

# Staircase patterns in words: subsequences, subwords, and separation number

Toufik Mansour\*    Reza Rastegar†    Alexander Roitershtein ‡

July 27, 2019; Revised: March 4, 2020

## Abstract

We revisit staircases for words and prove several exact as well as asymptotic results for longest left-most staircase subsequences and subwords and staircase separation number. The latter is defined as the number of consecutive maximal staircase subwords packed in a word. We study asymptotic properties of the sequence  $h_{r,k}(n)$ , the number of  $n$ -array words with  $r$  separations over alphabet  $[k]$  and show that for any  $r \geq 0$ , the growth sequence  $(h_{r,k}(n))^{1/n}$  converges to a characterized limit, independent of  $r$ . In addition, we study the asymptotic behavior of the random variable  $\mathcal{S}_k(n)$ , the number of staircase separations in a random word in  $[k]^n$  and obtain several limit theorems for the distribution of  $\mathcal{S}_k(n)$ , including a law of large numbers, a central limit theorem, and the exact growth rate of the entropy of  $\mathcal{S}_k(n)$ . Finally, we obtain similar results, including growth limits, for longest  $L$ -staircase subwords and subsequences.

*MSC2010:* Primary 05A15, 05A16, 68R15; Secondary 60C05, 60J10, 60J22.

*Keywords:*  $k$ -ary words, pattern occurrences, staircase patterns, generating functions, random words, Markov chains.

## 1 Introduction and statement of results

Analysis of patterns in random words is one of the central topics in computer science, statistics, and combinatorics, with various important applications in biology, technology, and physics [4, 16, 20, 25]. In this paper, we are primarily concerned with the properties of a specific class of patterns called *staircase* or *smooth* patterns. Staircase patterns and related topics, such as  $(a, b)$ -rectangle patterns in permutations, have been the subject of several studies in recent years [26–28, 30]. Our initial motivation for studying these patterns stems out of their relation with a certain growth model of statistical physics [31].

---

\*Department of Mathematics, University of Haifa, 199 Abba Khoushy Ave, 3498838 Haifa, Israel; e-mail: tmansour@univ.haifa.ac.il

†Occidental Petroleum Corporation, Houston, TX 77046 and Departments of Mathematics and Engineering, University of Tulsa, OK 74104, USA - Adjunct Professor; e-mail: reza\_rastegar2@oxy.com

‡Department of Statistics, Texas A&M University, College Station, TX 77843, USA; e-mail: alexander@stat.tamu.edu

We define *words* as finite arrays of elements (*letters*) from a linearly ordered set (*alphabet*)  $[k] := \{1, \dots, k\}$  for some given  $k \in \mathbb{N}$ . We refer to (not necessarily distinct) letters  $i$  and  $j$  in  $[k]$  as *neighbors* if  $|i - j| \leq 1$ . Note that in the linear setup, in contrast to the cyclic one, the letters 1 and  $k$  are not considered to be neighbors. Let  $\pi = \pi_1\pi_2 \cdots \pi_n \in [k]^n$  be a  $k$ -ary word of the length  $n$ . We say that  $\pi$  is a *staircase* if  $\pi_{i+1} - \pi_i \in \{-1, 0, 1\}$  for all  $i = 1, 2, \dots, n - 1$  [28] (such words are referred to as *smooth* in [27, 30]). In other words,  $\pi$  is a staircase if any pair of adjacent letters in the word are nearest neighbors in the linear alphabet  $[k]$ . Staircase words can be interpreted as a certain class of Motzkin paths with steps  $(1, 1)$ ,  $(1, -1)$  and  $(1, 0)$ , laying in a strip with vertex heights bounded within the interval from 0 to  $k - 1$  [28].

Our first object of interest is the *staircase separation number*. For an integer  $d \geq 1$ , we call a word  $\pi$  *d-separated* if it can be written as  $\pi^{(1)}\pi^{(2)} \cdots \pi^{(d+1)}$  such that  $\pi^{(i)}$ s are maximal staircase subwords in  $\pi$ , meaning while  $\pi^{(i)}$  is a staircase,  $\pi^{(i)}$  followed by the first letter in  $\pi^{(i+1)}$  is not a staircase subword. Staircase words are considered 0-separated. For a  $d$ -separated word  $\pi \in [k]^n$ , we refer to  $d$  as the staircase separation number of  $\pi$  and denote it by  $\mathcal{S}_k(\pi)$ . For instance, the word 12313242321 is a concatenation of 5 consecutive maximal staircase subwords 123, 1, 32, 4, 2321, and hence  $\mathcal{S}_k(12313242321) = 4$ . The frequency sequence  $h_{r,k}(n)$  is then defined as the number of words  $\pi$  in  $[k]^n$  with  $\mathcal{S}_k(\pi) = r$ , that is

