

JOURNAL

de Théorie des Nombres
de BORDEAUX

anciennement Séminaire de Théorie des Nombres de Bordeaux

Jason P. BELL

The upper density of an automatic set is rational

Tome 32, n° 2 (2020), p. 585-604.

<http://jtnb.centre-mersenne.org/item?id=JTNB_2020__32_2_585_0>

© Société Arithmétique de Bordeaux, 2020, tous droits réservés.

L'accès aux articles de la revue « Journal de Théorie des Nombres de Bordeaux » (<http://jtnb.centre-mersenne.org/>), implique l'accord avec les conditions générales d'utilisation (<http://jtnb.centre-mersenne.org/legal/>). Toute reproduction en tout ou partie de cet article sous quelque forme que ce soit pour tout usage autre que l'utilisation à fin strictement personnelle du copiste est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

cedram

Article mis en ligne dans le cadre du
Centre de diffusion des revues académiques de mathématiques
<http://www.centre-mersenne.org/>

The upper density of an automatic set is rational

par JASON P. BELL

RÉSUMÉ. Soient $k \geq 2$ un nombre naturel et S un ensemble k -reconnaissable de nombres naturels. On montre que la densité inférieure et la densité supérieure de S sont des nombres rationnels qui sont récursivement calculables, et on donne un algorithme pour les calculer. En outre, on montre que, pour $k \geq 2$ et $(\alpha, \beta) \in \mathbb{Q}^2$ tels que $0 < \alpha < \beta < 1$ ou $(\alpha, \beta) \in \{(0, 0), (1, 1)\}$, il existe un sous-ensemble k -reconnaissable de \mathbb{N} tel que la densité inférieure et la densité supérieure sont respectivement égales à α et β . De plus, on montre que ces valeurs sont précisément celles qui peuvent apparaître comme densités inférieure et supérieure d'un ensemble reconnaissable.

ABSTRACT. Given a natural number $k \geq 2$ and a k -automatic set S of natural numbers, we show that the lower density and upper density of S are recursively computable rational numbers and we provide an algorithm for computing these quantities. In addition, we show that for every natural number $k \geq 2$ and every pair of rational numbers (α, β) with $0 < \alpha < \beta < 1$ or with $(\alpha, \beta) \in \{(0, 0), (1, 1)\}$ there is a k -automatic subset of the natural numbers whose lower density and upper density are α and β respectively, and we show that these are precisely the values that can occur as the lower and upper densities of an automatic set.

1. Introduction

Given a subset S of the natural numbers, a natural question to ask is: *What proportion of natural numbers lie in S ?* To answer this, one lets $\pi_S(x)$ denote the number of elements of S that are less than x and one then studies how $\pi_S(x)/x$ behaves as x tends to infinity. In general, $\lim_{x \rightarrow \infty} \pi_S(x)/x$ need not exist, but when it does, we call the limit the *density* of the set of S . To deal with the fact that densities of sets of natural numbers need not exist, one can instead consider the *upper* and *lower densities* of the set S , given respectively by $\limsup_{x \rightarrow \infty} \pi_S(x)/x$ and $\liminf_{x \rightarrow \infty} \pi_S(x)/x$, which together provide a rough answer to the motivating question asked above.

The lower density and upper density of a set of natural numbers are real numbers α and β in $[0, 1]$ with $\alpha \leq \beta$, and given (α, β) satisfying these conditions there exists a set whose lower and upper densities are α and

Manuscrit reçu le 18 février 2020, révisé le 8 juillet 2020, accepté le 28 juillet 2020.

2020 *Mathematics Subject Classification*. 11B85, 68Q45.

Mots-clefs. upper density, automatic sets, Cobham's theorem, formal languages.

The author was supported by NSERC Grant RGPIN-2016-03632.

β , respectively. When one restricts one’s attention to so-called automatic sets (that is, subsets of the natural numbers whose elements are precisely those whose base- k expansions are accepted by a finite-state machine for some $k \geq 2$) the questions of density become significantly more constrained. For example, when S is an automatic set whose density exists, a result of Cobham [4] shows that the density is necessarily rational. We give further background on automatic sets and automata in Section 2.

In general, the density of an automatic set of natural numbers need not exist. As an example, let S denote the set of numbers whose base- k expansion has even length. If we let $\pi_S(x)$ denote the number of elements of S that are less than x , then

$$\pi_S(k^{2n}) = 1 + \sum_{j=1}^n (k^{2j} - k^{2j-1}) \sim \left(\frac{k}{k+1}\right) \cdot k^{2n},$$

while

$$\pi_S(k^{2n+1}) = 1 + \sum_{j=1}^n (k^{2j} - k^{2j-1}) \sim \left(\frac{1}{k+1}\right) \cdot k^{2n+1},$$

and so the lower density of S is at most $1/(k+1)$ and the upper density is at least $k/(k+1)$. (In Section 4 we give a more general construction that shows that $1/(k+1)$ and $k/(k+1)$ are respectively the lower and upper densities of S .) One can nevertheless ask what one can say about the upper and lower densities of an automatic set. In this paper we answer this question completely, showing the lower and upper densities are recursively computable rational numbers in $[0, 1]$ and, moreover, we characterize exactly which pairs (α, β) can be realized as the lower density and upper density of an automatic set.

The first part of this characterization is given by the following more general result concerning automatic sequences.

Theorem 1.1. *Let $k \geq 2$ be a natural number and let $h : \mathbb{N} \rightarrow \Delta \subseteq \mathbb{Q}_{\geq 0}$ be a k -automatic sequence and let $s(n) = \sum_{j < n} h(j)$. Then*

$$\limsup_{n \rightarrow \infty} s(n)/n$$

is a recursively computable rational number.

We make some remarks about what is meant by “recursively computable” in the statement of Theorem 1.1. Given a k -automatic sequence h , one can build a deterministic finite automaton with output Γ , which takes the base k -expansion of n as input, reading left to right, and gives $h(n)$ as output. We give an algorithm that allows one to determine the limsup of $s(n)/n$ from the machine Γ (see Section 4 for further details).

As an immediate corollary of Theorem 1.1, we obtain the following result.

Corollary 1.2. *Let $k \geq 2$ be a natural number and let $S \subseteq \mathbb{N}$ be a k -automatic set. Then the upper and lower densities of S are recursively computable rational numbers.*

We point out that Corollary 1.2 can be seen as an extension of the result of Cobham [4] mentioned earlier about the density of an automatic set of natural numbers, when it exists. Since the density exists if and only if the upper and lower densities coincide, this follows from Corollary 1.2. As the example given in which S is the set of numbers whose base k -expansion has even length illustrates, our result is strictly stronger than Cobham's result.

In light of Corollary 1.2, it is natural to ask which possible pairs $(\alpha, \beta) \in \mathbb{Q}^2$ can occur as the lower and upper densities of an automatic set. We are able to completely characterize which pairs can occur.

Theorem 1.3. *Let $k \geq 2$ be a positive integer and let (α, β) be a pair of rational numbers satisfying either $0 < \alpha \leq \beta < 1$ or $(\alpha, \beta) \in \{(0, 0), (1, 1)\}$. Then there is a k -automatic set S whose lower density and upper density are α and β respectively. Conversely, if S is a k -automatic set whose lower density is α and whose upper density is β then either $(\alpha, \beta) \in \{(0, 0), (1, 1)\}$ or α, β are rational numbers with $0 < \alpha \leq \beta < 1$.*

The outline of this paper is as follows. In Section 2, we give some of the basic background on automatic sequences and sets and related notation we will make use of. In Section 3, we give an overview of the strategy used to proof Theorem 1.1 and then prove Theorem 1.1. In Section 4 we provide an algorithm that allows one to compute the upper and lower density of an automatic set. In Section 5 we give examples that are used to demonstrate Theorem 1.3, and finally in Section 6 we give some concluding remarks and raise a question concerning possible extensions of Theorem 1.1 to morphic sequences.

