Periods and Borders of Random Words*

Štěpán Holub[1] and Jeffrey Shallit[2]

1 **Department of Algebra, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic**
`holub@karlin.mff.cuni.cz`

2 **School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1 Canada**
`shallit@cs.uwaterloo.ca`

Abstract

We investigate the behavior of the periods and border lengths of random words over a fixed alphabet. We show that the asymptotic probability that a random word has a given maximal border length k is a constant, depending only on k and the alphabet size ℓ. We give a recurrence that allows us to determine these constants with any required precision. This also allows us to evaluate the expected period of a random word. For the binary case, the expected period is asymptotically about $n - 1.641$. We also give explicit formulas for the probability that a random word is unbordered or has maximum border length one.

1998 ACM Subject Classification G.2.1 Combinatorics

Keywords and phrases random word, period, word border

Digital Object Identifier 10.4230/LIPIcs.STACS.2016.44

1 Introduction and Notation

A *word* is a finite sequence of letter chosen from a finite alphabet Σ. The periodicity of words is a classical and well-studied topic in both discrete mathematics and combinatorics on words, starting with the classic paper of Fine and Wilf [4] and continuing with the works of Guibas and Odlyzko [6, 7, 5]. For more recent work, see, for example, [8, 15, 12].

We say that a word w has period p if $w[i] = w[i + p]$ for all i that make the equation meaningful. (If $|w| = n$ and one indexes beginning at position 1, this would be for $1 \le i \le n - p$.) Trivially every word of length n has all periods of length $\ge n$, so we restrict our attention to periods $\le n$. The least period is sometimes called *the* period. For example, the French word `entente` has periods 3, 6, and 7.

Empirically, one quickly discovers that a randomly chosen word typically has a least period that is very close to its length. This readily follows from the fact that the number of words over a given alphabet grows exponentially as the length increases. It can also be seen as a particular case of the fact that most strings are not compressible.

In this paper, we quantify this basic observation and show that the expected least period of a string of length n over an ℓ-letter alphabet is $n - \alpha_\ell(n)$, where $\alpha_\ell(n)$ is $\mathcal{O}(1)$.

Another concept frequently studied in formal language theory is that of *border* of a word [13, 3, 14]. A word x has border w if w is both a prefix and a suffix of x. Normally we do not consider the trivial borders of length 0 or $n = |w|$. Thus, for example, the English word

* This work was partially supported by Czech Science Foundation grant number 13-01832S.

33rd Symposium on Theoretical Aspects of Computer Science (STACS 2016).
Editors: Nicolas Ollinger and Heribert Vollmer; Article No. 44; pp. 44:1–44:10
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

`ionization` has one border: `ion`. Less trivially, the word `alfalfa` has two borders: `a` and `alfa`. A word with no borders is *unbordered*.

There is an obvious connection between periods of a word and its borders: if w has a period p, then it has a border of length $|w| - p$. For example, the English word `abracadabra`, of length 11, has periods 7, 10, and 11, while it has borders of length 1 and 4.

Consequently, the least period of a word corresponds to the length of the longest border (and an unbordered word corresponds to least period n, the length of the word). The reader should be constantly aware of this duality, since it is often useful and more natural to think about periods in terms of borders. This can be seen from the announced result: it is more compact to speak directly about the expected maximum border length, which is $\alpha_\ell(n)$.

If P is a set of integers, we shall write $n - P$ for $\{n - p \mid p \in P\}$, and $P - n$ for $\{p - n \mid p \in P\}$.

By $\mathrm{pref}_i(v)$, we mean the prefix of length i of the word v.

2 Multiperiodic Words and the Average Border Length

We shall obtain our results by counting words with a given length n and a given finite set of periods $P \subseteq \{1, 2, \dots n\}$, or equivalently, with a given set of border lengths $n - P$. For technical reasons, in order to be able to deal with unbordered words, we shall always suppose that $n \in P$, that is, we shall say that every word has a border length zero.

