. Together with the divided occurrences, we find out the number of occurrences of s in c_i in terms of N;
8. Lastly, we will determine the number of occurrences of s in C^1 in terms of N.

The number of arbitrary strings s in the sequence c_i

We start the proof by concentrating on sequences c_i to find out how many times arbitrary strings s of length k occur in sequences c_i. Note that we place a comma between every member of the sequence, as previously mentioned, to make the sequences more clear. For example

$$c_2 = 00, 01, 02, \ldots, 98, 99. \tag{2}$$

We say that the string s occurs *divided* if there occurs a comma between two digits of s. If this is not the case, we say that s occurs *undivided*. Take for instance the occurrences of $s = 25$ in c_2. We have that 25 occurs undivided at $\ldots, 24, 25, 26, \ldots$ and divided at $\ldots, 52, 53, \ldots$.

Note that a string s cannot occur undivided if $k > i$. When $k \leq i$, it is easy to count how many times s occurs undivided in c_i. For example, the string $s = 25$ occurs undivided 300 times in the sequence c_4. The string can occur in three ways, namely in the left, the middle or the right of every member of c_4, i.e. $25\cdot\cdot$, $\cdot 25 \cdot$ and $\cdot\cdot 25$. Because the other digits of the member can be any digit, every case occurs a 10^2 times in c_4 (even when it happens that those digits again create the same string, which happens at 2525. It will be counted double, but that's okay since it contains the string $s = 25$ twice). Thus in general, any string s of length 2 occurs $3 \cdot 10^2 = 300$ times undivided in c_4.

The same argument can be used for all lengths of strings s in any sequence c_i. Every string s with length k in c_i appears undivided in $i-k+1$ different positions and in each position, the string s occurs 10^{i-k} times. Hence, any string s of any length k appears $(i-k+1)10^{i-k}$ times in c_i for all i.

For the divided occurrences of strings s in c_i we can say, given that there are $k-1$ places for a comma to divide a string s and the fact that there are $10^i - 1$ commas in c_i, that there can be at most $(k-1)(10^i - 1)$. Note that the length of the sequence c_i can be written as $L(c_i) = i10^i$. If we divide the supremum of divided occurrences of any string s of length k with the length of c_i, we get

$$\begin{aligned}&\frac{(k-1)10^i}{i10^i} = \frac{k-1}{i} \to 0, \qquad \text{as } i \to \infty \qquad \Rightarrow \\ &(k-1)10^i \in o\big(L(c_i)\big), \end{aligned} \tag{3}$$

using the little-o notation we have introduced earlier. Now, if we also consider the number of undivided occurrences of s in c_i in terms of $L(c_i)$, we get

$$\begin{aligned}(i-k+1)10^{i-k} &= i10^{i-k} - (k-1)10^{i-k} \\ &= 10^{-k}L(c_i) + o\big(L(c_i)\big). \end{aligned} \tag{4}$$

Observe that if $(k-1)10^i \in o\big(L(c_i)\big)$, then also $-(k-1)10^i \in o\big(L(c_i)\big)$, which justifies the last equality. If we combine the results of equation (3) and (4) we can represent the total occurrences of s in c_i as

$$v(c_i, s) = 10^{-k}L(c_i) + o\big(L(c_i)\big). \tag{5}$$

Note that we have used the same function v as in Definition 2.1 to notate the number of divided occurrences of s in c_i. Also note that any finite sum of elements of $o\big(L(c_i)\big)$ is again contained in $o\big(L(c_i)\big)$, since the limit of its definition still will go to zero.

The number of arbitrary strings s in the chain C^1_r

Now consider the chain of sequences $C^1 = c_1c_2c_3\ldots$. To make calculations easier, let us define C^1_r as the chain $c_1c_2\ldots c_r$. Later we will let r tend to infinity, resulting in the infinite chain C^1.

Observe that $L(C^1_r)$ and $L(c_1) + L(c_2) + \ldots + L(c_r)$ are equal. But if we compare $v(C^1_r, s)$ with $v(c_1, s) + v(c_2, s) + \ldots + v(c_r, s)$, we see that the second formula does not count the possibility that the string s can be divided by a comma separating c_i and c_{i+1}. This can happen at most the number of different places a comma can occur inside a string s of length k times the number of those commas. This is equal to $(k-1)\cdot(r-1)$ and thus that number is contained in $O(r)$. If we combine these results we get the following.

$$\begin{aligned}v(C^1_r, s) &= \sum_{i=1}^{r} v(c_i, s) + O(r) \\ &= \sum_{i=1}^{r} 10^{-k}L(c_i) + \sum_{i=1}^{r} o\big(L(c_i)\big) + O(r) \\ &= 10^{-k}L(C^1_r) + \sum_{i=1}^{r} o\big(L(c_i)\big) + O(r) \\ &= 10^{-k}L(C^1_r) + o(L(C^1_r)). \end{aligned} \tag{6}$$

To justify the last step, we need to prove that both the terms $\sum_{i=1}^{r} o\big(L(c_i)\big)$ and $O(r)$ belong to $o(L(C^1_r))$.

First, for every element of $O(r)$ there exists some finite constant K such that $O(r)/r \leq K$ and thus $O(r) \leq K \cdot r$ for all r. Now observe that

$$\frac{K \cdot r}{L(C_r^1)} \leq \frac{K \cdot r}{L(c_r)} = \frac{K \cdot r}{r10^r} = \frac{K}{10^r} \to 0, \qquad \text{as } r \to \infty,$$

hence we have that $O(r) \subset o\big(L(C_r^1)\big)$.

Remember that the term $\sum_{i=1}^r o\big(L(c_i)\big)$ was the sum of the divided occurrences and thus in this case is less or equal to $\sum_{i=1}^r (k-1)10^i$. If we divide by $L(C_r^1)$, we get

$$\frac{\sum_{i=1}^r (k-1)10^i}{L(C_r^1)} = \frac{(k-1)\sum_{i=1}^r 10^i}{\sum_{i=1}^r i10^i} \to 0, \qquad \text{as } r \to \infty,$$

which proves that $\sum_{i=1}^r o\big(L(c_i)\big)$ is contained in $o\big(L(C_r^1)\big)$. This justifies the equation (6).

The digits of the member of c_i that contains the Nth digit of c_i

Now we can proceed to look at the first N digits of sequences c_i and see how many times string s occurs undivided in those digits.

Before we can do this, we need to find out in which member of c_i the Nth digit of sequence c_i lies. For example, the 52nd digit of c_2 will be inside the member 25.

$$\begin{array}{rcccccccccc} c_2 = & 0 & 0, & 0 & 1, & \ldots & 2 & 4, & 2 & 5, & \ldots \\ & \text{1st} & \text{2nd} & \text{3rd} & \text{4th} & \ldots & \text{49th} & \text{50th} & \text{51st} & \text{52nd} & \ldots \end{array}$$

Let us create a function that determines the digits of the member that contains the Nth digit of c_i. If we take the example above and take that the input of the function is 52 (or 51), the output must be 25. With this function, we will notate the output member of c_i as $p_{i-1}p_{i-2}\ldots p_1p_0$, for instance the member 25 in c_2 has $p_1 = 2$ and $p_0 = 5$.

Because of the convenient ordering of the members of the sequences c_i, it is easy to find out the place of those members. If we take the same example again, we can easily predict where the member 25 occurs in c_2. It is namely the member after the 25th member in the sequence and because all members of c_2 have length 2, the digits of the member 25 will appear on the 51st and 52nd place. In terms of p_1 and p_0, we can write the positions as $2(p_1 10 + p_2) + 2q$, where $0 < q \leq 1$. For the general case of strings in c_i, this will yield the following.

$$N = i\sum_{t=0}^{i-1} p_t 10^t + iq, \qquad 0 < q \leq 1. \tag{7}$$

The position of strings s in members of c_i

Now we have formulated a function to link the members of c_i with their position N in c_i, we can proceed to find an expression for the number of times a certain string s occurs in the first N digits of sequence c_i.

Recall that the number of undivided occurrences of each string s with length k in sequences c_i is equal to $(i-k+1)10^{i-k}$. If we have a string s with length k smaller than i, then s can occur undivided in members the sequence c_i in different position. Let f denote the position of the first digit of string s in each member of the sequence c_i where the string s does appear.

To make the idea of the variable f more clear, the first occurrence of $s = 25$ in c_4 for each possible f is given in Figure 1.

$$\begin{array}{c} \qquad\qquad\quad f=3 \quad f=2 \quad f=1 \\ \qquad\qquad\quad \downarrow \qquad\; \downarrow \qquad\; \downarrow \\ c_4 = 0000, 0001, \dots, 0025, \dots, 0250, \dots, 2500, \dots, 9999 \end{array}$$

Figure 1: An example of three different places where $s = 25$ can occur in c_4.

Let us first denote the number of undivided occurrences of strings s in the sequence c_i and the first N digits of c_i as $v_u(c_i, s)$ and $v_u(c_i, N, s)$ respectively.

Observe that f needs to be between 1 and $i - k + 1$ and thus we can also write the number of undivided occurrences as $v_u(c_i, s) = \sum_{f=1}^{i-k+1} 10^{i-k}$.

Now, we want to create an expression for $v_u(c_i, N, s)$. If we take the previous example, but we only look at the first 7000 digits we get the following figure.

$$\begin{array}{c} \qquad\qquad\quad f=3 \quad f=2 \quad f=2 \quad f=3 \\ \qquad\qquad\quad \downarrow \qquad\; \downarrow \qquad\; \downarrow \qquad\; \downarrow \\ c_4 = 0000, 0001, \dots, 0025, \dots, 0250, \dots, 1259, \dots, 1725, \dots, 1749 \end{array}$$

Figure 2: Occurrences of $s = 25$ in the first 7000 digits of c_4.

As we can see in Figure 2, 25 appears 20 times in position $f = 2$ and 18 times in position $f = 3$ in c_4. Obviously, 25 does not appear in position $f = 1$, since the first time that that can happen is at the 2500th string, which is not reached in the first 7000 digits.

It is clear that the number of appearances per position depends on the last member, which is 1749 in the previous example. For example for $f = 3$, the first two digits of each string need to be 00 to 17 to have a chance to be contained in the first 7000 digits.

Possibilities of digits in front of the string s in a member of c_i

We know that the digits in front of the string s are important for counting the appearance of the string. With equation (7), we can determine which member in c_i is reached after N digits. Let us illustrate this, using the previous example. If $N = 7000$, then $(p_3, p_2, p_1, p_0) = (1, 7, 4, 9)$, with $q = 1$.

Using the same kind of formula, we can count the number of possibilities of the digits in front of s. Consider the following formula.

$$\sum_{t=i-f+1}^{i-1} p_t 10^{t+f-i-1} + q', \qquad 0 \leq q' \leq 1. \tag{8}$$

Let us use the same example to test this formula. Hence we view the occurrence of the string $s = 25$ in position $f = 3$ in the first 7000 digits of c_4. Thus $N = 7000$ and $(p_3, p_2, p_1, p_0) = (1, 7, 4, 9)$.

$$\begin{aligned}\sum_{t=4-3+1}^{4-1} p_t 10^{t+3-4-1} + q' =& \sum_{t=2}^{3} p_t 10^{t-2} + q' \\ =& p_2 10^0 + p_3 10^1 + q' \\ =& 7 + 10 + q', \qquad\qquad 0 \leq q' \leq 1.\end{aligned}$$

Ignoring non-natural numbers, the outcome can be 17 or 18. If we think about it, this result indeed makes sense since the formula does not take in account that the member we reach in 7000 digits is before or after the member 1725, which precisely makes the difference between counting 17 or 18 strings $s = 25$ in position $f = 3$. In the case of the previous example, this would be 18 because we stop at 1749. This confirms our first deductions.

Possibilities of digits after the string s in a member of c_i

Let us now consider the digits after a string s in some position f. These digits are less complicated to deal with since they follow after each other. Take for instance the previous example in the case that $f = 2$. Because there is one digit after the string 25, that string will appear in 10 consecutive members.

$$\ldots, 0249, 0250, 0251, \ldots, 0258, 0259, 0260, \ldots.$$

In the case of $f = 1$, the string $s = 25$ will appear in 100 consecutive members in c_4 since, in these members, there are 2 digits after the string.

The number of undivided occurrences of strings s in sequence c_i

Hence, to get the total number of occurrences of a string at a given position f, we need to multiply ten to the power of the number of digits in a member after the string with the number given by equation (8). If we take note that the number of digits after the string is equal to $i - k - f + 1$, the total number will yield $10^{i-k-f+1} \sum_{t=i-f+1}^{i-1} \left(p_t 10^{t+f-i-1} + q'\right)$.

If we sum this result over all possible positions of $f \in \{1, \ldots, i - k + 1\}$, we get the following.

$$\begin{aligned}v_u(c_i, N, s) &= \sum_{f=1}^{i-k+1} \left(10^{i-k-f+1} \sum_{t=i-f+1}^{i-1} \left(p_t 10^{t+f-i-1} + q'\right)\right) \\ &= \sum_{f=1}^{i-k+1} \left(10^{-k} \sum_{t=i-f+1}^{i-1} \left(p_t 10^t + q' 10^{i-f+1}\right)\right) \\ &= 10^{-k} \sum_{f=1}^{i-k+1} \sum_{t=i-f+1}^{i-1} p_t 10^t + 10^{-k} \sum_{f=1}^{i-k+1} (f-1) q' 10^{i-f+1}, \qquad 0 \leq q' \leq 1.\end{aligned}$$

To simplify the last equation, let us see what actually happens in the double summation. The result of this for various values of k is given in Table 1. In columns with single summations, all possible indexes of the summation are given. In the double summation, the index of the second summation can appear multiple times because of the first summation. The number of times the index appears is denoted before each index value t.

f	$\sum_{t=i-f+1}^{i-1}$	k	$\sum_{f=1}^{i-k+1}$	$\sum_{f=1}^{i-k+1}\sum_{t=i-f+1}^{i-1}$
i	$1 \leq t \leq i-1$	i	$f=1$	0
$i-1$	$2 \leq t \leq i-1$	$i-1$	$1 \leq f \leq 2$	$1 \times t = i-1$
$i-2$	$3 \leq t \leq i-1$	$i-2$	$1 \leq f \leq 3$	$2 \times t = i-1, 1 \times t = i-2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
3	$t = i-2$	3	$1 \leq f \leq i-2$	$i-3 \times t = i-1, \ldots, 1 \times t = 3$
2	$t = i-1$	2	$1 \leq f \leq i-1$	$i-2 \times t = i-1, \ldots, 1 \times t = 2$
1	0	1	$1 \leq f \leq i$	$i-1 \times t = i-1, \ldots, 1 \times t = 1$

Table 1: Index values over which the summation sums for each f and k.

To make the previous reasoning more clear, we can consider the following picture.

