# Approximation of Cantor Rational Cardinalities by Primitive Words

Tara Trauthwein

June 13, 2020

Supervisor: Prof. Alexander D. Rahm
University of Luxembourg

## 1 Introduction

This project is about numerical computations related to the Cantor ternary set (in the following just called the Cantor set). This set is created by iteratively deleting the open middle third from a set of line segments. One starts by deleting the open middle third $(\frac{1}{3}, \frac{2}{3})$ from the interval $[0, 1]$, leaving two line segments: $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. We can express these iterations by the following relations:

$$\begin{cases} C_0 := [0,1] \\ C_n := \left(\frac{1}{3} \cdot C_{n-1}\right) \cup \left(\frac{2}{3} + \frac{1}{3} \cdot C_{n-1}\right), \; n \geqslant 1, \end{cases}$$

where for a set $A \subset \mathbb{R}$ and $\lambda, \mu \in \mathbb{R}$ the set $\lambda \cdot A + \mu$ is defined as:

$$\lambda \cdot A + \mu := \{\lambda \cdot a + \mu \mid a \in A\}.$$

The Cantor set $\mathcal{C}$ contains all the points in the interval $[0, 1]$ that are not deleted in this process:

$$\mathcal{C} := \bigcap_{n=1}^{\infty} C_n.$$

The Cantor rationals, $\mathcal{C} \cap \mathbb{Q}$, are the rational points in the Cantor set. If we look only at those which have their denominator in a certain interval that we fix, then we are considering a finite set, and we can wonder how large that set is. To be precise, we represent each Cantor rational as a fraction $\frac{p}{q}$ such that $p, q \in \mathbb{Z}$ are co-prime. Then we define the set of Cantor rationals of denominator $q$ as

$$N_q := \left\{ \frac{p}{q} \in \mathcal{C} \,\middle|\, p, q \in \mathbb{N}_{>0}, \; p \text{ is co-prime to } q \right\}.$$

We then sum up over the cardinalities $\#N_q$ of the sets $N_q$ :

$$N(T) := \sum_{(1-c)T \leqslant q \leqslant T} \#N_q,$$

for some $0 < c < 1$. For small $T \in \mathbb{N}$, one can compute $N(T)$ directly, but how does it behave asymptotically?

The article [4] by Rahm, Solomon, Trauthwein and Weiss aims to give an asymptotic approximation formula based on a heuristic argument, as well as numerical evidence.

In this project we establish a way to compute said approximations. In section 2, we recall fundamental results from combinatorics of words and in section 3, we remind the reader of the Möbius and Euler totient function. Section 4 states and proves some useful facts about purely periodic rationals and their ternary expansion. In section 5 we detail ways to calculate the number of primitive even and odd words and in section 6, the approximation formulas are given. Some results of the ensuing computations are shortly presented in section 7 and the code used to carry them out is listed in section 8.

## 2 Recall from combinatorics of words

We recall some results and their proofs from [3].

**Definition 2.1.** An **alphabet** $A$ is a non-empty finite set of symbols, called **letters** (in our case, $A = \{0, 1\}$ or $A = \{0, 1, 2\}$).
A **word** is a finite sequence of symbols from $A$, for example $u = 00121$. The empty word $\mathbb{1}$ is the sequence of no symbols at all.
We denote by $A^*$ the set of all words over the alphabet $A$.
Taking the **product** of words is the operation defined as

$$a_1 \dots a_n \cdot b_1 \dots b_m = a_1 \dots a_n b_1 \dots b_m$$

for any two words $a_1 \dots a_n$ and $b_1 \dots b_m$ in $A^*$.
The **length** of a word $w$, denoted by $|w|$, is the total number of letters in $w$, for example $|u| = |00121| = 5$.
A word $u$ is a **factor** of another word $w$ (resp. **prefix** or **suffix**) if there exist words $x$ and $y$ such that

$$w = xuy \text{ (resp. } w = uy \text{ or } w = xu).$$

All these are called **proper** if they are different from $w$.
We denote by $\operatorname{pref}_k(w)$ the prefix of length $k$ of $w$ (or $w$ if $|w| < k$).
Similar notations are used for the suffixes, which we denote by $\operatorname{suf}_k(w)$.

For the following, we fix an alphabet $A$.

Define the cyclic permutation function $c : A^* \to A^*$ by:

$$\begin{cases} c(\mathbb{1}) = \mathbb{1} \\ c(w) = \operatorname{pref}_1^{-1}(w) w \operatorname{pref}_1(w), \text{ for } w \in A \setminus \{\mathbb{1}\}, \end{cases}$$

**Definition 2.2.**

- Two words $x$ and $y$ are called **conjugates**, denoted $x \sim y$, if they can be obtained from each other by a sequence of cyclic permutations $c$ as defined above, i.e. there exist $k, l \in \mathbb{N}$ s.t. $x = c^k(y)$ and $y = c^l(x)$. It is easy to see that this is an equivalence relation.

- Let $w = a_1 \dots a_n$ with $a_i \in A$ for all $i = 0, 1, ..., n$. The number $p$ is called **a period** of $w$ if

  $$a_i = a_{i+p} \text{ for } i = 1, ..., n - p.$$

  We call the smallest period of $w$ **the period** of $w$ and denote it as $p(w)$. The elements in the conjugacy class of $\operatorname{pref}_{p(w)}(w)$ are called **cyclic roots** of $w$.

- We say that a word $w \neq \mathbb{1}$ is **primitive** if it is not a proper integer power of any of its cyclic roots, i.e. if for any cyclic root $u$, we cannot write

  $$w = u^n = \underbrace{uu...u}_{n \text{ times}}$$

  for some $n \in \mathbb{N} \setminus \{0, 1\}$.

*Example* 2.3. Some examples of conjugates, periods and primitive words:
Take $w = 00121$. Its conjugates are given by: 00121, 10012, 21001, 12100 and 01210. We obtained these by applying a cyclic permutation each time. The only periods of this word are the numbers $n \geqslant 5 = |w|$, since for those the condition to be satisfied is an empty condition. Hence $p(w) = 5$ and the cyclic roots are given by the conjugates of $w$. Thus $w$ is not a proper integer power of any of its cyclic roots and we conclude that $w$ is primitive.
Note that the word $w = 01010$ is also primitive, but its period is $p(w) = 2$. Its cyclic roots are given by 01 and 10 and its conjugates are 01010, 00101, 10010, 01001 and 10100.
The word $w = 020202$ is not primitive since it can be written $w = (02)^3$. Its conjugates are given by 020202 and 202020.

**Proposition 2.4** ([3], Theorem 2, p.6)**.** *Let $u, v, x, y \in A^*$ s.t. $uv = xy$. Then there exists a unique word $t \in A^*$ s.t. one of the two following statements is true:*

*(i) $u = xt$ and $y = tv$*

*(ii) $x = ut$ and $v = ty$.*

*Proof.* For symmetry reasons, we can assume without loss of generality that $|u| \geqslant |x|$. Then since $uv$ and $xy$ represent the same word in $A^*$, $x$ must be a prefix of $u$, which implies that there is a unique word $t \in A^*$ (possibly empty), such that $u = xt$. Now we can write

$$xy = uv = xtv$$

and again since those words are the same, we must have $y = tv$. Hence we get (i). Similarly, in assuming $|x| \geqslant |u|$, we get statement (ii). ■

**Proposition 2.5** ([3], Theorem 3, p.6)**.** *Let $u, v \in A^*$. The following assertions are equivalent:*

*(i) $u$ and $v$ commute, i.e. $uv = vu$.*

*(ii) $u$ and $v$ satisfy a non-trivial relation, i.e. $\exists\, \alpha, \beta \in \{u, v\}^*$, the alphabet consisting of the letters $u$ and $v$, such that $\alpha \neq \beta$ seen as words with letters $u$ and $v$, but $\alpha = \beta$ seen as words with letters in $A$.*

*(iii) There exists a word $t \in A^*$ such that $u = t^n$ and $v = t^m$, where $n, m \in \mathbb{N}$.*

*Example* 2.6. A non-trivial relation between words $u$ and $v$ is for example given by $u^2 = v$, but not by $uv = uv$.