There are two basic types of requirements. Let

$$\mathcal{G}_\ell(P, n) = \{w \in \Sigma_\ell^n \mid \text{for each } p \text{ in } P,\ p \text{ is a period of } w\},$$

and let $G_\ell(P, n)$ be the cardinality of $\mathcal{G}_\ell(P, n)$. Similarly, let

$$\mathcal{F}_\ell(P, n) = \{w \in \mathcal{G}_\ell(P, n) \mid \min P \text{ is the least period of } w\},$$

and let $F(P, n)$ be the cardinality of $\mathcal{F}_\ell(P, n)$.

Words with many periods have been amply studied. In particular, there is a fast algorithm constructing a word of length n with periods P and maximal possible number of letters. Such a word, called an FW-word in the literature, is unique up to renaming of the letters. Let $\mathfrak{c}(P, n)$ denote the cardinality of the alphabet of the FW-word of length n and periods P.

▶ **Example 1.** Let $P = \{p, q\}$ and $d = \gcd(p, q)$. The well-known periodicity lemma (often called the Fine and Wilf theorem, which is the origin of the term FW-word) states that if a word of length at least $p + q - d$ has periods p and q, then it also has period d. Moreover, the bound $p + q - d$ is sharp; for all $p, q \geq 1$ there are words of length $p + q - d - 1$ with period p and q but not period d. This can be stated, using the just-introduced terminology, by the two assertions $\mathfrak{c}(\{p, q\}, p + q - d) = d$ and $\mathfrak{c}(\{p, q\}, p + q - d - 1) > d$.

The number $\mathfrak{c}(P, n)$ can be computed and the corresponding FW-word constructed using the algorithm of Tijdeman and Zamboni [16] (see [17] for an alternative presentation). The computation is summarized by the following formula:

$$\mathfrak{c}(P, n) = \begin{cases} 1, & \text{if } m = 1; \\ n, & \text{if } m \geq n; \\ \mathfrak{c}(Q, n - m), & \text{if } 2m \leq n; \\ \mathfrak{c}(Q, n - m) + 2m - n, & \text{if } m < n < 2m; \end{cases}$$

where $m = \min P$ and $Q = (P - m) \setminus \{0\} \cup \{m\}$.

Since each word having the periods in P (and possibly others) results from a coding (a letter-to-letter mapping) of the corresponding FW-word, we obtain

$$G_\ell(P, n) = \ell^{\mathfrak{c}(P,n)},$$

which is the starting point of our computation.

Note that $\mathcal{F}_\ell(\{p\}, n)$ is the set of words from Σ_ℓ^n having least period p. Equivalently, $\mathcal{F}_\ell(\{n-r\}, n)$ is the set of words with the longest border of length r. For $0 \le r < n$, let

$$\lambda_\ell(r, n) = \frac{F(\{n-r\}, n)}{\ell^n}$$

denote the relative number of such words. Our goal is to compute

$$\alpha_\ell(n) = \sum_{r=0}^{n-1} r \cdot \lambda_\ell(r, n),$$

which is the expected maximum border length for words in Σ_ℓ^n. We first show that this quantity converges as n approaches infinity. This fact was recently independently proved in [1, Appendix].

▶ **Lemma 2.** *For each $\ell \ge 2$ and each $0 \le r < n$,*

$$|\lambda_\ell(r, n+1) - \lambda_\ell(r, n)| \le \frac{1}{\ell^{\lfloor n/2 \rfloor}}.$$

Proof. Case 1: $r \ge \lfloor n/2 \rfloor$. Then

$$|\lambda_\ell(r, n+1) - \lambda_\ell(r, n)| = \left| \frac{F_\ell(\{n+1-r\}, n+1)}{\ell^{n+1}} - \frac{F_\ell(\{n-r\}, n)}{\ell^n} \right|.$$

Recall that $F_\ell(\{n+1-r\}, n+1)$ (resp., $F_\ell(\{n-r\}, n)$) counts the words with longest border length r from Σ_ℓ^{n+1} (resp., Σ_ℓ^n). First, note that $F_\ell(\{p\}, n) \le \ell^p$ for any p and n. This implies

$$|\lambda_\ell(r, n+1) - \lambda_\ell(r, n)| \le \frac{1}{\ell^r}$$

and we are done.

