

# Critical exponents of balanced words

Narad Rampersad

Department of Mathematics and Statistics

University of Winnipeg

(joint work with J. Shallit and É. Vandomme)

- ▶ This talk is about two things: repetitions in words and the balance property of words.
- ▶ Much (everything?) is known about these things over a binary alphabet.
- ▶ So we consider larger alphabets.
- ▶ But we need to understand the binary case first.

- ▶ let  $u$  be a finite word
- ▶ the alphabet is usually a finite subset of  $\{0, 1, 2, \dots\}$
- ▶ the length of  $u$  is  $|u|$
- ▶ the number of times the letter  $a$  appears in  $u$  is  $|u|_a$
- ▶ a word  $w$  (finite or infinite) over an alphabet  $A$  is **balanced** if for every  $a \in A$  and every pair  $u, v$  of factors of  $w$  with  $|u| = |v|$  we have

$$||u|_a - |v|_a| \leq 1.$$

- ▶ the word 0020010201 is not balanced since  $|00200|_0 = 4$   
and  $|10201|_0 = 2$
- ▶ the word 01201210210 is balanced

Over the binary alphabet, the class of infinite aperiodic balanced words coincides with the class of Sturmian words (Morse and Hedlund 1940). Sturmian words are first differences of irrational Beatty sequences.

Let  $\alpha$  and  $\rho$  be real numbers with  $0 < \alpha < 1$ . A (slow) Beatty sequence is a sequence of the form

$$(\lfloor n\alpha + \rho \rfloor)_{n \geq 1} \text{ or } (\lceil n\alpha + \rho \rceil)_{n \geq 1}.$$

We consider only the case where  $\alpha$  is irrational.

For  $\alpha = \sqrt{2} - 1$  and  $\rho = 0$  the Beatty sequence  $(\lfloor n\alpha + \rho \rfloor)_{n \geq 1}$  is the sequence

0, 0, 1, 1, 2, 2, 2, 3, 3, 4, 4, 4, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9, 9, 9, ...

The corresponding Sturmian sequence is the sequence of first differences:

0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, ...

Why is the Sturmian word balanced? An integer  $j$  appears in the Beatty sequence if and only if

$$\begin{aligned} j &= \lfloor i\alpha + \rho \rfloor \\ \Leftrightarrow \quad \frac{\rho - j}{\alpha} - \left\lfloor \frac{\rho - j}{\alpha} \right\rfloor &< \frac{1}{\alpha} \\ \Leftrightarrow \quad \left( \frac{\rho - j - 1}{\alpha}, \frac{\rho - j}{\alpha} \right] \cap \mathbb{Z} &\neq \emptyset. \end{aligned}$$



The number of distinct  $j$ 's of the form  $\lfloor i\alpha + \rho \rfloor$  for  $a \leq i < a + n$  is the number of integers in the interval

$$\left( \frac{\rho - a - n}{\alpha}, \frac{\rho - a}{\alpha} \right]$$

which is

$$\left\lfloor \frac{\rho - a}{\alpha} \right\rfloor - \left\lfloor \frac{\rho - a - n}{\alpha} \right\rfloor = \left\lfloor \frac{n}{\alpha} \right\rfloor \text{ or } \left\lceil \frac{n}{\alpha} \right\rceil.$$

- ▶ Any block of  $n$  consecutive terms in a Beatty sequence contains either  $\lfloor n/\alpha \rfloor$  or  $\lceil n/\alpha \rceil$  distinct terms.
- ▶ Any block of  $n - 1$  consecutive terms in the corresponding first difference sequence has either  $\lfloor n/\alpha \rfloor - 1$  or  $\lceil n/\alpha \rceil - 1$  1's.
- ▶ Hence this binary sequence is balanced.