*Proof.* (iii) ⇒ (i): This is clear since $uv = t^{n+m} = t^{m+n} = vu$.
(i) ⇒ (ii): The relation $uv = vu$ is non-trivial.
(ii) ⇒ (iii): The result is clear if $u$ or $v$ is empty. Indeed, then the non-trivial relation simplifies in the first case to an expression of the form $v^k = v^l$, $k, l \in \mathbb{N}$. If $k = l$, we can just take $t = v$, $n = 0$, $m = 1$ and if $k \neq l$, then $v$ must be empty, hence we take $t$ to be the empty word. Assume thus that $u$ and $v$ are non-empty and assume also without loss of generality that $|u| \geqslant |v|$. We use strong induction on $|u| + |v|$. In the case $|u| = |v| = 1$, it is clear since a non-trivial relation between letters $u$ and $v$ in $A$ implies that $u = v$, since we can compare the words $\alpha$ and $\beta$ letter by letter in $A$. Suppose that $|u| + |v| = k + 1 \geqslant 3$ and that the implication is true for any $u', v'$ satisfying $|u'| + |v'| \leqslant k$.
Assume thus that $\alpha = \beta$ in a non-trivial way. Up to removing some common prefixes consisting of $u$'s and $v$'s, we can assume without loss of generality that $\alpha = u\alpha_1$ and $\beta = v\beta_1$, where $\alpha_1, \beta_1 \in \{u, v\}^*$. By Proposition 2.4, there exists a word $w \in A^*$ such that $u = vw$. Denote by $\alpha', \beta', \alpha_1', \alpha_2'$ the words we obtain by replacing every $u$ with $vw$ in $\alpha, \beta, \alpha_1, \alpha_2$ respectively. We have now

$$\alpha' = \beta' \Leftrightarrow vw\alpha_1' = v\beta_1' \Leftrightarrow w\alpha_1' = \beta_1'.$$

The last equality describes a relation between $v$ and $w$ which is non-trivial since $\beta'_1$ must start with a $v$ and $w\alpha'_1$ starts with a $w$. We also have $|v|+|w|=|u|<|u|+|v|=k+1$ since $v$ is non-empty, and hence by the induction hypothesis, there exists a word $t$ such that $v=t^n$ and $w=t^m$ and hence $u=vw=t^{n+m}$. ■

The following proposition gives a very useful characterization of primitive words.

**Proposition 2.7** ([3], Theorem 1, p.4)**.** *A word $w \in A \setminus \{\mathbb{1}\}$ is primitive if and only if it satisfies:*

$$\forall\, z \in A^* : [w = z^n \Rightarrow n = 1,\ w = z].$$

*Proof.* If a non-empty word $w$ satisfies the property, then it is primitive since it cannot be a proper integer power of any cyclic root.
Assume that $w$ is primitive and $w = z^n$ with $n \geqslant 2$. Let $r = \mathrm{pref}_{p(w)}(w)$. Since $w$ is primitive, we get that $|r| \nmid |w|$, but by the hypothesis, $|z| \mid |w|$. In particular, $|r| \nmid |z|$. Since $w$ is an integer power of $z$, the length $|z|$ is a period of $w$ and thus by the two arguments combined, $|z| > p(w) = |r|$. We are in the following situation:

$$w = \underbrace{rr...r}_{m \text{ times}}\ x = \underbrace{zz...z}_{n \text{ times}},$$

where $x$ is a prefix of $r$ and $m \geqslant n$. Considering the first $z$, it can be seen that $r = \mathrm{pref}_{|r|}(z)$. On the other hand, considering the second $z$ and using that $|r| \nmid |z|$, we see that $\mathrm{pref}_{|r|}(z) = st$, where $s$ is a suffix of $r$ and $t = \mathrm{pref}_{|r|-|s|}(r)$, with $s, t \neq \mathbb{1}$. Thus we have $r = st = ts$ and by Proposition 2.5, this implies that $r$ is a proper integer power of a non-empty word $u$. This contradicts the fact that $|r|$ is the period of $w$. ■

Another very important property of primitive words says the following:

**Proposition 2.8** ([3], Corollary 2, p.7)**.** *Let $w \in A^* \setminus \{\mathbb{1}\}$. There exists a unique primitive word $\rho$ and a unique $n \geqslant 1$ such that $w = \rho^n$.*

*Proof.* The existence of at least one such $\rho$ is clear by Proposition 2.7. Indeed, if $w$ is primitive, there is nothing to show. If $w$ is not primitive, by the negation of the second condition in Proposition 2.7, there exists a word $z \neq w$ such that $w = z^n$, $n \geqslant 2$. If $z$ is primitive, we are done. If not, decompose $z$ in the same way. We continue in this fashion and will eventually find a $\rho$ since words consisting of one letter are always primitive.
For the uniqueness, assume $\rho_1, \rho_2$ are primitive with $w = \rho_1^n$ and $w = \rho_2^m$, $n, m \geqslant 1$. Then $\rho_1^n = \rho_2^m$ which is a non-trivial relation between $\rho_1$ and $\rho_2$. By Proposition 2.5, there exits a word $t$ such that both $\rho_1$ and $\rho_2$ are powers of $t$. Since both are primitive, by Proposition 2.7 we must have $\rho_1 = t = \rho_2$ and thus also $n = m$. ■

**Definition 2.9.** Let $w$ be a non-empty word. The unique primitive word $\rho(w)$ such that $w = \rho(w)^n$ for some $n \geqslant 1$ is called the **primitive root** of $w$. The unique number $n$ is called the **exponent** of $w$.

# 3 Möbius and Euler totient function

**Definition 3.1.** Define the **Möbius function** $\mu : \mathbb{N}_{>0} \to \{-1, 0, 1\}$ by:

$$\mu(n) := \begin{cases} (-1)^k, & \text{if } n \text{ is the product of } k \text{ distinct prime numbers} \\ 0, & \text{if there exists a prime number } p \text{ such that } p^2 \mid n. \end{cases}$$

We also recall the following widely-known result and recall a short proof for the implication.

**Proposition 3.2** (Möbius Inverse Formula)**.** *Let $\alpha, \beta : \mathbb{N}_{>0} \to \mathbb{Z}$ be functions. Then*

$$\alpha(n) = \sum_{d|n} \beta(d), \ \forall\, n \geqslant 1$$

*implies*

$$\beta(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) \alpha(d), \ \forall\, n \geqslant 1.$$

The proof is a recall from [5], Theorem 2 (Note), p.4.

*Proof.* Let $n \geqslant 1$. Then:

$$\begin{aligned}
\sum_{d|n} \mu\left(\frac{n}{d}\right) \alpha(d) &= \sum_{d|n} \mu\left(\frac{n}{d}\right) \sum_{k|d} \beta(k) \\
&= \sum_{k|n} \beta(k) \sum_{\substack{k|d \\ d|n}} \mu\left(\frac{n}{d}\right) \\
&= \sum_{k|n} \beta(k) \sum_{\substack{k|\frac{n}{e} \\ e|n}} \mu\left(e\right) \\
&= \sum_{k|n} \beta(k) \sum_{e|\frac{n}{k}} \mu\left(e\right) \\
&= \beta(n)
\end{aligned}$$

using the fact that $k \mid \frac{n}{e}$, $e \mid n \Leftrightarrow k \mid n$, $e \mid \frac{n}{k}$ and the well-known result

$$\sum_{d|n} \mu(d) = \begin{cases} 0 & \text{if } n \neq 1 \\ 1 & \text{if } n = 1. \end{cases}$$

■

**Definition 3.3.** Define the **Euler totient function** $\Phi : \mathbb{N}_{>0} \to \mathbb{N}$ by:

$$\Phi(n) := \#\{m \in \mathbb{N}_{>0} \mid m \leqslant n, \, \gcd(m, n) = 1\}.$$

# 4 The ternary expansion and purely periodic rationals

Denote by $\mathcal{O}(3, q)$ the **multiplicative order** of 3 in $\mathbb{Z}/q\mathbb{Z}$, i.e. the smallest $\ell$ such that $3^\ell \equiv 1 \mod q$, for any $q \in \mathbb{N}_{>0}$.
Then recall the following well-known fact about the ternary number system.
For any $x \in \mathbb{R}^+$, there exists a unique $n \in \mathbb{N}$ and a unique sequence $(a_i)_{i \in \mathbb{N}} \in \{0, 1, 2\}^{\mathbb{N}}$ s.t.:

$$x = \sum_{i=-\infty}^{n} a_{n-i} \cdot 3^i.$$

We make the choice of representing the series as a finite sum whenever it is possible, in order to obtain uniqueness. (Note that for purposes of distinguishing Cantor rationals, the opposite choice of representing the series with an infinite sequence of digits "2" is usually made. But we do not need to distinguish Cantor rationals in the proof of Lemma 4.1, which works for ternary expansions in general.)
The coefficients above are of course determined by $x$, but we suppress this from the notation.
The sequence $a_0 a_1 a_2 \dots a_n.a_{n+1} a_{n+2} \dots$ is called the **ternary expansion** of $x$ (if $n = 0$ and $a_0 = 0$, we say that $a_1 a_2 \dots$ is the ternary expansion of $x$).