$$h_{r,k}(n) = \#\{\pi \in [k]^n : \mathcal{S}_k(\pi) = r\}.$$

We denote by  $\mathcal{S}_k(n)$  the staircase separation number of a random word  $\pi$  chosen uniformly from  $[k]^n$ . The distribution of the random variable  $\mathcal{S}_k(n)$  is related to the sequences  $h_{r,k}(n)$  through the identities

$$P_{k,n}^{(s)}(r) := P(\mathcal{S}_k(n) = r) = \frac{1}{k^n} h_{r,k}(n),$$

where  $P$  is a probability law that induces the uniform distribution on  $[k]^n$ . It will be technically convenient to define  $P$  on the union  $\cup_n [k]^n$  rather than on a single  $[k]^n$  with a specific  $n$ , and thus to have all random variables under consideration defined in the same probability space. Therefore, we consider  $P$  as the probability law of a given sequence of independent random variables  $(w_n)_{n \in \mathbb{N}}$ , each distributed uniformly over the alphabet set  $[k]$ , and refer to  $W_n = w_1 \cdots w_n$  as a *random word on  $[k]^n$* . We record this for future reference:

$$w_n, n \in \mathbb{N}, \text{ are independent, } P(w_i = w) = \frac{1}{k} \forall w \in [k], \quad W_n = w_1 \cdots w_n. \quad (1)$$

For example, in this notation,  $\mathcal{S}_k(n) = \mathcal{S}_k(W_n)$ .

Another object of interest is the longest  $L$ -staircase pattern, which we study in both subsequence and subword contexts. We say that  $\pi' = \pi_{i_1}\pi_{i_2} \cdots \pi_{i_r}$  is the *longest  $L$ -staircase* or *left-most staircase* subsequence of  $\pi$  if  $\pi'$  is a staircase word such that  $i_1 = 1$ , the differences  $i_2 - i_1, \dots, i_r - i_{r-1}$  are all minimal, and  $r$  is maximal under this rule. For a fixed  $\pi \in [k]^n$ , we denote this maximal  $r \leq n$  by  $\mathcal{J}_k(\pi)$ . That is the sequence  $(i_j)_{j \geq 1}$  is defined by setting  $i_1 = 1$  and then, recursively,

$$i_j = \inf \{m > i_{j-1} : m \leq n \text{ and } |\pi_m - \pi_{i_{j-1}}| \leq 1\}, \quad j \in \mathbb{N},$$

where, as usual, we assume that  $\inf \emptyset = +\infty$ . Formally,

$$\mathcal{J}_k(\pi) = \max\{r \in \mathbb{N} : j_r < \infty\}.$$

For instance, the 4-ary word 141321 is not a  $L$ -staircase word, however, its longest  $L$ -staircase subsequence is 1121 with  $\mathcal{J}_4(141321) = 4$ . The frequency sequence  $f_{r,k}(n)$  is then defined as the number of words  $\pi$  in  $[k]^n$  with  $\mathcal{J}_k(\pi) = r$ , that is

$$f_{r,k}(n) = \#\{\pi \in [k]^n : \mathcal{J}_k(\pi) = r\}.$$

Furthermore, we define  $\mathcal{J}_k(n)$  as the length of the longest  $L$ -staircase subsequence of a random word distributed uniformly over  $[k]^n$ . That is,

$$P_{k,n}^{ls}(r) := P(\mathcal{J}_k(n) = r) = \frac{1}{k^n} f_{r,k}(n),$$

where  $P$  is the uniform distribution on  $[k]^n$ .