Acknowledgments. I thank Jeffrey Shallit for bringing this problem to my attention and for making many useful comments and for bringing my attention to the work of Kemp. I also thank the referee for reading the paper carefully and giving numerous helpful comments and suggestions.

2. Background on automata and automatic sets

In this section we give the necessary background on finite-state automata and k -automatic sequences and sets.

Let Σ be a nonempty finite set and let Σ^* denote the free monoid on Σ . A *deterministic finite automaton with output* (DFAO) is a 6-tuple

$$\Gamma = (Q, \Sigma, \delta, q_0, \Delta, \tau),$$

where Q is a finite set of states, Σ is a finite input alphabet, δ is the transition function from $\Sigma \times Q$ to Q , $q_0 \in Q$ is the initial state, Δ is an

output alphabet, and τ is the output function from Q to Δ . Less formally, a DFAO is simply a directed graph in which the vertices are the elements of Q ; for each vertex $q \in Q$ and each $x \in \Sigma$ we have a directed arrow with label x from q to the state $\delta(x, q)$. Given a word $w \in \Sigma^*$, the DFAO gives us an output in Δ as follows: we begin at the initial state q_0 and then, reading w from left to right, we obtain a path in this directed graph by moving vertex to vertex as we read the letters of w . After reading w we end up at some state $q \in Q$ and we then apply τ to obtain an output in Δ . Thus we can associate a map $f : \Sigma^* \rightarrow \Delta$ with a DFAO Γ .

We give an example of a DFAO in Figure 1 that generates the map f from $\{0, 1\}^*$ to $\{0, 1\}$ and f is 1 precisely when the string w is either of the form 0^a or of the form $0^a 1u$ where u has odd length. In particular, f induces a well-defined map $h : \mathbb{N} \rightarrow \{0, 1\}$ given by taking the binary expansion of n and applying f ; then $h(n)$ is 1 precisely when the number of digits in the binary expansion of n is even and is 0 otherwise, where we take the binary expansion of 0 to be the empty word.

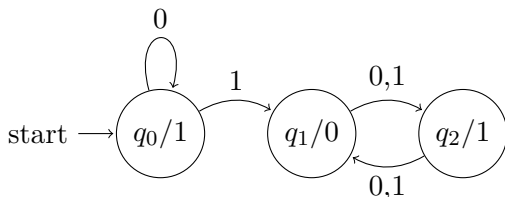


FIGURE 2.1. The DFAO generating the sequence $h(n)$.

Let $k \geq 2$ be a natural number and let Σ_k be the alphabet $\{0, 1, \dots, k-1\}$. For every natural number n , there is a word $w = (n)_k \in \Sigma_k^*$, which is the base- k expansion of n , where we define $(0)_k$ to be the empty word; conversely, given a non-empty word $w \in \Sigma_k^*$ with no leading zeros there is a natural number $n = [w]_k$, which is the natural number whose base- k expansion is w . In the case when w is the empty word, we take $[w]_k = 0$. A sequence $a : \mathbb{N} \rightarrow \Delta$ is called k -automatic if there exists a DFAO $\Gamma = (Q, \Sigma_k, \delta, q_0, \Delta, \tau)$ such that for each $n \in \mathbb{N}$, $a(n)$ can be computed from Γ by feeding the word $(n)_k$ into Γ , reading the digits from left to right. We then say that a subset $S \subseteq \mathbb{N}$ is a k -automatic set if the characteristic function of S , $\chi_S : \mathbb{N} \rightarrow \{0, 1\}$ defines a k -automatic sequence. It is worth noting that some authors define k -automatic sequences using the convention that one reads the base- k expansion of a number n from right to left. As it turns out, this definition is equivalent to the one we give (cf. [1, Corollary 4.3.5]).

An important property of an automatic sequence $f : \mathbb{N} \rightarrow \Delta$, when regarded as a map from Σ_k^* to Δ with the property that $f(0w) = f(w)$ for

every word $w \in \Sigma_k^*$, is that there is a finite set of maps $f = f_1, \dots, f_d : \Sigma_k^* \rightarrow \Delta$ with the property that for each $i \in \Sigma_k^*$ and each $j \in \{1, 2, \dots, d\}$ there is some $\ell \in \{1, 2, \dots, d\}$ such that $f_j(iw) = f_\ell(w)$ for every word $w \in \Sigma_k^*$, where iw is the concatenation of i and w . We call these maps the (left) k -kernel of f .

We make use of the following fact about automatic sequences, which we suspect is well-known, although we are unaware of a proof in the literature.

Proposition 2.1. *Let $k \geq 2$ be a natural number, let $h : \mathbb{N} \rightarrow \mathbb{Q}_{\geq 0}$ be a k -automatic sequence, and let $s(n) = \sum_{j < n} h(j)$. Then there exist $\beta \in (0, k)$, $C > 0$, $a \geq 1$, and rational numbers c_j for $j \in \{0, 1, \dots, a - 1\}$ such that*

$$|s(k^{an+j}) - c_j k^{an+j}| < C\beta^{an}$$

for every $n \geq 0$. Moreover, a and the rational numbers c_0, \dots, c_{a-1} are recursively computable and β can be effectively determined.

Proof. We have an automatic map f from Σ_k^* to \mathbb{Q} with $f(0w) = f(w)$ and $f((n)_k) = h(n)$, and we let $f = f_1, \dots, f_s$ denote the left k -kernel of f . Then there are $s \times s$ matrices A_0, \dots, A_{k-1} with entries in $\{0, 1\}$ such that each A_j has exactly one 1 in each row and such that for each $w \in \Sigma_k^*$ and each $i \in \{0, \dots, k - 1\}$ we have

$$[f_1(iw), \dots, f_s(iw)]^T = A_i[f_1(w), \dots, f_s(w)]^T.$$

Let $B = A_0 + \dots + A_{k-1}$ and let $v_0 = [f_1(\epsilon), \dots, f_s(\epsilon)]^T$. Then

$$\begin{aligned} e_1^T B^n v_0 &= \sum_{0 \leq i_0, \dots, i_{n-1} < k} e_1^T A_{i_{n-1}} \dots A_{i_0} v_0 \\ &= \sum_{\{w: |w|=n\}} f(w) \\ &= s(k^n). \end{aligned}$$

Now B is a sum of k matrices, each of which have exactly one 1 in each row, and the entries of B are nonnegative. In particular, the row sums of $k^{-1} \cdot B$ are all 1 and the entries are nonnegative. By a result of Karpelevič [6] (see also Higham and Lin [5]) we have:

- (i) k is an eigenvalue of B ;
- (ii) each eigenvalue of B of modulus k is of the form $k\omega$ with ω a root of unity;
- (iii) each eigenvalue of B has modulus at most k .