Case 2: $r < \lfloor n/2 \rfloor$. There is a useful correspondence between Σ_ℓ^n and Σ_ℓ^{n+1}, given by the insertion of a letter in the middle of the shorter word. The basic observation, already used in [9, 11], is that this insertion does not influence borders of length at most $\lfloor n/2 \rfloor$. Define

$$\mathcal{F} = \mathcal{F}_\ell(\{n+1-r\}, n+1),$$
$$\mathcal{B} = \{w_1 a w_2 \mid a \in \Sigma_\ell,\ |w_1| = \lfloor n/2 \rfloor,\ |w_2| = \lceil n/2 \rceil,\ w_1 w_2 \in \mathcal{F}_\ell(\{n-r\}, n)\}.$$

Then $|\mathcal{B}| = \ell \cdot F_\ell(\{n-r\}, n)$. Let $w \in \mathcal{F} \setminus \mathcal{B}$ and write $w = w_1 a w_2$ with $a \in \Sigma_\ell$, $|w_1| = \lfloor n/2 \rfloor$, and $|w_2| = \lceil n/2 \rceil$. The words w and $w_1 w_2$ have the same borders up to length $\lfloor n/2 \rfloor$. Since $w_1 w_2 \notin \mathcal{F}_\ell(\{n-r\}, n)$, we deduce that $w_1 w_2$ has a border of length at least $\lfloor n/2 \rfloor + 1$, that is, a period at most $\lceil n/2 \rceil - 1$. This implies

$$|\mathcal{F} \setminus \mathcal{B}| \le \ell \cdot \sum_{j=0}^{\lceil n/2 \rceil - 1} \ell^j < \ell^{\lceil n/2 \rceil + 1}. \tag{1}$$

Similarly, a word $w \in \mathcal{B} \setminus \mathcal{F}$ has period at most $\lceil n/2 \rceil$, and so

$$|\mathcal{B} \setminus \mathcal{F}| \leq \sum_{j=0}^{\lceil n/2 \rceil} \ell^j < \ell^{\lceil n/2 \rceil + 1}. \tag{2}$$

We thus obtain

$$|\lambda_\ell(r, n+1) - \lambda_\ell(r, n)| = \frac{1}{\ell^{n+1}} \Big| |\mathcal{B} \setminus \mathcal{F}| - |\mathcal{F} \setminus \mathcal{B}| \Big| < \frac{1}{\ell^{\lfloor n/2 \rfloor}}.$$ ◀

▶ **Theorem 3.** *For each $\ell \geq 2$ and $r \geq 0$, the limits*

$$\alpha_\ell := \lim_{n\to\infty} \alpha_\ell(n) \quad \text{and} \quad \lambda_\ell(r) = \lim_{n\to\infty} \lambda_\ell(r, n)$$

exist. Furthermore, the convergence is exponential.

Proof. Follows directly from the definition of $\alpha_\ell(n)$ and Lemma 2. ◀

3 Recurrences

From the estimates of the previous section, we know that $\alpha_\ell(n)$ and $\lambda_\ell(r, n)$ both converge quickly to α_ℓ and $\lambda_\ell(r)$, respectively. Thus, they can be estimated to a few digits by explicit enumeration (see [1, Appendix] for α_2 and α_3).

In order to evaluate $\alpha_\ell(n)$ to dozens of decimal places, however, we need a more efficient way to calculate $F_\ell(\{p\}, n)$. This can be done using the recurrence formulas that we derive below. They are reformulations and generalizations of formulas given by Harborth [9] for sets of periods.

We first prove the following auxiliary claim.

