Logarithmic density in morphic sequences

par Jason P. Bell

Résumé. Nous répondons affirmativement à une question d’Allouche et Shallit en montrant l’existence de la fréquence logarithmique des lettres et des mots dans une suite morphique.

Abstract. We study the logarithmic frequency of letters and words in morphic sequences and show that this frequency must always exist, answering a question of Allouche and Shallit.

Key words: Logarithmic frequency, morphic sequences, automatic sequences.
Primary classification: 68R15
Secondary classification: 68Q45, 11B85

1. Introduction

Let \( \Sigma \) be a finite alphabet. Given a sequence \( s = s(1), s(2), s(3), \ldots \) taking values in \( X \), it is natural to ask the question: With what frequency does a letter/word \( a \in \Sigma \) appear in the sequence \( s \)?

To make this question precise, we must carefully define what we mean by frequency. Given a sequence \( s \), the ordinary frequency of a letter \( a \in \Sigma \) in \( s \) is simply the limit
\[
\lim_{x \to \infty} \frac{\# \{ n \leq x \mid s(n) = a \}}{x},
\]
if it exists. More generally, if \( w \) is a word of length \( d \) on \( \Sigma \), then the ordinary frequency of \( w \) in \( s \) is the limit
\[
\lim_{x \to \infty} \frac{\# \{ n \leq x \mid s(n)s(n+1)\cdots s(n+d-1) = w \}}{x},
\]
if it exists.

The main problem with the ordinary frequency is that for many interesting sequences of classes, it cannot be guaranteed that it exists. For this reason, the logarithmic frequency is often a better measure of frequency.

Definition. Given a finite alphabet \( \Sigma \) and a sequence \( s = s(1), s(2), s(3), \ldots \) taking values in \( \Sigma \), we define the logarithmic density of a letter \( a \in \Sigma \) to be the limit
\[
\lim_{x \to \infty} \frac{1}{\log x} \sum_{\{ n \leq x \mid s(n) = a \}} \frac{1}{n}
\]
provided the limit exists. Similarly, we define the logarithmic density of a word \( w \in \Sigma^* \) to be the limit
\[
\lim_{x \to \infty} \frac{1}{\log x} \sum_{\{ n \leq x \mid s(n)s(=a) \}} \frac{1}{n}
\]
provided the limit exists, where \( d \) is the length of \( w \).

I thank NSERC for its generous support.
The main advantage of logarithmic density over ordinary density is that it exists for large classes of sequences; for example, Cobham showed that the logarithmic frequency of letters in automatic sequences always exists [2], while the ordinary frequency need not exist for automatic sequences. Logarithmic frequency generalizes ordinary frequency in the following sense: if the ordinary frequency of a letter/word in a sequence exists, then the logarithmic frequency will also exist and be equal to the ordinary frequency. Automatic sequences are a subset of the collection of morphic sequences. We define morphic sequences here.

Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \rightarrow \Sigma^*$ be a morphism. We say that a letter $a_i \in \Sigma$ is mortal if $\phi^j(a_i) = \epsilon$, the empty word, for some $j \geq 1$. Suppose that $\phi(a_1) = a_1 x$ for some non-empty word $x$ containing a non-mortal letter. Then we say that $\phi$ is prolongable on $a_1$. Then we can form the right infinite word

$$\phi^\omega(a_1) := a_1 x \phi(x) \phi^2(x) \cdots .$$

Note that the right-infinite word $\phi^\omega(a_1)$ is a fixed point of $\phi$. Words defined in this manner are called pure morphic words. In general, a word $u$ on a finite alphabet $\Delta$ is morphic if there is a set map $f : \Sigma \rightarrow \Delta$ such that $u = f(w)$ for some pure morphic word $w$ on $\Sigma$. We note that a word can be thought of as an infinite sequence of letters, and it is this interpretation that we use when we talk of letter frequency.

Since automatic sequences form a subset of the collection of morphic sequences, it is natural to ask whether the logarithmic frequency of a letter in a morphic sequence exists. This is a question of Allouche and Shallit [1, Section 8.8, p. 282]. Our main theorem is the following.

**Theorem 1.1.** Let $s$ be a morphic sequence on a finite alphabet $\Sigma$. Then the logarithmic frequency of a word $w \in \Sigma^*$ in $s$ exists.

We note that the ordinary frequency in morphic sequences has been studied by many authors [3, 4, 5, 6].