Then

$$s(k^n) = e_1^T B^n v_0 = \sum_{i=1}^m p_i(n) \gamma_i^n,$$

where $\gamma_1, \dots, \gamma_m \in \overline{\mathbb{Q}}$ are the eigenvalues of B and $p_1, \dots, p_m \in \overline{\mathbb{Q}}[x]$ are polynomials of degree at most one less than the largest Jordan block with

eigenvalue γ_i occurring in the Jordan form of B . We may assume that $\gamma_1, \dots, \gamma_t$ are the eigenvalues of modulus k (and hence $\gamma_i = k\omega_i$ with ω_i a root of unity for $i = 1, \dots, t$) and we may pick $\beta \in (0, k)$ such that $\gamma_{t+1}, \dots, \gamma_m$ have modulus strictly less than β . Now we pick $a > 0$ so that $\omega_1^a = \dots = \omega_t^a = 1$. For $j = 0, \dots, a - 1$ we have

$$s(k^{an+j}) = q_j(n)k^{an+j} + \sum_{i=t+1}^m p_{i,j}(n)\gamma_i^{an+j}$$

for some polynomials $q_j, p_{i,j}$ in $\overline{\mathbb{Q}}[x]$. A priori we only know that $q_j(x)$ is a polynomial with coefficients in $\overline{\mathbb{Q}}$, but we shall show that it is in fact a constant polynomial, where the constant is rational.

Let K be the Galois closure of the number field generated by the coefficients of p_1, \dots, p_m and $\gamma_1, \dots, \gamma_m$, and let G be the Galois group of K . If σ is in G , then σ permutes $\gamma_1, \dots, \gamma_m$, but since σ fixes k and takes roots of unity to other roots of unity, we see that σ permutes $\gamma_1, \dots, \gamma_t$; consequently, σ permutes $\gamma_{t+1}, \dots, \gamma_m$. From these facts, we first get

$$\sum_{\sigma \in G} \sigma(s(k^{an+j})) = |G|s(k^{an+j}),$$

as $s(k^{an+j})$ is rational. On the other hand, this is equal to

$$\left(\sum_{\sigma \in G} \sigma(q_j(n)) \right) k^{an+j} + \sum_{\sigma \in G} \sum_{i=t+1}^m \sigma(p_i(n))\sigma(\gamma_i)^n.$$

Let $h_j(n) := \sum_{\sigma \in G} \sigma(q_j(n))$. Then by construction $h_j(n)$ is a polynomial in n and by construction it is fixed by every element of G and hence $h_j(n)$ is a rational number for every n ; moreover, since σ permutes $\gamma_{t+1}, \dots, \gamma_m$, they each have modulus strictly less than β .

Thus

$$s(k^{an+j}) = |G|^{-1}h_j(n)k^{an+j} + O(\beta^{an+j}).$$

Finally observe that $|s(k^{an+j})| \leq C_0k^{an+j}$ where C_0 is the maximum of the absolute values of the range of f . It follows that $h_j(n)$ is a rational constant for $j = 0, \dots, a - 1$, and we let $c_j \in \mathbb{Q}$ denote the value $|G|^{-1}h_j(0)$. We have thus proved the rationality of the constants c_j .

We now make remarks concerning the computability of a and c_0, \dots, c_{a-1} . Observe that the $d \times d$ matrix B can be computed from the DFAO giving f , and if ω is a b -th root of unity such that $k\omega$ is a root of the characteristic polynomial of B , then the field extension $[\mathbb{Q}(\omega) : \mathbb{Q}]$ has degree at most d . It follows that $\phi(b) \leq d$, where $\phi(n)$ is Euler's ϕ -function. If $\phi(b) \leq d$ then each prime power factor p^r of b must satisfy $p^{r-1}(p-1) \leq d$, and so b divides $\prod_{p \leq d} p^{\lfloor \log_p(d) \rfloor}$. Then we can take a to be the quantity $\prod_{p \leq d} p^{\lfloor \log_p(d) \rfloor}$.

We next explain how to compute the rational numbers c_0, \dots, c_{a-1} . By our choice of a , the matrix B^a has an eigenvalue k^a and the remaining

eigenvalues are strictly less than k^a in modulus. Moreover, we have shown that the Jordan blocks in B^a associated to the eigenvalue k^a are all of size one and so the minimal polynomial of B^a has k^a as a simple root. We can compute the characteristic polynomial of B^a and we can find the factorization of this polynomial in $\mathbb{Q}[x]$ into irreducible factors (see [8] for details) and hence we can find all of the finitely many monic rational polynomial factors of the characteristic polynomial of B^a . Then by applying each of these factors to B^a , we can compute the minimal polynomial of B^a .

As mentioned before, the minimal polynomial has k^a as a simple root and so if $q(x)$ denotes the minimal polynomial of B^a then we have $q(x) = (x - k^a)q_0(x)$, where all roots of $q_0(x)$ are strictly less than k^a in modulus. Since k^a is not a root of $q_0(x)$, we can then use the division algorithm for polynomials to compute a rational polynomial $t(x)$ and a nonzero rational constant λ such that $q_0(x) = t(x)(x - k^a) + \lambda$. For $j \in \{0, 1, \dots, a - 1\}$ we have

$$s(k^{an+j}) = e_1^T B^{an+j} v_0.$$

To compute c_j , we let $v'_j := \lambda^{-1}q_0(B)B^j v_0$ and we let $v''_j := -\lambda^{-1}t(B)(B - k^a I)B^j v_0$, both of which are computable. By construction, $(B - k^a I)v'_j = 0$ since $(x - k^a)q_0(x)$ is the minimal polynomial of B . Similarly, $q_0(B)v''_j = 0$; moreover, $v'_j + v''_j = B^j v_0$ by construction. Then v''_j lies in the sum of the eigenspaces corresponding to the roots of $q_0(x)$ and since these are all strictly less than k^a in modulus, we see that

$$e_1^T B^{an} v''_j = o(k^{an})$$

as $n \rightarrow \infty$. On the other hand, $B^a v'_j = k^a v_j$ and so

$$e_1^T B^{an} v'_j = k^{an} e_1^T v'_j.$$

Then we can compute the rational number $c_j := e_1^T v'_j$ and by construction

$$e_1^T B^{an+j} v_0 = e_1^T B^{an} (v'_j + v''_j) = c_j k^{an} + o(k^{an})$$

and so we have shown that the c_j are recursively computable. Finally, since we can numerically calculate the eigenvalues of B to arbitrary precision, we can determine some $\beta \in (0, k)$ that is strictly larger than the eigenvalues that are strictly less than k in modulus. □

3. Proof of rationality

In this section we prove Theorem 1.1. To do so, we need a technical lemma, which we now prove.

Lemma 3.1. *Let $k \geq 2$ be a natural number, let γ be a positive real number, let s_n, s'_n be sequences of positive numbers, let u', v', u, v be positive real*

numbers, and let b and c be positive integers. If

$$\frac{(v'k^{b+c} + u'k^c + s'_n)}{(vk^{b+c} + uk^c + s_n)} \rightarrow \gamma$$

as $n \rightarrow \infty$ and

$$\limsup_{n \rightarrow \infty} \frac{(v'k^c + s'_n)}{(vk^c + s_n)} \leq \gamma,$$

then

$$\limsup_{n \rightarrow \infty} \frac{(v'k^{2b+c} + u'(k^{b+c} + k^c) + s'_n)}{(vk^{b+2c} + u(k^{b+c} + k^c) + s_n)} \geq \gamma.$$

Proof. Let

$$\begin{aligned} X_n &= v'k^c + s'_n, & X'_n &= v'k^{b+c} + u'k^c + s'_n, \\ & & \text{and } X''_n &= v'k^{2b+c} + u'(k^{b+c} + k^c) + s'_n. \end{aligned}$$

Similarly, we let

$$\begin{aligned} Y_n &= vk^c + s_n, & Y'_n &= vk^{b+c} + uk^c + s_n, \\ & & \text{and } Y''_n &= vk^{2b+c} + u(k^{b+c} + k^c) + s_n. \end{aligned}$$

Then $X''_n = (k^b + 1)X'_n - k^bX_n$ and $Y''_n = (k^b + 1)Y'_n - k^bY_n$ and our goal is to show that $\limsup_{n \rightarrow \infty} X''_n/Y''_n \geq \gamma$.