▶ **Lemma 4.** *Let a word w have a period $p < |w|$ and let u be the prefix of w of length $|w| - p$. Then w has a period $q > p$ if and only if u has a period $q - p$.*

Proof. Note that u is a border of w. The following conditions are easily seen to be equivalent:

- w has a period q,
- w has a border of length $|w| - q$,
- u has a border of length $|w| - q$,
- u has a period $|u| - (|w| - q)$.

Since $|u| - (|w| - q) = (|w| - p) - (|w| - q) = q - p$, the proof is completed. ◀

▶ **Theorem 5.** *Let P be a set of periods with $m = \min P$ and $\max P < n$. Then*

$$F_\ell(P, n) = G_\ell(P, n) - \sum_{p=\lceil m/2 \rceil}^{m-1} H_\ell(P, p, n)\,, \tag{3}$$

where

$$H_\ell(P, p, n) := \begin{cases} F_\ell\big((P - p) \cup \{p\}, n - p\big), & \text{if } p < \lceil n/2 \rceil; \\ \ell^{2p-n} \cdot F_\ell(P - p, n - p), & \text{if } p \geq \lceil n/2 \rceil\,. \end{cases} \tag{4}$$

Proof. From $G_\ell(P, n)$ we have to subtract the number of words from Σ_ℓ^n that have periods P but also a period smaller than m. We define, for each $1 \leq p < m$, the set

$$\mathcal{H}_\ell(P, p, n) = \{w \in \Sigma_\ell^n \mid w \text{ has periods } P \cup \{p\}, \text{and no period } p' \text{ with } p < p' < m\}\,.$$

If $p < \lceil m/2 \rceil$ then $\mathcal{H}_\ell(P,p,n)$ is empty, since a word $w \in \mathcal{H}_\ell(P,p,n)$ also has a period $2p$, and $p < 2p < m$ contradicts the definition of $\mathcal{H}_\ell(P,p,n)$. Moreover, the sets $\mathcal{H}_\ell(P,p,n)$ are pairwise disjoint, and

$$\mathcal{G}_\ell(P,n) \setminus \mathcal{F}_\ell(P,n) = \bigcup_{p=\lceil m/2 \rceil}^{m-1} \mathcal{H}_\ell(P,p,n)\,.$$

It remains to show that $H_\ell(P,p,n)$ is the cardinality of $\mathcal{H}_\ell(P,p,n)$ for each $\lceil m/2 \rceil \le p < m-1$.

Let $p < \lceil n/2 \rceil$. We claim that $w \mapsto \mathrm{pref}_{n-p} w$ is a one-to-one mapping of $\mathcal{H}_\ell(P,p,n)$ to $\mathcal{F}_\ell\big((P-p) \cup \{p\}, n-p\big)$. Let $w \in \mathcal{H}_\ell(P,p,n)$. By Lemma 4, the word $\mathrm{pref}_{n-p} w$ has periods $P - p$ and no period $p' - p$ with $p < p' < m$, that is, no period less than $m - p$. Since $m - p = \min\big((P-p) \cup \{p\}\big)$ and since $\mathrm{pref}_{n-p} w$ also has a period p, we have $\mathrm{pref}_{n-p} w \in \mathcal{F}_\ell\big((P-p)\cup\{p\}, n-p\big)$. Similarly, one can verify that if $v \in \mathcal{F}_\ell\big((P-p)\cup\{p\}, n-p\big)$, then $w_v := (\mathrm{pref}_p v)^{n/p} \in \mathcal{H}_\ell(P,p,n)$ and $\mathrm{pref}_{n-p} w_v = v$.

Let $p \ge \lceil n/2 \rceil$. Again, using Lemma 4, it is straightforward to verify that

$$\mathcal{H}_\ell(P,p,n) = \{vuv \mid v \in \mathcal{F}_\ell(P-p, n-p), u \in \Sigma_\ell^{2p-n}\}\,.$$ ◀

If $\min P$ is small, then we can formulate a more explicit formula that uses the Möbius μ-function.