Our approach to proving Theorem 1.1 is to obtain how many occurrences of a word $w$ appear in certain “blocks” of $\phi^\omega(a_1)$. We then use the fact that $\phi$ is a morphism to obtain sufficiently small blocks that we can in fact compute the logarithmic frequency.

The outline of this paper is as follows. In Section 2, we discuss incidence matrices and basic results about nonnegative matrices. In Section 3, we give notation that will be used in obtaining the proof of our main theorem. In Section 4, we give asymptotic results that will be necessary for the proof. In Section 5 we prove Theorem 1.1.

**2. Incidence Matrices**

In this section we discuss the relationship between matrices and morphisms and Perron-Frobenius theory.

Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \rightarrow \Sigma^*$ be a morphism. We define the incidence matrix $M(\phi)$ of $\phi$ to be the $d \times d$ matrix whose $(i, j)$ entry is the number of occurrences of $a_i$ in $\phi(a_j)$. We note that $M(\phi^m) = M(\phi)^m$. From this we get the useful fact.

**Proposition 2.1.** Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \rightarrow \Sigma^*$ be a morphism. If $w$ is a finite word on $\Sigma$ then

$$\text{length}(\phi^m(w)) = [1 \ 1 \ \ldots \ 1] M(\phi)^m v,$$

where $v$ is the $d \times 1$ column vector whose $i$th coordinate is the number of occurrences of the letter $a_i$ in the word $w$. 

Proof. See Corollary 8.2.4 of Allouche and Shallit [1]. □

We note that if $\Sigma$ is a finite alphabet and $\phi : \Sigma^* \to \Sigma^*$ is a morphism, then $M(\phi)$ is a nonnegative matrix; that is, its entries are all nonnegative real numbers. Of fundamental importance in the study of nonnegative matrices is the Perron-Frobenius theorem.

**Theorem 2.2.** (Perron-Frobenius) Let $A$ be a nonnegative matrix. Then there is a nonnegative real number $\alpha$ such that if $\lambda$ is an eigenvalue of $A$ then either $|\lambda| < \alpha$ or $\lambda = \alpha \omega$ for some root of unity $\omega$.

The eigenvalue $\alpha$ is called a Perron-Frobenius eigenvalue of $A$. We note that if $\alpha$ is a Perron-Frobenius eigenvalue of $A$ then any other eigenvalue of $A$ on the circle $|z| = \alpha$ satisfies $\lambda^d = \alpha$ for some $d$. We thus obtain the following remark.

**Remark 1.** If $A$ is a nonnegative matrix then there is some number $d$ such that $A^d$ if $\lambda$ is not a Perron-Frobenius eigenvalue of $A$ then $|\lambda| < \alpha$.

**Definition.** Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a morphism. We say that $\phi$ is strongly Perron-Frobenius if $M(\phi)$ if every eigenvalue of $A$ other than the Perron-Frobenius eigenvalue is strictly less than the Perron-Frobenius eigenvalue of $A$ in absolute value.

**Lemma 2.3.** Let $A$ be a $d \times d$ nonnegative matrix and let $v, w \in \mathbb{R}^d$ be two vectors with nonnegative entries. Suppose $\alpha$ is the Perron-Frobenius eigenvalue of $A$ and $\beta < \alpha$ is such that all other eigenvalues of $A$ are strictly less than $\beta$ in absolute value. Then there is a (possibly zero) polynomial $p(x)$ such that $w^T A^n v = p(n) \alpha^n + O(\beta^n)$.

**Proof.** There exists an invertible complex $d \times d$ matrix $S$ such that

$$S^{-1} A S = J$$

is in Jordan form. The $i, j$ entry of $J^n$ can be expressed as $q_{i,j}(n) \alpha^n + O(\beta^n)$ for some polynomial $q_{i,j}$, possibly zero. Let $w_0 = w S$ and $v_0 = J S^{-1} v$. Then

$$w^T A^n v = w_0^T J^n v_0.$$

Since the entries of $J^n$ are all of the form $q(n) \alpha^n + O(\beta^n)$, we obtain the desired result. □

We are considering nonnegative matrices with integer entries. In this case an interesting trichotomy arises: the Perron-Frobenius eigenvalue is either strictly larger than 1, is equal to 1, or is 0. (It cannot be strictly between 0 and 1 since the product of the nonzero roots of the characteristic polynomial of $M$ is an integer.) If the Perron-Frobenius eigenvalue is zero, then $M$ is nilpotent. In particular, if $M = M(\phi)$ for some morphism $\phi$ of a finite alphabet, then every letter is mortal with respect to this morphism. If the Perron-Frobenius eigenvalue is 1 and $M$ is strongly Perron-Frobenius, the only eigenvalues of $A$ are 0 and 1. In particular, the $O(\beta^n)$ term can be eliminated in the statement of Lemma 2.3 and all quantities involved become polynomials in $n$. The case that the Perron-Frobenius eigenvalue is strictly larger than 1 is where the bulk of the difficulty lies.