Now suppose that the conclusion to the statement of the lemma does not hold. Then there exists some $\epsilon > 0$ with $\gamma - \epsilon > 0$ such that

$$X''_n/Y''_n < \gamma - \epsilon$$

for all n sufficiently large. In other words, for n large we have

$$(k^b + 1)X'_n - k^bX_n < (\gamma - \epsilon)((k^b + 1)Y'_n - k^bY_n).$$

We may rewrite this inequality as

$$(3.1) \quad (k^b + 1)X'_n + (\gamma - \epsilon)k^bY_n < (\gamma - \epsilon)(k^b + 1)Y'_n + k^bX_n,$$

for n sufficiently large. By assumption, $X'_n/Y'_n \rightarrow \gamma$ for n large. Thus

$$(3.2) \quad X'_n \geq (\gamma - \epsilon/(3k^b))Y'_n$$

for all n sufficiently large. Similarly, since $\limsup X_n/Y_n \leq \gamma$, we have

$$(3.3) \quad X_n \leq (\gamma + \epsilon/(3k^b))Y_n$$

for n sufficiently large. Then using Equations (3.2) and (3.3) along with Equation (3.1), we have

$$\begin{aligned} (k^b + 1)(\gamma - \epsilon/(3k^b))Y'_n + k^b(\gamma - \epsilon)Y_n & \\ & \leq (k^b + 1)X'_n + k^b(\gamma - \epsilon)Y_n \\ & < (\gamma - \epsilon)(k^b + 1)Y'_n + k^b X_n \\ & < (\gamma - \epsilon)(k^b + 1)Y'_n + k^b(\gamma + \epsilon/(3k^b))Y_n. \end{aligned}$$

for n sufficiently large. Then by rearranging the inequality

$$\begin{aligned} (k^b + 1)(\gamma - \epsilon/(3k^b))Y'_n + k^b(\gamma - \epsilon)Y_n & \\ & < (\gamma - \epsilon)(k^b + 1)Y'_n + k^b(\gamma + \epsilon/(3k^b))Y_n, \end{aligned}$$

we see

$$(k^b + 1)\epsilon(1 - 1/(3k^b))Y'_n < k^b Y_n \epsilon(1 + 1/(3k^b))$$

for all n sufficiently large. Since $\epsilon > 0$, we then must have

$$(3.4) \quad (k^b + 1)(1 - 1/(3k^b))Y'_n < k^b(1 + 1/(3k^b))Y_n$$

for all sufficiently large n .

To complete the proof, observe that since b is a positive integer,

$$(k^b + 1)(1 - 1/(3k^b)) = k^b + 2/3 - 1/(3k^b) \geq k^b + 1/3 = k^b(1 + 1/(3k^b)),$$

and so using Equation (3.4), we see

$$k^b(1 + 1/(3k^b))Y'_n \leq (k^b + 1)Y'_n(1 - 1/(3k^b)) < k^b(1 + 1/(3k^b))Y_n,$$

or equivalently that $Y'_n \leq Y_n$ for all sufficiently large n . But it is immediate that $Y'_n \geq Y_n + uk^c > Y_n$ for all n , and so we get a contradiction. The result follows. \square

This result is key to obtaining the proof of Theorem 1.1. We will find it useful to make use of certain assumptions and to fix notation for the remainder of the proof.

Notation 3.2. Throughout the remainder of the paper we make the following assumptions and notation.

- (1) We let $\Sigma_k = \{0, 1, \dots, k - 1\}$ and for $w \in \Sigma_k^*$, we let $|w|$ denote the length of w and we let \prec denote the lexicographic order on Σ_k^* where $0 < 1 < \dots < k - 1$.
- (2) We let $h : \mathbb{N} \rightarrow \mathbb{Q}$ be k -automatic and let $f : \Sigma_k^* \rightarrow \mathbb{Q}$ be a map associated to a DFAO with the property that $f(0w) = f(w)$ and $f(w) = h([w]_k)$.
- (3) We let $f = f_1, \dots, f_d : \Sigma_k^* \rightarrow \mathbb{Q}$ denote the maps in the left k -kernel of f .

- (4) For each $i \in \{1, \dots, d\}$ and each $u \in \Sigma_k^*$, we let $\delta(i, u) \in \{1, \dots, d\}$ denote the number j with the property that $f_i(uw) = f_j(w)$ for every $w \in \Sigma_k^*$.
- (5) For $w \in \Sigma_k^*$ we let $s_i(w) = \sum_{\{v \prec w, |v|=|w|\}} f_i(v)$ for $i = 1, \dots, d$ and we let $s(n) = s_1((n)_k)$.
- (6) We let γ denote $\limsup_{n \rightarrow \infty} s(n)/n$.
- (7) Appealing to Proposition 2.1, we have $a \geq 1$, rational numbers $c_{i,j}$, $1 \leq i \leq d$, $0 \leq j < a$, and $\beta \in (0, k)$ be such that

$$s_i((k^{an+j})_k) = c_{i,j}k^{an+j} + O(\beta^{an}).$$

- (8) For each word $u \in \Sigma_k^*$ with no leading zeros, whose length is a multiple of a , and each $j \in \{0, \dots, a - 1\}$, we let

$$\gamma_j(u) := \limsup_{\{w: a \mid (|w|-j)\}} s([uw]_k)/[uw]_k.$$

Using the assumption and notation above, our goal in Lemma 5.1 is to show that one can decide when γ is zero, so for the remainder of this section we assume that γ is strictly positive. We pick words $w_1, w_2, \dots \in \Sigma_k^*$ with the property

$$s([w_m]_k)/[w_m]_k \rightarrow \gamma.$$

There is some $j \in \{0, 1, \dots, a - 1\}$ such that $|w_m| \equiv j \pmod{a}$ for infinitely many m , and by refining our collection if necessary, we may assume that $|w_m| \equiv j \pmod{a}$ for every m . In other words $\gamma = \gamma_j(\epsilon)$, where ϵ is the empty word. Observe that for each $\ell \geq 0$ there is at least one word of length ℓa that is a prefix of infinitely many w_m , and for such words we have $\gamma_j(u) = \gamma$.

Notation 3.3. In addition to the items in Notation 3.2, we make the following assumptions and fix additional notation.

- (9) We assume $j \in \{0, 1, \dots, a - 1\}$ and $u_1, u_2, \dots, u_{2d+1}$ are words of length a in Σ^* such that u_1 has no leading zeros and such that $\gamma_j(u_1 \dots u_{2d+1}) = \gamma$.
- (10) We let \mathcal{S} denote the set of words u with no leading zeros and whose length is a multiple of a such that $\gamma_j(u) = \gamma$.
- (11) We let $p_1, p_2, p_3 \in \{1, \dots, 2d + 1\}$ with $p_1 < p_2 < p_3$ be such that $\ell := \delta(1, u_1 \dots u_{p_1}) = \delta(1, u_1 \dots u_{p_2}) = \delta(1, u_1 \dots u_{p_3})$. (Such p_1, p_2, p_3 exist by the pigeonhole principle.)
- (12) We let $U = u_1 \dots u_{p_1}$, $V = u_{p_1+1} \dots u_{p_2}$, and $W = u_{p_2+1} \dots u_{p_3}$.