If $x \in \mathbb{Q}^+$, this sequence becomes periodic after an initial finite sequence of coefficients, since when doing an integer division in any number system, there are only a finite number of possible rests.

Next, we need some definitions and facts for this specific situation. We call a rational $\frac{p}{q} \in \mathbb{Q} \cap (0,1)$ **purely periodic** if its ternary expansion is periodic from the start. As $\frac{p}{q} \in (0,1)$, one can always write uniquely (up to the above convention):

$$\frac{p}{q} = \sum_{i=1}^{\infty} a_i \cdot 3^{-i}$$

with $a_i \in \{0,1,2\}^{\mathbb{N}}$ for all $i = 1, 2, \ldots$. Hence if $\frac{p}{q}$ is purely periodic, it is periodic starting from $a_1$.

**Lemma 4.1.** *Let* $\frac{p}{q} \in \mathbb{Q} \cap (0,1)$ *with* $\gcd(p,q) = 1$*.*

*(i) The rational* $\frac{p}{q}$ *is purely periodic if and only if* $3 \nmid q$*.*

*(ii) If* $\frac{p}{q}$ *is purely periodic, the period length* $\ell$ *of* $\frac{p}{q}$ *depends only on* $q$ *and is given by* $\mathcal{O}(3,q)$*. Moreover,* $\ell \mid \Phi(q)$*, where* $\Phi$ *is the Euler totient function.*

*Proof.* (i) The rational $\frac{p}{q}$ can be written uniquely as:

$$\frac{p}{q} = \frac{1}{3}b_1 + \ldots + \frac{1}{3^n}b_n + \frac{1}{3^{n+1}}a_1 + \ldots + \frac{1}{3^{n+\ell}}a_\ell + \frac{1}{3^{n+\ell+1}}a_1 + \ldots \tag{1}$$

where the ternary expansion of $\frac{p}{q}$ is given by

$$b_1 \ldots b_n\overline{a_1 \ldots a_\ell},$$

i.e. $b_1 \ldots b_n$ is the non-periodic part and $a_1 \ldots a_\ell$ is the repeating sequence. We can assume here that $n$ is minimal. If $n = 0$, then $\frac{p}{q}$ is purely periodic.
Rewrite (1) in a different form:

$$\begin{aligned}
&\frac{p}{q} - \frac{1}{3}b_1 - \ldots - \frac{1}{3^n}b_n = \frac{1}{3^{n+1}}a_1 + \ldots + \frac{1}{3^{n+\ell}}a_\ell + \frac{1}{3^{n+\ell+1}}a_1 + \ldots \\
\Leftrightarrow &3^n\frac{p}{q} - \sum_{k=1}^{n} 3^{n-k}b_k = \frac{1}{3}a_1 + \ldots + \frac{1}{3^\ell}a_\ell + \frac{1}{3^{\ell+1}}a_1 + \ldots \qquad (2)
\end{aligned}$$

Define

$$Q := 3^n\frac{p}{q} - \sum_{k=1}^{n} 3^{n-k}b_k. \tag{3}$$

Then by (2):

$$\begin{aligned}
Q &= \frac{1}{3}a_1 + \ldots + \frac{1}{3^\ell}a_\ell + \frac{1}{3^{\ell+1}}a_1 + \ldots \\
\Rightarrow 3^\ell Q &= 3^{\ell-1}a_1 + \ldots + 3^0 a_\ell + \frac{1}{3}a_1 + \ldots + \frac{1}{3^\ell}a_\ell + \frac{1}{3^{\ell+1}}a_1 + \ldots \\
\Rightarrow 3^\ell Q &= 3^{\ell-1}a_1 + \ldots + 3^0 a_\ell + Q \\
\Rightarrow \left(3^\ell - 1\right) Q &= 3^{\ell-1}a_1 + \ldots + 3^0 a_\ell.
\end{aligned}$$

Replacing $Q$ again by the expression in (3) and multiplying by $q$, we get:

$$\begin{aligned}
&\left(3^n p - q\sum_{k=1}^{n} 3^{n-k}b_k\right)\left(3^\ell - 1\right) = q\left(3^{\ell-1}a_1 + \ldots + 3^0 a_\ell\right) \\
\Leftrightarrow &3^n p\left(3^\ell - 1\right) = q\left(\sum_{i=1}^{\ell} 3^{\ell-i}a_i + \left(3^\ell - 1\right)\sum_{k=1}^{n} 3^{n-k}b_k\right) \qquad (4)
\end{aligned}$$

Assume $3 \mid q$. Then 3 must divide the left hand side of (4). As $3 \nmid (3^\ell - 1)$ (since $\ell > 0$) and $3 \nmid p$ (since $\gcd(p,q) = 1$), we must have $3 \mid 3^n$, which implies $n \geqslant 1$ and $\frac{p}{q}$ is not purely periodic.

Assume $3 \nmid q$. If $n \geqslant 1$, we have that 3 divides the left hand side of (4). As $3 \nmid q$, we must have that 3 divides

$$\begin{aligned} &\sum_{i=1}^{\ell} 3^{\ell-i} a_i + (3^\ell - 1) \sum_{k=1}^{n} 3^{n-k} b_k \\ =&\sum_{k=1}^{n} 3^{n+\ell-k} b_k + \sum_{i=1}^{\ell} 3^{\ell-i} a_i - \sum_{m=1}^{n} 3^{n-m} b_m. \end{aligned}$$

This expression can only be divisible by 3 if $a_\ell - b_n = 0$, since all other terms are already divisible by 3 and $a_i, b_j \in \{0,1,2\}$. But if $a_\ell = b_n$, the ternary expansion can be written:

$$b_1 \dots b_{n-1} a_\ell a_1 \dots a_{\ell-1} a_\ell a_1 \dots$$

and thus it is periodic starting from the $n^{\text{th}}$ coefficient with periodic part $a_\ell a_1 \dots a_{\ell-1}$. This contradicts the assumption that $n$ is minimal. Thus if $3 \nmid q$, then $n = 0$.
(ii) Assume $\frac{p}{q}$ to be purely periodic. Equation (4) with $n = 0$ gives:

$$p(3^\ell - 1) = q \left( \sum_{k=1}^{\ell} 3^{l-k} a_k \right).$$

Since $\gcd(p,q) = 1$, we must have $q \mid (3^\ell - 1)$.
Let $m = \mathcal{O}(3,q)$. Then $q \mid 3^m - 1$ and by basic group theory, $m \mid \ell$. Moreover, since $\frac{p}{q} < 1$, we have:

$$p \cdot \frac{3^m - 1}{q} < 3^m - 1 < 3^m$$

and thus $\exists\, c_1, \dots, c_m \in \{0,1,2\}$ s.t.

$$p \cdot \frac{3^m - 1}{q} = 3^{m-1} c_1 + \ldots + 3^0 c_m \tag{5}$$

Recall that

$$\frac{p}{q} = \frac{1}{3} a_1 + \ldots + \frac{1}{3^\ell} a_\ell + \frac{1}{3^{\ell+1}} a_1 + \ldots$$

and calculate:

$$(3^m - 1)\frac{p}{q} = 3^{m-1} a_1 + 3^{m-2} a_2 + \ldots + 3^0 a_m + \frac{1}{3} a_{m+1} + \ldots + \frac{1}{3^{\ell-m}} a_\ell - \frac{p}{q} + \underbrace{3^{m-\ell}\frac{p}{q}}_{<3^{m-\ell}}.$$

Comparing this with (5) and by uniqueness of the ternary expansion, we get:

$$\begin{gathered} a_1 = c_1, \dots, a_m = c_m \\ \text{and } a_{m+k} = a_k \;\forall\, k = 1, \dots \ell - m. \end{gathered}$$

This implies (keeping in mind that $m \mid \ell$) that the ternary expansion of $\frac{p}{q}$ is $m$-periodic and equal to

$$\overline{c_1 \dots c_m}.$$

As we assumed that $\ell$ is the period, i.e. the smallest number such that $a_i = a_{i+\ell}$ for all $i \in \mathbb{N}_{>0}$, we must have $\ell = m$.
As $\ell = \mathcal{O}(3,q)$ and $\#(\mathbb{Z}/q\mathbb{Z}) = \Phi(q)$, by basic group theory we must have $\ell \mid \Phi(q)$. ∎

# 5 Computing the number of primitive words

From what we established above, we now know that a purely periodic Cantor rational can be written uniquely in ternary expansion and that the expansion is periodic from the beginning. The periodic part $a_1 \dots a_\ell$ can be seen as a primitive word over the alphabet $\{0, 1, 2\}$. Indeed, it is easy to see that $a_1 \dots a_\ell = (b_1 \dots b_m)^n$ implies $n = 1$ and $a_1 \dots a_\ell = b_1 \dots b_m$, thus, using the characterization in Proposition 2.7, $a_1 \dots a_\ell$ is primitive. This motivates a closer look at primitive words, in particular at how to calculate their number effectively.