### 3. Notation

In this section we introduce notation that will be used throughout our proof that logarithmic density of words in morphic sequences exists.

Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and suppose that $\phi : \Sigma^* \to \Sigma^*$ is a strongly Perron-Frobenius morphism prolongable on $a_1$. 
We introduce \(d\) functions
\[
   f_j(n; \phi) = \text{length}(\phi^n(a_j)) \quad \text{for } 1 \leq j \leq d.
\]
In addition to these \(d\) functions, to each word \(w \in \Sigma^*\) we define \(d\) functions
\[
   g_{j,w}(n; \phi) = \# \text{ of occurrences of } w \text{ in } \phi^n(a_j)
\]
for \(1 \leq i, j \leq d\). Note that
\[
   f_j(n; \phi) \geq g_{j,w}(n; \phi)
\]
for \(n \geq 1\).

Write \(\phi(a_1) = a_1x\) with \(x \in \Sigma^*\) and let \(m\) be a natural number. Write
\[
   \phi^m(x) = a_{i_1}a_{i_2} \cdots a_{i_{t(m)}},
\]
where
\[
   t(n; \phi) = f_1(n + 1; \phi) - f_1(n; \phi).
\]
We note that if
\[
   \phi^m(x) = a_{i_1}a_{i_2} \cdots a_{i_{t(m)}},
\]
then
\[
   \phi^{n+m}(x) = \phi^n(a_{i_1}) \cdots \phi^n(a_{i_{t(m)}})
\]
has length
\[
   t(m; \phi) = \sum_{j=1}^{t(m; \phi)} f_{i_j}(n; \phi).
\]
We define \(t(m; \phi)\) functions
\[
   h_{j,m}(n; \phi) = f_1(n + m; \phi) + \sum_{k=1}^{j} f_{i_k}(n; \phi) \quad \text{for } 0 \leq j \leq t(m; \phi).
\]
Note that in the expression of \(h_{j,m}(n; \phi)\), \(f_1(n + m; \phi)\) is just the length of \(\phi^{n+m}(a_1)\),
and the sum is the length of \(\phi^n(a_{i_1} \cdots a_{i_j})\). Thus \(h_{j,m}(n; \phi)\) is the length of an initial
subword of \(\phi^{n+m+1}(a_1)\) that contains \(\phi^{n+m}(a_1)\). Let \(w \in \Sigma^*\). We now define two new functions,
which will be important in getting upper and lower bounds when estimating
the logarithmic density.
\[
   A_{m,w}(n; \phi) = \sum_{j=1}^{t(m; \phi)} \log(h_{j,m}(n; \phi) + g_{j,w}(n; \phi) - 1) - \log(h_{j,m}(n; \phi) - 1)
\]
and
\[
   B_{m,w}(n; \phi) = \sum_{j=1}^{t(m; \phi)} \log(h_{j+1,m}(n; \phi)) - \log(h_{j+1,m}(n; \phi) - g_{j,w}(n; \phi)).
\]
In the study of logarithmic frequency of a word \(w \in \Sigma^*\) in a morphic sequence, it is of
course necessary to know the positions at which \(w\) occur.