Now using the notation above, we explain how the remainder of the proof goes. Using the fact that $\ell = \delta(1, U) = \delta(1, UV) = \delta(1, UVW)$, we see $\ell = \delta(\ell, V) = \delta(\ell, W)$. By construction $\gamma = \gamma_j(UVW)$ and $\gamma \geq \gamma_j(UW)$. We use Lemma 3.1 to show that $\gamma_j(UV^2W) \geq \gamma_j(UVW) \geq \gamma_j(UW)$. But since $\gamma_j(UV^2W) \leq \gamma$, we get that $\gamma_j(UV^2W) = \gamma$. We then use an

induction argument to show that $\gamma_j(UV^mW) = \gamma$ for every $m \geq 1$. Finally, we show that as $m \rightarrow \infty$, $\gamma_j(UV^mW)$ tends to a recursively computable rational number, which then gives the result. Since there are only finitely many possibilities for words U and V with the length of UV bounded by $(2d + 1)a$, we can compute the values $\gamma_i(AB)$ over all possible triples (A, B, i) , with A, B non-trivial words in Σ^* such that A has no leading zeros, $|A| + |B| \leq (2d + 1)a$ and $i \leq d$, and the maximum will be the limsup of $s(n)/n$ as $n \rightarrow \infty$.

A key step in this strategy is the following lemma, which is inspired by work of Schaeffer and Shallit [10].

Lemma 3.4. *Adopt the assumptions and notation of Notation 3.2 and Notation 3.3. Suppose that $A, B, C \in \Sigma_k^*$ are non-trivial words whose lengths are multiples of a such that A has no leading zeros, $ABC \in \mathcal{S}$, and $e := \delta(1, A) = \delta(1, AB) = \delta(1, ABC)$. Then AB^2C in \mathcal{S} and*

$$\gamma(AB^2C) = \gamma(ABC) = \gamma.$$

Proof. We have $\gamma_j(ABC) = \gamma$ and $\gamma_j(AC) \leq \gamma$. We claim that $\gamma(AB^2C) \geq \gamma(ABC)$, which will then give the result since $\gamma(AB^2C) \leq \gamma$. By assumption, there is a sequence of words w_n with $|w_n| \equiv j \pmod{a}$ and $|w_n| \rightarrow \infty$ such that

$$s([ABCw_n]_k)/[ABCw_n]_k \rightarrow \gamma.$$

Then

$$(3.5) \quad s([ABCw_n]_k) = \sum_{\{v: |v|=|A|, v \prec A\}} \sum_{\{|w|=|BCw_n|\}} f(vw) + \sum_{\{w: |w|=|BCw_n|, w \preceq BCw_n\}} f(Aw),$$

where \preceq is (pure) lexicographic ordering on $\{0, 1, \dots, k\}^*$ and we use \prec for strict inequality. By Proposition 2.1 there is a rational number $\kappa, \beta \in (0, k)$, and a positive constant C_0 such that

$$(3.6) \quad \left| \sum_{\{v: |v|=|A|, v \prec A\}} \sum_{|w|=|BCw_n|} f(vw) - \kappa k^{|BCw_n|} \right| < C_0 \beta^{|w_n|}$$

for every n , where

$$(3.7) \quad \kappa = \sum_{\{v: |v|=|A|, v \prec A\}} c_{\delta(1,v),j} \in \mathbb{Q}.$$

Now from the definition of e we have

$$f(Aw) = f(ABw) = f(ABCw) = f_e(w)$$

for every w . Furthermore $f_e(Bw) = f_e(w)$ for every w . We have

$$\begin{aligned} & \sum_{\{w: |w|=|BCw_n|, w \preceq BCw_n\}} f(Aw) \\ &= \sum_{\{w: |w|=|BCw_n|, w \preceq BCw_n\}} f_e(w) \\ &= \sum_{\{v: |v|=|B|, v \prec B\}} \sum_{|w|=|Cw_n|} f_e(vw) + \sum_{\{w: |w|=|Cw_n|, w \preceq Cw_n\}} f_e(Bw) \\ &= \sum_{\{v: |v|=|B|, v \prec B\}} \sum_{|w|=|Cw_n|} f_e(vw) + \sum_{\{w: |w|=|Cw_n|, w \preceq Cw_n\}} f_e(w). \end{aligned}$$

Using Proposition 2.1 again, there is some nonnegative rational κ' and some positive constant C'_0 and some $\beta' \in (0, k)$ such that

$$\left| \sum_{\{v: |v|=|B|, v \prec B\}} \sum_{|w|=|Cw_n|} f_e(vw) - \kappa' k^{|Cw_n|} \right| < C'_0 (\beta')^{|w_n|},$$

where

$$(3.8) \quad \kappa' = \sum_{\{v: |v|=|B|, v \prec B\}} c_{\delta(e,v),j} \in \mathbb{Q}.$$

By increasing either β or β' , we may assume that $\beta = \beta'$. Using this fact along with Equations (3.5) and (3.6), we see there is some $C_1 > 0$ such that

$$(3.9) \quad \left| s([ABCw_n]_k) - (\kappa k^{|BC|} + \kappa' k^{|C|}) k^{|w_n|} - s_e(Cw_n) \right| < C_1 \beta^{|w_n|}.$$

We similarly get positive constants C_2 and C_3 such that

$$(3.10) \quad \left| s([ACw_n]_k) - \kappa k^{|C|} k^{|w_n|} - s_e(Cw_n) \right| < C_2 \beta^{|w_n|}$$

and

$$(3.11) \quad \left| s([ABBCw_n]_k) - (\kappa k^{|BBC|} + \kappa' k^{|BC|} + \kappa' k^{|C|}) k^{|w_n|} - s_e(Cw_n) \right| < C_3 \beta^{|w_n|}.$$

We next let $r_n = [Cw_n]_k$ and we let $u = [B]_k$ and $v = [A]_k$. By assumption

$$s([ABCw_n]_k) / [ABCw_n]_k \rightarrow \gamma \quad \text{as} \quad n \rightarrow \infty,$$

and hence Equation (3.9) gives

$$\left(\kappa k^{|BC|} k^{|w_n|} + \kappa' k^{|C|} k^{|w_n|} + s_e([Cw_n]_k) \right) / \left(vk^{|BCw_n|} + uk^{|Cw_n|} + r_n \right) \rightarrow \gamma$$

as $n \rightarrow \infty$. Now since $\limsup s([ACw_n]_k) / [ACw_n]_k \leq \gamma$, we have

$$\limsup \left(\kappa k^{|C|} k^{|w_n|} + s_e([Cw_n]_k) \right) / (vk^{|Cw_n|} + r_n) \leq \gamma.$$

Now we apply Lemma 3.1 with $u' = \kappa', v' = \kappa, s'_n = s_e([Cw_n]_k)k^{-|w_n|}$, $b = |B|, c = |C|, s_n = r_n k^{-|w_n|}$ and u, v as above to deduce that

$$\limsup \frac{(\kappa k^{|BBC|} k^{|w_n|} + \kappa' k^{|BC|} k^{|w_n|} + \kappa' k^{|C|} k^{|w_n|} + s_e([Cw_n]_k))}{(u k^{|BBCw_n|} + v k^{|BCw_n|} + v k^{|Cw_n|} + r_n)}$$

is greater than or equal to γ , and so $\gamma_j(ABBC) = \gamma$, which gives the result. \square

As an immediate consequence we get the following.