**Definition 5.1.** Let $A$ be an alphabet of size $a = \#A$ and let $\ell \in \mathbb{N}$.
Then the function m is defined by

$$\mathrm{m}(\ell, a) := \#\{w \in A^* \mid w \text{ is primitive and } |w| = \ell\}.$$

Assume $A = A_1 \sqcup A_2$, where the letters in $A_1$ are said to be even and the letters in $A_2$ are said to be odd. A word $w \in A^*$ is said to be **even** if $w$ contains an even number of letters from $A_2$ and **odd** if it contains an odd number of letters from $A_2$. Assume for the sake of simplicity that $\#A_1 = \left\lceil \frac{\#A}{2} \right\rceil$ and $\#A_2 = \left\lfloor \frac{\#A}{2} \right\rfloor$.
Then, we define

$$\mathrm{e}(\ell, a) := \#\{w \in A^* \mid w \text{ is primitive, even and } |w| = \ell\}$$

and

$$\mathrm{o}(\ell, a) := \#\{w \in A^* \mid w \text{ is primitive, odd and } |w| = \ell\}.$$

*Remark* 5.2. The assumption $\#A_1 = \left\lceil \frac{\#A}{2} \right\rceil$ is the natural way of defining the number of even and odd letters, since in general we consider $A = \{0, 1, 2, ..., n\}$ for some $n \in \mathbb{N}_{>0}$.

## 5.1 Computing the Number of Primitive Words

The following proposition is again a recall from [3]. We adapted it very slightly to better suit our needs.

**Proposition 5.3** ([3], Lemma 3, p.14)**.** *For an alphabet $A$ with $\#A = a$, for all $\ell \in \mathbb{N}_{>0}$, we have:*

$$a^\ell = \sum_{d \mid \ell} \mathrm{m}(d, a).$$

*Proof.* By Proposition 2.8, there is a 1-to-1 correspondence between a word $x \in A^*$ and the pair $(\rho, e)$, where $\rho$ and $e$ are the primitive root and exponent of $x$ respectively. Hence the number of all words of length $\ell$ can be written on the one hand as $a^\ell$ and on the other hand by the number of all such pairs. The only condition on the exponent is that it divides $\ell$ and for each exponent, the number of possible primitive $\rho$ is exactly given by $\mathrm{m}(\frac{\ell}{e}, a)$. Hence we have:

$$a^\ell = \sum_{e \mid \ell} m\left(\frac{\ell}{e}, a\right) = \sum_{d \mid \ell} m(d, a)$$

which was the claim. ∎

**Proposition 5.4** ([3], Theorem 7, p.16)**.** *For an alphabet $A$ with $\#A = a$, for all $\ell \in \mathbb{N}_{>0}$, we have:*

$$\mathrm{m}(\ell, a) = \sum_{d \mid \ell} \mu\left(\frac{\ell}{d}\right) a^d.$$

*Proof.* Apply the Möbius Inverse Formula (see Proposition 3.2) to the expression in Proposition 5.3. ∎

This formula gives a much better way to calculate the number of primitive words over a given length and a given alphabet. Instead of computing them all and counting, we just need to compute the divisors of $\ell$ and sum over them.

## 5.2 Computing the Number of Primitive Even Words

In the following we are going to elaborate two different but equivalent formulas for the number of primitive even words. This subsection is not directly based on [3] but is inspired from its proof strategies.

### 5.2.1 The First Formula

For the first one, denote by $\mathrm{val}_p(n)$ for a prime $p$ and $n \in \mathbb{N}_{>0}$ the $p$-valuation of $n$, i.e.

$$\mathrm{val}_p(n) = \max\{k \in \mathbb{N} : p^k \mid n\}.$$

Then we have:

**Proposition 5.5.** *Let $\ell \in \mathbb{N}_{>0}$ and let $A$ be an alphabet with $a = \#A$ and the number of even elements in $A$ be given by $\lceil \frac{a}{2} \rceil$. Then*

$$\mathrm{e}(\ell, a) = \sum_{d \mid \ell} \mu\left(\frac{\ell}{d}\right) \left[ a^d + (\mathrm{val}_2(d) + 1) \frac{1 - (-1)^a}{4} - \frac{1}{2} \sum_{k=0}^{val_2(d)} a^{d/2^k} \right].$$

This will be an immediate consequence of the following proposition:

**Proposition 5.6.** *Let $\ell \in \mathbb{N}_{>0}$ and let $A$ be an alphabet with $a = \#A$ and the number of even elements in $A$ is given by $\lceil \frac{a}{2} \rceil$. Then*

$$\sum_{d \mid \ell} \mathrm{e}(\ell, a) = a^\ell + (\mathrm{val}_2(\ell) + 1) \frac{1 - (-1)^a}{4} - \frac{1}{2} \sum_{k=0}^{\mathrm{val}_2(\ell)} a^{\ell/2^k}. \tag{6}$$

To show this, we first need a lemma.

**Lemma 5.7.** *Under the assumptions of Proposition 5.6, we have:*

$$\left\lceil \frac{a^\ell}{2} \right\rceil = \frac{a^\ell}{2} + \frac{1 - (-1)^a}{4} = \sum_{\substack{d \mid \ell \\ \ell/d \text{ even}}} \mathrm{m}(d, a) + \sum_{\substack{d \mid \ell \\ \ell/d \text{ odd}}} \mathrm{e}(d, a). \tag{7}$$

*This expression gives exactly the number of all even words of length $\ell$.*

*Proof.* **Step 1.** We are going to show that the left hand side of equation (7) is exactly the number of all even words of length $\ell$. We prove this by induction. The formula is true for $\ell = 1$, since then the number of even words of length 1 is given by the number of even letters, which is

$$\left\lceil \frac{a}{2} \right\rceil = \frac{a}{2} + \frac{1 - (-1)^a}{4}.$$

Assume it is true for words of length $n$. Then the number of even words of length $n+1$ is given by:

$$\#\{\text{even words of length } n\} \cdot \#\{\text{even letters}\} + \#\{\text{odd words of length } n\} \cdot \#\{\text{odd letters}\}.$$

In formulas:

$$
\begin{aligned}
&\left(\frac{a^n}{2}+\frac{1-(-1)^a}{4}\right)\cdot\left(\frac{a}{2}+\frac{1-(-1)^a}{4}\right)+\left(a^n-\frac{a^n}{2}-\frac{1-(-1)^a}{4}\right)\cdot\left(\frac{a}{2}-\frac{1-(-1)^a}{4}\right)\\
=&a^n\left(\frac{a}{2}-\frac{1-(-1)^a}{4}\right)+\left(\frac{a^n}{2}+\frac{1-(-1)^a}{4}\right)\cdot\left(\frac{a}{2}+\frac{1-(-1)^a}{4}-\frac{a}{2}+\frac{1-(-1)^a}{4}\right)\\
=&\frac{a^{n+1}}{2}+\frac{(1-(-1)^a)^2}{8}\\
=&\frac{a^{n+1}}{2}+\frac{1-(-1)^a}{4}
\end{aligned}
$$

which proves the claim of Step 1.

**Step 2.** For the right hand side of equation (7), we consider Proposition 2.8, which states that there is a 1-to-1 correspondence between words $x \in A^*$ and pairs $(\rho, e)$, where $\rho$ is primitive. Every even word must satisfy one of the following conditions:

- its primitive root $\rho$ is even;
- its exponent $e$ is even.