▶ **Lemma 6.** *Let P be a set of periods with $m = \min P \le \lfloor n/2 \rfloor + 1$. Then*

$$F_\ell(P,n) = \sum_{d|m} \mu\left(\frac{m}{d}\right) G_\ell\left(P \cup \{d\}, n\right). \tag{5}$$

Proof. Let w be a word of length n with a period m and let p be the least period of w. Then, by the periodicity lemma, we have that p divides m, since $p < m$ implies $p + m - 1 \le n$. Therefore, for each divisor p of m,

$$G_\ell\left(P \cup \{p\}, n\right) = \sum_{d|p} F_\ell(P \cup \{d\}, n),$$

and the claim follows from Möbius inversion. ◀

4 Explicit Formulas

In this section we derive explicit formulas for $\lambda_\ell(0)$ and $\lambda_\ell(1)$, which are the asymptotic probabilities that a random word is unbordered, or has longest border of length one, respectively. These are two cases in which Theorem 5 yields a relatively simple expression, since $\lceil m/2 \rceil \ge \lfloor n/2 \rfloor$.

4.1 Unbordered Words

The number of unbordered words satisfies a well known recurrence formula (see, e.g., [9, p. 143, Eq. (34)] for the binary case and [11] for the general case). The formula can be verified using Theorem 5 but we shall give an elementary proof. In this section, let u_n denote $F_\ell(\{n\}, n)$, and let $t(n)$ denote $\lambda_\ell(0,n)$.

▶ **Theorem 7.**

$$u_n = \begin{cases} \ell, & \text{if } n = 1; \\ \ell(\ell-1) & \text{if } n = 2; \\ \ell \cdot u_{n-1}, & \text{if } n \geq 3 \text{ is odd;} \\ \ell \cdot u_{n-1} - u_{n/2}, & \text{if } n \geq 4 \text{ is even.} \end{cases}$$

Proof. For $k = 1, 2$, the verification is straightforward. Let x and y be nonempty words with $|x| = |y|$ and consider words xy, xay and $xaby$ where a and b are letters.

Since the shortest border of xay has length at most $|x|$, the word xy is unbordered if and only if xay is. This proves $u_n = \ell \cdot u_{n-1}$ if n is odd.

On the other hand, $xaby$ can have the shortest border of length $|x| + 1$. Therefore, $xaby$ is unbordered if and only if (i) xy is unbordered and (ii) $xa \neq by$. Since the shortest border is itself unbordered, we obtain $u_n = \ell^2 \cdot u_{n-2} - u_{n/2} = \ell \cdot u_{n-1} - u_{n/2}$ if n is even. ◀

Theorem 7 directly yields, for each $n \geq 1$,

$$t(2n+1) = t(2n) = t(2n-1) - t(n)\ell^{-n} .$$

Therefore

$$t(2n) = t(1) + \sum_{i=2}^{2n} (t(i) - t(i-1)) = 1 - \sum_{j=1}^{n} t(j)\ell^{-j} .$$

Defining the generating function $L_0(x) = \sum_{n \geq 1} t(n)x^n$, we get

$$\lim_{n\to\infty} \lambda_\ell(0, n) = 1 - L_0\left(\frac{1}{\ell}\right) . \tag{6}$$

The next step is to obtain a functional equation for $L_0(x)$:

$$\begin{aligned} L_0(x)(1-x) &= t(1)x + \sum_{k\geq 2} (t(k) - t(k-1))x^k = \\ &= t(1)x + \sum_{j \geq 1} (t(2j) - t(2j-1))x^{2j} = \\ &= t(1)x - \sum_{j\geq 1} t(j)\ell^{-j}x^{2j} = x - L_0(x^2/\ell) . \end{aligned}$$