Proposition 3.5. *Adopt the assumptions and notation of Notation 3.2 and Notation 3.3. Suppose that $A, B, C \in \Sigma_k^*$ are non-trivial words whose lengths are multiples of a such that A has no leading zeros, $ABC \in \mathcal{S}$, $\delta(1, A) = \delta(1, AB) = \delta(1, ABC)$, and $\gamma_j(ABC) = \gamma$. Then $\gamma_j(AB^s C) = \gamma$ for every $s \geq 1$.*

Proof. By assumption, the result holds for $s = 1$ and Lemma 3.4 gives $\gamma_j(AB^2 C) = \gamma$. Now suppose that the result holds whenever $s \leq m$. Then $\gamma_j(AB^m C) = \gamma$ and we may apply Lemma 3.4, leaving A and B unchanged and using $B^{m-1} C$ for C and we get $\gamma_j(AB^2 B^{m-1} C) = \gamma$. The result now follows by induction. \square

Proof of the Theorem 1.1. Without loss of generality we may scale the set Δ by a positive rational number and then assume that $\Delta \subseteq \mathbb{Q} \cap [0, 1]$. In particular, the maps f_i take values in $[0, 1] \cap \mathbb{Q}$ for $i = 1, \dots, d$. Applying Proposition 3.5 to $A = U, B = V$, and $C = W$, we see $\gamma_j(AB^m) = \gamma$ for every $m \geq 0$. If we take w to be a word whose length is congruent to $j \pmod a$, then the argument in Lemma 3.4 shows that

$$s([AB^m w]_k) = \kappa k^{|w|+m|B|} + \kappa'(1 + k^{|B|} + \dots + k^{(m-1)|B|})k^{|w|} + s_\ell([w]_k),$$

with $\kappa = \kappa_{A,j}$ and $\kappa' = \kappa'_{B,j}$ given as in Equations (3.7) and (3.8); that is,

$$(3.12) \quad \kappa = \kappa_{A,j} = \sum_{\{v: |v|=|A|, v \prec A\}} c_{\delta(1,v),j}$$

and

$$(3.13) \quad \kappa' = \kappa'_{B,j} = \sum_{\{v: |v|=|B|, v \prec V\}} c_{\delta(\ell,v),j}.$$

Then since f_ℓ takes values in $[0, 1]$, we see that $0 \leq s_\ell([w]_k) \leq k^{|w|}$.

Hence

$$(3.14) \quad \begin{aligned} \kappa k^{m|B|} + \frac{\kappa'(k^{m|B|} - 1)}{(k^{|B|} - 1)} &\leq s([AB^m w]_k)/k^{|w|} \\ &\leq \kappa k^{m|B|} + \kappa' \frac{(k^{m|B|} - 1)}{(k^{|B|} - 1)} + 1. \end{aligned}$$

On the other hand,

$$\begin{aligned}
 [A]_k k^{m|B|} + [B]_k \frac{(k^{m|B|} - 1)}{(k^{|B|} - 1)} &\leq [AB^m w]_k / k^{|w|} \\
 &\leq [A]_k k^{m|B|} + [B]_k \frac{(k^{m|B|} - 1)}{(k^{|B|} - 1)} + 1.
 \end{aligned}$$

Combining these two equations, we see

$$\begin{aligned}
 \left| s([AB^m w]_k) / [AB^m w]_k - \left(\kappa + \kappa' / (k^{|B|} - 1) \right) / \left([A]_k + [B]_k / (k^{|B|} - 1) \right) \right| \\
 = O(1/k^{m|B|}),
 \end{aligned}$$

where the implied constant in O is independent of m . Since κ and κ' are rational,

$$(3.15) \quad \alpha := \left(\kappa + \kappa' / (k^{|B|} - 1) \right) / \left([A]_k + [B]_k / (k^{|B|} - 1) \right)$$

is a rational number. We claim that $\gamma = \alpha$. To see this, observe that for every m , there are words $w_{m,n}$, $n \geq 1$, such that

$$s([AB^m w_{m,n}]_k) / [AB^m w_{m,n}]_k \rightarrow \gamma \quad \text{as} \quad n \rightarrow \infty.$$

On the other hand, we have shown that there is a fixed constant $C > 0$ such that

$$|s([AB^m w_{m,n}]_k) / [AB^m w_{m,n}]_k - \alpha| < Ck^{-m|B|}.$$

Taking the limit as n tends to infinity then gives that $|\gamma - \alpha| < Ck^{-m|B|}$. Since $|B| > 0$ and since this holds for every $m > 0$, $\gamma = \alpha$. Finally, we note that γ is recursively computable, since Proposition 2.1 gives that a and the rational constants $c_{i,j}$ are recursively computable and hence $\kappa_{A,j}$ and $\kappa'_{B,j}$ given in Equations (3.12) and (3.13) are recursively computable for each pair of words (A, B) and each $j \in \{0, \dots, a - 1\}$ and since the limsup of $s(n)/n$ is of the form $\gamma_i(AB^m)$ for every $m \geq 1$ with some pair of words (A, B) with $|A| + |B| \leq (2d + 1)a$ and some $i \in \{0, 1, \dots, a - 1\}$, we can proceed as follows. The limit of $\gamma_i(AB^m)$ as $m \rightarrow \infty$ is recursively computable by Equation (3.15), and letting A and B range over non-trivial words with $|AB| \leq (2d + 1)a$ and A having no leading zeros and letting i range over $0, \dots, a - 1$, and computing these limits and taking the maximum of these values, we see we can compute the limsup of $s(n)/n$ as $n \rightarrow \infty$. \square

Proof of Corollary 1.2. Define $f : \mathbb{N} \rightarrow \{0, 1\}$ and $g : \mathbb{N} \rightarrow \{0, 1\}$ via the rules $f(n) = 1$ if $n \in S$ and $f(n) = 0$ if $n \notin S$ and $g(n) = 1$ if $n \notin S$ and $f(n) = 0$ if $n \in S$. We let $s(n) = \sum_{i < n} f(i)$ and we let $t(n) = \sum_{i < n} g(i)$. Then $\gamma := \limsup s(n)/n$ and $\gamma' := \limsup t(n)/n$ are recursively computable rational numbers by Theorem 1.1. Notice that γ is

just the upper density of S . On the other hand, $s(n) + t(n) = n$ and if we let β denote the lower density of S then

$$\begin{aligned} \beta &= \liminf s(n)/n = -\limsup(-s(n)/n) \\ &= -\limsup(t(n) - n)/n = -(-1 + \limsup t(n)/n) = 1 - \gamma', \end{aligned}$$

which is a recursively computable rational number. The result follows. \square

4. An algorithm for computing the limsup

In this section, we give an algorithm to compute $\limsup s(n)/n$, where $s(n)$ is the n -th partial sum of a $\mathbb{Q}_{\geq 0}$ -valued automatic sequence $h(n)$. We note that this algorithm essentially falls out of the proof of Theorem 1.1, but we record it here explicitly as it may be of interest to people working with a given automatic sequence or set. Then there is an automatic map $f : \Sigma_k^* \rightarrow \mathbb{Q}_{\geq 0}$ satisfying $f(0w) = f(w)$ and $f((n)_k) = h(n)$.