**Definition.** Let \(\Sigma = \{a_1, \ldots, a_d\}\) be a finite alphabet and suppose that \(\phi : \Sigma^* \to \Sigma^*\) is a
strongly Perron-Frobenius morphism prolongable on \(a_1\). We define
\[
   s_w(n) = s_w(n; \phi) := \begin{cases} 1 & \text{if there is an occurrence of } w \text{ at the } n\text{th position of } \phi^w(a_1), \\ 0 & \text{otherwise}. \end{cases}
\]

**Remark 2.** When the morphism \(\phi\) is understood, we drop the argument \(\phi\) in all functions
described above.
4. Asymptotics

In this section, we prove several asymptotic results that we will need to prove Theorem 1.1

**Proposition 4.1.** Let \( \Sigma = \{a_1, \ldots, a_d\} \) be a finite alphabet and let \( \phi : \Sigma^* \to \Sigma^* \) be a strongly Perron-Frobenius morphism prolongable on \( a_1 \) and suppose that every letter \( a \in \Sigma \) appears in \( \phi^k(a_1) \) for some \( k \). Then there are polynomials \( p_1, \ldots, p_d \) with \( p_1 \neq 0 \) and real numbers \( \alpha, \beta \) with \( \alpha > \beta > 0 \) such that

\[
\text{length}(\phi^n(a_j)) \sim p_j(n)\alpha^n + \beta^n.
\]

**Proof.** Let \( \alpha \) denote the largest eigenvalue of the matrix \( M(\phi) \). Choose \( \beta < \alpha \) such that the remaining eigenvalues of \( M(\phi) \) are strictly less than \( \beta \) in absolute value. Then by Lemma 2.3, there exist polynomials \( p_1, \ldots, p_d \) such that

\[
f_j(n) = p_j(n)\alpha^n + O(\beta^n).
\]

To see why \( p_1 \neq 0 \), note that \( M(\phi) \) has \( \alpha \) as an eigenvalue. Hence there is some nonzero vector \( \mathbf{v} \) such that \( M(\phi)\mathbf{v} = \alpha\mathbf{v} \). By Proposition 2.1, it follows that \( p_j \) is nonzero for some \( j \). By hypothesis, \( f_1(n) \geq f_j(n - k) \), and so \( p_1 \) must also be nonzero. \( \square \)

We note that by the definition of the function \( h_{j,m}(n) \), they must also have asymptotics as described in the statement of Proposition 4.1

**Proposition 4.2.** Let \( \Sigma = \{a_1, \ldots, a_d\} \) be a finite alphabet and let \( \phi : \Sigma^* \to \Sigma^* \) be a strongly Perron-Frobenius morphism prolongable on \( a_1 \). If \( w \in \Sigma^* \), and \( \alpha \) is the Perron-Frobenius eigenvalue of \( M(\phi) \) then there are polynomials \( q_1, \ldots, q_d \) such that \( g_{j,w}(n) = q_j(n)\alpha^n(1 + o(1)) \).

**Proof.** Let \( m \) denote the length of \( w \). By relabeling if necessary, we may assume that \( \{a_e, \ldots, a_d\} \) are the mortal letters with respect to \( \phi \). We note that in any word of the form \( \phi^n(a_j) \) there is a uniform bound on the size of the “gaps” between consecutive appearances of non-mortal letters. Hence are distinct natural numbers \( b \) and \( c \) with \( b < c \) such that

1. the shortest prefix of \( \phi^b(a_j) \) containing \( m \) non-mortal letters is the same as the shortest prefix of \( \phi^c(a_j) \) containing \( m \) non-mortal letters for \( 1 \leq j \leq d \);
2. the shortest suffix of \( \phi^b(a_j) \) containing \( m \) non-mortal letters is the same as the shortest suffix of \( \phi^c(a_j) \) containing \( m \) non-mortal letters for \( 1 \leq j \leq d \).

Pick \( k \) such that \( \phi^k(a_j) = \varepsilon \) for \( e \leq j \leq d \). Then by construction, the first \( d \) letters of \( \phi^n(a_j) \) are the same as the first \( d \) letters of \( \phi^{n+c-b}(a_j) \) for \( 1 \leq j \leq d \) and \( n \geq b + k \).

Write

\[
\phi(a_j) = a_{i_1,j} \cdots a_{i_m,j},
\]

for \( 1 \leq j \leq d \). Then

\[
\phi^{n+1}(a_j) = \phi^n(a_{i_1,j}) \cdots \phi^{b}(a_{i_m,j}).
\]