Therefore

$$L_0(x) = \frac{x}{1-x} - \frac{L_0(x^2/\ell)}{1-x} .$$

Successively substituting $x = 1/\ell$, $x = 1/\ell^3$, $x = 1/\ell^7$, ..., we get

$$\begin{aligned} L_0\left(\frac{1}{\ell}\right) &= \frac{1}{\ell - 1} - \left(1 + \frac{1}{\ell-1}\right) L_0\left(\frac{1}{\ell^3}\right), \\ L_0\left(\frac{1}{\ell^3}\right) &= \frac{1}{\ell^3 - 1} - \left(1 + \frac{1}{\ell^3-1}\right) L_0\left(\frac{1}{\ell^7}\right), \\ &\vdots \\ L_0\left(\frac{1}{\ell^{2^i-1}}\right) &= \frac{1}{\ell^{2^i-1} - 1} - \left(1 + \frac{1}{\ell^{2^i-1}-1}\right) L_0\left(\frac{1}{\ell^{2^{i+1}-1}}\right) . \end{aligned}$$

Since it is easy to see that

$$\lim_{n\to\infty} \frac{1}{\ell^{2^n-1}-1} \prod_{i=1}^{n-1} \left(1 + \frac{1}{\ell^{2^i-1}-1}\right) L_0\left(\frac{1}{\ell^{2^n-1}}\right) = 0\,,$$

we obtain

$$L_0\left(\frac{1}{\ell}\right) = \sum_{n\geq 1} \left(\frac{(-1)^{n+1}}{\ell^{2^n-1}-1} \prod_{i=1}^{n-1} \left(1 + \frac{1}{\ell^{2^i-1}-1}\right)\right).$$

A similar analysis was given previously by [2], although our analysis is slightly cleaner.

4.2 Words With Longest Border of Length One

There is also a relatively simple recurrence for $F_\ell(\{n-1\}, n)$, that is, for words with the longest border of length 1. The particular case $\ell = 2$ was previously given by Harborth [9, p. 143, Eq. (36)]. In this section, we let v_n denote $F_\ell(\{n-1\}, n)$, and let $s(n)$ denote $\lambda_\ell(1, n)$.

▶ **Theorem 8.**

$$v_n = \begin{cases} 0, & \textit{if } n = 1; \\ \ell & \textit{if } n = 2; \\ \ell \cdot v_{n-1} - v_{(n+1)/2}, & \textit{if } n \geq 3 \textit{ is odd}; \\ \ell \cdot v_{n-1} - (\ell-1)v_{n/2}, & \textit{if } n \geq 4 \textit{ is even.} \end{cases}$$

Proof. Verify that $v_1 = 0$ and $v_2 = \ell$, and let x and y be nonempty words with $|x| = |y|$. Consider words $cxyc$, $cxayc$ and $cxabyc$ where a, b, c are (not necessarily distinct) letters.

The letter c is the longest border of the word $cxayc$ if and only if (i) c is the longest border of $cxyc$ and (ii) $cxa \neq ayc$. Moreover, (i') c is the shortest border of $cxyc$, and (ii') $cxa = ayc$ $(= cxc)$ if and only if c is the shortest border of cxc. This implies $v_n = \ell \cdot v_{n-1} - v_{(n+1)/2}$ for $n \geq 3$ odd.

Similarly, c is the shortest border of $cxabyc$ if and only if (i) c is the longest border of $cxyc$ and (ii) $cxa \neq byc$. As above, we have to subtract the number of words cxc with the longest border c. It follows that $v_n = \ell^2 \cdot v_{n-2} - v_{n/2} = \ell v_{n-1} + (\ell-1)v_{n/2}$ for $n \geq 4$ even. ◀

From Theorem 8, we deduce

$$s(2n) - s(2n-2) = -s(n)\ell^{-n}, \qquad n \geq 2, \tag{7}$$
$$s(2n) - s(2n-1) = (\ell-1)s(n)\ell^{-n}, \qquad n \geq 2, \tag{8}$$
$$s(2n+1) - s(2n) = -s(n+1)\ell^{-n}, \qquad n \geq 1\,. \tag{9}$$

Using (7), we obtain

$$s(2n) = s(2) + \sum_{r=j}^{n}(s(2j) - s(2j-2)) = 1/\ell - \sum_{j=1}^{n} s(j)\ell^{-j}\,.$$