We now use the assumptions and notation of Notation 3.2 and Notation 3.3. We assume that we have a DFAO that accepts $w \in \Sigma_k^*$ as input, reading left to right, and gives $f(w)$ as output. From the DFAO, we can construct automatic sequences $f = f_1, \dots, f_d$ that make up the left k -kernel of f . By Proposition 2.1 there is a recursively computable natural number a and recursively computable rational numbers $c_{i,j}$ with $1 \leq i \leq d, 0 \leq j < a$ such that

$$s_i((k^{an+j})_k)/k^{an+j} = c_{i,j} + o(1).$$

The first step is compute a and the values $c_{i,j}$. Using this terminology we get that there are words A and B whose lengths are multiples of a with $|A| + |B| \leq (2d + 1)a$ and some $\ell \in \{0, 1, \dots, a - 1\}$ such that

$$\limsup_n s(n)/n = (\kappa_{A,j} + \kappa'_{B,j}/(k^{|B|} - 1)) / ([A]_k + [B_k]/(k^{|B|} - 1))$$

where $\kappa_{A,j}$ and $\kappa'_{B,j}$ are as given in Equations (3.12) and (3.13). Moreover, each of the values of this form, by construction, occurs as a limit point of the sequence $s(n)/n$. Consequently, one can take the maximum of the numbers of the form

$$(\kappa_{A,j} + \kappa'_{B,j}/(k^{|B|} - 1)) / ([A]_k + [B_k]/(k^{|B|} - 1)),$$

as one lets A and B range over the set of non-trivial words whose lengths are multiples of a with $|A| + |B| \leq (2d + 1)a$ such that A has no leading zeros and lets j range over $\{0, \dots, a - 1\}$, and the maximum of these values will be $\limsup_n s(n)/n$.

As an example of how one can apply this in practice, we let $h(n)$ be the 3-automatic sequence whose value is 1 if the most significant ternary digit of n is equal to 1 and is zero otherwise, and we let $s(n)$ denote the n -th partial sum of $\{h(j)\}_{j \geq 0}$. Then there is $f : \Sigma_3^* \rightarrow \{0, 1\}$ with the property that $f(0w) = f(w)$ for $w \in \Sigma_3^*$ and $f((n)_3) = h(n)$. If we look at the left kernel

of f , it consists of the $f = f_1, f_2, f_3$ where f_2 is the constant function 0 and f_3 is the constant function 1, and we have the rules $f_1([0w]_k) = f_1([w]_k)$, $f_1([1w]_k) = f_3([w]_k)$, $f_1([2w]_k) = f_2([w]_k)$ and $f_2([iw]_k) = f_2([w]_k)$ and $f_3([iw]_k) = f_3([w]_k)$ for $w \in \{0, 1, 2\}^*$ and $i \in \{0, 1, 2\}$. In this case, we find we can take $a = 1$ in Proposition 2.1 and that $f_1(3^n) \sim (1/2) \cdot 3^n$, $f_2(3^n) = 0$, $f_3(3^n) = 3^n$. By the algorithm described above, the limsup of $s(n)/n$ is the maximum over words $A, B \in \{0, 1, 2\}^*$ such that A has no leading zeros and $|A| + |B| \leq 6$. We let $i \in \{1, 2\}$ denote the first letter of A and we write $A = iA'$. Using Equations (3.12) and (3.13) we see

$$\kappa_{A,0} = k^{|A|-1}/2 + \delta_{i,2}k^{|A|} + \delta_{i,1}[A']_k,$$

and

$$\kappa'_{B,0} = \delta_{i,1}[B]_k,$$

regardless of what A' and B are. Thus Equation (3.15) gives $\gamma_0(AB^m)$ tends to

$$(\kappa_{A,0} + \kappa'_{B,0}/(k^{|B|} - 1))/([A]_k + [B]_k/(k^{|B|} - 1))$$

as $m \rightarrow \infty$. Checking these values for the allowable A and B , we see this is maximized when A is the one-letter word 1 and B is the one-letter word 2, in which case one gets a limit of $3/4$.

5. Proof of Theorem 1.3

In this section we give the proof Theorem 1.3. Given a set of natural numbers S , we let $\pi_S(x)$ denote the number of elements in S that are less than x . To get the final part of this characterization we need a simple lemma.

Lemma 5.1. *Let $k \geq 2$ be a natural number and let S be a k -automatic set. Then*

$$\liminf_{N \rightarrow \infty} \pi_S(N)/N = 0 \implies \limsup_{N \rightarrow \infty} \pi_S(N)/N = 0$$

and

$$\limsup_{N \rightarrow \infty} \pi_S(N)/N = 1 \implies \liminf_{N \rightarrow \infty} \pi_S(N)/N = 1.$$

Proof. By Proposition 2.1, there exist $\beta \in (0, k)$, a positive integer a , and nonnegative rational numbers c_0, \dots, c_{a-1} such that $\pi_S(k^{an+j}) = c_j k^{an+j} + O(\beta^{an})$. We claim that if $\liminf_{N \rightarrow \infty} \pi_S(N)/N = 0$ then c_0, \dots, c_{a-1} . To see this, suppose that this is not the case. Then there is some i such that $c_i > 0$. Then for every N sufficiently large there is some n such that $k^{an+i} \leq N < k^{a(n+1)+i}$. Hence $\pi_S(N)/N \geq \pi_S(k^{an+i})/k^{a(n+1)+i} \sim c_i/k^a$ as $n \rightarrow \infty$. It follows that if some $c_i > 0$ then $\liminf \pi_S(N)/N > 0$, and so the claim follows. Thus if $\liminf_{N \rightarrow \infty} \pi_S(N)/N = 0$ then we have $c_0 = \dots = c_{a-1} = 0$. We claim that this then gives that $\limsup \pi_S(N)/N = 0$. To see this,

suppose that $\limsup \pi_S(N)/N = \gamma > 0$. Then there are infinitely many N such that $\pi_S(N) > \gamma N/2$. For such N we have an n , depending upon N , such that $k^{an} \leq N < k^{an+a}$ and so

$$\pi_S(k^{an+a})/k^{an+a} > \pi_S(N)/(k^a N) > \gamma/(2k^a),$$

contradicting the fact that $c_0 = 0$. Thus we have shown that

$$\liminf \pi_S(N)/N = 0 \implies \limsup \pi_S(N)/N = 0.$$

Similarly, if we let T denote the complement of S then T is automatic and $\limsup \pi_S(N)/N = 1$ if and only if $\liminf \pi_T(N)/N = 0$, and so if $\limsup \pi_S(N)/N = 1$ then $\liminf \pi_T(N)/N = 0$ and hence $\limsup \pi_T(N)/N = 0$ and thus $\liminf \pi_S(N)/N = 1$. This completes the proof. \square

Proof of Theorem 1.3. By Lemma 5.1 and Corollary 1.2, if S is a k -automatic having lower density and upper density α and β respectively then either $(\alpha, \beta) \in \{(0, 0), (1, 1)\}$ or α, β are rational numbers with $0 < \alpha \leq \beta < 1$. We notice that the empty set has upper and lower density 0 and \mathbb{N} has upper and lower density 1, and so to complete the proof, it suffices to show that whenever α, β are rational numbers satisfying $0 < \alpha \leq \beta < 1$ then there is a k -automatic set whose lower density is α and whose upper density is β . Since a set is k -automatic if and only if it is k^m -automatic for each positive integer m , we may replace k with a power and assume that $k\alpha \geq \beta$ and that $k\beta < (k - 1)$. We let

$$\alpha' = (k\alpha - \beta)/(k - 1) \quad \text{and} \quad \beta' = (k\beta - \alpha)/(k - 1).$$

Since $k\alpha > \beta \geq \alpha$, we see $\alpha', \beta' > 0$. Also, since $k\beta, k\alpha < k - 1$ we have $\alpha', \beta' < 1$. We have

$$(5.1) \quad k\alpha' + \beta' = (k + 1)\alpha \quad \text{and} \quad k\beta' + \alpha' = (k + 1)\beta.$$