Looking at both sides and counting occurrences of \( w \), we see

\[
g_{j,w}(n + 1) = g_{i_1,w}(n) + \cdots + g_{i_m,w}(n) + C(j, n),
\]

where \( C(j, n) \) counts the number of occurrences of \( w \) that “overlap” at least two words of the form \( \phi^{n}(a_{i,j}) \). By the above remarks, we have \( C(j, n) = C(j, n + c - b) \) for \( n \geq b + k \). Let

\[
G_{j,w}(n) = g_{j,w}(n) - g_{j,w}(n - c + b).
\]
Then we obtain the vector equation:

\[
\begin{bmatrix}
G_{1,w}(n+1) \\
\vdots \\
G_{d,w}(n+1)
\end{bmatrix}
= M(\phi)
\begin{bmatrix}
G_{1,w}(n) \\
\vdots \\
G_{d,w}(n)
\end{bmatrix}
\]

for \( n \geq b + k \). It follows from Lemma 2.3 that there are polynomials \( Q_1, \ldots, Q_d \) such that

\[
G_{j,w}(n) = Q_j(n)\alpha^n + O(\beta^n).
\]

We note that if \( \alpha = 1 \), then by the remarks at the end of Section 2, we can ignore the \( O(\beta^n) \) term and by telescoping we see that

\[
g_{j,w}(n) = \left( \sum_{k=0}^{\lfloor (n-b-k)/(c-b) \rfloor} Q_j(n-k(c+b)) \right) \alpha^n + O(\beta^n),
\]

which is asymptotic to some polynomial \( q_j(n) \). If \( \alpha > 1 \), then we can pick \( \beta > 1 \). Then

\[
g_{j,w}(n) = \left( \sum_{k=0}^{\lfloor (n-b-k)/(c-b) \rfloor} Q_j(n-k(c+b)) \right) \alpha^n + O(\beta^n),
\]

which is again asymptotic to \( q_j(n)\alpha^n \) for some polynomial \( q_j(n) \).

We note that we do not say anything about the error term. We could get a stronger error term, however, by replacing \( \phi \) by a suitable iterate.

5. Estimates

In this section we prove Theorem 1.1. To this end we need the following estimates.

**Estimate 1.** Let \( a \) and \( b \) be integers with \( b > a > 2 \). Then

\[
\log(b) - \log(a) \leq \sum_{n=a}^{b-1} \frac{1}{n} \leq \log(b-1) - \log(a-1).
\]

**Proof.** This is a straightforward consequence of the fact that the function \( 1/x \) is decreasing on \([1, \infty)\), using upper and lower Riemann sums. \( \square \)

**Estimate 2.** Let \( a, b, c \) be positive real numbers with \( b > a \) and let \( F(x) = \log(x+c) - \log(x) \). Then

\[
|F(b) - F(a)| \leq (b-a)c/a^2.
\]

**Proof.** This is an easy consequence of the Mean Value Theorem. \( \square \)

**Lemma 5.1.** Let \( \Sigma = \{a_1, \ldots, a_d\} \) be a finite alphabet and let \( \phi : \Sigma^* \to \Sigma^* \) be a strongly Perron-Frobenius morphism prolongable on \( a_1 \). Then

\[
B_{m,w}(n) \leq \sum_{j=f_1(n+m)}^{f_1(n+m+1)-1} s_w(j)/j \leq t(m)\text{length}(w)/f_1(n+m) + A_{m,w}(n)
\]