We can thus write the number of all even words of length $\ell$ as the following sum:

$$\sum_{\substack{d\mid\ell\\ \ell/d \text{ even}}} \mathrm{m}(d,a)+\sum_{\substack{d\mid\ell\\ \ell/d \text{ odd}}} \mathrm{e}(d,a).$$

This proves equation (7). ■

*Proof of Proposition 5.6.* **Step 1.** Assume $\ell$ is odd. Then by Lemma 5.7 we have:

$$\frac{a^\ell}{2}+\frac{1-(-1)^a}{4}=\sum_{\substack{d\mid\ell\\ \ell/d \text{ even}}} \mathrm{m}(d,a)+\sum_{\substack{d\mid\ell\\ \ell/d \text{ odd}}} \mathrm{e}(d,a)=\sum_{d\mid\ell}\mathrm{e}(d,a)$$

since $\frac{\ell}{d}$ is always odd as $\ell$ is odd. This is exactly equation (6), since $\mathrm{val}_2(\ell)=0$.
**Step 2.** To prove the statement of the proposition for general $\ell$, we use induction on $\mathrm{val}_2(\ell)$.
**Base Case:** $\mathrm{val}_2(\ell)=0$, i.e. $\ell$ is odd. This case was dealt with in Step 1.
**Induction step:** Assume that equation (6) is true for all $\ell=n\in\mathbb{N}$ where $\mathrm{val}_2(n)=p$, for some $p\geqslant 0$. Assume that $\mathrm{val}_2(\ell)=p+1$. This implies that $\ell$ is even, i.e. $\ell=2\cdot\ell'$, for some $\ell'\in\mathbb{N}$.
By Lemma 5.7 we have for $\ell=2\ell'$:

$$
\begin{aligned}
\frac{a^\ell}{2}+\frac{1-(-1)^a}{4}&=\frac{a^{2\ell'}}{2}+\frac{1-(-1)^a}{4}\\
&=\sum_{\substack{d\mid 2\ell'\\ 2\ell'/d \text{ even}}} \mathrm{m}(d,a)+\sum_{\substack{d\mid 2\ell'\\ 2\ell'/d \text{ odd}}} \mathrm{e}(d,a)\\
&=\sum_{d\mid\ell'}\mathrm{m}(d,a)+\sum_{d\mid 2\ell'}\mathrm{e}(d,a)-\sum_{d\mid\ell'}\mathrm{e}(d,a)\\
&=a^{\ell'}+\sum_{d\mid 2\ell'}\mathrm{e}(d,a)-\sum_{d\mid\ell'}\mathrm{e}(d,a)
\end{aligned}
$$

where we use Proposition 5.3 and the fact that ($d\mid 2\ell'$ and $\frac{2\ell'}{d}$ even ) $\Leftrightarrow d\mid\ell'$.

Now by the definition of $\ell'$, we have $\mathrm{val}_2(\ell') = p$ and we can apply the induction hypothesis:

$$\sum_{d|\ell'} \mathrm{e}(d,a) = a^{\ell'} + \big(\mathrm{val}_2(\ell')+1\big)\,\frac{1-(-1)^a}{4} - \frac{1}{2}\sum_{k=0}^{\mathrm{val}_2(\ell')} a^{\ell'/2^k}.$$

Thus we can calculate:

$$\begin{aligned}
\sum_{d|\ell} \mathrm{e}(d,a) &= \sum_{d|2\ell'} \mathrm{e}(d,a) \\
&= \frac{a^{2\ell'}}{2} + \frac{1-(-1)^a}{4} - a^{\ell'} + a^{\ell'} + \big(\mathrm{val}_2(\ell')+1\big)\,\frac{1-(-1)^a}{4} - \frac{1}{2}\sum_{k=0}^{\mathrm{val}_2(\ell')} a^{\ell'/2^k} \\
&= a^{2\ell'} + \big(\mathrm{val}_2(2\ell')+1\big)\,\frac{1-(-1)^a}{4} - \frac{1}{2}a^{2\ell'} - \frac{1}{2}\sum_{k=1}^{\mathrm{val}_2(2\ell')} a^{2\ell'/2^k} \\
&= a^{2\ell'} + \big(\mathrm{val}_2(2\ell')+1\big)\,\frac{1-(-1)^a}{4} - \frac{1}{2}\sum_{k=0}^{\mathrm{val}_2(2\ell')} a^{2\ell'/2^k} \\
&= a^{\ell} + \big(\mathrm{val}_2(\ell)+1\big)\,\frac{1-(-1)^a}{4} - \frac{1}{2}\sum_{k=0}^{\mathrm{val}_2(\ell)} a^{\ell/2^k}.
\end{aligned}$$

Thus we have shown that the equation (6) is true for all $\ell \in \mathbb{N}$ with $\mathrm{val}_2(\ell) = p+1$. By induction, this concludes the proof of the proposition. ■

*proof of Prop. 5.5.* Apply the Möbius Inverse Formula (see Proposition 3.2) to the relation in Proposition 5.6. ■

### 5.2.2 The Second Formula

Here the author would like to thank Noam Solomon for the idea of this formula. It goes as follows:

**Proposition 5.8.** *Let $\ell \in \mathbb{N}_{>0}$ and let $A$ be an alphabet with $a = \#A$ and the number of even elements in $A$ be given by $\lceil \frac{a}{2} \rceil$. Then*

$$\begin{aligned}
\mathrm{e}(\ell,a) &= \sum_{\substack{d|\ell \\ \ell/d \text{ even}}} \mu\left(\frac{\ell}{d}\right) a^d + \sum_{\substack{d|\ell \\ \ell/d \text{ odd}}} \mu\left(\frac{\ell}{d}\right)\left(\frac{a^d}{2} + \frac{1-(-1)^a}{4}\right) \\
&= \sum_{\substack{d|\ell \\ \ell/d \text{ even}}} \mu\left(\frac{\ell}{d}\right) a^d + \sum_{\substack{d|\ell \\ \ell/d \text{ odd}}} \mu\left(\frac{\ell}{d}\right)\left\lceil \frac{a^d}{2} \right\rceil.
\end{aligned}$$

Recall that $\mathrm{o}(\ell,a)$ denotes the number of odd primitive words of length $\ell$ over an alphabet $A$ with $a = \#A$. In particular, we have

$$\mathrm{m}(\ell,a) = \mathrm{e}(\ell,a) + \mathrm{o}(\ell,a).$$

**Lemma 5.9.** *Let $\ell \in \mathbb{N}_{>0}$ and let $A$ be an alphabet with $a = \#A$ and the number of even elements in $A$ is given by $\lceil \frac{a}{2} \rceil$. Then*

$$\left\lfloor \frac{a^\ell}{2} \right\rfloor = \frac{a^\ell}{2} - \frac{1-(-1)^a}{4} = \sum_{\substack{d|\ell \\ \ell/d \text{ odd}}} \mathrm{o}(d,a). \tag{8}$$

*Proof.* We are going to prove this by using a double counting argument and showing that both sides of the equation give the number of odd words of length $\ell$.
**Step 1.** By Lemma 5.7, the number of all even words of length $\ell$ is given by:

$$\frac{a^\ell}{2} + \frac{1-(-1)^a}{4},$$

therefore the number of all odd words of length $\ell$ is:

$$a^\ell - \left(\frac{a^\ell}{2} + \frac{1-(-1)^a}{4}\right) = \frac{a^\ell}{2} - \frac{1-(-1)^a}{4}.$$

**Step 2.** We use again Proposition 2.8, which states that there is a 1-to-1 correspondence between words $x \in A^*$ and pairs $(\rho, e)$, where $\rho$ is primitive. A word is odd if and only if the following two statements are true:

- its primitive root $\rho$ is odd;
- its exponent $e$ is odd.

We can thus write the number of all odd words of length $\ell$ as the following sum:

$$\sum_{\substack{d \mid \ell \\ \ell/d \text{ odd}}} \mathrm{o}(d, a),$$

which proves the Lemma. ∎

To prove Proposition 5.8, we need a modified version of the Möbius Inverse Formula.

**Proposition 5.10.** *Let $\alpha, \beta : \mathbb{N}_{>0} \to \mathbb{Z}$ be functions. Then*

$$\alpha(n) = \sum_{\substack{d \mid n \\ n/d \text{ odd}}} \beta(d), \ \forall\, n \geqslant 1$$

*implies*

$$\beta(n) = \sum_{\substack{d \mid n \\ n/d \text{ odd}}} \mu\left(\frac{n}{d}\right) \alpha(d), \ \forall\, n \geqslant 1.$$

*Proof.* Let $n \geqslant 1$. Then:

$$\begin{aligned}
\sum_{\substack{d \mid n \\ n/d \text{ odd}}} \mu\left(\frac{n}{d}\right) \alpha(d) &= \sum_{\substack{d \mid n \\ n/d \text{ odd}}} \sum_{\substack{k \mid d \\ d/k \text{ odd}}} \mu\left(\frac{n}{d}\right) \beta(k) \\
&= \sum_{\substack{k \mid n \\ n/k \text{ odd}}} \beta(k) \sum_{\substack{k \mid d \\ d \mid n}} \mu\left(\frac{n}{d}\right) \\
&= \sum_{\substack{k \mid n \\ n/k \text{ odd}}} \beta(k) \sum_{e \mid \frac{n}{k}} \mu\left(e\right) \\
&= \beta(n)
\end{aligned}$$

using the facts that

$$d \mid n,\ k \mid d,\ \frac{n}{d} \text{ odd},\ \frac{d}{k} \text{ odd} \Leftrightarrow k \mid n,\ k \mid d,\ d \mid n,\ \frac{n}{k} \text{ odd}$$

and

$$k \mid \frac{n}{e},\, e \mid n \Leftrightarrow k \mid n,\, e \mid \frac{n}{k}$$

as well as the well-known result

$$\sum_{d|n} \mu(d) = \begin{cases} 0 & \text{if } n \neq 1 \\ 1 & \text{if } n = 1. \end{cases}$$