Defining the generating function $L_1(x) = \sum_{k\geq 1} s(k)x^k$, we then get

$$\lambda_\ell(1) = \frac{1}{\ell} - L_1\left(\frac{1}{\ell}\right).$$

A functional equation for L_1 is obtained as follows:

$$\begin{aligned}L_1(x)(1-x) &= s(1)x + \sum_{k\geq 2}(s(k)-s(k-1))x^k = \\ &= \frac{1}{\ell}x^2 + \sum_{i\geq 1}(s(2i+1)-s(2i))x^{2i+1} + \sum_{i\geq 2}(s(2i)-s(2i-1))x^{2i} = \\ &= \frac{1}{\ell}x^2 - \sum_{i\geq 1} s(i+1)\ell^{-i}x^{2i+1} + \sum_{i\geq 2}(\ell-1)s(i)\ell^{-i}x^{2i} = \\ &= \frac{1}{\ell}x^2 - \frac{\ell}{x}L_1\left(\frac{x^2}{\ell}\right) + (\ell-1)L_1\left(\frac{x^2}{\ell}\right).\end{aligned}$$

We have

$$L_1(x) = \frac{x^2}{\ell(1-x)} + \frac{\ell-1-\ell/x}{1-x}L_1\left(\frac{x^2}{\ell}\right),$$

and

$$L_1\left(\frac{1}{\ell^{2^i-1}}\right) = \frac{1}{\ell^{2^i}(\ell^{2^i-1}-1)} - \ell^{2^i-1}\frac{\ell^{2^i}-\ell+1}{\ell^{2^i-1}-1}L_1\left(\frac{1}{\ell^{2^{i+1}-1}}\right).$$

From here, we deduce

$$L_1\left(\frac{1}{\ell}\right) = \sum_{n\geq 1}\frac{(-1)^{n+1}}{\ell^{n+1}}\prod_{i=1}^{j}\frac{\ell^{2^{i-1}}-\ell+1}{\ell^{2^i-1}-1}.$$

We do not know how to obtain similar expressions for other border lengths.

5 Particular Values

Theorem 5, as well as explicit formulas from the previous section allow fast computer evaluation of $\alpha_\ell(n)$ and $\lambda_\ell(r,n)$ for large n, and therefore also evaluation of $\lambda_\ell(r)$ and α_ℓ with high precision. We list some rounded values in the following tables.

ℓ	α_ℓ
2	1.64116491178296695613
3	0.68587617299708343978
4	0.42195659003603599699
5	0.30201601806282253073
y 10	0.12233344445364555354
50	0.02081648979722449000

r	$\lambda_2(r)$
0	0.26778684021788911238
1	0.30042007151830329926
2	0.19891874779036456415
3	0.11216079483159432642
5	0.03044609816129782975
10	0.00097577734413168807

And some values of $\lambda_\ell(r)$ rounded to four decimal digits:

$\lambda_\ell(r)$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 10$
$r = 0$	0.55698	0.68775	0.76006	0.89000
$r = 1$	0.28270	0.23024	0.19034	0.09890
$r = 2$	0.10547	0.06126	0.03961	0.00999
$r = 3$	0.03641	0.01555	0.00798	0.00100

For example, we see that a long binary word chosen randomly has about 27% chance to be unbordered. A bit more probable, at 30%, is that such a word will have its longest border of