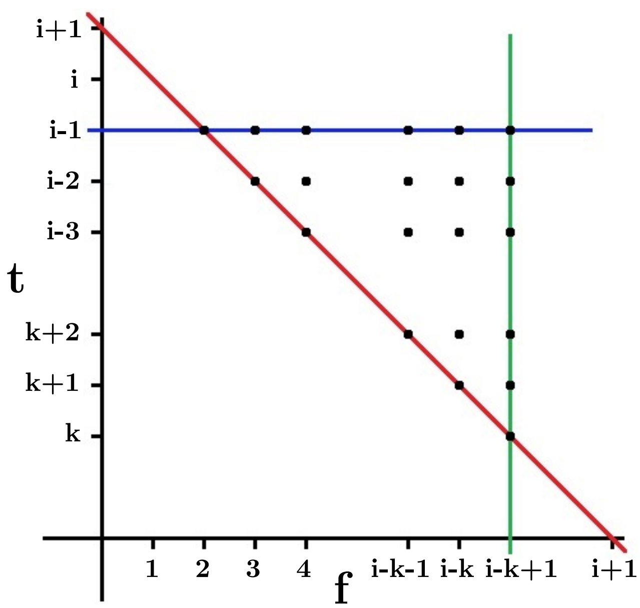

Figure 3: A visual representation of the double summation.

The first summation, the summation with f as its index, is represented as the f-axis which goes from 1 to $1-k+1$. The upper limit is given by the green line in Figure 3. The second summation, represented as the t-axis, goes from $i-f+1$ to $i-1$. These limits are given by the red and blue line respectively.

These three lines, which are given by the limits of the summations, create a triangle which denotes the combination of values of f and t over which the double summation sums. Note that this confirms the results given in Table 1.

We will be able to create a single summation out of the double summation. Since the term inside the double summation only uses the index t, we also will be using this index for the single summation.

From the t-axis of Figure 3 it is obvious that the summation must be from k to $i-1$. The number of times a given t must be evaluated is dependent of the number of f values, which solely depends on k. If we read from the graphic of Figure 3, we see that the number of times is exactly $t-k+1$. If we add these observations, we note that the double summation is equal to a summation from $t=k$ to $t=i-1$ where the term inside must be multiplied with $t-k+1$. Hence we can simplify the $v_u(c_i,N,s)$ to the following equation.

$$\begin{aligned} v_u(c_i,N,s) &= 10^{-k}\sum_{f=1}^{i-k+1}\sum_{t=i-f+1}^{i-1} p_t 10^t + 10^{-k}\sum_{f=1}^{i-k+1}(f-1)q'10^{i-f+1} \\ &= 10^{-k}\sum_{t=k}^{i-1}(t-k+1)p_t10^t + 10^{-k}\sum_{f=1}^{i-k+1}(f-1)q'10^{i-f+1}, \qquad 0\le q'\le 1. \end{aligned} \tag{9}$$

The number of occurrences of s in c_i in terms of N

In the last steps to complete the counting function, we will make estimations in terms of the number of digits N. For the divided occurrences, remember that any string s can occur divided at most $(k-1)(10^i-1)$ times, since s can be divided in $k-1$ ways by 10^i-1 different commas. If we divide by the number of digits, we see that the number of divided occurrences belongs to $o(N)$. Note that N tends to infinity if and only if i tends to infinity.

$$\frac{(k-1)(10^i-1)}{N} = \frac{(k-1)(10^i-1)}{i\sum_{t=0}^{i-1}p_t10^t+iq} \le \frac{(k-1)10^i}{i10^{i-1}} = \frac{(k-1)10}{i} \to 0, \text{ as } i\to\infty. \tag{10}$$

Also note that the second term of $v_u(c_i,N,s)$, namely $\sum_{f=1}^{i-k+1}(f-1)q'10^{i-f+1}$, is also contained in $o(N)$. We can see this if we view two separate cases, namely when $f\in o(i)$ and when $f\in O(i)$.

$$\begin{aligned} \frac{\sum_{f=1}^{i-k+1}(f-1)q'10^{i-f+1}}{i\sum_{t=0}^{i-1}p_t10^t+iq} &= \frac{\sum_{f\in o(i)}(f-1)q'10^{i-f+1}+\sum_{f\in O(i)}(f-1)q'10^{i-f+1}}{i\sum_{t=0}^{i-1}p_t10^t+iq} \\ &\le \frac{\sum_{f\in o(i)}mq'10^{i-1}+\sum_{f\in O(i)}iq'10^{wi}}{i\sum_{t=0}^{i-1}p_t10^t}, \quad \begin{array}{l} m=\sup\big(o(i)\big), \\ 0<w<1, \end{array} \\ &\le \frac{\sum_{f\in o(i)}mq'10^{i-1}}{i\sum_{t=0}^{i-1}p_t10^t} + \frac{\sum_{f\in O(i)}iq'10^{wi}}{i\sum_{t=0}^{i-1}p_t10^t} \\ &\le \frac{m\sum_{f\in o(i)}10^{i-1}}{i\sum_{t=0}^{i-1}10^t} + \frac{\sum_{f\in O(i)}10^{wi}}{\sum_{t=0}^{i-1}10^t} \\ &\le \frac{m^2 10^{i-1}}{i\sum_{t=0}^{i-1}10^t} + \frac{i10^{wi}}{10^{i-1}} \\ &\le \frac{m}{i} + \frac{i}{10^{(1-w)i-1}} \to 0+0, \qquad \text{as } i\to\infty. \end{aligned} \tag{11}$$

Now consider the first term of $v_u(c_i, N, s)$, the term $10^{-k}\sum_{t=k}^{i-1}(t-k+1)p_t 10^t$, in terms of the number of digits N.

$$\frac{10^{-k}\sum_{t=k}^{i-1}(t-k+1)p_t10^t}{i\sum_{t=0}^{i-1}p_t10^t+iq} = \frac{10^{-k}\sum_{t=k}^{i-1}tp_t10^t}{i\sum_{t=0}^{i-1}p_t10^t+iq} + \frac{10^{-k}(1-k)\sum_{t=k}^{i-1}p_t10^t}{i\sum_{t=0}^{i-1}p_t10^t+iq},$$

$$\frac{10^{-k}(1-k)\sum_{t=k}^{i-1}p_t10^t}{i\sum_{t=0}^{i-1}p_t10^t+iq} \le \frac{10^{-k}(1-k)}{i} \to 0, \qquad \text{as } i\to\infty, \tag{12}$$

$$\begin{aligned}\frac{10^{-k}\sum_{t=k}^{i-1}tp_t10^t}{i\sum_{t=0}^{i-1}p_t10^t+iq} &= 10^{-k}\frac{(i-1)p_{i-1}10^{i-1}+(i-1)p_{i-2}10^{i-2}+\ldots+kp_k10^k}{ip_{i-1}10^{i-2}+ip_{i-2}10^{i-2}+\ldots+ip_1 10+ip_0+iq} \\ &\to 10^{-k}, \qquad \text{as } i\to\infty. \end{aligned}\tag{13}$$

If we put the results together of equations (10), (11), (12) and (13), we get the formula for the number of appearances of string s in the first N digits of c_i.

$$\begin{aligned} 10^{-k}\sum_{t=k}^{i-1}(t-k+1)p_t10^t &= 10^{-k}i\sum_{t=0}^{i-1}p_t10^t+iq+o(N), \\ v_u(c_i,N,s) &= 10^{-k}N+o(N). \\ \text{and hence} \qquad v(c_i,N,s) &= 10^{-k}N+o(N). \end{aligned}\tag{14}$$

The number of occurrences of s in C^1 in terms of N

Now that we have a formulation for counting strings s in the first N digits of c_i, we can go back to C^1. First suppose that Nth digit of C^1 appears as the Mth digit of the rth sequence of C^1, that is sequence c_r. With these assumptions, we can formulate the equation of counting strings s in the first N digits of C^1 as follows.

$$v(C^1,N,s) = v(C^1_{r-1},s)+v(c_r,M,s).$$

Since we already have found expressions for both equations, namely equation (6) and (14), we can simply add those results to get an expression for $v(C^1,N,s)$. Note that the difference between N and M is exactly the length of C^1_{r-1}, which is given by $L(C^1_{r-1})$. This also directly implies that $o(L(C^1_{r-1}))\subset o(N)$ and $o(M)\subset o(N)$.

$$\begin{aligned} v(C^1,N,s) &= 10^{-k}L(C^1_{r-1})+o(L(C^1_{r-1}))+10^{-k}M+o(M) \\ &= 10^{-k}\big(L(C^1_{r-1})+M\big)+o(L(C^1_{r-1}))+o(M) \\ &= 10^{-k}N+o(N). \end{aligned}\tag{15}$$

Let us use this result and fill in the equation given in Definition 2.1.

$$\lim_{N\to\infty}\frac{v(x,N,s)}{N} = \lim_{N\to\infty}\frac{10^{-k}N+o(N)}{N} = \lim_{N\to\infty}10^{-k}+\frac{o(N)}{N} = 10^{-k} = \frac{1}{g^k}.$$

And thus, by Definition 2.1, we have proven that the decimal number $0.C^1$ is indeed normal in base ten. □

2.1.2 Normality proof of $0.C^j$

As we did with the chain C^1, we first suppose that the Nth digit of the chain C^j appears in the tth of j consecutive sequences c_r. Further assume that the Nth digit appears as the Mth digit of that specific sequence.

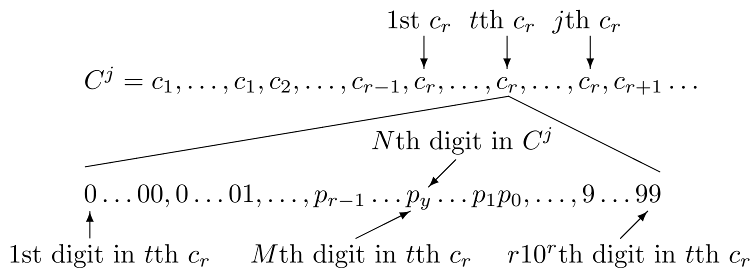

Figure 4: The Nth digit of C_j, which is the Mth digit of the tth c_r.

For the occurrences of arbitrary strings s in C^j we can use $v(C^1, s)$, $v(c_i, s)$ and $v(c_i, N, s)$, which are given by equation (6), (5) and (14) respectively.

If we compare the members of C^j_{r-1}, that is the part of C^j from the first c_1 to the last sequence c_{r-1} (see Figure 4), we see that all those members appear in the same number as j times C^1_{r-1}. The fact that the value of divided occurrences in C^1 is based on the number of commas and not on specific digits means that j times $v(C^1_r, s)$ will be equal to $v(C^j_{r-1}, s)$ (with the exception for the divided occurrences of $j-1$ commas, but that is easily contained in $O(k) \subset o(L(C^1_{r-1}))$, see equation (6)).

To finish the equation for $v(C^j, N, s)$ we just need to add $t-1$ times $v(c_r, s)$ and $v(c_r, M, s)$, given by equation (5) and (14), to j times $v(C^1_r, s)$. Again, the divided occurrences of the t commas that are not counted with these terms are contained in $O(k) \subset o(L(c_{r-1}))$.

$$
\begin{aligned}
v(C^j, N, s) &= jv(C^1_{r-1}, s) + (t-1)v(c_{r-1}, s) + v(c_{r-1}, M, s) \\
&= j\Big(10^{-k} L(C^1_{r-1}) + o\big(L(C^1_{r-1})\big)\Big) \\
&\quad\quad + (t-1)\Big(10^{-k} L(c_{r-1}) + o\big(L(c_{r-1})\big)\Big) + 10^{-k} M + o(M) \\
&= j10^{-k} L(C^1_{r-1}) + (t-1)10^{-k} L(c_{r-1}) + 10^{-k} M + o(N) \quad (16) \\
&= 10^{-k}\big(jL(C^1_{r-1}) + (t-1)L(c_{r-1}) + M\big) + o(N) \\
&= 10^{-k} N + o(N) \quad (17)
\end{aligned}
$$

To explain the step at equation (16) where we added all the o-notation terms to $o(N)$ we can use the same argument we used for equation (6). Intuitively, if we add all the values of the o-notation terms we get N and thus it seems obvious this will hold. Again, the justification of equation (6) explains a likewise situation.

The addition we did resulting in equation (17) (slightly spoiled by the previous observation) is directly justified if we again take a look at the first N digits of C^j in Figure 4. We can see in the figure that the terms we have added sum up to N.

Now we have stated $v(C^j, N, s)$, we can fill in the decimal number $0.C^j$ into Definition 2.1 like we did with $0.C^1$. From this we can conclude that the decimal number $0.C^j$ is normal in base ten. $\square$

2.1.3 Normality proof of the constant of Champernowne

To prove this theorem, we will use a specific case of Theorem 2.2 to prove that the constant of Champernowne is normal in base ten, namely the fact that C^9 is normal.

Notice that when we put a specific single digit after each comma in C^9 we can construct the sequence C. These digits are given in red in the figure underneath. The blue digits and commas are from the chain C^9.

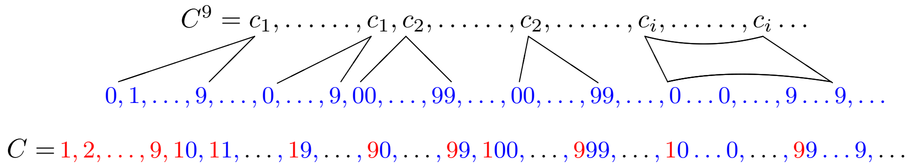

Figure 5: The sequence C, created by adding a digit after each comma of C^9.

Assume that the Nth digit of the chain C^9 is contained in a member of a sequence c_r, that is, the Nth digits is contained in a member of r digits. Recall that every sequence c_i in C^9 contains $10^i - 1$ commas (excluding the comma separating consecutive sequences). Let us denote the number of commas, for example after N digits in C^9, as $\text{com}(C^9, N)$ and the number of commas in a sequence, for example c_i, as $\text{com}(c_i)$.

$$\text{com}\big(C^9, L(C^9_{r-1})\big) + 1 \le \text{com}(C^9, N) \le \text{com}\big(C^9, L(C^9_r)\big),$$
$$\left(\sum_{i=1}^{r-1} 9 \cdot \big(\text{com}(c_i) + 1\big)\right) + 1 \le \text{com}(C^9, N) \le \sum_{i=1}^{r} 9 \cdot \big(\text{com}(c_i) + 1\big),$$
$$\left(\sum_{i=1}^{r-1} 9 \cdot 10^i\right) + 1 \le \text{com}(C^9, N) \le \sum_{i=1}^{r} 9 \cdot 10^i,$$
$$10^r \le \text{com}(C^9, N) \le 10^{r+1} - 1,$$

We will now observe the value of $\text{com}(C^9, N)$ in terms of N. Also note that N will go to infinity if and only if r goes to infinity.

$$N > L(C^9_{r-1}) = 9 \cdot L(C^1_{r-1}) = 9 \sum_{i=1}^{r-1} L(c_i) = 9 \sum_{i=1}^{r-1} i10^i > 9(r-1)10^{r-1}.$$
$$\frac{\text{com}(C^9, N)}{N} < \frac{10^{r+1}}{9(r-1)10^{r-1}} = \frac{100}{9(r-1)} \to 0, \qquad \text{as } r \to \infty.$$

Hence $\text{com}(C^9, N)$ is contained in $o(N)$.

Aside from the first 9 digits of C, the difference in digits between C^9 and C is equal to the difference in commas. Therefore we can write the relation between N and M as follows.

$$M = N + \text{com}(C^9, N) + 9 = N + o(N). \tag{18}$$

If we would add a digit after a comma of some sequence, it can cause that an arbitrary string s will not occur anymore in that place, or it can happen that it will appear where it first didn't appear. This can happen at most the length of the string minus one, that is $k - 1$, for each comma.

Thus it can be said that the effect of an extra digit after each comma in terms of appearances of arbitrary strings s is contained in $O\big(\text{com}(C^9, N)\big)$ and therefore will be contained in $o(N)$.