■

*proof of Proposition 5.8.* Apply the modified Möbius Inverse Formula (Proposition 5.10) to the equation (8) in Lemma 5.9 to get:

$$\mathrm{o}(\ell, a) = \sum_{\substack{d|\ell \\ \ell/d}} \mu\left(\frac{\ell}{d}\right)\left(\frac{a^d}{2} - \frac{1-(-1)^a}{4}\right).$$

Using that every primitive word is either odd or even, we get:

$$\begin{aligned} \mathrm{e}(\ell, a) &= \mathrm{m}(\ell, a) - \mathrm{o}(\ell, a) \\ &= \sum_{d|\ell} \mu\left(\frac{\ell}{d}\right) a^d - \sum_{\substack{d|\ell \\ \ell/d \text{ odd}}} \mu\left(\frac{\ell}{d}\right)\left(\frac{a^d}{2} - \frac{1-(-1)^a}{4}\right) \\ &= \sum_{\substack{d|\ell \\ \ell/d \text{ even}}} \mu\left(\frac{\ell}{d}\right) a^d + \sum_{\substack{d|\ell \\ \ell/d \text{ odd}}} \mu\left(\frac{\ell}{d}\right)\left(\frac{a^d}{2} + \frac{1-(-1)^a}{4}\right). \end{aligned}$$

■

# 6 Approximation formulas

For $q \in \mathbb{N}_{>0}$, denote by $\ell$ the period length in ternary expansion of a purely periodic rational in $(0,1)$ with denominator $q$ (or equivalently, $\ell = \mathcal{O}(3,q)$) and define:

$$\begin{aligned} \mathrm{MLO}(q) &:= \mathrm{round}\left(\frac{\Phi(q)\,\mathrm{m}(\ell,2)}{\mathrm{e}(\ell,3)}\right) \\ \mathrm{FP}(q) &:= \mathrm{round}\left(\left(\frac{2}{3}\right)^{\ell(q)} \cdot 2 \cdot \Phi(q)\right), \end{aligned}$$

where $\Phi$ is the Euler totient function.
These are the functions that we will compare to the number of Cantor rationals with denominator $q$, denoted by $\#N_q$, for $q$ ranging from 1 to $3^{10}$. We do comparisons only for those $q$ which are not multiples of 3, i.e. we only look at purely periodic Cantor rationals. For a heuristic justification for the choice of these functions, see paper [4].

*Remark* 6.1. If $q$ is such that its period length $\ell$ is prime and at least 3, then $\mathrm{MLO}(q)$ takes a special form. Indeed, in this case the divisors of $\ell$ are just 1 and $\ell$, which implies that $\mu(\ell) = -1$. Moreover, $\ell$ is odd. We get:

$$\begin{aligned} \mathrm{m}(\ell, a) &= \sum_{d|\ell} \mu\left(\frac{\ell}{d}\right) a^d \\ &= a^\ell - a \end{aligned}$$

and

$$
\begin{aligned}
\mathrm{e}(\ell, a) &= \sum_{\substack{d|\ell \\ \ell/d \text{ even}}} \mu\left(\frac{\ell}{d}\right) a^d + \sum_{\substack{d|\ell \\ \ell/d \text{ odd}}} \mu\left(\frac{\ell}{d}\right)\left(\frac{a^d}{2} + \frac{1-(-1)^a}{4}\right) \\
&= \frac{a^\ell}{2} + \frac{1-(-1)^a}{4} - \frac{a}{2} - \frac{1-(-1)^a}{4} \\
&= \frac{a^\ell - a}{2}.
\end{aligned}
$$

We thus deduce that $\mathrm{MLO}(q)$ takes the form:

$$
\mathrm{MLO}(q) = \mathrm{round}\left(2 \cdot \Phi(q) \cdot \frac{2^\ell - 2}{3^\ell - 3}\right).
$$

This implies that for $\ell$ prime and big enough, we get $\mathrm{MLO}(q) = \mathrm{FP}(q)$ due to rounding.

For general $q$, we have that

$$
\frac{\mathrm{MLO}(q)}{\mathrm{FP}(q)} = \frac{\mathrm{m}(\ell, 2)/\,\mathrm{e}(\ell, 3)}{2 \cdot (2/3)^\ell} \overset{\ell\to\infty}{\longrightarrow} 1.
$$

Indeed, by the computations done in Section 5, we get the following:

$$
\frac{\mathrm{m}(\ell, 2)}{\mathrm{e}(\ell, 3)} \cdot \left(\frac{3}{2}\right)^\ell = \frac{\sum_{d|\ell} \mu\left(\frac{\ell}{d}\right) 2^d}{\sum_{d|\ell} \mu\left(\frac{\ell}{d}\right) 3^d - \sum_{\substack{d|\ell \\ \ell/d \text{ odd}}} \mu\left(\frac{\ell}{d}\right) \frac{3^d - 1}{2}} \cdot \left(\frac{3}{2}\right)^\ell \sim \frac{2^\ell}{3^\ell - \frac{3^\ell}{2}} \cdot \left(\frac{3}{2}\right)^\ell \overset{\ell\to\infty}{\longrightarrow} 2.
$$

## 7 Results

In the following are some results of numerical computations to compare the values of the functions MLO and FP with the actual number of Cantor rationals with denominator $q$, where $3 \nmid q$.

Define for $c \in (0, 1)$ and for $T \in \mathbb{R}_{>0}$:

$$
\begin{aligned}
I_T &:= [(1-c)T, T] \\
N(T) &:= \sum_{\substack{q \in I_T \\ 3 \nmid q}} \# N_q \\
F(T) &:= \sum_{\substack{q \in I_T \\ 3 \nmid q}} \mathrm{FP}(q) \\
M(T) &:= \sum_{\substack{q \in I_T \\ 3 \nmid q}} \mathrm{MLO}(q).
\end{aligned}
$$

*Remark* 7.1. Note that the function $M(T)$ used here is different from $M(T)$ used in [4]. Indeed, in paper [4] one restricts the sum further to those $q$ such that $q$ is a divisor of $\frac{1}{2}\left(3^{\ell(q)} - 1\right)$, where $\ell(q)$ denotes the period length of $q$. This makes $M(T)$ significantly smaller.

We did the computation for $c = \frac{1}{2}$ and $c = \frac{2}{3}$, summing over the intervals $[\frac{1}{2} \cdot T, T]$ for $T \in \{57 \cdot 2^k : 1 \leqslant k \leqslant 10\}$ and $[\frac{1}{3} \cdot T, T]$ for $T \in \{3^k : 1 \leqslant k \leqslant 10\}$ respectively. In this way, the intervals are overlapping only at the points $T$. See table 1 as well as figures 1 and 2 for results with $c = \frac{1}{2}$ and table 2 as well as figures 3 and 4 for results with $c = \frac{2}{3}$.

From these computations, we can draw the following preliminary conclusions:

- The functions MLO and FP seem to give a decent approximation of the number of purely periodic Cantor rationals. Indeed, the ratios between the projected value and the actual value decrease to numbers between 0.8 and 1.5.

- We get better results when $c = \frac{1}{2}$ as when $c = \frac{2}{3}$. This comes from the fact that both approximation functions may sometimes switch the correct values, i.e. $\mathrm{MLO}(q_1)$ is sometimes a good approximation of $N_{q_2}$ and $\mathrm{MLO}(q_2)$ of $N_{q_1}$. If both $q_1$ and $q_2$ lie in the same interval, this mistake does not matter since we consider the sum over the interval. In the case $c = \frac{2}{3}$ this seems to happen for $q_1 = 3^n - 1$ and $q_2 = 3^n + 1$. Here $q_1$ and $q_2$ are not in the same interval, hence the discrepancies.

- The functions MLO and FP give very similar results. This is natural, considering that

$$\frac{\mathrm{m}(\ell, 2)/\,\mathrm{e}(\ell, 3)}{2 \cdot (2/3)^{\ell}} \xrightarrow{\ell \to \infty} 1,$$

as explained in Section 6, and hence

$$\frac{M(T)}{N(T)} \xrightarrow{T \to \infty} 1.$$

However, $M(T)$ seems to give slightly better results.