- ▶ We are interested in repetitions in balanced words.
- ▶ Let  $u$  be a finite word and write  $u = u_0u_1 \cdots u_{n-1}$ , where the  $u_i$  are letters.
- ▶ A positive integer  $p$  is a **period** of  $u$  if  $u_i = u_{i+p}$  for all  $i$ .
- ▶ Let  $e = |u|/p$  and let  $z$  be the prefix of  $u$  of length  $p$ .
- ▶  $z$  is a **fractional root** of  $u$ .
- ▶ We say that  $u$  has **exponent**  $e$  and write  $u = z^e$ .
- ▶ e.g.,  $01011010 = (01011)^{8/5}$
- ▶ A **square** (resp. **cube**) is a repetition with exponent 2 (resp. 3)

The **critical exponent** of an infinite word  $w$  is

$$E(w) = \sup\{r \in \mathbb{Q} : \text{there is a finite, non-empty factor of } w \\ \text{with exponent } r\}.$$

For example, the word

$$w = 012021012102012021 \dots$$

obtained by iterating the substitution

$$0 \rightarrow 012, \quad 1 \rightarrow 02, \quad 2 \rightarrow 1$$

contains no squares, but has repetitions with exponents arbitrarily close to 2, so  $E(w) = 2$ .

## Dejean's Theorem

Given an alphabet  $A$  of size  $k$ , the least critical exponent among all infinite words over  $A$  is

$$\begin{cases} 7/4, & k = 3 \\ 7/5, & k = 4 \\ k/(k-1), & k = 2 \text{ or } k \geq 5. \end{cases}$$

However, imposing the balance property makes it much harder to avoid repetitions.

- ▶ To understand the repetitions in a Sturmian word, we need another equivalent definition of Sturmian words.
- ▶ Let  $\alpha$  be an irrational real number between 0 and 1, called the **slope**.
- ▶ Suppose  $\alpha$  has continued fraction expansion  $\alpha = [d_0, d_1, d_2, d_3, \dots]$ .

The **characteristic Sturmian word with slope  $\alpha$**  is the infinite word  $c_\alpha$  obtained as the limit of the sequence of **standard words**  $s_n$  defined by

$$s_0 = 0, \quad s_1 = 0^{d_1-1}1, \quad s_n = s_{n-1}^{d_n} s_{n-2}, \quad n \geq 2.$$

For  $n \geq 2$ , we also define the **semi-standard words**

$$s_{n,t} = s_{n-1}^t s_{n-2},$$

where  $1 \leq t < d_n$ .



One characteristic Sturmian word is of particular significance. Let  $\phi = (1 + \sqrt{5})/2$ . The **Fibonacci word** is the characteristic Sturmian word

$$c_\theta = 0100101001001010010100100100100 \dots$$

with slope  $\theta := 1/\phi^2 = [0, 2, \overline{1}]$ . We call the corresponding standard words the **finite Fibonacci words**:

$$f_0 = 0, \quad f_1 = 01, \quad f_2 = 010, \quad \dots$$

- ▶ the word  $f_i$  has length  $F_{i+2}$  (the  $(i+2)$ -th Fibonacci number) and has  $F_{i+1}$  0's and  $F_i$  1's
- ▶ Mignosi and Pirillo (1992) showed that  $E(c_\theta) = 2 + \phi$ .
- ▶ The more general results of Damanik and Lenz (2002) and Justin and Pirillo (2001) show that this is minimal over all Sturmian words (i.e., all aperiodic balanced binary words).
- ▶ What about balanced words over larger alphabets?

- ▶ An infinite word  $y$  has the **constant gap** property if, for each letter  $a$ , there is some number  $d$  such that the distance between successive occurrences of  $a$  in  $y$  is always  $d$ .
- ▶ This is stronger than being periodic.
- ▶  $(0120)^\omega$  is periodic but is not a constant gap word (contains both 00 and 0120)
- ▶  $(0102)^\omega$  is a constant gap word

## Theorem (Graham 1973; Hubert 2000)

A recurrent aperiodic word  $x$  is balanced if and only if  $x$  is obtained from a Sturmian word  $u$  over  $\{0, 1\}$  by the following procedure: replace the positions containing 0's in  $u$  by a periodic sequence  $y$  with constant gaps over some alphabet  $A$  and replace the positions containing 1's in  $u$  by a periodic sequence  $y'$  with constant gaps over some alphabet  $B$ , disjoint from  $A$ .

e.g., take the Sturmian word

$$u = 0101001010010101001010010101001010010100 \dots,$$

$y = (01)^\omega$  and  $y' = (2324)^\omega$ , then

$$x = 0213012041021302104120130214012031021401 \dots$$

is balanced.