Given this previous result, we can formulate the occurrence of arbitrary strings s in C using $v(C^9, N)$ from equation (17).

$$v(C, M, s) = v(C^9, N, s) + o(N) = 10^{-k}N + o(N). \tag{19}$$

The last equation seems weird since the occurrence of arbitrary strings in the first M digits of C depends solely on the number N and the length of the strings. Mainly because we want to prove that $0.C$ is also normal, like $0.C^9$, it seems strange that $v(C, M, s)$ and $v(C^9, N, s)$ are equal while they evaluate over M and N digits respectively where $M > N > 0$.

This, however, is possible because of equation (18) which tells us that M and N can be considered equal when N tends to infinity. Since equations like equation (19) only say something about the case that N tends to infinity, we can safely say the following about the sequence C.

$$v(C, M, s) = 10^{-k}M + o(M). \tag{20}$$

If we put the value of $v(C, M, s)$ into Definition 2.1 again. We can conclude that $0.C$ and therefore the constant of Champernowne is indeed normal in base ten. $\square$

2.2 The combinatorial method

The combinatorial method of normality proofs, as stated in the paper of Pollack and Vandehey [2, p.2], is a result of Besicovitch's definition of (ϵ, k)-normal. The combinatorial method is a theorem which is stated below. The definition of (ϵ, k)-normality will be given when we explain the theorem.

Theorem 2.4. *Consider a sequence $\{a_n\}_{n=1}^{\infty}$ and suppose that the lengths of members a_n are growing on average and the lengths do not dominate each other. Mathematically, this is given as*

$$m \in o\Big(\sum_{n=1}^{m} L(a_n)\Big) \qquad \textit{and} \qquad m \cdot \max_{1\le n\le m} L(a_n) \in O\Big(\sum_{n=1}^{m} L(a_n)\Big),$$

$$\textit{i.e.} \qquad \frac{m}{\sum_{n=1}^{m} L(a_n)} \to 0 \quad \textit{and} \quad \frac{m \cdot \max_{1\le n\le m} L(a_n)}{\sum_{n=1}^{m} L(a_n)} \le K,$$

for some constant K, as m tends to infinity.
In addition, suppose that for any fixed $\epsilon > 0$ and $k \in \mathbb{N}$, the number of members $a_n, n \le m$ that are not (ϵ, k)-normal must be a number contained in $o(m)$. Then $x = 0.a_1a_2a_3\ldots$ is normal. [2, p.2]

Now that we have stated the theorem, it might be welcome to clarify the notions mentioned in this theorem.

The lengths of members a_n are growing on average

With the term "growing on average", we mean that the sequence of members globally always need to increase in length. When the length of the members keeps on growing, larger arbitrary strings can occur undivided. As N tends to infinity, the length of members also will tend to infinity. Only in this case, all possible strings have a chance to occur undivided. Recall that in the normality proof by Champernowne, only the undivided occurrences are relevant since the divided occurrences are negligible in terms of N.

To make it more clear why the term "growing on average" is linked to its mathematical statement, let us first see what happens when the length doesn't grow on average. If we assume that the length would not grow on average, for example $L(a_n)$ would tend to a constant B. Then, for every real number $B > \epsilon > 0$, there exists a natural number N such that the average length of the members after the Nth member is at least $B - \epsilon$. This will result in the following for all $n' > N$.

$$\begin{aligned}
\lim_{m\to\infty} \frac{m}{\sum_{n=1}^{m} L(a_n)} &= \lim_{m\to\infty} \frac{m}{\sum_{n=1}^{n'} L(a_n) + \sum_{n=n'}^{m} L(a_n)} \\
&\ge \lim_{m\to\infty} \frac{m}{\sum_{n=1}^{n'} L(a_n) + \sum_{n=n'}^{m} (B-\epsilon)} \\
&= \lim_{m\to\infty} \frac{m}{\sum_{n=1}^{n'} \big(L(a_n) - B + \epsilon\big) + \sum_{n=1}^{m} (B-\epsilon)} \\
&= \lim_{m\to\infty} \frac{m}{\sum_{n=1}^{n'} \big(L(a_n) - B + \epsilon\big) + m(B-\epsilon)} \\
&= \frac{1}{B-\epsilon} \qquad \Rightarrow \qquad m \notin o\Big(\sum_{n=1}^{m} L(a_n)\Big).
\end{aligned}$$

If we have a sequence $\{a_n\}_{n=1}^{\infty}$ such that $m \in o\big(\sum_{n=1}^{m} L(a_n)\big)$ does hold, we know that there cannot exist such a constant B and hence $L(a_n)$ will grow on average.

The length of members a_n do not dominate each other

The notion that lengths should not dominate can be explained the easiest by considering the notion in the form of the inequality. If we examine it we get the following.

$$\exists K \in \mathbb{R} \qquad \text{such that} \qquad \frac{m \cdot \max_{1\leq n\leq m} L(a_n)}{\sum_{n=1}^{m} L(a_n)} \leq K$$

$$\Rightarrow \qquad \max_{1\leq n\leq m} L(a_n) \leq K \cdot \frac{\sum_{n=1}^{m} L(a_n)}{m}, \qquad \text{as } m \to \infty.$$

From the equation above we see that the inequality indicates that there must exist a constant such that the maximum length of the members is at most the average length of the members times that constant. Hence the maximum length of a member must be proportional to the average.

The number of a_n's that are not (ϵ, k)-normal is contained in $o(m)$

This criterion is important since it fails to hold for sequences of members where a significant number of members have the same structure. For example, when a significant number of the members are alike each other or when a digit doesn't occur in a significant number of members. The consequence will be that those members will fail the (ϵ, k)-normality property for a certain ϵ and k.

Before we can analyze this criterion, we first need to know what it means to be (ϵ, k)-normal. The definition of (ϵ, k)-normality is given as follows.

Definition 2.2. *A string a is said to be (ϵ, k)-normal in base g if, for every string s consisting of k digits of base g, we have*

$$\left| \frac{v(a,s)}{L(a)} - \frac{1}{g^k} \right| \leq \epsilon, \tag{21}$$

where $v(a,s)$ denotes the number of times string s occurs in string a considered in base g. [2, p.2]

Unsurprisingly, this statement is very much alike the definition of normality. A member a is (ϵ, k)-normal for some fixed $\epsilon > 0$ and $k \in \mathbb{N}$ when the equation (21) holds for any string s with k digits. The main difference with normality is that this definition is that we can choose ϵ and k. Intuitively, it determines whether a member comes somewhat close to being normal for a certain length of arbitrary strings.

Let us make this notion more clear with an example. Let us choose $\epsilon = 1/20$ and $k = 1$ and check whether certain members are (ϵ, k)-normal. All the possible arbitrary strings when $k = 1$ are given by all single digit strings $0, 1, \ldots, 9$. If we consider $a = 123456789012345$, we see that every single digit string occurs once or twice. Hence the value inside the the absolute value bars equals $-1/30$ or $1/30$ and thus the left hand side is less or equal to $\epsilon = 1/20$. Therefore $a = 123456789012345$ is (ϵ, k)-normal in this case. If we consider $a = 123456789012$, which also has that every single digit string occurs once or twice, we get the values $-1/60$ or $1/15$ between the absolute value bars. Note that $1/15 > 1/20$ and therefore $a = 123456789012$ is not (ϵ, k)-normal in this case.

Observe that if a member a_n is (ϵ, k)-normal for every $\epsilon > 0$ and $k \in \mathbb{N}$, it becomes arbitrarily close to being normal according to definition 2.1.

If we want to prove that a sequence of members satisfies this criterion, we must prove that the number of members a_n of the sequence $\{a_n\}_{n=1}^{\infty}$ which are not (ϵ, k)-normal for every $\epsilon > 0$ and $k \in \mathbb{N}$ is a number which is contained in $o(m)$. In other words, almost all members need to be (ϵ, k)-normal for every $\epsilon > 0$ and $k \in \mathbb{N}$ for the criterion to hold for the sequence $\{a_n\}_{n=1}^{\infty}$.

2.2.1 Normality proof, using the combinatorial method

Consider the constant of Champernowne and let us prove that it is normal using Theorem 2.4. We can represent the constant as $0.C$ like we did earlier. By the way the sequence C is defined, we can say that the nth member has the value of the nth natural number, thus we can say that $a_n = n$. Hence, in terms of Theorem 2.4, we have the sequence $\{a_n\}_{n=1}^{\infty} = \{n\}_{n=1}^{\infty}$.

Sum of the length of members in terms of m

Since the first two criteria of the theorem both use the term $\sum_{n=1}^{m} L(a_n)$ we will first express this in terms of the number of considered members m. In base ten, there exists a number l such that we can write m as $b_l 10^l + b_{l-1}10^{l-1} + \ldots + b_1 10 + b_0$ with $0 \leq b_i \leq 9$ for $0 \leq i < l$ and $1 \leq b_l \leq 9$. Note that the index of the sum n has length $L(n) = i + 1$ when it has a value between 10^i and 10^{i+1}.

$$\begin{aligned}
\sum_{n=1}^{m} L(a_n) &= \sum_{n=1}^{m} L(n) \\
&= L(1) + \ldots + L(9) + L(10) + \ldots + L(99) + \ldots \\
&\qquad + L(10^{l-1}) + \ldots + L(10^l - 1) + L(10^l) + \ldots + L(m) \\
&= 1 + \ldots + 1 + 2 + \ldots + 2 + \ldots + l + \ldots + l + (l+1) + \ldots + (l+1) \\
&= 9 \cdot 1 + 9 \cdot 10 \cdot 2 + \ldots + 9 \cdot 10^{l-1} l + (m - 10^l + 1)(l+1).
\end{aligned}$$

Because $10^l \leq m < 10^{l+1}$, we know that $L(m) = l + 1$. More specifically, we know that $L(n) = l + 1$ if $10^l \leq n \leq m$. Hence, we have $m - 10^l + 1$ times that the number n has length $L(n) = l + 1$.

Observe that $\lfloor \log_{10} m \rfloor = l$ and $m10^{-1} \leq 10^{\lfloor \log_{10} m \rfloor} \leq m$. We can rewrite the previous inequality in the following way.

$$\begin{aligned}
\sum_{n=1}^{m} L(n) &\geq 9 \cdot 1 + 9 \cdot 10 \cdot 2 + \ldots + 9 \cdot 10^{l-1} l \\
&= 9 \cdot 1 + 9 \cdot 10 \cdot 2 + \ldots + 9 \cdot 10^{\lfloor \log_{10} m \rfloor - 1} \lfloor \log_{10} m \rfloor. \\
&\geq 9 \cdot 1 + 9 \cdot 10 \cdot 2 + \ldots + 9 \cdot m10^{-2} \lfloor \log_{10} m \rfloor. \qquad (22)
\end{aligned}$$

Proof that members of sequence C grow on average

Now let us proceed in proving the criteria of Theorem 2.4. The first criterion says that the length of members a_n in base ten needs to grow on average.

In our case, for every member $a_n = n$ which is represented by only 9's in base ten, the next member $a_{n+1} = n + 1$ will increase in length. In all other cases $n + 1$ will have the same length. Thus the length of members will indeed grow on average.

To show this mathematically, let us use the expressions used in the theorem and the inequality shown in equation (22).

$$
\begin{aligned}
\frac{m}{\sum_{n=1}^{m} L(a_n)} &\leq \frac{m}{9\cdot 1 + 9\cdot 10\cdot 2 + \ldots + 9\cdot m10^{-2}\lfloor \log_{10} m\rfloor} \\
&= \frac{1}{\frac{9+180+\ldots}{m} + 9\cdot 10^{-2}\lfloor \log_{10} m\rfloor} \to 0, \qquad \text{as } m\to\infty \qquad\qquad \Rightarrow \\
m &\in o\left(\sum_{n=1}^{m} L(a_n)\right). \qquad (23)
\end{aligned}
$$

With this, we have shown mathematically that members of the sequence C grow on average.

Proof that length of members of sequence C do not dominate

We now advance to prove the second criterion, namely that no lengths of members $a_n = n$ dominate. We recall the expression used in the theorem.

$$
m\cdot \max_{1\leq n\leq m} L(a_n) = O\left(\sum_{n=1}^{m} L(a_n)\right) \quad \Rightarrow \quad \frac{m\cdot \max_{1\leq n\leq m} L(a_n)}{\sum_{n=1}^{m} L(a_n)} \leq K,
$$

for some constant K, as m tends to infinity. Let us prove that there indeed exist such a constant. To prove this, we will again use equation (22). Also note that $\max_{1\leq n\leq m} L(a_n) = L(m) = l+1 = \lfloor \log_{10} m\rfloor + 1$.

$$
\begin{aligned}
\frac{m\cdot \max_{1\leq n\leq m} L(a_n)}{\sum_{n=1}^{m} L(a_n)} &\leq \frac{m\cdot L(m)}{9\cdot 1 + 9\cdot 10\cdot 2 + \ldots + 9\cdot m10^{-2}\lfloor \log_{10} m\rfloor} \\
&= \frac{\lfloor \log_{10} m\rfloor + 1}{\frac{9+180+\ldots}{m} + 9\cdot 10^{-2}\lfloor \log_{10} m\rfloor}
\end{aligned}
$$

$$
\begin{aligned}
\frac{m\cdot \max_{1\leq n\leq m} L(a_n)}{\sum_{n=1}^{m} L(a_n)} &< \frac{\lfloor \log_{10} m\rfloor}{9\cdot 10^{-2}\lfloor \log_{10} m\rfloor} + \frac{1}{9\cdot 10^{-2}\lfloor \log_{10} m\rfloor} \\
&\leq \frac{1}{9\cdot 10^{-2}} + \frac{1}{9\cdot 10^{-2}\cdot 1}, \qquad \text{if } m\geq 10 \\
&= \frac{200}{9} \qquad (24)
\end{aligned}
$$

If $1\leq m\leq 9$, we have that $\frac{m\cdot \max_{1\leq n\leq m} L(a_n)}{\sum_{n=1}^{m} L(a_n)} = \frac{m}{m} = 1 < \frac{200}{9}$. Thus we have proven that there indeed exist a constant K and that constant can be at most $\frac{200}{9}$.

Proof that the number of members of the sequence C which are not (ϵ, k)-normal is $o(m)$

Now we will prove the third criterion of Theorem 2.4. To proof this, we need to count the number of members which are (ϵ, k)-normal for some given $\epsilon > 0$ and $k\in\mathbb{N}$ and show that this number is contained in $o(m)$. Thus let us take $\epsilon > 0$ and $k\in\mathbb{N}$ arbitrarily and count how many members of C are (ϵ, k)-normal in that case.

Since the only kind of occurrences of arbitrary strings s in a member n are by definition the undivided occurrences, we only have to consider those occurrences. Also note that each of the members n is also a member of the sequence c_i when $i = L(n)$. Assume that the first digit of the member appears as the $Z+1$st digit of c_i. Hence the digits of the member n are covered by the $Z+1$st to the $Z+i$th digit of c_i.

Recall from the proof of Champernowne, that the number of undivided occurrences of arbitrary strings s in the sequence c_i is given by equation (9).

$$v_u(c_i, N, s) = 10^{-k} \sum_{t=k}^{i-1} (t-k+1) p_t 10^t + 10^{-k} \sum_{f=1}^{i-k+1} (f-1) q' 10i - f + 1. \tag{9}$$

If we subtract the number of undivided occurrences in the first Z digits of c_i from those in the first $Z+i$ digits of c_i, we will get the number of occurrences of arbitrary strings in the member we are looking for.