for \( m, n \geq 1 \).
Proof. Write $\phi(a_1) = a_1x$ and $\phi^m(x) = a_{i_1} \cdots a_{i_{t(m)}}$. Then
\begin{equation}
\phi^{m+n}(x) = \phi^n(a_{i_1}) \cdots \phi^n(a_{i_{t(m)}}).
\end{equation}
Using the notation of Section 3, we have
\begin{equation}
f_1(n+m+1) - 1 \leq \sum_{j=1}^{t(m) h_{j+1,m}(n) - 1} \sum_{k=h_{j,m}(n)}^{t(m) h_{j+1,m}(n) - 1} 1/k \leq \sum_{j=1}^{t(m) h_{j+1,m}(n) - 1} \sum_{k=h_{j,m}(n)}^{t(m) h_{j+1,m}(n) - 1} 1/k.
\end{equation}
By assumption, there are exactly $g_{i_j,w}(n)$ occurrences of the word $w$ in $\phi^n(a_{i_j})$. This accounts for all occurrences of $w$ in $\phi^{m+n}(x)$ except for those which intersect at least two words of the form $\phi^n(a_{i_k})$ that appear in the right-hand side of equation (9). Since there are $t(m)$ such words, this leaves at most
\begin{equation}
t(m)\text{length}(w)
\end{equation}
unaccounted for occurrences of $w$. Therefore, we see that
\begin{equation}
t(m)\text{length}(w)/f_1(n+m) + \sum_{j=1}^{t(m) h_{j,m}(n+1)+g_{i_j,w}(n) - 1} \sum_{k=h_{j,m}(n)}^{t(m) h_{j+1,m}(n) - 1} 1/k \geq \sum_{j=1}^{t(m) h_{j+1,m}(n) - 1} \sum_{k=h_{j,m}(n)}^{t(m) h_{j+1,m}(n) - 1} 1/k.
\end{equation}
We now use the estimates from Remark 1 to obtain:
\begin{equation}
\sum_{j=1}^{t(m) h_{j+1,m}(n) - 1} \sum_{k=h_{j+1,m}(n)-g_{i_j,w}(n)}^{t(m) h_{j+1,m}(n) - 1} 1/k \geq \sum_{j=1}^{t(m) h_{j+1,m}(n) - 1} \log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - g_{i_j,w}(n))
\end{equation}
\begin{equation}
= B_{m,w}(n).
\end{equation}
Similarly,
\begin{equation}
\sum_{j=1}^{t(m) h_{j,m}(n)+g_{i_j,w}(n) - 1} \sum_{k=h_{j,m}(n)}^{t(m) h_{j+1,m}(n) - 1} 1/k \leq \sum_{j=1}^{t(m) h_{j,m}(n) + g_{i_j,w}(n) - 1} - \log(h_{j,m}(n) - 1)
\end{equation}
\begin{equation}
= A_{m,w}(n).
\end{equation}
The result follows. \hfill \square

Lemma 5.2. Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a strongly Perron-Frobenius morphism prolongable on $a_1$. Let $m$ be a natural number and $w \in \Sigma^*$, then
\begin{equation}
\lim_{n \to \infty} A_{m,w}(n)
\end{equation}
and
\begin{equation}
\lim_{n \to \infty} B_{m,w}(n)
\end{equation}
exist and are finite.
Proof. Fix $m$ and write $\phi(a_1) = a_1x$ and $\phi^m(x) = a_{i_1} \cdots a_{i_{t(m)}}$. Consider $A_{m,w}(n)$. Since $t(m)$ is fixed it is sufficient to show that
\begin{equation}
\log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - g_{i_j,w}(n))
\end{equation}
and
\begin{equation}
\log(h_{j,m}(n) + g_{i_j,w}(n) - 1) - \log(h_{j,m}(n + 1) - 1)
\end{equation}

tend to a finite limit for each \( j \). By the remarks immediately following Proposition 4.1 and by Proposition 4.2, both \( h_{j+1,m}(n) \) and \( h_{j+1,m}(n) - g_{i,j,w}(n) \) can be expressed in the form \( p(n)\alpha^n(1 + o(1)) \), for suitable polynomials \( p(x) \), where \( \alpha \) is a Perron-Frobenius eigenvalue of \( M(\phi) \). Thus

\[
\log \left( \frac{h_{j+1,m}(n)}{h_{j+1,m}(n) - g_{i,j,w}(n)} \right) = \log p(n)/q(n) + o(1),
\]

for some polynomials \( p \) and \( q \). We note that \( p \) and \( q \) are nonzero and satisfy \( p(n) \geq q(n) \) for sufficiently large \( n \), since

\[
h_{j+1} \geq h_{j+1,m}(n) - g_{i,j,w}(n) \geq h_0(n) \geq f_1(n + m),
\]

and \( f_1(n) \) is asymptotic to \( Cn^k\alpha^n \) for some positive constant \( C \) by Proposition 4.1.

We note that if \( p \) and \( q \) have the same degree then the limit exists and is finite, so the only way a problem can arise is if the degree of \( p \) is strictly greater than the degree of \( q \). But if this occurs the limit must be infinite, and by construction we have

\[
\log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - g_{i,j,w}(n)) \leq \sum_{j=f_1(n+m)}^{f_1(n+m+1)} 1/j \sim \log \alpha,
\]

and so the limsup cannot exceed \( \alpha \). A similar argument works for \( B_{m,w}(n) \). The result follows.