From the construction it is clear that to understand the repetitions in the word  $x$  one has to understand the repetitions in the underlying Sturmian word  $u$ .

A word  $x$  is a **conjugate** of a word  $y$  if we can write  $x = uv$  and  $y = vu$  for some words  $u$  and  $v$ .

## Theorem (Damanik and Lenz 2002)

Let  $w$  be a primitive factor of a Sturmian word  $u$ .

- ▶ If  $w^2$  is a factor of  $u$  then  $w$  is a conjugate of either a standard word  $s_k$  or a semi-standard word  $s_{k,t}$ .
- ▶ If  $w^3$  is a factor of  $u$  then  $w$  is a conjugate of a standard word  $s_k$ .



- ▶ To create a balanced word with repetitions of small exponent, we try to choose the periodic words  $y$  and  $y'$  in such a way that the Hubert construction “breaks up” the repetitions in  $u$ .
- ▶ The analysis depends on understanding the number of 0's and 1's in the standard and semi-standard words modulo the periods of  $y$  and  $y'$ .

## Proposition

There is an infinite ternary balanced word  $x_3$  with critical exponent

$$E(x_3) = 2 + \frac{\sqrt{2}}{2} \approx 2.7071.$$

- ▶ Let  $\alpha = \sqrt{2} - 1 = [0, \overline{2}]$ .
- ▶ Let  $c_\alpha$  be the characteristic Sturmian word with slope  $\alpha$ .
- ▶ So  $c_\alpha$  is the limit of the standard words  $s_k$  defined by

$$s_0 = 0, \quad s_1 = s_0 1, \quad s_k = s_{k-1}^2 s_{k-2}, \quad k \geq 2.$$

Define  $x_3$  by replacing the 0's in  $c_\alpha$  by  $(01)^\omega$  and by replacing the 1's with 2's. We have

$$s_1 = 01, \quad s_2 = 01010, \quad s_3 = (01010)^2 01, \dots$$

and

$$c_\alpha = 01010010100101010010100101010010100101001010010 \dots$$

and

$$x_3 = 021201202102120210212012021201202102120 \dots .$$

- ▶ The critical exponent of  $c_\alpha$  is  $3 + \sqrt{2}$ .
- ▶ Our goal is to show that the critical exponent of  $x_3$  is  $2 + \frac{\sqrt{2}}{2}$ .

- ▶ Let  $(z')^e$  be a repetition of exponent  $e \geq 2$  in  $x_3$  ( $e \in \mathbb{Q}$ ).
- ▶ Apply the morphism that sends  $\{0, 1\} \rightarrow 0$  and  $2 \rightarrow 1$  to  $x_3$ .
- ▶ We see that there is a corresponding repetition  $z^e$  of the same length in  $c_\alpha$ .

0212012 0212012  $\rightarrow$  0101001 0101001

- ▶ Suppose that  $z$  is primitive.
- ▶  $z$  is either a conjugate of one of the standard words  $s_k$  defined above or a conjugate of one of the semi-standard words

$$s_{k,1} = s_{k-1}s_{k-2}, \quad k \geq 2.$$



The lengths of the standard and semi-standard words are given in terms of  $q_n$ , defined by:

$$\frac{p_n}{q_n} = [d_0, d_1, d_2, d_3, \dots, d_n],$$

where

$$p_{-2} = 0, \quad p_{-1} = 1, \quad p_n = d_n p_{n-1} + p_{n-2} \text{ for } n \geq 0;$$

$$q_{-2} = 1, \quad q_{-1} = 0, \quad q_n = d_n q_{n-1} + q_{n-2} \text{ for } n \geq 0.$$

$$|s_n| = q_n \text{ and } |s_{n,1}| = q_{n-1} + q_{n-2}$$

The convergents have the following approximation property:

$$\left| \alpha - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}. \quad (1)$$

The following fact is classical:

$$\frac{q_{n+1}}{q_n} = [d_{n+1}, d_n, \dots, d_1]. \quad (2)$$

- ▶ We return to a repetition  $(z')^e$  in  $x_3$  and the corresponding  $z^e$  in  $c_\alpha$ .
- ▶ Suppose that  $z$  is a conjugate of a standard word  $s_k$ .
- ▶ Note that  $|s_k|_0$  is odd for every  $k \geq 1$ .
- ▶ Hence  $|z|_0$  is odd.
- ▶ Recall:  $z'$  is obtained by replacing each 0 in  $z$  with 0 or 1 alternately and the 1's with 2's.
- ▶ It follows that  $z'z'$  cannot occur in  $x_3$  so there is no repetition  $(z')^e$  in  $x_3$ .