Remember that the number q' in equation (9), introduced in equation (8), can be between zero and one. The first part, the part without q', in equation (9) determines the number of occurrences of string s before the members that contains the Nth digit. The number q' indicates which part of the strings occurring in the member with the Nth digit occurs within these N digits. Since the Zth and $Z+i$th digit are the last one of their member, the number q' will be one in both cases.

Because we have different values for the p_t's, we will denote them as functions of the position of the last digit of the member to make a distinction between them.

$$\begin{aligned} v_u(c_i, n, s) &= v_u(c_i, Z+i, s) - v_u(c_i, Z, s) \\ &= 10^{-k} \sum_{t=k}^{i-1} (t-k+1) p_t(Z+i) 10^t + 10^{-k} \sum_{f=1}^{i-k+1} (f-1) 10^{i-f+1} \\ &\qquad - 10^{-k} \sum_{t=k}^{i-1} (t-k+1) p_t(Z) 10^t - 10^{-k} \sum_{f=1}^{i-k+1} (f-1) 10^{i-f+1} \\ &= 10^{-k} \sum_{t=k}^{i-1} \big((t-k+1) p_t(Z+i) 10^t - (t-k+1) p_t(Z) 10^t \big) \\ &= 10^{-k} \sum_{t=k}^{i-1} \Big((t-k+1) \big(p_t(Z+i) - p_t(Z) \big) 10^t \Big) \end{aligned}$$

The members given by the numbers $p_t(Z)$ and $p_t(Z+i)$ follow directly after each other. Therefore, there is a T such that $p_T(Z) + 1 = p_T(Z+i)$ and for all $t < T$ we have that $p_t(Z) = 9$ and $p_t(Z+i) = 0$. In the case that $t > T$ we have that $p_t(Z) = p_t(Z+i)$. To understand what happens here, note that $T = 2$ when we consider the members 2499 and 2500 and $T = 0$ when we consider the members 2500 and 2501.

Let us implement this in the previous equation.

$$
\begin{aligned}
v_u(n,s) &= 10^{-k}\sum_{t=k}^{i-1}\Big((t-k+1)\big(p_t(Z+i)-p_t(Z)\big)10^t\Big) \\
&= 10^{-k}\Big((T-k+1)10^T-\sum_{t=k}^{T-1}(t-k+1)9\cdot 10^t\Big) \\
&= (T-k+1)10^{T-k}-\sum_{u=0}^{T-k-1}(u+1)9\cdot 10^u, && \text{with } u=t-k, \\
&= (T-k+1)10^{T-k}-\sum_{u=0}^{T-k-1}(u+1)10\cdot 10^u+\sum_{v=0}^{T-k-1}(v+1)10^v, \\
&= \sum_{v=0}^{T-k}(v+1)10^v-\sum_{u=0}^{T-k-1}(u+1)10^{u+1}, \\
&= 1+\sum_{v=1}^{T-k}(v+1)10^v-\sum_{u=0}^{T-k-1}(u+1)10^{u+1}, \\
&= 1+\sum_{w=0}^{T-k-1}(w+2)10^{w+1}-\sum_{u=0}^{T-k-1}(u+1)10^{u+1}, && \text{with } v=w+1, \\
&= 1+\sum_{u=0}^{T-k-1}\Big((u+2)10^{u+1}-(u+1)10^{u+1}\Big), && \text{with } w=u, \\
&= 1+\sum_{u=0}^{T-k-1}10^{u+1}, \\
&= 1+\sum_{v=1}^{T-k}10^v, && \text{with } v=u+1 \\
&= \sum_{v=0}^{T-k}10^v,
\end{aligned}
$$

The occurrences of arbitrary strings s apparently depends on the number T which differs per member n. Let us find out what this means for the average of the members of C. In any case, we will have that T is at least 0. If $T=0$, the index of the sum will go from 0 to $-k$, where k is strictly positive, and thus the sum will be equal to 0.

In case that the last digit of the member before the member n was a 9, the number T would be at least 1. This occurs exactly one tenth of the time when n tends to infinity. In this case, the sum will only have a non-zero result if $k=1$. When k is indeed equal to 1, the sum will also be equal to 1. For simplicity, we will denote the values of T we did not evaluate yet as the other terms. With this said, we can write the following.

$$v_u(n,s) = 10^{-1}\mathbf{1}_{k=1} + \text{other terms}.$$

The case that there are two 9's before the member n happens one hundredth of the time. In that case the sum will yield 10 if $k = 1$ and 1 if $k = 2$. Hence we get

$$\begin{aligned} v_u(n,s) &= 10^{-1}\mathbf{1}_{k=1} + 10^{-2}\cdot 10\cdot \mathbf{1}_{k=1} + 10^{-2}\mathbf{1}_{k=2} + \text{other terms},\\ &= 2\cdot 10^{-1}\mathbf{1}_{k=1} + 10^{-2}\mathbf{1}_{k=2} + \text{other terms}. \end{aligned}$$

For the case of 3 9's before the member n, which occurs one thousandth of the time, the sum will yield 100, 10 and 1 when k is equal to 1, 2 and 3 respectively.

$$v_u(n,s) = 3\cdot 10^{-1}\mathbf{1}_{k=1} + 2\cdot 10^{-2}\mathbf{1}_{k=2} + 10^{-3}\mathbf{1}_{k=3} + \text{other terms}.$$

We can go on with this until the maximum value of T, which is equal to $i-1$.

$$v_u(n,s) = (i-1)10^{-1}\mathbf{1}_{k=1} + (i-2)10^{-2}\mathbf{1}_{k=2} + \ldots + 2\cdot 10^{2-i}\mathbf{1}_{k=i-2} + 10^{1-i}\mathbf{1}_{k=i-1}.$$

The only possible value of k that is not mentioned in the previous equation is $k = i$. This, however, will be not very surprising since $i - k = 0$ when $k = i$. Observe the following if we use the value of k inside the terms of the previous equation.

$$\begin{aligned} &v_u(n,s) = (i-k)10^{-k}\mathbf{1}_{k=1} + (i-k)10^{-k}\mathbf{1}_{k=2} + \ldots + (i-k)\cdot 10^{-k}\mathbf{1}_{k=i-2} + (i-k)10^{-k}\mathbf{1}_{k=i-1},\\ &v_u(n,s) = (i-k)10^{-k}. \end{aligned}$$

Now let us substitute this into the equation of the definition of (ϵ,k)-normality, which is given by equation (21). When we elaborate this equation, we get a requirement for members n such that they will be (ϵ,k)-normal.

$$\left|\frac{v(n,s)}{L(n)} - \frac{1}{10^k}\right| = \left|\frac{(i-k)10^{-k}}{i} - 10^{-k}\right| = \frac{k10^{-k}}{i} \le \epsilon \qquad \Rightarrow \qquad i \ge \frac{k}{\epsilon}10^{-k}.$$

When the contrary is true, that is when $i = L(n) < (k/\epsilon)\cdot 10^{-k}$, we have a requirement for the members n such that they are possibly not (ϵ,k)-normal. Now let us count those members.

$$L(n) < \frac{k}{\epsilon}\cdot 10^{-k} \qquad \Rightarrow \qquad n < 10^{k\epsilon^{-1}10^{-k}}. \tag{25}$$

Note that if we consider the number of members n which are lower than a certain number, then that number of those members is lower than that same number since the sequence of members n yield the natural number. In other words, the number of natural numbers lower than some number K is obviously lower than K since it is equal to $K-1$.

We will now consider the number of members n that are not (ϵ,k)-normal in terms of the total number of observed members m. If we take note of the first m members and the previous observation, then we know that the number of members $n \le m$ that are possibly not (ϵ,k)-normal is smaller than $10^{k\epsilon^{-1}10^{-k}}$. Hence we can consider the number of those members in terms of the total number of members m.

$$\begin{aligned} \forall\epsilon > 0, k\in\mathbb{N}, \qquad &\frac{\#\{n\le m\,|\,n \text{ is not } (\epsilon,k)\text{-normal}\}}{m} < \frac{10^{k\epsilon^{-1}10^{-k}}}{m} \to 0, \qquad \text{as } m\to\infty\\ \Rightarrow \qquad &\#\{n\le m\,|\,n \text{ is not } (\epsilon,k)\text{-normal}\} \in o(m). \end{aligned}$$

For every choice of $\epsilon > 0$ and $k \in \mathbb{N}$, the number of members n that are not (ϵ,k)-normal is always smaller than a number that depends on ϵ and k. After we have chosen $\epsilon > 0$ and $k \in \mathbb{N}$ the number of member that are not (ϵ, k)-normal does not change when we let m tend to infinity and hence we have that the number of members n which are not (ϵ, k)-normal is always contained in $o(m)$.

Now that we have proven all 3 criteria, we have proven that the constant of Champernowne is indeed normal in base ten according to Theorem 2.4. □

2.2.2 Proof of the combinatorial method, in the case of the constant of Champernowne

The statement of Theorem 2.4 is very general, but since we are primarily interested in the constant of Champernowne in base ten, we will be proving the theorem only for that specific case. The general proof can be found in the paper of Pollack and Vandehey [2, p.4].

Recall the equations (1) and (21) that are used to denote when a constant is normal and (ϵ, k)-normal respectively. Note that if we fix $\epsilon > 0$ for every k and we prove that the constant of Champernowne is (ϵ, k)-normal, we can let N tend to infinity at the end of the proof which makes ϵ insignificant and prove that it is indeed normal in base ten.

Let us start the proof with a given integer N and let $m = m(N)$ be such that the Nth digit of the constant lies in the mth member of C, that is the member a_m, which is equal to m in the case of the sequence C. For example, if we take $N = 25$, then $m(25) = 17$.

$$\begin{array}{rcccccccccc} C = & 1, & 2, & \dots & 4, & 1 & 5, & 1 & 6, & 1 & 7, & \dots \\ & \text{1st} & \text{2nd} & \dots & \text{19th} & \text{20th} & \text{21th} & \text{22th} & \text{23th} & \text{24th} & \text{25th} & \dots \end{array}$$

The occurrences of s in C in terms of the members of C

First we will consider the number of occurrences of arbitrary strings s in the member where the Nth digit occurs, $v(m, s)$. We can say that it is at most equal to $L(m)$ since the number of occurrences of any string s cannot exceed the length of m. Note that the difference between $v(C, N, s)$ and $v(1, 2, \dots, m; s)$ is at most $L(m) + k$. This is because the term $v(C, N, s)$ compared to $v(1, 2, \dots, m; s)$ only evaluates occurrences of arbitrary strings in some of the digits of m. The occurrences of strings inside these digits is at most $v(m, s)$ which is at most $L(m)$. The strings that occur in both the first N digits and in the missing digits can be at most the length of the string s minus 1, which is $k - 1$. Therefore we can say that $|v(C, N, s) - v(1, 2, \dots, m; s)| \leq L(m) + k - 1$.

Now we need the terms undivided and divided again which we used in the proof of Champernowne in section 2.1. We have that $v(1, 2, \dots, m; s)$ counts both divided and undivided occurrences of strings s. Note that the divided occurrences are at most $(k - 1) \cdot (m - 1)$ since a string s has $k - 1$ places where a comma can appear and there are $m - 1$ comma inside the sequence $1, 2, \dots, m$. Therefore the divided occurrences can be written as $(k - 1) \cdot (m - 1)$. Observe that the undivided occurrences of string s can be covered precisely by $v(1, s) + v(2, s) + \dots + v(m, s) = \sum_{n=1}^{m} v(n, s)$. Hence we have that $v(1, 2, \dots, m; s) = \sum_{n=1}^{m} v(n, s) + (k - 1) \cdot (m - 1)$.

Using the previous statements, we will able to formulate the occurrences of arbitrary strings s in the first

N digits of C in terms of the occurrences of arbitrary strings s in the first m members of C.

$$\begin{aligned}
|v(C,N,s) - v(1,2,\ldots,m;s)| &\leq L(m) + k - 1,\\
\Big|v(C,N,s) - \sum_{n=1}^{m} v(n,s)\Big| &\leq L(m) + k - 1 + (k-1)\cdot(m-1)\\
&= L(m) + m(k-1)\\
&\leq m\cdot k, \qquad\qquad \text{since } L(m) \leq m. \qquad (26)
\end{aligned}$$

Note that the number $\sum_{n=1}^{m} L(n)$, is equal to the position of the last digit of m in C. Because the Nth digit lies in string m we can say that

$$\sum_{n=1}^{m-1} L(n) < N \leq \sum_{n=1}^{m} L(n) \qquad \Rightarrow \qquad \frac{N}{\sum_{n=1}^{m} L(n)} \to 1, \qquad \text{as } m \to \infty. \tag{27}$$

which can be expressed as $N \sim \sum_{n=1}^{m} L(n)$. To proceed, we need the first two criteria of Theorem 2.4, given by equation (23) and (24).

$$\frac{m \cdot \max_{1\leq n\leq m} L(a_n)}{\sum_{n=1}^{m} L(a_n)} < \frac{200}{9}, \qquad \Rightarrow \qquad \frac{m\cdot k}{\sum_{n=1}^{m} L(n)} < \frac{200k}{9L(m)} \to 0, \qquad \text{as } m \to \infty.$$

From the result from the equations above and the result that $N \sim \sum_{n=1}^{m} L(n)$ it follows directly that

$$m \cdot k \in o\Bigg(\sum_{n=1}^{m} L(n)\Bigg) \equiv o(N). \tag{28}$$

When we put the result of equation (26) and (28) together, we can get the following equation.

$$v(C,N,s) = \sum_{n=1}^{m} v(n,s) + o(N). \tag{29}$$

The occurrences of s inside the members of C which are not (ϵ, k)-normal

For this part we need the last criterion of the theorem. Let $T \subset \mathbb{N}$ be the set of integers n such that the members a_n are not (ϵ, k)-normal. By the last criterion, the number of members of the first m members, which are not (ϵ, k)-normal must be contained in $o(m)$, hence $\#\{n \in T \,|\, n \leq m\} \in o(m)$. Also note that $v(n,s)$ can't be more that the length of the member n and thus the sum of those $v(n,s)$ over some set of n's can't be more than the sum of the lengths of those members. Hence $\sum_{n\in U} v(n,s) \in O\big(\sum_{n\in U} L(n)\big)$. Consider the sum of the occurrences of arbitrary strings s inside the first m members of C which are not

(ϵ, k)-normal.

$$\sum_{n\in T}^{m} v(n,s) \in O\Big(\sum_{n\in T}^{m} L(n)\Big) \subset O\Big(\max_{1\le n\le m} L(n)\cdot \sum_{n\in T}^{m} 1\Big) \equiv O\big(L(m)\cdot \#\{n\in T \,|\, n\le m\}\big),$$

$$\forall K\in\mathbb{R}, \qquad K\cdot L(m)\cdot\#\{n\in T\,|\,n\le m\} \in O\big(L(m)\cdot \#\{n\in T\,|\,n\le m\}\big)$$

$$\frac{K\cdot L(m)\cdot\#\{n\in T\,|\,n\le m\}}{L(m)\cdot m} = \frac{K\cdot\#\{n\in T\,|\,n\le m\}}{m} \to 0, \qquad \text{as } m\to\infty, \qquad \text{since } \#\{n\in T\,|\,n\le m\}\in o(m$$

$$O\big(L(m)\cdot\#\{n\in T\,|\,n\le m\}\big) \subset o\big(L(m)\cdot m\big)$$

$$\sum_{n\in T}^{m} v(n,s) \in o\big(L(m)\cdot m\big) \equiv o(N). \tag{30}$$

The latter is true since 90% to 99% of the members inside the first N digits of C have a length of $L(m)-1$ or $L(m)$. Therefore the terms $L(m)\cdot m$ and N will both tend to infinity with the same speed.