**Lemma 5.3.** Let \( \Sigma = \{a_1, \ldots, a_d\} \) be a finite alphabet and let \( \phi : \Sigma^* \to \Sigma^* \) be a strongly Perron-Frobenius morphism prolongable on \( a_1 \). Then there is a positive constant \( K \) and a natural number \( k \) such that for all \( m \) sufficiently large,

\[
\limsup_{n \to \infty} |A_{m,w}(n) - B_{m,w}(n)| \leq \begin{cases} Km^k\alpha^{-m/2} & \text{if } \alpha > 1 \\ 0 & \text{if } \alpha = 1 \end{cases}
\]

**Proof.** Using Estimate 2, we have

\[
A_{m,w}(n) - B_{m,w}(n) = \sum_{j=1}^{t(m)} \left( \log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - g_{i,j,w}(n)) \\
- \log(h_{j,m}(n + 1) + g_{i,j,a}(n)) + \log(h_{j,m}(n + 1)) \right) \\
\leq \sum_{j=1}^{t(m)} \left| h_{j+1,m}(n) - h_{j,m}(n) - g_{i,j,w}(n) \right| g_{i,j,w}(n) \right|/h_{j,m}(n)^2 \\
\leq \sum_{j=1}^{t(m)} |f_{i,j}(n)/h_{j,m}(n)|^2.
\]

We now divide the proof into cases.

**CASE I:** \( \alpha > 1 \). By Remark 4, it is no loss of generality to assume that there exist natural numbers \( e_1, \ldots, e_d \) such that

\[
f_j(n) \geq f_1(n - e_j)
\]
for all $n$; furthermore, $h_{j,m}(n) \geq h_{0,m}(n) = f_1(n+m)$. Pick $e \geq e_1, \ldots, e_d$. By Proposition 4.1 there exist positive constants $C_0$ and $C_1$ and a natural number $k$ such that

$$C_0 n^k \alpha^n \leq f_1(n) \leq C_1 n^k \alpha^n$$

for all sufficiently large $n$. Hence

$$|A_{m,w}(n) - B_{m,w}(n)| \leq t(m) \frac{f_1(n+e)}{f_1(n+m)}^2$$

$$\leq t(m) \left( \frac{C_1(n+e) \alpha^{n+k}}{C_0(n+m) \alpha^{n+m}} \right)^2$$

for all $n$ sufficiently large, Thus

$$|A_{m,w}(n) - B_{m,w}(n)| \leq t(m) C_1^2 C_0^{-2} \alpha^{-2m+e}$$

for $n$ sufficiently large. Since $t(m) \leq f_1(m) \leq C_1 m^k \alpha^{m}$ for $m$ sufficiently large, we see that

$$|A_{m,w}(n) - B_{m,w}(n)| \leq C_1^3 C_0^{-2} m^k \alpha^{-m+e}$$

whenever $m$ and $n$ are both sufficiently large. The result follows. \hfill \Box

**CASE II:** $\alpha = 1$. In this case, note that $f_{j_1}(n) \leq h_{j+1,m}(n) - h_{j,m}(n) \leq f_1(n+m+1) - f_1(n+m)$ and so $f_{j_1}(n)/h_{j,m}(n) \leq f_1(n+m+1)/f_1(n+m) - 1 \leq K/(n+m)$ as $n \to \infty$ for some positive constant $K$, since $f_1$ is a polynomial. Thus

$$\sum_{j=1}^{t(m)} |f_{j_1}(n)/h_{j,m}(n)|^2 \leq K t(m)/(n+m)^2.$$

Since $t(m) = O(m^k)$, the result follows. \hfill \Box

We are almost ready to prove Theorem 1.1. We first make a few simple remarks.

**Remark 3.** If $\Sigma$ and $\Delta$ are two finite alphabets and and $f : \Sigma \to \Delta$. Suppose $w$ is a pure morphic word on $\Sigma$ such that the logarithmic frequency of every subword exists. Then the logarithmic frequency of every subword of $f(w)$ exists.

This is a consequence of the fact that if $x$ is a subword of $f(w)$ then $f^{-1}(x)$ is a finite collection of subwords of $w$ and the logarithmic frequency of $x$ in $f(w)$ is simply the sum of the logarithmic frequencies of the subwords in $f^{-1}(x)$.