010100101001 010100101001  $\rightarrow$  021201202102 120210212012

- ▶ Now suppose that  $z$  is a conjugate of a semi-standard word.
- ▶ Then  $|z| = q_{k-2} + q_{k-1}$  for some  $k \geq 2$ .
- ▶ Justin and Pirillo (2001) gave precise technical results about the exponents of certain kinds of repetitions in Sturmian words.
- ▶ One of these results is that the longest factor of  $c_\alpha$  with this period has length  $2(q_{k-2} + q_{k-1}) + q_{k-1} - 2$ .

It follows that for a repetition  $z^e$  where  $z$  is a conjugate of a semi-standard word,

$$\begin{aligned} e &\leq \frac{2(q_{k-2} + q_{k-1}) + q_{k-1} - 2}{q_{k-2} + q_{k-1}} \\ &= 2 + \frac{q_{k-1} - 2}{q_{k-2} + q_{k-1}} \\ &= 2 + \frac{q_{k-1}/q_{k-2} - 2/q_{k-2}}{1 + q_{k-1}/q_{k-2}}. \end{aligned}$$

Now by (2) we have that  $q_{k-1}/q_{k-2}$  converges to  $[2, \bar{2}] = \sqrt{2} + 1$ , and by (1) we have

$$q_{k-1}/q_{k-2} < \sqrt{2} + 1 + 1/q_{k-2}^2.$$



Thus, we have

$$e < 2 + \frac{\sqrt{2} + 1 + 1/q_{k-2}^2 - 2/q_{k-2}}{\sqrt{2} + 2 - 1/q_{k-2}^2}.$$

The fraction on the right clearly tends to

$(\sqrt{2} + 1)/(\sqrt{2} + 2) = \sqrt{2}/2$  as  $k \rightarrow \infty$ , and is increasing for  $k \geq 3$ , so the convergence is from below. Thus  $e < 2 + \sqrt{2}/2$ .

Indeed, for every  $k \geq 2$ , there are repetitions  $z^e$  in the Sturmian word  $c_\alpha$  with exponent

$$e = 2 + \frac{(q_{k-1} - 2)/q_{k-2}}{1 + q_{k-1}/q_{k-2}} \xrightarrow{k \rightarrow \infty} 2 + \frac{\sqrt{2}}{2},$$

where the convergence is from below.

Now if  $z$  is the conjugate of a semi-standard word  $s_{k,1}$ , we note that  $|s_{k,1}|_0$  is even for every  $k \geq 2$  and so every such repetition  $z^e$  in  $c_\alpha$  gives rise to a repetition  $(z')^e$  in  $x_3$ , since  $|z|_0$  in this case is even.

0101001 0101001 010  $\rightarrow$  0212012 0212012 021

- ▶ Finally, suppose that  $z^e$  is a repetition where  $z$  is not primitive.
- ▶ From existing results we can calculate that the critical exponent of  $c_\alpha$  is  $3 + \sqrt{2}$ .
- ▶ So  $z$  cannot have exponent  $\geq 3$ .
- ▶ Thus  $z$  is a square and we have

$$e < \frac{3 + \sqrt{2}}{2} < 2 + \frac{\sqrt{2}}{2}.$$

- ▶ We conclude that  $E(x_3) = 2 + \frac{\sqrt{2}}{2}$ .

For larger alphabets, we construct balanced words according to this table.