The occurrences of s inside the members of C which are (ϵ, k)-normal

The set of integers n such that numbers n are (ϵ, k)-normal is complementary to the set T and is therefore equal to $\mathbb{N}\backslash T$. Let this set be S. Naturally, this set S of integers n have that the members a_n of C, which are equal to n, are (ϵ, k)-normal. This results in the following.

$$\left|\frac{v(a_n,s)}{L(a_n)} - 10^{-k}\right| \le \epsilon \quad \Rightarrow \quad L(n)(10^{-k}-\epsilon) \le v(n,s) \le L(n)(10^{-k}+\epsilon)$$

$$\Rightarrow \quad v(n,s) = L(n)\big(10^{-k}+O(\epsilon)\big), \qquad \text{for } n\in S=\mathbb{N}\backslash T. \tag{31}$$

We will combine the results we got from equations (29), (30) and (31). From this, we will get an expression for $v(C,N,s)$ we can use in equation (21) of the definition of (ϵ, k)-normality, Definition 2.2. Before we do this, note that since $\#\{n\in T\,|\,n\le m\}\in o(m)$ and $\#\{n\in T+S\,|\,n\le m\}=m$, we have that $\#\{n\in S\,|\,n\le m\}\sim m$ as m tends to infinity and recall that $N\sim\sum_{n=1}^{m} L(n)$ which is shown in equation (27).

$$\begin{aligned}
v(C,N,s) &= \sum_{n=1}^{m} v(n,s) + o(N) \\
&= \sum_{n\in S}^{m} v(n,s) + \sum_{n\in T}^{m} v(n,s) + o(N) \\
&= \sum_{n\in S}^{m} L(n)\big(10^{-k}+O(\epsilon)\big) + o(N) && \text{since } \sum_{n\in T}^{m} v(n,s)\in o(N), \\
&= \sum_{n=1}^{m} L(n)\big(10^{-k}+O(\epsilon)\big) + o(N), && \text{since } \#\{n\in S\,|\,n\le m\}\sim m, \\
&= N\big(10^{-k}+O(\epsilon)\big) + o(N), && \text{since } \sum_{n=1}^{m} L(n)\sim N.
\end{aligned}$$

Let us fill this expression in the equation of the definition of (ϵ, k)-normality, equation (21). Take note that $N^{-1}o(N) = o(1)$ indeed makes sense since $o(N)$ denotes a term that tends to zero when it is divided by N. Hence, the term we get when we divide $o(N)$ with N is contained in $o(1)$.

$$\begin{aligned}\left|\frac{v(a,s)}{L(a)} - \frac{1}{g^k}\right| = \left|\frac{v(C,N,s)}{N} - \frac{1}{10^k}\right| &= \left|\frac{N\big(10^{-k} + O(\epsilon)\big) + o(N)}{N} - 10^{-k}\right| \\ &= \left|10^{-k} + O(\epsilon) + N^{-1}o(N) - 10^{-k}\right| = O(\epsilon) + o(1)\end{aligned}$$

Since ϵ was chosen arbitrarily, the term will indeed tend to zero as N tends to infinity and hence we have proven Theorem 2.4 in the case of the constant of Champernowne. □

3 Transcendence

Since 1937 it is known that the constant of Champernowne is, in addition to normal, also transcendent. In that year, Kurt Mahler composed a proof that showed that a whole set of constants, including the constant of Champernowne, are transcendental [3]. Let us first state what it means for a constant to be transcendent.

Definition 3.1. *A constant x is said to be transcendental when it cannot be a zero of some polynomial with integer coefficients.*

While the notion of transcendence is easy to state, the proofs that show that a specific constant is transcendental are quite involved. In fact, to give an idea of the complexity of the notion, it created an entire branch of number theory on its own: Transcendental number theory.

To get more feeling for what it means for a number to be transcendental, we can see that it is a subset of the irrational numbers. This can be illustrated by taking any rational constant $z = a/b$. That constant is namely the solution of the equation $bx = a$. The subset is even a strict subset since every n-root of rational numbers also cannot be transcendental since it would be the solution of the polynomial $bx^n = a$.

The study of transcendental numbers practically started with Liouville, who has stated the first criterion whereby transcendental numbers can be constructed [5, p.1]. The transcendental numbers that can be constructed in this way are called Liouville numbers. Examples of transcendental numbers which are not Liouville numbers are e and π.

3.1 Transcendence proof by Mahler

As previously stated, Mahler proved in his paper the transcendence of a certain set of constants, more specifically, special decimal numbers which are not Liouville numbers. We will investigate his proof in the case of the constant of Champernowne. The general proof for the set for which Mahler proved transcendence can be found in his paper [3].

The set of constants for which Mahler has proved transcendence can be described as follows. Let $f(k)$ be a strictly increasing polynomial with natural numbers as outcome for all natural numbers k. Now consider the sequence $f(1), f(2), f(3), \ldots$ and let us call each $f(k)$ a member of that sequence. We define the constant x to be the decimal number we get if we concatenate the sequence of members $f(k)$ after the period. Thus the constant is defined as $x = 0.f(1), f(2), f(3), \ldots$ for some given polynomial $f(k)$.

If we, for example, take the function $f(k) = (k^2 + k)/2$, the resulting constant will be given by the decimal number $x = 0.1, 3, 6, 10, 15, 21, 28, 36, \ldots$. It is obvious that constant of Champernowne belongs to this set of constants since that constant is created by simply taking the polynomial $f(k) = k$.

The theorem Mahler proved in his paper is given by the following theorem.

Theorem 3.1. *Let x be a decimal number given by*

$$x = 0.f(1), f(2), f(3), \ldots$$

where $f(\cdot)$ is a strictly increasing polynomial with $f(k) \in \mathbb{N}$ for all $k \in \mathbb{N}$. Then x is transcendental, without being a Liouville number.

The theorem we are going to prove will only consider the case of $f(k) = k$ since we are only interested in the transcendence of the constant of Champernowne.

Theorem 3.2. *The constant of Champernowne in base ten is transcendental.*

With the statement that the set of constants are not Liouville numbers, the first theorem implies that transcendence of those constants can not be shown to be transcendental in the same way as it is proven for Liouville numbers. Mahler shows that the constants which are described in Theorem 3.1 cannot be a Liouville number in the last part of his paper [3]. We also will proof that the constant of Champernowne is not a Liouville number in Section 3.1.2.

The results of Theorem 3.1 are similar for decimal numbers in every base q greater or equal to 2, hence Mahler left the base as a variable. We, on the other hand, are only interested in base ten and thus, in our case, q will be the fixed number 10.

The theorem will be proven with the use of a result by Schneider, like Mahler did in his paper. Details on the theorem of Schneider can be found by another paper by Mahler, which he published in 1936 [4]. The intuition behind the theorem is that a transcendental constant can be approximated very accurately by a sequence of fractions. By constructing that specific sequence of fractions and subtracting the constant by those fractions, we can prove that the remainder term can always be made so small such that there can not exist a polynomial with integer coefficients having that constant as an solution.

Following this method, the proof will concentrate on creating this fraction and prove the transcendence of the constant of Champernowne by using the theorem of Schneider. Without further ado, we will start with proving the theorem.

3.1.1 The argument of Mahler in the case of the constant of Champernowne

We will be proving the transcendence of the constant of Champernowne with the use of the following steps.

1. First we create a workable representation of the constant.
2. Next, we consider the construction of the constant in terms of its members;
3. Thereafter, we will construct a sequence of fractions that approximate the constant;
4. After subtracting the fractions from the constant, we will investigate the remainder;
5. Finally, we complete the proof with the use of the theorem of Schneider;

Representation of the constant of Champernowne in digits $Z_{k,l}$

Let the constant of Champernowne be given by c and let f be the first order polynomial $f(x) = x$. Let us again call each natural number $f(k) = k$ a member of the constant c and let each member k be a decimal number which has v digits. Furthermore, let us denote the digits as $Z_{k,l}$ where the second subscript stands for the position of the digit inside the member k. Hence we can represent each member k as

$$k = Z_{k,0}Z_{k,1}\dots Z_{k,v-2}Z_{k,v-1} = 10^{v-1}Z_{k,0} + 10^{v-2}Z_{k,1} + \dots + 10Z_{k,v-2} + Z_{k,v-1} = \sum_{l=0}^{v-1} Z_{k,l}10^{v-l}.$$

With this expression for k we write c as

$$c = 0.1, 2, 3, \dots, 25, \dots, k, \dots = 0.Z_{1,0}, Z_{2,0}, Z_{3,0}, \dots, Z_{25,0}Z_{25,1}, \dots, Z_{k,0}Z_{k,1}\dots Z_{k,v-1}, \dots$$

Because it is difficult to investigate the properties of this expression, we will simplify representation of c.

Construction of the constant c in terms of its members $f(k)$

Assume again that member k has v digits, from this we know that $10^{v-1} \leq k \leq 10^v - 1$. Let us now create the sets K_v of all members k which have exactly v digits. Observe the number of members k contained in the set K_v, thus $\#K_v$, is equal to $(10^v - 1) - (10^{v-1} - 1) = 9 \cdot 10^{v-1}$. Also note that the total number of digits of all members k with v digits together will be $v \cdot 9 \cdot 10^{v-1}$. If we count the total number of digits of all members with at most $v - 1$ digits, we get

$$\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}.$$

The total number of members with v digits, up to the member k is equal to $k - (10^{v-1} - 1)$. Therefore the total number of digits of those members is equal to $v(k - 10^{v-1} + 1)$. If we add this number to the number of digits of all members up to $v - 1$ digits, we get the total amount of digits of the 1st to the kth member,

$$\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} + v(k - 10^{v-1} + 1).$$

Note that this value is equal to the position of the last digit of the member k in the constant c. Therefore, the contribution of this member to c is equal to

$$k \cdot 10^{-\left(\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} + v(k - 10^{v-1} + 1)\right)}. \tag{32}$$

For the contribution of all members that are contained in the constant c, which sums up to c again, we sum the result of equation (32) for every natural number k.

$$\begin{aligned} c &= \sum_{k=1}^{\infty} k \cdot 10^{-\left(\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} + v(k - 10^{v-1} + 1)\right)} \\ &= \sum_{k=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} + v(10^{v-1} - 1)} k \cdot 10^{-vk} \end{aligned}$$

Note that the summation over the members k is equivalent to summing over the lengths of members v and then summing over members k with a certain length v. The latter will be a summation from the member $k = 10^{v-1}$ to the member $k = 10^v - 1$ since those are the first and last member with v digits. Hence, let us substitute the sum by the sums we just mentioned.

$$\begin{aligned} c &= \sum_{v=1}^{\infty} \sum_{k=10^{v-1}}^{10^v - 1} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} + v(10^{v-1} - 1)} k \cdot 10^{-vk} \\ &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} + v(10^{v-1} - 1)} \sum_{k=10^{v-1}}^{10^v - 1} k \cdot 10^{-vk}. \end{aligned} \tag{33}$$

Simplifying the summations

To proceed, we need the a formula from difference calculus that can be found in a book by Ernesto Cesaro and Gerhard Kowalewski [6], which is given as follows.

$$\sum_{z=0}^{\infty} F(z)x^z = \sum_{h=0}^{\infty} \frac{x^h \partial^h F(0)}{(1-x)^{h+1}}. \tag{34}$$

In this formula, the function $F(z)$ is defined for all non-negative integers z and has the following as its hth differential for a given number z.

$$\partial^h F(z) = \sum_{i=0}^{h} \binom{h}{i} (-1)^i F(z+h-i).$$

We can rewrite the sum at the end of equation (33) in such a way that we can use the formula of equation (34). Let us call this term S_v and let us denote 10^{-v} as x.

$$\begin{aligned}
S_v &= \sum_{k=10^{v-1}}^{10^v-1} k \cdot 10^{-vk} = \sum_{k=10^{v-1}}^{\infty} k \cdot 10^{-vk} - \sum_{l=10^v}^{\infty} l \cdot 10^{-vl}, \qquad \text{take } z = k - 10^{v-1} = l - 10^v, \\
&= \sum_{z=0}^{\infty} (z+10^{v-1}) 10^{-v(z+10^{v-1})} - \sum_{z=0}^{\infty} (z+10^v) 10^{-v(z+10^v)} \\
&= 10^{-v10^{v-1}} \sum_{z=0}^{\infty} (z+10^{v-1}) 10^{-vz} - 10^{-v10^v} \sum_{z=0}^{\infty} (z+10^v) 10^{-vz} \\
&= 10^{-v10^{v-1}} \sum_{z=0}^{\infty} (z+10^{v-1}) x^z - 10^{-v10^v} \sum_{z=0}^{\infty} (z+10^v) x^z.
\end{aligned} \tag{35}$$

Note that the polynomials used in the sums of equation (35), $F_1(z) = z + 10^{v-1}$ and $F_2(z) = z + 10^v$, are first order and become equal to 1 after differentiating once and thus it is only necessary for h to sum over 0 and 1. By using equation (34) we can rewrite S_v with the use of these polynomials.

$$\begin{aligned}
S_v &= 10^{-v(10^{v-1})} \sum_{h=0}^{1} \frac{x^h \partial^h F_1(0)}{(1-x)^{h+1}} - 10^{-v10^v} \sum_{h=0}^{1} \frac{x^h \partial^h F_2(0)}{(1-x)^{h+1}} \\
&= 10^{-v(10^{v-1})} \sum_{h=0}^{1} \frac{10^{-vh} \partial^h F_1(0)}{(1-10^{-v})^{h+1}} - 10^{-v10^v} \sum_{h=0}^{1} \frac{10^{-vh} \partial^h F_2(0)}{(1-10^{-v})^{h+1}} \\
&= 10^{-v(10^{v-1}-1)} \sum_{h=0}^{1} \frac{\partial^h F_1(0)}{(10^v-1)^{h+1}} - 10^{-v(10^v-1)} \sum_{h=0}^{1} \frac{\partial^h F_2(0)}{(10^v-1)^{h+1}} \\
&= 10^{-v(10^{v-1}-1)} A_v - 10^{-v(10^v-1)} B_v,
\end{aligned}$$

where the terms A_v and B_v are given by

$$A_v = \sum_{h=0}^{1} \frac{\partial^h F_1(0)}{(10^v - 1)^{h+1}} = \frac{10^{v-1}}{(10^v - 1)^1} + \frac{1}{(10^v - 1)^2} = \frac{10^{v-1}(10^v - 1) + 1}{(10^v - 1)^2} = \frac{10^{2v-1} - 10^{v-1} + 1}{10^{2v} - 2 \cdot 10^v + 1}$$

$$B_v = \sum_{h=0}^{1} \frac{\partial^h F_2(0)}{(10^v - 1)^{h+1}} = \frac{10^{v}}{(10^v - 1)^1} + \frac{1}{(10^v - 1)^2} = \frac{10^{v}(10^v - 1) + 1}{(10^v - 1)^2}. = \frac{10^{2v} - 10^{v} + 1}{10^{2v} - 2 \cdot 10^v + 1}.$$

We can rewrite equation (33) using this new expression for S_v in terms of A_v and B_v.

$$\begin{aligned} c &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} + v(10^{v-1}-1)} \left(10^{-v(10^{v-1}-1)} A_v - 10^{-v(10^v - 1)} B_v\right) \\ &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} \left(A_v - 10^{-v(10^v - 10^{v-1})} B_v\right) \\ &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} \left(A_v - B'_v\right). \end{aligned} \tag{36}$$

Using the expressions for A_v and B_v we can write out the expression for c.

$$\begin{aligned} c &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} \left(\frac{10^{2v-1} - 10^{v-1} + 1}{10^{2v} - 2 \cdot 10^v + 1} - 10^{-v(10^v - 10^{v-1})} \frac{10^{2v} - 10^v + 1}{10^{2v} - 2 \cdot 10^v + 1} \right) \\ &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} \left(10^{-1} \frac{10^{2v} - 10^v + 10}{10^{2v} - 2 \cdot 10^v + 1} - 10^{-v(10^v - 10^{v-1})} \left(1 - \frac{-10^v}{(1 - 10^v)^2}\right) \right) \\ &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} \left(10^{-1} \left(1 - \frac{-10^v + 9}{(1 - 10^v)^2}\right) - 10^{-v(10^v - 10^{v-1})} \left(1 - \frac{-10^v}{(1 - 10^v)^2}\right) \right) \\ &= \sum_{v=1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} E_v. \end{aligned} \tag{37}$$

We leave the expression for c as it is and we will denote the term between the brackets as E_v. Note that this term is approximately equal to one over ten since the terms between the smaller brackets will tend to one very fast and the second factor will already be very small when v equals 1.