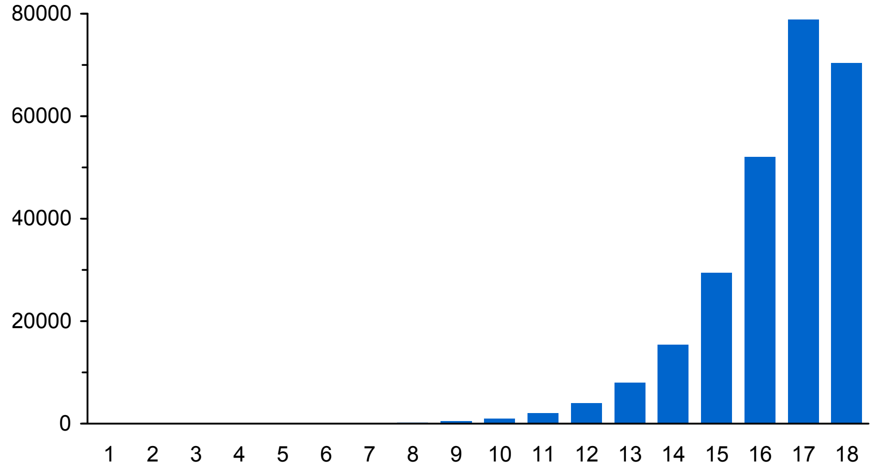

Figure 1 Distribution of lengths of shortest period for binary words of length 18.

length one. Over a five-letter alphabet, more than three words out of four are unbordered, on average.

Figure 1 shows the distribution of lengths of the shortest period for binary words of length $n = 18$.

Our original motivation was a question about the average period of a binary word. The answer is, that the border of a binary word has asymptotically constant expected length, namely

$$\alpha_2 \doteq 1.64116491178296695612774416940082554065953687825771543\ldots .$$

6 Final Remarks

Recently there has been some interest in computing the expected value of the largest unbordered factor of a word [10]. This is a related, but seemingly much harder, problem.

References

1 A. Alatabbi, A. S. S. Islam, M. S. Rahman, J. Simpson, and W. F. Smyth. Enhanced covers of regular & indeterminate strings using prefix tables. *CoRR*, abs/1506.06793, 2015. URL: http://arxiv.org/abs/1506.06793.
2 G. Blom. Overlapping binary sequences (problem 94-20). *SIAM Review*, 37:619–620, 1995.
3 A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. *Discrete Math.*, 26:101–109, 1979.
4 N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.*, 16:109–114, 1965.
5 L. Guibas. Periodicities in strings. In A. Apostolico and Z. Galil, editors, *Combinatorial Algorithms on Words*, volume F12 of *NATO ASI*, pages 257–269. Springer-Verlag, 1985.
6 L. J. Guibas and A. M. Odlyzko. Periods in strings. *J. Combin. Theory. Ser. A*, 30:19–42, 1981.
7 L. J. Guibas and A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory. Ser. A*, 30:183–208, 1981.
8 V. Halava, T. Harju, and L. Ilie. Periods and binary words. *J. Combin. Theory. Ser. A*, 89:298–303, 2000.

9 H. Harborth. Endliche 0-1–Folgen mit gleichen Teilblöcken. *Journal für die reine und angewandte Mathematik*, 271:139–154, 1974.

10 A. Loptev, G. Kucherov, and T. Starikovskaya. On maximal unbordered factors. In F. Cicalese et al., editor, *CPM 2015*, volume 9133 of *Lecture Notes in Computer Science*, pages 343–354. Springer-Verlag, 2015.

11 P. T. Nielsen. A note on bifix-free sequences. *IEEE Trans. Inform. Theory*, IT-19:704–706, 1973.

12 E. Rivals and S. Rahmann. Combinatorics of periods in strings. *J. Combin. Theory. Ser. A*, 104:95–113, 2003.

13 D. M. Silberger. Borders and roots of a word. *Progress in Mathematics*, 30:191–199, 1971.

14 D. M. Silberger. How many unbordered words? *Comment. Math. Prace Mat.*, 22:143–145, 1980.

15 R. Tijdeman and L. Zamboni. Fine and Wilf words for any periods. *Indag. Math.*, 14:135–147, 2003.

16 R. Tijdeman and L. Q. Zamboni. Fine and Wilf words for any periods II. *Theor. Comput. Sci.*, 410(30-32):3027–3034, 2009.

17 Š. Holub. On an algorithm for multiperiodic words. *Acta Polytechnica*, 53(4):344–346, 2013.