$k$	$\alpha$	c.f.	$y$	$y'$
3	$\sqrt{2} - 1$	$[0, \overline{2}]$	$(01)^\omega$	$2^\omega$
4	$1/\phi^2$	$[0, 2, \overline{1}]$	$(01)^\omega$	$(23)^\omega$
5	$\sqrt{2} - 1$	$[0, \overline{2}]$	$(0102)^\omega$	$(34)^\omega$
6	$(78 - 2\sqrt{6})/101$	$[0, 1, 2, 1, 1, \overline{1, 1, 1, 2}]$	$0^\omega$	$(123415321435)^\omega$
7	$(63 - \sqrt{10})/107$	$[0, 1, 1, 3, \overline{1, 2, 1}]$	$(01)^\omega$	$(234526432546)^\omega$
8	$(23 + \sqrt{2})/31$	$[0, 1, 3, 1, \overline{2}]$	$(01)^\omega$	$(234526732546237526432576)^\omega$
9	$(23 - \sqrt{2})/31$	$[0, 1, 2, 3, \overline{2}]$	$(01)^\omega$	$(234567284365274863254768)^\omega$
10	$(109 + \sqrt{13})/138$	$[0, 1, 4, 2, \overline{3}]$	$(01)^\omega$	$(234567284963254768294365274869)^\omega$

**Table:** Periodic words  $y$  and  $y'$  for the construction of  $x_k$

We have shown

$$E(x_3) = 2 + \frac{\sqrt{2}}{2} \approx 2.7071$$

We can also prove that

$$E(x_4) = 1 + \frac{\phi}{2} \approx 1.8090.$$

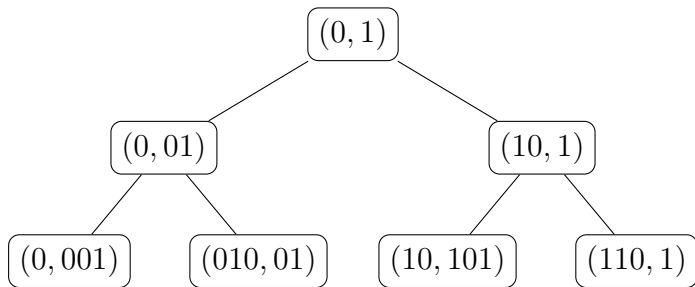
For  $k \geq 5$ , computer calculations suggest that

$$E(x_k) = \frac{k-2}{k-3}.$$

For values of  $k$  up to 9 (except for 4), we can show that these critical exponents are the smallest possible over  $k$  letters.

- ▶ We show optimality by a backtracking computer search over the **tree of standard pairs**.
- ▶ This is the tree of pairs with root  $(0, 1)$  and each vertex  $(u, v)$  has children  $(u, uv)$  and  $(vu, v)$ .





- ▶ All finite balanced binary words appear in this tree.
- ▶ As we search, we take the longer of  $u$ ,  $v$  and try all possible replacements of the 0's and 1's with constant gap sequences.
- ▶ If every replacement results in a word with too large a critical exponent, we backtrack.

- ▶ Why can't we prove the claimed critical exponents for  $k \geq 5$ ?
- ▶ We have less information for repetitions  $z^e$  in Sturmian words that have exponent  $e < 2$ .
- ▶ We need to know the number of 0's and 1's in  $z$  modulo the periods of  $y$  and  $y'$  respectively.
- ▶ If  $e \geq 2$ , then  $z$  is an integer power of a conjugate of a standard or semi-standard word and we can count the number of 0's and 1's in such words.

- ▶ But if  $e < 2$  then we cannot be sure that  $z$  is an integer power of a conjugate of a standard or semi-standard word.
- ▶ Saari showed that this is the case if  $z$  is the **minimal fractional root** of the repetition.
- ▶ But we have no information about non-minimal periods of factors of Sturmian words.
- ▶ So the main open problem is to prove the conjectured critical exponents for  $k \geq 5$ .

We can overcome this difficulty for the 4-letter alphabet with the following technical lemma, but this approach fails for larger alphabets.

## Lemma

Let  $w$  be a factor of the Fibonacci word. Write  $w = x^f$ , where  $f \in \mathbb{Q}$  and  $|x|$  is the least period of  $w$ . Suppose that  $w$  has another representation  $w = z^e$ , where  $e \in \mathbb{Q}$ ,  $z$  is primitive, and  $|x| < |z|$ . Then  $e < 1 + \phi/2$ .

So the main open problem is to prove the conjectured critical exponents for  $k \geq 5$ .

# The End