The fraction that approximates the constant

Using the result from equation (37), we can create a sequence of fractions, which we will denote as P_s/Q_s, that will converge to the constant c very quickly. This sequence will make it possible to eventually prove the transcendence of c. Let D_s denote the least common multiple of the numbers $10 - 1, 10^2 - 1, \ldots, 10^s - 1$ for

$s \geq 1$. Let the denominator and the numerator of the fractions be

$$Q_s = D_s^2 10^{\sum_{m=1}^{s-1} m \cdot 9 \cdot 10^{m-1}} \qquad \text{and} \qquad P_s = Q_s \left(\sum_{v=1}^{s} \left(10^{10 - \sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} E_v \right) + B_s' \right) \tag{38}$$

and let us have a remainder term which is given as follows.

$$R_s = \sum_{v=s+1}^{\infty} \left(10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} E_v \right) - B_s'. \tag{39}$$

Note that the terms Q_s, P_s and R_s are created in such a way that we have

$$c - \frac{P_s}{Q_s} = R_s. \tag{40}$$

To give an idea of these sequences we will work out the first four term of every sequence.

s or v	1	2	3	4
D_s^2	81	9 801	120 758 121	1 231 853 592 321
Q_s	81	9 801 000 000 000	$1.207 \cdot 10^{197}$	$1.231 \cdot 10^{2901}$
A_v	$\frac{10}{81}$	$\frac{991}{9801}$	$\frac{99901}{998001}$	$\frac{9999001}{99980001}$
B_v'	$10^{-9}\frac{91}{81}$	$10^{-180}\frac{9901}{9801}$	$10^{-2700}\frac{999001}{998001}$	$10^{-36000}\frac{99990001}{99980001}$
E_v	$\frac{10}{81} - 10^{-9}\frac{91}{81}$	$\frac{991}{9801} - 10^{-180}\frac{9901}{9801}$	$\frac{99901}{998001} - 10^{-2700}\frac{999001}{998001}$	$\frac{9999001}{99980001} - 10^{-36000}\frac{99990001}{99980001}$
P_s	10	1 209 999 989 980	$1.490 \cdot 10^{196}$	$1,519 \cdot 10^{2900}$
$\frac{P_s}{Q_s}$	0.12345679012	$0.1\ldots979900$	$0.1\ldots997999000$	$0.1\ldots999799990000$
R_s	$1.018 \cdot 10^{-9}$	$9.101 \cdot 10^{-190}$	$9.010 \cdot 10^{-2890}$	$9.001 \cdot 10^{-38890}$

Table 2: The first four terms of the sequences $D_s^2, Q_s, A_v, B_v', E_s, P_s, \frac{P_s}{Q_s}$ and R_s.

It is obvious that Q_s is a natural number and by Table 2 we see that P_s is also natural since D_s^2 is a multiple of the divisor of E_s and the addition of B_s' removes the remaining negative term of 10 to a negative power of the last E_v term of the sum inside P_s. Since both Q_s and P_s are natural numbers, we may use Q_s as the denominator of the fraction. This was, for example, not the case if P_s was rational instead of an integer.

The value of the remainder term R_s in terms of Q_s

By increasing the number s, the fraction given by equation (40) will approximate the constant better and the remainder term R_s will get smaller. However, for the constant to be transcendent, the remainder term needs to decrease faster than the extent to which the denominator increases. Thus let us consider the construction of the fraction and determine the remainder term in terms of the denominator.

To show this, we will derive asymptotic formulas for $\log Q_s$ and $\log R_s$. To obtain them, we need the following formula, denoting the number of digits of all member with at most $v-1$ digits.

$$\begin{aligned} i_v &= \sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1} = \sum_{m=1}^{v-1} m\big(10^m - 10^{m-1}\big) = \sum_{m=1}^{v-1} m10^m - \sum_{n=0}^{v-2}(n+1)10^n, \qquad \text{with } n = m-1, \\ &= \sum_{m=1}^{v-1} m10^m - \sum_{n=1}^{v-1}(n+1)10^n + v10^{v-1} - 1 = v10^{v-1} - \sum_{m=1}^{v-1} 10^m - 1 = (v-1)10^{v-1} + O(10^{v-1}). \end{aligned}$$

First note that D_s, the least common multiple of $10^t - 1$ where $1 \leq t \leq s$, is surely smaller than just the product of these 10^t. With that in mind, observe the following representations of Q_s using the formula of i_v.

$$\begin{aligned} Q_s = D_s^2 10^{\sum_{m=1}^{s-1} m \cdot 9 \cdot 10^{m-1}} &< \left(\prod_{t=1}^{s} 10^t\right)^2 10^{i_s} = 10^{s(s+1)} 10^{(s-1)10^{s-1} + O(10^{s-1})} && \Rightarrow \\ \log Q_s &\sim (s-1)10^{s-1} \log 10 && \Rightarrow \quad (41) \\ \lim_{s\to\infty} \frac{\log Q_{s+1}}{\log Q_s} &= \lim_{s\to\infty} \frac{(s)10^s \log 10}{(s-1)10^{s-1}\log 10} = 10. && (42) \end{aligned}$$

Secondly, we will represent R_s in the following way. First we will determine $R_s - B'_s$. The second expression is justified since every following term in the sum will be 10 times smaller than the previous term and the denominator in the fraction between the brackets is similar to 1 since 10^{-v} is negligible compared to 1.

$$\begin{aligned} R_s - B'_s &= \sum_{v=s+1}^{\infty} 10^{-\sum_{m=1}^{v-1} m \cdot 9 \cdot 10^{m-1}} E_v = \sum_{v=s+1}^{\infty} 10^{-i_v} E_v && \Rightarrow \\ R_s - B'_s &\sim 10^{-i_{s+1}} && \Rightarrow \\ &\sim 10^{-s10^s} && \Rightarrow \end{aligned}$$

To determine R_s, we will also need to evaluate B'_s, which can be obtained from equation (36).

$$\begin{aligned} B'_s &= 10^{-s(10^s - 10^{s-1})} B_v \sim 10^{-s10^s} \\ R_s &= (R_s - B'_s) + B'_s \sim 10^{-s10^s} && \Rightarrow \\ \log |R_s| &\sim \log \left|10^{-s10^s}\right| = -s10^s \log 10. && (43) \end{aligned}$$

We can observe from the previous equation that, if s is large enough, the term R_s will not vanish. Multiplying this term with the denominator will provide us with a notion of how well the fraction approximates the constant. If the absolute value of that term gets smaller if we increase s linearly, we know that the accuracy of the approximation increases more that linear. Using equation (41) and (43) we get

$$\begin{aligned}
\log|Q_sR_s| &= \log Q_s + \log|R_s| \\
&\sim (s-1)10^{s-1}\log 10 - s10^s\log 10 \\
&\sim s(10^{-1}-1)10^s\log 10 \qquad \Rightarrow \\
|Q_sR_s| &< 1, && (44)\\
|Q_sR_s| &\sim 10^{-s(1-10^{-1})10^s} && (45)\\
&= 10^{-1}10^{-(s-1)(1-10^{-1})10^1 10^{s-1}} \\
&= 10^{-1}(10^{-(s-1)(1-10^{-1})10^{s-1}})^{10} \\
|Q_sR_s| &\sim 10^{-1}|Q_{s-1}R_{s-1}|^{10}. && (46)
\end{aligned}$$

From equations (44) and (46) we see that the term $|Q_sR_s|$ is similar to 10 times the previous step to the power 10 while $|Q_sR_s|$ is smaller than 1 for all sufficiently large s. Therefore we can conclude the following.

$$|Q_sR_s| < |Q_{s-1}R_{s-1}|. \tag{47}$$

And thus we can conclude that the fraction from equation (40) indeed tends to the constant c very quickly.

Completing the proof by using the proof of Schneider

To prove transcendence with the results we previously have obtained, we need the following theorem which is proven by Schneider.

Theorem 3.3. *For the real number x, if there is a constant $k > 1$ for a sequence of fractions*

$$\frac{p_1}{q_1}, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \dots, \frac{p_s}{q_s}, \dots, \qquad p_i \in \mathbb{Z},\, q_i \in \mathbb{N},\, \gcd(p_i, q_i) = 1,\, \forall i \in \mathbb{N},$$

which has for an increasing s that the denominator q_s is a power of a natural number q and increases strictly in such a way that

$$0 < \left| x - \frac{p_s}{q_s} \right| \leq q_s^{-k} \qquad \text{and} \qquad \lim_{s\to\infty} \sup_s \frac{\log q_{s+1}}{\log q_s} < \infty.$$

Then the constant x is transcendent.

This theorem is proven in an paper [4] which Mahler also produced, a year before he published the paper that proves that the constant of Champernowne is transcendental.

To prove that the constant c is transcendent, we actually need to modify the theorem a bit. If we look to the proof of the theorem, we see that it shows effortlessly that the constant x, which is specified in the theorem, is indeed transcendent. But the proof will also hold if we take a weaker requirement for q_s, namely, take $q_s = q'_s q''_s$ with q'_s a natural number such that

$$\lim_{s\to\infty} \frac{\log q'_s}{\log q_s} = 0$$

and q''_s a power of q. Therefore, the actual theorem which we are going to use to prove transcendence of the constant c is given as follows.

Theorem 3.4. *For the real number x, if there is a constant $k > 1$ for a sequence of fractions*

$$\frac{p_1}{q_1}, \frac{p_2}{q_2}, \frac{p_3}{q_3}, \ldots, \frac{p_s}{q_s}, \ldots, \qquad p_i \in \mathbb{Z},\ q_i \in \mathbb{N},\ \gcd(p_i, q_i) = 1,\ \forall i \in \mathbb{N},$$

which has for an increasing s that the denominator $q_s = q_s' q_s''$, where q_s' is a natural number and q_s'' is a power of a natural number q, strictly increases in such a way that the following three requirements hold:

$$0 < \left| x - \frac{p_s}{q_s} \right| \leq q_s^{-k}, \qquad \lim_{s\to\infty} \sup_s \frac{\log q_{s+1}}{\log q_s} < \infty \qquad \textit{and} \qquad \lim_{s\to\infty} \frac{\log q_s'}{\log q_s} = 0.$$

Then the constant x is transcendent.

To prove that the constant c is transcendental according to Theorem 3.4, we will be considering the chain of fractions P_s/Q_s, where the numerator and denominator are given by the equations (38) for all natural numbers s. Therefore, let us see if Q_s satisfies the three requirements of the theorem.

Fulfilling the requirements of Theorem 3.4

First we need to prove that Q_s fits the requirements. If we observe the way Q_s is defined in equation (38), we see that it is indeed a product of a natural number and a power of some natural number. Thus let us have that

$$q_s' = D_s^2, \qquad q = 10 \qquad \text{and} \qquad q_s'' = 10^{\sum_{m=1}^{s-1} m\cdot 9\cdot 10^{m-1}}.$$

Now consider the first requirement and let us fill in the constant c and its corresponding variables.

$$0 < \left| c - \frac{P_s}{Q_s} \right| = \left| R_s \right| \leq Q_s^{-k}. \tag{48}$$

To fulfill the first requirement, R_s may not be equal to zero and $|R(s)|$ needs to be less or equal to some negative power of Q_s. The first inequality already holds by the definition of R_s, give by equation (39), since it is not possible for R_s to be zero. For the second inequality, we need equations (41) and (43).

$$\left| R_s \right| = Q_s^{-a} \qquad \Rightarrow \qquad -a = \frac{\log |R_s|}{\log Q_s},$$

$$\lim_{s\to\infty} \frac{\log |R_s|}{\log Q_s} = \lim_{s\to\infty} \frac{-s10^s \log 10}{(s-1)10^{s-1} \log 10} = -10.$$

From this, we may conclude that, for a sufficiently large s, the second inequality in equation (48) will hold for any number k with $k < 10$.

That the second requirement holds can easily be shown with equation (42) since it shows that the fraction will tend to 10 in stead of infinity when s tends to infinity. For all other s, the fraction will just yield a rational number and thus the supremum will always be below infinity.

For the third requirement, we substitute the same Q_s from equation (38). Recall that D_s^2 is smaller than $10^{s(s+1)}$.

$$\begin{aligned}\lim_{s\to\infty}\frac{\log D_s^2}{\log Q_s} &= \lim_{s\to\infty}\frac{s(s+1)}{s(s+1)+\sum_{m=1}^{s-1} m\cdot 9\cdot 10^{m-1}}\\ &< \lim_{s\to\infty}\frac{s(s+1)}{s(s+1)+(s-1)9\cdot 10^{s-2}}\\ &\leq \lim_{s\to\infty}\frac{s+4}{s+4+\frac{9}{10^2}\cdot 10^s}\to 0, \qquad \text{since } s(s+1)\leq (s-1)(s+4) \text{ for } s>1.\end{aligned}$$

Now that we have shown that the sequence of fractions P_s/Q_s that approximate the constant c indeed meet the requirements of Theorem 3.4, we have proven that the constant of Champernowne c is indeed transcendent. □

3.1.2 Proof that the constant of Champernowne is not a Liouville number

We have proven that the constant c is indeed transcendent. To finish the proof given by Mahler, we still have to prove that this constant is not a trivial example of a transcendental number, which means that it may not be a Liouville number.

To be able to prove that a constant is not a Liouville number, we need the definition of a Liouville number.

Definition 3.2. *Real number x which possess a sequence of distinct rational approximations*

$$\frac{p_n}{q_n},\qquad n\in\mathbb{N}\qquad \textit{such that}\qquad \left|x-\frac{p_n}{q_n}\right|<\frac{1}{q_n^{w_n}},\qquad \textit{where } w_n \textit{ is such that}\qquad \limsup w_n=\infty,$$

are termed as Liouville numbers. [5, p.2]

To prove that the constant c is not a Liouville number, we will be proving the contrary. We will be proving that for any fraction approximation of c, there will be non-infinity values for w_n such that q^{-w_n} is smaller ot equal to the difference between c and the fraction.