**Remark 4.** Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : X^* \to X^*$ be a non-erasing morphism prolongable on $a_1$. It is no loss of generality to assume that for each $i \leq d$, there is some number $m$ such that $a_i$ appears as a letter in $\phi^m(a_1)$, otherwise we can just delete the letter $a_i$ from $\Sigma$. That is,

$$\text{(10)} \quad \text{For each } i, \text{ there exists } e_i \text{ such that } \phi^{e_i}(a_1) \text{ has an occurrence of } a_i.$$  

Consequently, there is some $e_i$ such that

$$\text{length}(\phi^n(a_i)) \leq \text{length}(\phi^{n+e_i}(a_1)).$$

**Proof of Theorem 1.1:** By Remark 3, we may assume that our morphic sequence is pure. Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi$ be a morphism from $\Sigma^*$ to itself that is prolongable on $a_1$. By Remark 1, we can assume that $\phi$ is strongly Perron-Frobenius by replacing it with a suitable iterate.

Write $\phi(a_1) = a_1 x$ with $x \in \Sigma^*$. Let $m$ be a natural number and write

$$\phi^n(x) = a_{i_1} a_{i_2} \cdots a_{i_{t(m,\phi)}},$$
We note that the word
\[
\phi^{n+m}(x) = \phi^n(a_{i_1}) \cdots \phi^n(a_{i_{(m)}})
\]
has length
\[
t(m) \sum_{j=1}^{t(m)} f_{ij}(n).
\]
Since for fixed \(m\), \(t(n) \text{length}(w)/f_1(m + n) \to 0\) as \(n \to \infty\), we see by Lemma 5.1 that
\[
\left| \limsup_{n \to \infty} \sum_{j=1}^{t(m) h_{j+1,m}(n)-1} s_w(k)/k - \liminf_{n} \sum_{j=1}^{t(m) h_{j+1,m}(n)-1} s_w(k)/k \right|
\]
\[
\leq \left| \limsup_{n \to \infty} A_{m,w}(n) - \liminf_{n} B_{m,w}(n) \right|
\]
for all \(m\). By Lemmas 5.2 and 5.3, \(\lim_{n \to \infty} A_{m,w}(n)\) and \(\lim_{n} B_{m,w}(n)\) exist, and
\[
\lim_{m \to \infty} \lim_{n \to \infty} |A_{m,w}(n) - B_{m,w}(n)| = 0,
\]
we see that the limit
\[
\lim_{m \to \infty} \lim_{n \to \infty} \sum_{j=1}^{t(m) h_{j+1,m}(n)-1} s_w(k)/k
\]
exists. But this is the same as saying
\[
\lim_{m \to \infty} \lim_{n \to \infty} \sum_{f_1(m+n) \leq j \leq f_1(m+n+1)} s_w(j)/j
\]
eexists, which implies
\[
\lim_{n \to \infty} \sum_{f_1(n) \leq j \leq f_1(n+1)} s_w(j)/j
\]
eexists. We let \(\gamma\) denote this limit. We claim that the logarithmic frequency of the subword \(w\) in the pure morphic sequence \(s\) is \(\gamma\). Let \(\varepsilon > 0\) and pick \(N > 0\) such that
\[
\left| \sum_{j=f_1(n)+1}^{f_1(n+1)} s_w(j)/j - \gamma \right| < \varepsilon
\]
for all \(n \geq N\). Then for \(f_1(M) \leq x < f_1(M + 1)\), we use Proposition 4.1 to infer that there are positive constants \(C_0\) and \(C_1\) such that
\[
C_0 \alpha^k \alpha^n \leq f_1(n) \leq C_1 \alpha^k \alpha^n
\]
for \(n\) sufficiently large. To obtain the estimate
\[
\frac{1}{\log x} \sum_{j \leq x} s_w(j)/j
\]
\[
\leq \frac{1}{\log x} \left( \sum_{i=1}^{f_1(N)} 1/i + \sum_{i=N}^{f_1(i)+1} \sum_{j=f_1(i)+1}^{1/j} 1/j \right)
\]
\[
\leq \frac{1}{M \log x + \log(C_0) + k \log M} \left( \log(f_1(N)) + 1 + (M - N + 1)(\gamma + \varepsilon) \right)
\]
\[
\to \frac{\gamma + \varepsilon}{\log \alpha}
\]
as $M \to \infty$. Similarly,

$$\liminf_{x \to \infty} \frac{1}{\log x} \sum_{1 \leq j \leq x} s_w(j)/j \geq \frac{(\gamma - \varepsilon)}{\log(\alpha)}.$$ 

The result follows. \qed

Acknowledgments

I thank Jean-Paul Allouche for many helpful comments and suggestions.

References