Table 1: Values for $c = \frac{1}{2}$

| $T$ | $N(T)$ | $F(T)$ | $M(T)$ | $F(T)/N(T)$ | $M(T)/N(T)$ |
|---|---|---|---|---|---|
| 57 | 8 | 20 | 20 | 2.5 | 2.5 |
| 114 | 28 | 43 | 41 | 1.536 | 1.464 |
| 228 | 70 | 73 | 71 | 1.043 | 1.014 |
| 456 | 86 | 96 | 93 | 1.116 | 1.081 |
| 912 | 170 | 213 | 204 | 1.253 | 1.2 |
| 1824 | 220 | 395 | 388 | 1.795 | 1.764 |
| 3648 | 488 | 432 | 422 | 0.8852 | 0.8648 |
| 7296 | 654 | 800 | 783 | 1.223 | 1.197 |
| 14592 | 1250 | 1344 | 1326 | 1.075 | 1.061 |
| 29184 | 1258 | 1530 | 1508 | 1.216 | 1.199 |
| 58368 | 1936 | 2519 | 2490 | 1.301 | 1.286 |

Table 2: Values for $c = \frac{2}{3}$

| $T$ | $N(T)$ | $F(T)$ | $M(T)$ | $F(T)/N(T)$ | $M(T)/N(T)$ |
|---|---|---|---|---|---|
| 3 | 2 | 2 | 3 | 1 | 1.5 |
| 9 | 2 | 9 | 9 | 4.5 | 4.5 |
| 27 | 10 | 30 | 26 | 3 | 2.6 |
| 81 | 14 | 39 | 38 | 2.786 | 2.714 |
| 243 | 98 | 128 | 124 | 1.306 | 1.265 |
| 729 | 122 | 207 | 197 | 1.697 | 1.615 |
| 2187 | 378 | 632 | 622 | 1.672 | 1.646 |
| 6561 | 758 | 926 | 903 | 1.222 | 1.191 |
| 19683 | 2046 | 2516 | 2479 | 1.23 | 1.212 |
| 59049 | 2758 | 3936 | 3879 | 1.427 | 1.406 |