First we will make use of equation (46) and create a variable Q in such a way that

$$|Q_sR_s|<\frac{1}{2Q}\leq |Q_{s-1}R_{s-1}|. \tag{49}$$

With the use of equation (40) we will have the following for some variable P.

$$c-\frac{P}{Q}=\left(\frac{P_s}{Q_s}-\frac{P}{Q}\right)+R_s\qquad \text{and thus}\qquad c-\frac{P}{Q}=R_s\qquad \text{if}\qquad \frac{P_s}{Q_s}=\frac{P}{Q}.$$

If we cannot choose such a P, let P be such that $|P_sQ - PQ_s| \geq 1$ which results in

$$\begin{aligned}
\left|c - \frac{P}{Q}\right| &= \left|\left(\frac{P_s}{Q_s} - \frac{P}{Q}\right) + R_s\right| \\
&\geq \left|\frac{P_sQ - PQ_s}{QQ_s}\right| - |R_s| \\
&\geq \frac{1}{QQ_s} - |R_s| \\
&> \frac{2|Q_sR_s|}{Q_s} - |R_s|, \qquad \text{using equation (49),} \\
&= |R_s|. \qquad (50)
\end{aligned}$$

Thus in both cases, we will have that $|c - P/Q| \geq |R_s|$. From the equations (45) and (49) we get

$$\frac{1}{Q} \leq 2|Q_{s-1}R_{s-1}| \qquad \Rightarrow \qquad \frac{1}{Q} \leq 10^{-(s-1)(1-10^{-1})10^{s-1}} \qquad \text{when } s \text{ is large enough.}$$

And from the equations (50) and (43) together with the previous result we finally get

$$\left|c - \frac{P}{Q}\right| \geq |R_s| \qquad \Rightarrow \qquad \left|c - \frac{P}{Q}\right| \geq 10^{-s10^s} \geq 10^{-2(s-1)(1-10^{-1})10^{s-1}} \geq Q^{-2}.$$

With the last equation we have proven that there indeed exist non-infinity values for w_n such that $|c - P/Q| \geq Q^{-w_n}$ and thus the constant c is not a Liouville number. □

4 Continued fractions

A continued fraction is a way to express a number as a sequence of nested fractions. The way to obtain a continued fraction expansion of a number is very much alike the Euclidean algorithm, which is an algorithm that can be used to find the greatest common divisor of two integers.

The process of the Euclidean algorithm is given by subtracting the smallest of the two integers a whole number of times from the other number such that the remainder is the smallest possible positive number. This process is repeated for the smallest number and the remainder until the remainder vanishes. The greatest common divisor will be the last remainder term. Let us illustrate this process with the integers 273 and 119.

$$\begin{aligned} 273 &= 2 \cdot 119 + 35 \\ 119 &= 3 \cdot 35 + 14 \\ 35 &= 2 \cdot 14 + 7 \\ 14 &= 2 \cdot 7. \end{aligned}$$

Hence, the greatest common divisor of 273 and 119 is 7 since that is the last remainder.

The process of constructing the continued fraction expansion of a number goes as follows. First we express the number as the addition of the whole part and the decimal part of the number, which needs to be a non-negative number smaller than one. Thereafter we express the decimal part as a fraction with one as the numerator and we repeat the process for its denominator. More general, the procedure of the continued fraction expansion of a real number $x = x_0$ is given as follows.

$$x_i = \lfloor x_i \rfloor + \{x_i\}, \qquad \text{where, when } \{x_i\} \neq 0, \text{ we write } \{x_i\} = \frac{1}{x_{i+1}}.$$

Hence, a continued fraction expansion of a real number x will have the following form.

$$x_0 = \lfloor x_0 \rfloor + \cfrac{1}{\lfloor x_1 \rfloor + \cfrac{1}{\lfloor x_2 \rfloor + \cfrac{1}{\lfloor x_3 \rfloor + \cdots}}}.$$

To illustrate this procedure we will take $\frac{273}{119}$ as example to clarify the procedure of the continued fraction expansion and to show its similarity to the Euclidean algorithm.

$$\begin{aligned} \frac{273}{119} &= 2 + \frac{35}{119} = 2 + \frac{1}{\frac{119}{35}} \\ \frac{119}{35} &= 3 + \frac{14}{35} = 3 + \frac{1}{\frac{35}{14}} \\ \frac{35}{14} &= 2 + \frac{7}{14} = 2 + \frac{1}{\frac{14}{7}} \\ \frac{14}{7} &= 2 \end{aligned}$$

Note the whole part of each fraction is similar to the factors in the Euclidean algorithm, these numbers are the terms of the continued fraction. The neat thing of this process is that we can put each fraction in the expression of the previous fraction, which creates the sequence of nested fraction that we mentioned earlier. Therefore, we can express the fraction of the previous example as

$$\frac{273}{119} = 2 + \frac{1}{3 + \frac{1}{2+\frac{1}{2}}}.$$

The two procedures are completely the same if we multiply the equations of the continued fraction procedure with the denominator of the fraction at the left hand side. This means that the continued fraction of a rational number is equivalent to the Euclidean algorithm with integers. Since every two integers have a greatest common divisor of at least one, the Euclidean algorithm will eventually end. Thus every fraction of two integers, that is every rational number, will also have a finite continued fraction expansion.

For irrational numbers, the latter is impossible and thus they will have infinite continued fraction expansions. Hence, this will also be the case for the constant of Champernowne.

4.1 The continued fraction expansion of the constant of Champernowne

In the beginning of this paper we have mentioned that the constant of Champernowne has a peculiar continued fraction expansion. With the principle of continued fractions explained, we can find out what is so special about the continued fraction of the constant of Champernowne. Since the constant is irrational, its continued fraction will not end. Let us therefore start with the first few terms.

$$\begin{aligned}
c = 0.1234567891\ldots &= 0 + \cfrac{1}{8.1000000670\ldots} \\
&= 0 + \cfrac{1}{8 + \cfrac{1}{9.9999932924\ldots}} \\
&= 0 + \cfrac{1}{8 + \cfrac{1}{9 + \cfrac{1}{1.00000067076\ldots}}} \\
&= 0 + \cfrac{1}{8 + \cfrac{1}{9 + \cfrac{1}{1 + \cfrac{1}{149083.6457209038\ldots}}}} \\
&= 0 + \cfrac{1}{8 + \cfrac{1}{9 + \cfrac{1}{1 + \cfrac{1}{149083 + \cfrac{1}{1.5486566936\ldots}}}}}
\end{aligned}$$

The forth term inside the fraction is 149083, which seems unusually large. This could already be seen from the first fraction, since it has six consecutive zeroes in the decimals of the denominator.

The fact that there is such a large term in the continued fraction means that the constant can be approximated relatively good by a relatively simple fraction. We can find the fraction that approximates the constant relatively good if we only consider the finite continued fraction with the terms before the large term.

$$\frac{1}{8+\frac{1}{9+\frac{1}{1}}} = \frac{1}{8+\frac{1}{10}} = \frac{10}{81} = 0.12345679012. \tag{51}$$

The approximation only fails after the seventh digit. Considering that the denominator is only a 2 digit number we can conclude that this approximation is indeed remarkably good.

Rational approximations of constants which are given by the finite continued fraction of the first n terms of its continued fraction, are called the *convergents* of that constant. In the previous example, we were considering the third convergent of the constant of Champernowne.

The fourth term of the continued fraction of the constant of Champernowne is not the only one which is really large. This can be seen in Table 3. In this table, a_4, a_{18} and a_{40} are given by 149083, $457\ldots987$ and $445\ldots423$ where the last two have 166 and 2504 digits respectively.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
a_n	8	9	1	a_4	1	1	1	4	1	1	1	3	4	1	1	1	15	a_{18}	6	1
n	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
a_n	1	21	1	9	1	1	2	3	1	7	2	1	83	1	156	4	58	8	54	a_{40}

Table 3: The first 40 terms of the continued fraction expansion of the constant of Champernowne

4.2 General notions for continued fractions

To investigate these large terms in the continued fraction of the constant of Champernowne, we first need to introduce some notions that will help us analyze the continued fraction.

4.2.1 The set of convergents

When we consider a continued fraction expansion of some real number, we can create a sequence of rational approximations of that real number. These approximations are given by the convergents of its continued fraction. We already have shown one of those convergents for the constant of Champernowne in equation (51). In general, we can write the continued fraction of some real number x in the following way.

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ldots}}}. \tag{52}$$

The set of terms of the continued fraction expansion of the constant x is given by $\{a_0; a_1, a_2, a_3, \ldots\}$. We will consider the term a_1 as the first term of the continued fraction expansion since it is the first term inside the nested fraction.

The convergents of the real number x are given by $\frac{p_n}{q_n}$. The subscript indicates that $\frac{p_n}{q_n}$ is the n^{th} convergent of x. This convergent is given by the finite continued fraction expansion with the terms $a_0, a_1, \ldots a_n$.

$$\frac{p_n}{q_n} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ldots + \frac{1}{a_{n-1} + \frac{1}{a_n}}}} \qquad \text{with} \qquad p_n \in \mathbb{Z}, \quad q_n \in \mathbb{N}, \quad \gcd(p_n, q_n) = 1.$$

The set of convergents of x will be denotes as CFA(x). The first four elements of this set are given by the convergents with $n = 0, 1, 2, 3$.

$$\begin{aligned} \frac{p_0}{q_0} &= a_o = 0 = \frac{0}{1}, & \frac{p_1}{q_1} &= a_o + \frac{1}{a_1} = 0 + \frac{1}{8} = \frac{1}{8}, \\ \frac{p_2}{q_2} &= a_o + \frac{1}{a_1 + \frac{1}{a_2}} = 0 + \frac{1}{8 + \frac{1}{9}} = \frac{9}{73}, & \frac{p_3}{q_3} &= a_o + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3}}} = 0 + \frac{1}{8 + \frac{1}{9 + \frac{1}{1}}} = \frac{10}{81}. \end{aligned} \tag{53}$$

Hence the set CFA(c) is given by $\{0, \frac{1}{8}, \frac{9}{73}, \frac{10}{81}, \ldots\}$.

4.2.2 Euler's rule

Convergents of a real number logically tend to go to that real number. An interesting thing to know would be the actual difference between the convergent and the real number x. To know this, we need to know how the convergents relate to the terms a_i of the continued fraction of the real number x.

The actual value of a convergent is most of the time hard to determine at first sight, since they are given by some finite continued fraction. It would be more clear if a convergent can be given by just a simple fraction. We could obtain this fraction by working out the continued fraction, but that calculation can get involved when the continued fraction gets larger.

If we look at the continued fraction of a convergent $\frac{p_n}{q_n}$, we see that the two integers p_n and q_n depend on the terms $a_0, a_1, \ldots, a_n$ and $a_1, a_2, \ldots, a_n$ respectively. The exact relation between these terms and the integers p_n and q_n was found by Euler [7, p.72]. The calculation of both p_n and q_n goes as follows for both terms:

First take the product of all the terms. Then take every product that can be obtained by omitting any pair of consecutive terms. Then take every product that can be obtained by omitting any two separate pairs of consecutive terms, and so on. This procedure will be notated as $[a_0, a_1, \ldots a_n]$. To clarify this procedure, let us consider an example.

$$[a_0, a_1, a_2, a_3] = a_0a_1a_2a_3 + a_0a_1 + a_0a_3 + a_2a_3 + 1$$

When we have an even number of terms, like above, we will get end with a product where all terms are omitted, which results in a one.

As a result of Euler's rule, we can write a finite continued fraction or a convergent as one fraction in stead of a large fraction of nested fractions. In general, the n^{th} convergent of a real number x can be written in the following way.

$$\frac{p_n}{q_n} = \frac{[a_0, a_1, a_2, \ldots, a_n]}{[a_1, a_2, \ldots, a_n]}. \tag{54}$$

We can illustrate this by calculating the third convergent of the constant of Champernowne again.

$$\frac{p_3}{q_3} = \frac{[a_0, a_1, a_2, a_3]}{[a_1, a_2, a_3]} = \frac{a_0a_1a_2a_3 + a_0a_1 + a_0a_3 + a_2a_3 + 1}{a_1a_2a_3 + a_1 + a_3} = \frac{0 \cdot 8 \cdot 9 \cdot 1 + 0 \cdot 8 + 0 \cdot 1 + 9 \cdot 1 + 1}{8 \cdot 9 \cdot 1 + 8 + 1} = \frac{10}{81}.$$

This value indeed corresponds with the value obtained in equations (51) and (53).

4.2.3 The difference between a real number and its convergents

A way to be able to find out the difference to a real number x and its convergents is by first writing our its convergent in terms of its previous convergent. This is possible by using Euler's rule. Let us consider the numerator of some convergent of the real number x, given by $p_n = [a_0, a_1, \dots, a_n]$. By writing out p_n in terms of its previous terms p_{n-1} and p_{n-2}, we will get the following [7, p.74].

If $n =$ even,

$$\begin{aligned} p_n &= a_0a_1\dots a_n + a_0a_1\dots a_{n-3}a_{n-2} + \dots + a_2a_3\dots a_{n-1}a_n + \dots + a_0a_1 + \dots + a_{n-1}a_n + 1,\\ p_{n-1} &= a_0a_1\dots a_{n-1} + a_0a_1\dots a_{n-4}a_{n-3} + \dots + a_2a_3\dots a_{n-2}a_{n-1} + \dots + a_0 + \dots + a_{n-1},\\ p_{n-2} &= a_0a_1\dots a_{n-2} + a_0a_1\dots a_{n-5}a_{n-4} + \dots + a_2a_3\dots a_{n-3}a_{n-2} + \dots + a_0a_1 + \dots + a_{n-3}a_{n-2} + 1,\\ p_n &= p_{n-1}a_n + p_{n-2}. \end{aligned}$$

If $n =$ odd,

$$\begin{aligned} p_n &= a_0a_1\dots a_n + a_0a_1\dots a_{n-3}a_{n-2} + \dots + a_2a_3\dots a_{n-1}a_n + \dots + a_0 + \dots + a_n,\\ p_{n-1} &= a_0a_1\dots a_{n-1} + a_0a_1\dots a_{n-4}a_{n-3} + \dots + a_2a_3\dots a_{n-2}a_{n-1} + \dots + a_0a_1 + \dots + a_{n-2}a_{n-1} + 1,\\ p_{n-2} &= a_0a_1\dots a_{n-2} + a_0a_1\dots a_{n-5}a_{n-4} + \dots + a_2a_3\dots a_{n-3}a_{n-2} + \dots + a_0 + \dots + a_{n-2},\\ p_n &= p_{n-1}a_n + p_{n-2}. \end{aligned}$$

The same holds for the denominator q_n. This can be seen by omitting a_0 in every equation above. Therefore, for all n, the n^{th} convergent can indeed be written in terms of the two previous convergents.

$$\frac{p_n}{q_n} = \frac{p_{n-1}a_n + p_{n-2}}{q_{n-1}a_n + q_{n-2}}. \tag{55}$$

To go on with finding the difference between the real number x and its convergents, we need some new notation. Let us denote the continued fraction of x from the n^{th} term as x_n. With this we mean the following.