Now let A, B , and C be positive integers with $C > A, B$ and $A/C = \alpha'$ and $B/C = \beta'$. We let T_0 be the set of natural numbers that are either $0, 1, \dots, A - 2$, or $A - 1 \pmod C$; we let T_1 be the set of natural numbers that are either $0, 1, \dots, B - 2$, or $B - 1 \pmod C$. We let U_0 be the set of natural numbers whose base- k expansion has even length and we let U_1 be the set of natural numbers whose base- k expansion has odd length. We let $S = (U_0 \cap T_0) \cup (U_1 \cap T_1)$. Since each of U_0, U_1, T_0, T_1 are k -automatic sets, so is S as such sets are closed under finite intersections and unions. For an interval $I \subseteq [k^{2n}, k^{2n+1})$ the number of elements in $S \cap I$ satisfies

$$(5.2) \quad \frac{B \cdot \#(I \cap \mathbb{N})}{C} - C \leq \#(S \cap I) \leq \frac{B \cdot \#(I \cap \mathbb{N})}{C} + C$$

Similarly, for an interval $I \subseteq [k^{2n+1}, k^{2n+2})$ the number of elements in $S \cap I$ satisfies

$$(5.3) \quad \frac{A \cdot \#(I \cap \mathbb{N})}{C} - C \leq \#(S \cap I) \leq \frac{A \cdot \#(I \cap \mathbb{N})}{C} + C$$

From these inequalities, it is straightforward to deduce that $\pi_S(k^{2n})$ is asymptotic to

$$(A/C)(1 + (k^2 - k) + \dots + (k^{2n} - k^{2n-1})) + (B/C)((k - 1) + (k^3 - k^2) + \dots + (k^{2n-1} - k^{2n-2})),$$

which gives

$$(A/C)(k^{2n+1} + 1)/(k + 1) + (B/C)(k^{2n} - 1)/(k + 1).$$

From this we obtain the asymptotic result

$$(5.4) \quad \pi_S(k^{2n}) \sim \alpha' k^{2n+1}/(k + 1) + \beta' k^{2n}/(k + 1) = \alpha k^{2n}.$$

Similarly,

$$(5.5) \quad \pi_S(k^{2n+1}) \sim \beta' k^{2n+2}/(k + 1) + \alpha' k^{2n+1}/(k + 1) = \beta k^{2n+1}$$

as $n \rightarrow \infty$. It follows that for $N \in [k^{2n+1} + 1, k^{2n}]$ we have

$$\pi_S(N)/N \sim \beta k^{2n+1}/N + \beta' - \beta' k^{2n+1}/N$$

as $n \rightarrow \infty$. If we fix n and let N range over the interval $[k^{2n+1} + 1, k^{2n}]$, then since the function $\beta k^{2n+1}/x + \beta' - \beta' k^{2n+1}/x$ has derivative of the form κx^{-2} , it is monotonic on the interval $[k^{2n+1}, k^{2n+2}]$ and hence the maximum and minimum are attained at the end points. When $N = k^{2n+1}$, $\beta k^{2n+1}/N + \beta' - \beta' k^{2n+1}/N$ is equal to β , and at $N = k^{2n+2}$ it is equal to $\beta/k + \beta'(1 - 1/k) = \alpha$ and so

$$\alpha = \liminf_{n \rightarrow \infty} \inf_{N \in [k^{2n+1}, k^{2n}]} \pi_S(N)/N$$

and

$$\beta = \limsup_{n \rightarrow \infty} \sup_{N \in [k^{2n+1}, k^{2n}]} \pi_S(N)/N.$$

Similarly,

$$\alpha = \liminf_{n \rightarrow \infty} \inf_{N \in [k^{2n}, k^{2n+1}]} \pi_S(N)/N$$

and

$$\beta = \limsup_{n \rightarrow \infty} \sup_{N \in [k^{2n}, k^{2n+1}]} \pi_S(N)/N.$$

It follows that α and β are respectively the liminf and limsup of $\pi_S(N)/N$, as claimed. □

6. Concluding remarks

We have established that the upper and lower densities of an automatic set are recursively computable rational numbers. It is natural to ask whether similar results hold when one looks at larger classes of sets. A generalization of automatic sets is sets associated to a morphic word on the alphabet $\{0, 1\}$. A result of Cobham [4] shows that automatic sets are precisely those corresponding to *uniform* morphisms, and so these morphic sets form a strictly larger class. In this case, the density of morphic sets need not be rational, if it exists. For example, the Fibonacci word, which is the right-infinite word $01001\dots$ that is the unique fixed point of the morphism $0 \mapsto 01$ and $1 \mapsto 0$ whose first letter is 0, corresponds to the set $\{1, 3, 4, \dots\}$ (i.e., the elements of the set are the positions where the 0's occur in the sequence). This set has density $1/\rho^2$, where $\rho = (\sqrt{5} + 1)/2$ [9, Proposition 2.1.10]. We suspect that the densities, when they exist, and the upper and lower densities of morphic sets should be algebraic numbers. We believe the techniques in this paper along with those given in [2] might be useful in establishing these facts, although we leave this as a question for others to think about.

Another interesting generalization of automatic sets are those produced via push-down automata; i.e., context-free subsets of \mathbb{N} , where we once again assume that the subset is formed by taking the natural numbers whose base- k expansions form a context-free sublanguage of $\{0, 1, \dots, k-1\}^*$ for some $k \geq 2$. Here there is an interesting dichotomy that arises: unambiguous context-free and ambiguous context-free. The former case is much better behaved and work of Chomsky and Schützenberger [3], along with basic asymptotic results for algebraic functions, shows that if the density of such a set of natural numbers exists then it is necessarily an algebraic number. On the other hand, the question of whether upper and lower densities of unambiguous context-free subsets of \mathbb{N} are algebraic or not is apparently open. For ambiguous context-free sets, the behaviour can be much more pathological. Work of Kemp [7], with some small additional amount of arguing, shows that such densities can be transcendental.

References

- [1] J.-P. ALLOUCHE & J. SHALLIT, *Automatic sequences. Theory, applications, generalizations*, Cambridge University Press, 2003.
- [2] J. P. BELL, “Logarithmic frequency in morphic sequences”, *J. Théor. Nombres Bordeaux* **20** (2008), no. 2, p. 227-241.
- [3] N. CHOMSKY & M.-P. SCHÜTZENBERGER, “The algebraic theory of context-free languages”, in *Computer programming and formal systems*, Studies in Logic and the Foundations of Mathematics, North-Holland, 1963, p. 118-161.
- [4] A. COBHAM, “Uniform tag sequences”, *Math. Syst. Theory* **6** (1972), p. 164-192.
- [5] N. J. HIGHAM & L. LIN, “On p th roots of stochastic matrices”, *Linear Algebra Appl.* **435** (2011), no. 3, p. 448-463.

- [6] F. I. KARPELEVICH, “On the characteristic roots of matrices with nonnegative elements”, *Izv. Akad. Nauk SSSR, Ser. Mat.* **15** (1951), p. 361-383.
- [7] R. KEMP, “A note on the density of inherently ambiguous context-free languages”, *Acta Inf.* **14** (1980), no. 3, p. 295-298.
- [8] A. K. LENSTRA, H. W. J. LENSTRA & L. LOVÁSZ, “Factoring polynomials with rational coefficients”, *Math. Ann.* **261** (1982), no. 4, p. 515-534.
- [9] M. LOTHAIRE, *Algebraic Combinatorics on Words*, Encyclopedia of Mathematics and Its Applications, vol. 90, Cambridge University Press, 2002.
- [10] L. SCHAEFFER & J. SHALLIT, “The critical exponent is computable for automatic sequences”, *Int. J. Found. Comput. Sci.* **23** (2012), no. 8, p. 1611-1626.

Jason P. BELL
Department of Pure Mathematics
University of Waterloo
Waterloo, ON N2L 3G1, Canada
E-mail: jpbell@uwaterloo.ca
URL: <http://www.math.uwaterloo.ca/~jpbell>