Figure 1: Values of $N(T)$, $M(T)$ and $F(T)$ for $c = \frac{1}{2}$

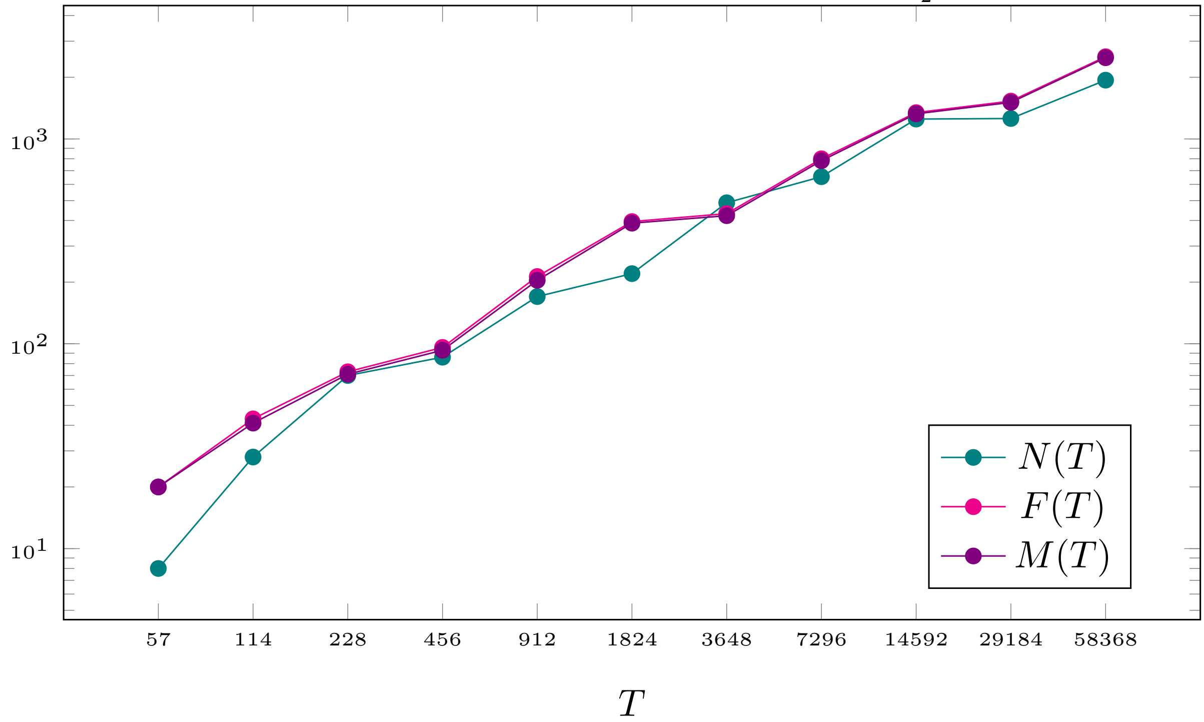



Figure 2: Ratios $M(T)/N(T)$ and $F(T)/N(T)$ for $c = \frac{1}{2}$

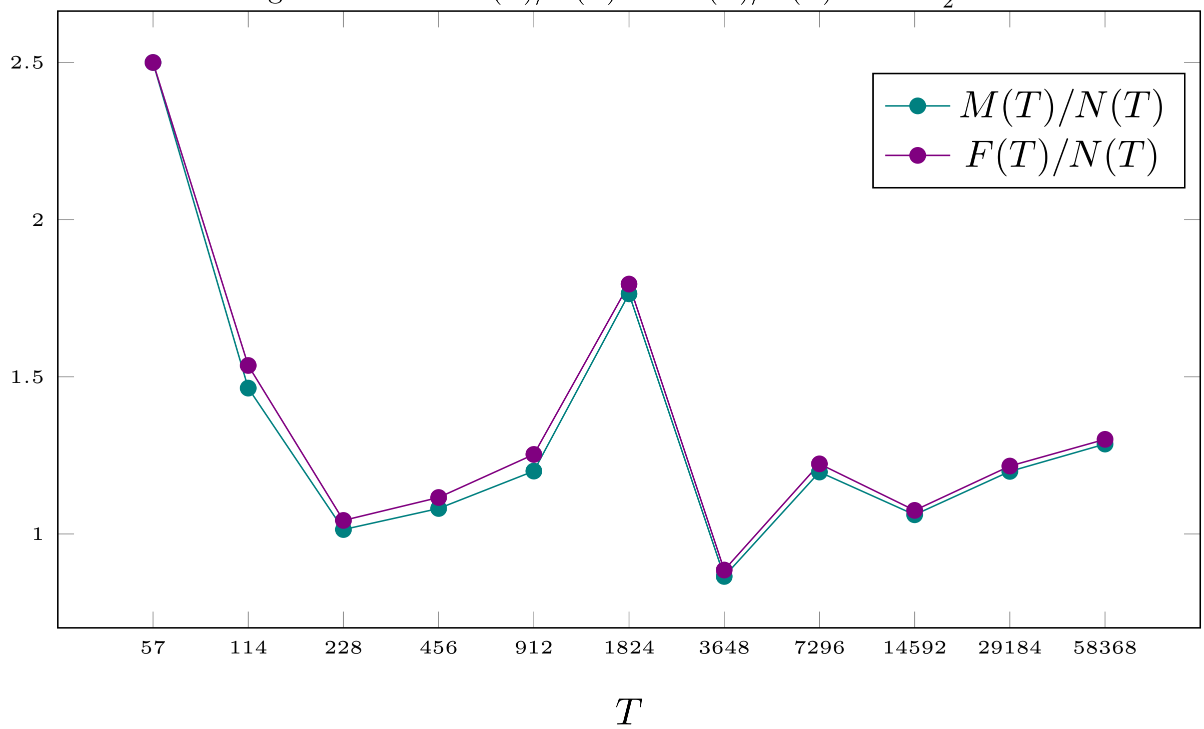


Figure 3: Values of $N(T)$, $M(T)$ and $F(T)$ for $c = \frac{2}{3}$

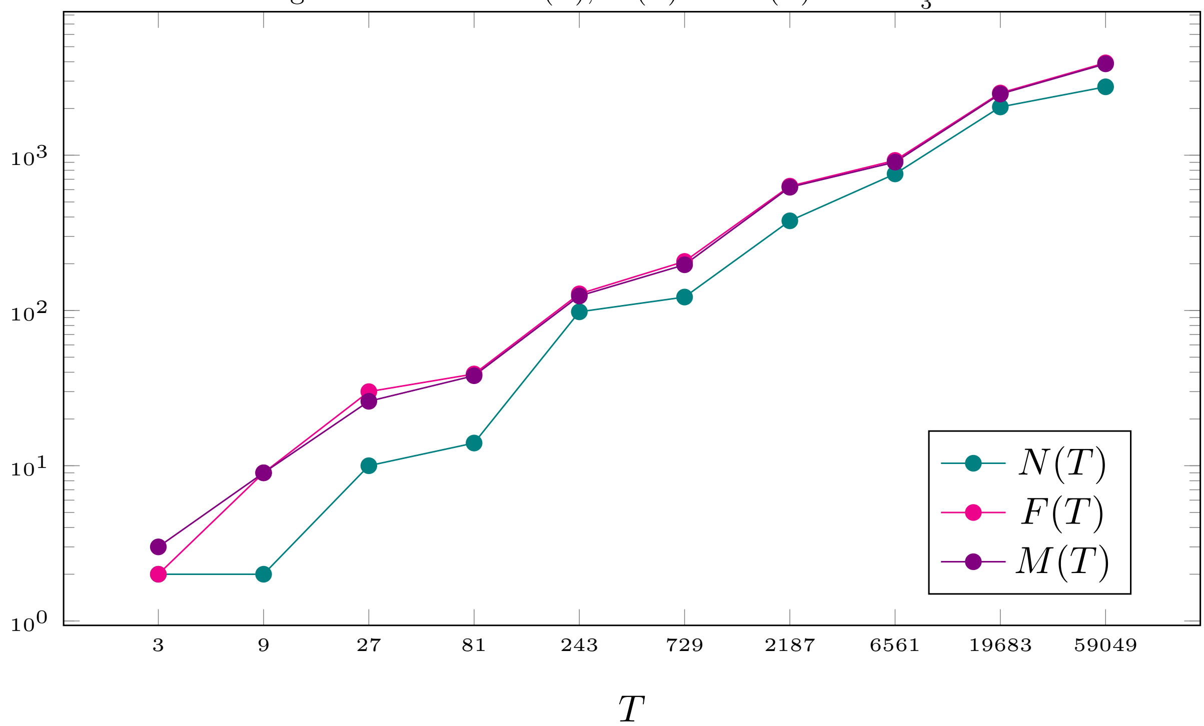
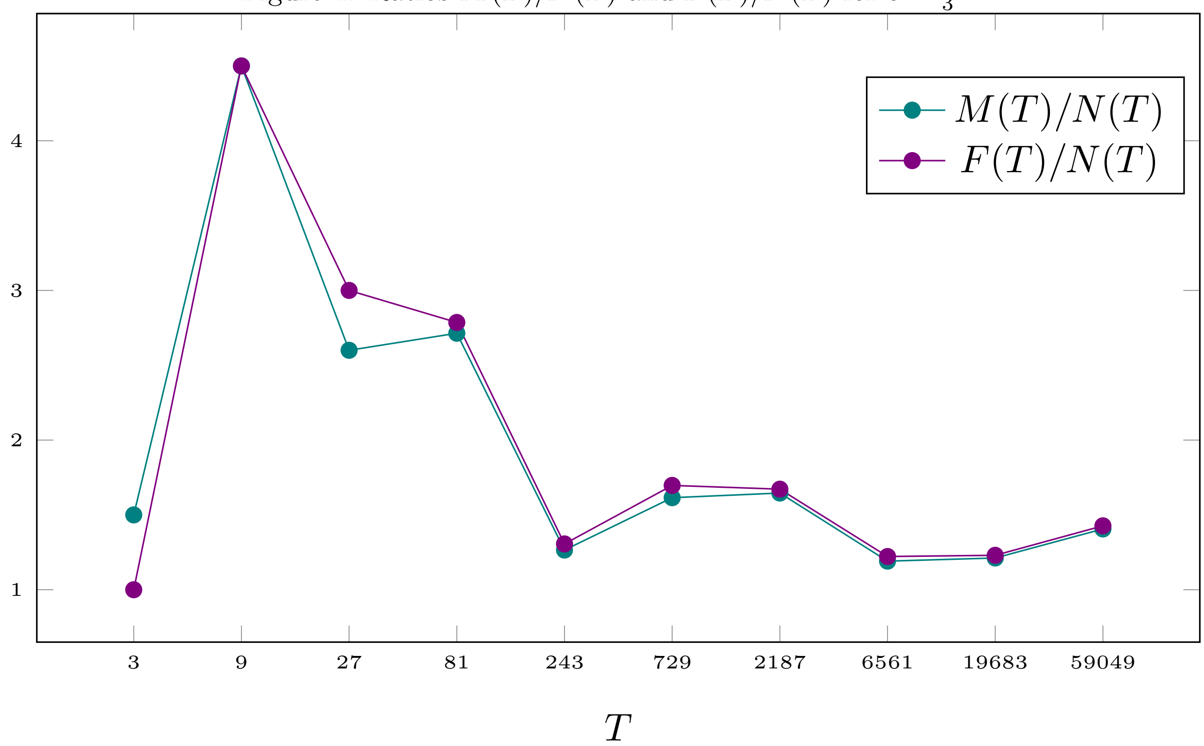



Figure 4: Ratios $M(T)/N(T)$ and $F(T)/N(T)$ for $c = \frac{2}{3}$



## 8 Code for the computation

The following Python code was used to carry out the computations. The algorithm for modular exponentiation is a well-known algorithm taken from [2]. We also use the `labmath` package, see [1].

```
"""
@author: Tara Trauthwein
"""

import math
from decimal import Decimal
import labmath as lm

# returns the p-valuation of a number n, where p is a prime
def val(p,n):
    count=0
    while(n%p==0):
        count+=1
        n/=p
    return count

# returns the number of primitive words
# p: number of different letters
# n: length of a word
# l: list of divisors of n
def m(p,n,l):
    return sum(lm.mobius(int(n/d)) * Decimal(p)**Decimal(d) for d
        in l)

# returns the number of primitive even words over the alphabet
   {0,1,2} following the first formula
# n: length of a word
# l: list of divisors of n
def e(n,l):
    psum=0
    for d in l:
        val2=val(2,d)
        internal= Decimal('3')**Decimal(d)
        internal-=sum(Decimal('3')**(Decimal(d)/(Decimal('2')**
           Decimal(k))) \
                      for k in range(1,val2+1))
        internal+=val2+1
        psum+= lm.mobius(int(n/d)) * Decimal(0.5) * internal
    return psum

# returns the number of primitive even words over the alphabet
   {0,1,2} following the second formula
# n: length of a word
# l: list of divisors of n
def e_2(n,l):
```

```python
    s=0
    for d in l:
        if (int(n/d)%2==0):
            a=Decimal(3)**Decimal(d)
        else:
            a=(Decimal(3)**Decimal(d)+1)*Decimal('0.5')
        s+=lm.mobius(int(n/d))*a
    return s

# modular exponentiation
# returns (base**expo) mod modu
def mod_exp(base,expo,modu):
    if (modu==1):
        return 0
    exp=bin(expo)[2:]
    res=1
    base %= modu
    for i in exp[::-1]:
        if (i=='1'):
            res *= base
            res %= modu
        base *= base
    return res

# returns the period length of a rational p/q
# or equivalently, the order of 3 in Z/qZ
# q: denominator
# toti: Euler totient of q
# uses the fact that the order divides toti
# calculates (3**d) mod q for d dividing toti
def length(q,toti):
    list_div=list(lm.divisors(toti))
    list_div.sort()

    for i in range(1,len(list_div)-1):
        p=list_div[i]
        if (p<math.log(q,3)):
            continue
        if (mod_exp(3,p,q)==1):
            return p
    return toti

# returns MLO(q) if 3 does not divide q
# return -1 if 3 divides q
def MLO(q):
    if (q==1):
        return 2
    if (q%3==0):
        return -1
    toti=lm.totient(q,1)
    l=length(q,toti)
```

```python
    listl=list(lm.divisors(l))
    res=Decimal(toti)/Decimal(e(l,listl))*Decimal(m(2,l,listl))
    return round(res)

# returns FP(q)
def FP(q):
    toti=lm.totient(q,1)
    order=length(q,toti)
    return (Decimal('2')/Decimal('3'))**order*toti*2
```

## References

[1] Lucas Brown. *labmath*. Version 1.1.0. URL: https://pypi.org/project/labmath/.

[2] Cormen, Leiserson, Rivest, and Stein. *Algorithmique*. 3rd ed. Paris: Dunod, 2010. ISBN: 978-2-10-054526-1.

[3] Juhani Karhumäki. *Combinatorics of Words. Lecture notes.* University of Turku. URL: https://www.utu.fi/en/units/sci/units/math/staff/Documents/karhumaki/combwo.pdf.

[4] Alexander Rahm, Noam Solomon, Tara Trauthwein, and Barak Weiss. *The distribution of rational numbers on Cantor's middle thirds set.* 2019. arXiv: 1909.01198 [math.NT].

[5] Zvezdelina Stankova-Frenkel. *Möbius Inversion Formula. Multiplicative Functions. Berkley Math Circle 1998-99.* UC Berkeley, 1998-1999. URL: https://math.berkeley.edu/~stankova/MathCircle/Multiplicative.pdf.