$$x_n = a_n + \frac{1}{a_{n+1} + \frac{1}{a_{n-2}+\dots}}.$$

These terms x_n are called *complete quotients* of x. With this, we can write the real number x in terms of the first $n-1$ terms of the continued fraction and x_n and we can write every x_n as a function of a_n and x_{n-1}.

$$x = x_0, \qquad x = a_0 + \frac{1}{x_1}, \qquad x = a_0 + \frac{1}{a_1 + \frac{1}{x_2}}, \dots \quad \& \quad x_n = a_n + \frac{1}{x_{n+1}}.$$

This enables us to write an infinite continued fraction of an irrational number as a finite continued fraction. With equation (54), we have a way to write a continued fraction as a normal fraction. Hence, every real number can be written out as a normal fraction [7, p.79].

$$x = \frac{[a_0, a_1, a_2, \dots, a_{n-1}, x_n]}{[a_1, a_2, \dots, a_{n-1}, x_n]}.$$

Using the result from equation (55) we get

$$x = \frac{[a_0, a_1, a_2, \ldots, a_{n-1}]x_n + [a_0, a_1, a_2, \ldots, a_{n-2}]}{[a_1, a_2, \ldots, a_{n-1}]x_n + [a_1, a_2, \ldots, a_{n-2}]} = \frac{p_{n-1}x_n + p_{n-2}}{q_{n-1}x_n + q_{n-2}}.$$

Now that we have written the real number x in terms of two convergents and a complete quotient, let us inspect the difference between the real number x and its n^{th} convergent.

$$\begin{aligned}
\left| x - \frac{p_n}{q_n} \right| &= \left| \frac{p_n x_{n+1} + p_{n-1}}{q_n x_{n+1} + q_{n-1}} - \frac{p_n}{q_n} \right| \\
&= \left| \frac{q_n(p_n x_{n+1} + p_{n-1})}{q_n(q_n x_{n+1} + q_{n-1})} - \frac{(q_n x_{n+1} + q_{n-1})p_n}{(q_n x_{n+1} + q_{n-1})q_n} \right| \\
&= \left| \frac{q_n p_{n-1} - q_{n-1}p_n}{q_n(q_n x_{n+1} + q_{n-1})} \right|
\end{aligned}$$

We can further simplify this expression by using equation (55) and Euler's rule to its numerator.

$$\begin{aligned}
q_n p_{n-1} - q_{n-1}p_n &= (a_n q_{n-1} + q_{n-2})p_{n-1} - q_{n-1}(a_n p_{n-1} + p_{n-2}) \\
&= -(q_{n-1}p_{n-2} - q_{n-2}p_{n-1}) \\
&= (-1)^{n-1}(q_1 p_0 - q_0 p_1) \\
&= (-1)^{n-1}([a_1][a_0] - [1][a_0, a_1]) \\
&= (-1)^{n-1}(a_1 a_0 - (a_0 a_1 + 1)) \\
&= (-1)^n.
\end{aligned}$$

And hence we have that the difference between the real number x and its n^{th} convergent is equal to

$$\left| x - \frac{p_n}{q_n} \right| = \left| \frac{(-1)^n}{q_n(q_n x_{n+1} + q_{n-1})} \right| < \frac{1}{q_n^2 x_{n+1}}. \tag{56}$$

4.2.4 Approximation coefficients of general fractions

Another notion we will be introducing is the approximation coefficient. The approximation coefficient is defined as follows.

Definition 4.1. *The approximation coefficient of a rational number $\frac{A}{B}$ with respect to a real number x, is defined by*

$$ac\left(x, \frac{A}{B}\right) := B|Bx - A|. \quad \textit{[8, p.82]}$$

The aim of the approximation coefficient of a rational approximation $\frac{A}{B}$ with respect to the real number x is to quantify how well the rational number $\frac{A}{B}$ can approximate x with respect to its denominator. The smaller the value of $\text{ac}(x, \frac{A}{B})$ is, the better the approximation. Obviously, the approximation coefficient is zero if and only if $x = \frac{A}{B}$.

Let us illustrate the notion of approximation coefficients by applying it to the convergents of the constant of Champernowne, which are given by equations (53). The results are given in Table 4.

n	1	2	3	4
$\frac{p_n}{q_n}$	$\frac{0}{1}$	$\frac{1}{8}$	$\frac{9}{73}$	$\frac{10}{81}$
$\mathrm{ac}(c, \frac{p_n}{q_n})$	$0.12345\ldots$	$0.09876\ldots$	$0.90122\ldots$	$0.0000067\ldots$

Table 4: Approximation coefficients of the first four continued fraction approximations of c.

4.2.5 Approximation coefficients of convergents

Now that we know the difference between a real number and its convergents and the notion of approximation coefficients, we can determine the approximation coefficient of any convergent of any real number. Note that we can write the approximation coefficient of the convergent $\frac{p_n}{q_n}$ of the real number x in the following way, using Definition 4.1.

$$\mathrm{ac}\left(x, \frac{p_n}{q_n}\right) = q_n|q_n x - p_n| = q_n^2\left|x - \frac{p_n}{q_n}\right|.$$

Observe that the latter expression has an upper bound if we also consider equation (56).

$$\mathrm{ac}\left(x, \frac{p_n}{q_n}\right) = q_n^2\left|x - \frac{p_n}{q_n}\right| < q_n^2\frac{1}{q_n^2 x_{n+1}} = \frac{1}{x_{n+1}}.$$

Recall that the $n+1^{\text{th}}$ complete quotient of the real number x is given as $a_{n+1} + \frac{1}{x_{n+1}}$. If x would be negative, only a_0 would be negative, since the decimal part of x needs to have a non-negative value smaller than one. Because $\frac{1}{x_{n+1}}$ is by definition smaller than one, x_{n+1} needs to be greater than one. Hence the $n+1^{\text{th}}$ complete quotient is always larger than a_{n+1}. This results in the following.

$$\mathrm{ac}\left(x, \frac{p_n}{q_n}\right) < \frac{1}{x_{n+1}} < \frac{1}{a_{n+1}}. \tag{57}$$

For the difference between the real number x and one of its convergent $\frac{p_n}{q_n}$ we get

$$\left|x - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2 a_{n+1}}. \tag{58}$$

4.2.6 The theorem of Legendre

In the book "Essai sur la théorie des nombres", Legendre provides a theorem that states that certain rational approximations of some real number x need to be in its set of continued fraction approximations CFA(x). This theorem is given as follows.

Theorem 4.1. *If a rational approximation $\frac{A}{B}$ with respect to the real number x has an approximation coefficient smaller than one half, then $\frac{A}{B}$ belongs to the set of continued fraction approximations of x. I.e.*

$$ac\left(x, \frac{A}{B}\right) < \frac{1}{2} \qquad \Rightarrow \qquad \frac{A}{B} \in CFA(x). \quad [8, p.82]$$

Another result Legendre stated is that there are only rational approximations with an approximation coefficient of at most 1 inside the set of continued fraction approximations. This result follows easily from equation (57) since the terms of a continued fraction a_n are at least 1.

These two statements both turn out to be best possible. That is, for every $\epsilon > 0$, there exist a rational approximation $\frac{A}{B}$ to the real number x with $\mathrm{ac}(x, \frac{A}{B}) < \frac{1}{2} + \epsilon$ and $\frac{A}{B} \notin \mathrm{CFA}(x)$. The similar also holds for the second statement [8, p.2].

4.3 The relation between high watermarks and transcendence of c

Previously, we have observed that the constant of Champernowne contains terms which seem to be really large. We will be calling those terms high watermarks. There are more terms which are large but they are not always considered high watermarks. These terms and the high watermarks are both mentioned in the paper of Sikora [9]. These other large terms are called n^{th} generation high watermarks.

With equation (58), we have determined an upper bound for the difference between the real number x and one of its convergents $\frac{p_n}{q_n}$. Note that that upper bound depends is a product of a_{n+1}^{-1}. This implies that the upper bound will be really small if a_{n+1} is a high watermark, which implies that the convergent $\frac{p_n}{q_n}$ will be a good approximation.

Recall that in the proof of Mahler in Chapter 3, we also proved that there must be a sequence of fractions that approximate the constant very well. Let us again consider this upper bound, given by equation (48) and recall that we have proven this was true for all natural numbers $k < 10$ when s is sufficiently large and that $\lim_{s\to\infty} \log|R_s|/\log Q_s = -10$.

$$0 < \left|c - \frac{P_s}{Q_s}\right| = \left|R_s\right| \leq Q_s^{-k}. \tag{48}$$

We have that $Q_s > 2$ for all $s \geq 1$. Therefore, the approximation coefficient of the sequence of approximations in the proof of Mahler for a sufficiently large s have

$$\mathrm{ac}\left(c - \frac{P_s}{Q_s}\right) = Q_s^2\left|c - \frac{P_s}{Q_s}\right| \leq Q_s^2 Q_s^{-9} = Q_s^{-7} < 2^{-7} < \frac{1}{2}.$$

From the theorem of Legendre, Theorem 4.1, it follows from the previous result that the sequence of fractions $\frac{P_s}{Q_s}$ must be contained in $\mathrm{CFA}(c)$ for all sufficiently large s. Hence, there must be a sequence of convergents that have that $\lim_{s\to\infty} \log|R_s|/\log Q_s = -10$.

Since the sequence of fraction in the proof of Mahler and the sequence of convergents related to high watermarks are both proved to be remarkably small, it might be possible that they coincide. To confirm this theory, we will be comparing two characteristics of the sequence of fractions in the proof of Mahler with the properties of the high watermarks of the constant of Champernowne, given in a paper by John Sikora [9, p.5].

The first property of the sequence of fractions in the proof of Mahler we are going to compare has just been mentioned. Namely, that $\lim_{s\to\infty} \log|R_s|/\log Q_s = -10$. To confirm this, we will calculate $\log_{q_n}|c - \frac{p_n}{q_n}|$ of the first 8 high watermarks, beginning with the high watermark at $n = 4$. If the property holds for the high watermarks of the constant of champernowne, then $\log_{q_n}|c - \frac{p_n}{q_n}|$ has to tend to -10.

For the second property we are going to compare is the denominators. The denominator Q_s is given in equation (38). If the fractions of the proof of Mahler coincide with the convergents of the high watermarks, then the denominator of the continued fraction q_n is a divisor of Q_s and thus $\frac{Q_s}{q_n}$ needs to yield a whole number. This is true since the greatest common divisor of the fractions of Mahler are not necessarily equal to 1.

n	4	18	40	162
q_n	81	$4.9 \cdot 10^{11}$	$4.99 \cdot 10^{192}$	$4.999 \cdot 10^{2893}$
$\left\lvert c - \frac{p_n}{q_n} \right\rvert$	$1.0 \cdot 10^{-9}$	$9.1 \cdot 10^{-190}$	$9.01 \cdot 10^{-2890}$	$9.001 \cdot 10^{-38890}$
$\log_{q_n} \left\lvert c - \frac{p_n}{q_n} \right\rvert$	-4.715	-16.170	-14.992	-13.439
n	526	1708	4838	13522
q_n	$4.9999 \cdot 10^{38894}$	$4.99999 \cdot 10^{488895}$	$4.999999 \cdot 10^{5888896}$	$4.9999999 \cdot 10^{68888897}$
$\left\lvert c - \frac{p_n}{q_n} \right\rvert$	$9.0001 \cdot 10^{-488890}$	$9.00001 \cdot 10^{-5888890}$	$9.000001 \cdot 10^{-68888890}$	$9.0000001 \cdot 10^{-788888890}$
$\log_{q_n} \left\lvert c - \frac{p_n}{q_n} \right\rvert$	-12.569	-12.045	-11.698	-11.451

Table 5: The denominator and difference between c of the high watermarks of c.

In the table above, Table 5, we see that $\log_{q_n} |c - \frac{p_n}{q_n}|$ indeed seems to tend to -10, which confirms the deductions of the proof by Mahler.

The convergents relative to the terms considered in the generations after the high watermarks, for example the second generation [9, p.18], fail to hold the equation $\lim_{s\to\infty} \log |R_s| / \log Q_s = -10$ and thus are indeed not part of the sequence of fractions Mahler was talking about.

s, n	1,4	2,18	3,40	4,162
Q_s	81	$9.801 \cdot 10^{12}$	$1.207 \cdot 10^{197}$	$1.231 \cdot 10^{2901}$
$\frac{q_n}{Q_s}$	1	20	24200	24642000
s, n	5,526	6,1708	7,4838	8,13522
Q_s	$1.520 \cdot 10^{38909}$	$1.259 \cdot 10^{488913}$	$8.098 \cdot 10^{5888934}$	$8.099 \cdot 10^{68888942}$
$\frac{q_n}{Q_s}$	$3.041 \cdot 10^{14}$ *	$2.518 \cdot 10^{17}$ *	$1.619 \cdot 10^{38}$ *	$1.619 \cdot 10^{45}$ *

Table 6: The denominator of the fractions of the proof by Mahler and the integer fractions $\frac{q_n}{Q_s}$.

From Table 6 we see that the denominators Q_s are indeed multiples of q_n. The denominators of the first fractions of both the proof of Mahler and the convergents of the high watermarks even coincide. Although it is not visible in the table, the starred values are also integers. Since no other group of fractions meet the first property, it is indeed true that the convergents of the high watermarks are indeed the set of fractions that contains the sequence of fractions for which Mahler proved transcendence.

5 Conclusion

Three properties of the constant of Champernowne are normality, transcendence and its peculiar continued fraction. The goal was to state and explain these properties, clearly work out the proofs of the first two properties of the constant and investigate its continued fraction expansion. The constant is proved to be normal, using the argument of Champernowne and using the combinatorial method of proofs. For the proof that the constant is transcendental is, the argument of Mahler was used. The continued fraction expansion of the constant is shown to be quite remarkable with the use of some general notions for continued fraction expansions and thereafter a link was made between the sequence of fractions used to proof transcendence and the convergents of the continued fraction expansion of the constant.

References

[1] D. G. Champernowne, *The construction of decimals normal in the scale of ten*, Journal of the Londen Mathimatical Society, Volume s1-8, Issue 4 (1933), page 254-260.

[2] P. Pollack & J. Vandehey, *Besicovitch, bisection, and the normality of 0.(1)(4)(9)(16)(25)...*, The American Mathematical Monthly, Issue 122:8 (2015), page 757-765.

[3] K. Mahler, *Arithmetische Eigenschaften einer Klasse von Dezimalbrüchen*, Noord-Hollandsche Uitgevers Maatschappij (1937), page 421-428.

[4] K. Mahler, *Ein Analogen zu einem Schneiderschen Satz*, Koninklijke Akademie van Wetenschappen te Amsterdam, Volume 39, Issue 6 (1936).

[5] A. Baker, *Transcendental Number Theory*, Camebridge, Camebridge University Press (1975).

[6] E. Cesàro & G. Kowalewski, *Elementares Lehrbuch Der Algebraischen Analysis Und Der Infinitesimalrechnung: Mit Zahlreichen Übungsbeispielen*, Leipzig, B.G. Teubner (1904).

[7] H. Davenport, *The higher arithmetic* (8th ed.), Camebridge, Camebridge University Press (2008).

[8] D. Barbolosi & H. Jager, *On a theorem of Legendre in the theory of continued fractions*, Journal de Théorie des Nombres de Bordeaux, Volume 6, Issue 1 (1994), page 81-94.

[9] J. K. Sikora, *On the High Water Mark Convergents of Champernowne's Constant in Base Ten*, https://arxiv.org/ftp/arxiv/papers/1210/1210.1263.pdf (2012).