Similarly, we define the  $L$ -staircase subword or left-most staircase subword in a word  $\pi \in [k]^n$  as  $\pi' = \pi_1\pi_2 \cdots \pi_r$ , where  $\pi'$  is a staircase word and  $r \leq n$  is maximal. For instance, the 4-ary word 112141321 is not a staircase word, its  $L$ -staircase subword is 1121. For  $\pi \in [k]^n$  we set  $\pi_0 = \pi_1$  and define  $\mathcal{I}_k(\pi)$  as

$$\mathcal{I}_k(\pi) = \max\{r \in \mathbb{N} : r \leq n \text{ and } |\pi_i - \pi_{i-1}| \leq 1 \ \forall i \leq r-1\}.$$

We then define  $g_{r,k}(n)$  as the number of words  $\pi$  in  $[k]^n$  such that  $\mathcal{I}_k(\pi) = r$ . That is,

$$g_{r,k}(n) = \#\{\pi \in [k]^n : \mathcal{I}_k(\pi) = r\}.$$

In addition, we define  $\mathcal{I}_k(n)$  as the length of the longest  $L$ -staircase subword of a random word sampled uniformly from  $[k]^n$ . Note that for all  $r \in \mathbb{N}_0$ ,

$$P_{k,n}^{lw}(r) := P(\mathcal{I}_k(n) = r) = \frac{1}{k^n} g_{r,k}(n). \quad (2)$$

Given that  $\mathcal{S}_1(\pi) = \mathcal{S}_2(\pi) = 0$  and  $\mathcal{J}_1(\pi) = \mathcal{J}_2(\pi) = \mathcal{I}_1(\pi) = \mathcal{I}_2(\pi) = n$  for all  $\pi \in [k]^n$  and  $n \in \mathbb{N}$ , in what follows we assume that  $k \geq 3$ .

To state our results we first introduce a few notations. Throughout this paper,  $a_n \sim b_n$ ,  $a_n = O(b_n)$ , and  $a_n = o(b_n)$  for sequences  $a_n$  and  $b_n$  with elements that might depend on  $r, k$ , and other parameters, means that, respectively,  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$ ,  $\limsup_{n \rightarrow \infty} \left| \frac{a_n}{b_n} \right| < \infty$ , and  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$  for all feasible values of the parameters when the latter are fixed. As usual,  $a_n = \Theta(b_n)$  indicates that both  $a_n = O(b_n)$  and  $b_n = O(a_n)$  hold true. For a random variable  $X$ , we denote its probability distribution by  $\mathcal{L}(X)$ , and its mean and variance by, respectively,  $E(X)$  and  $\sigma^2(X)$ . We use the notation  $\lim_{n \rightarrow \infty} \mathcal{L}(X_n) = X$  to indicate the convergence in distribution of a sequence of random variables  $X_n$  to a random variable  $X$ , as  $n$  tends to infinity. We denote by  $\mathcal{N}(0, 1)$  a standard normal random variable.

Recall that the sequence of Chebyshev polynomials  $U_n(t)$  of the second kind can be defined as the unique solution to the recursion equation

$$U_n(t) = 2tU_{n-1}(t) - U_{n-2}(t), \quad n \geq 2, \quad (3)$$

with initial conditions

$$U_0(t) = 1 \quad \text{and} \quad U_1(t) = 2t. \quad (4)$$

The polynomials  $U_n$  can be explicitly expressed as

$$U_n(t) = \frac{(t + \sqrt{t^2 - 1})^{n+1} - (t - \sqrt{t^2 - 1})^{n+1}}{2\sqrt{t^2 - 1}}. \quad (5)$$

A standard reference for basic properties of Chebyshev's polynomial, which are frequently invoked in this paper, is [36]. Notice the resemblance of (3) to the Fibonacci recurrence and of (5) to Binet's formula for Fibonacci numbers. The intimate relation between Chebyshev polynomials and Fibonacci, Lucas, and more generally Horadam's numbers (i. e., solutions to a general second-order linear recursion with constant coefficients) has been intensively explored, among other areas, in enumerative combinatorics for computing generating functions (see, for instance, [20, 27, 30]), number theory for representation of integer sequences (see, for instance, an early work [6] and recent [33, 35]), and applied linear algebra for computing determinants and inverse matrices (see, for instance, [13, 14, 17, 40]). The formula (3) also provides a natural link between the theory of Chebyshev polynomials and a discrete Poisson equation and its Sturm-Liouville theory [7, 13–15, 22], and subsequently to the harmonic functions and first passage time of random walks (see, for instance, the classical [24, 29] and more recent [10, 12, 18]). Altogether, the link fundamentally bridges between analytic and probabilistic methods exploited in this paper.

Our first result deals with the first and second moments of  $\mathcal{S}_k(n)$ ,  $\mathcal{J}_k(n)$  and  $\mathcal{I}_k(n)$ . We remark that in the context of a “cyclic order” alphabet, namely when 1 and  $k$  are considered as a neighbor pair, the above quantities are relatively easy to analyze. For instance, in the cyclic setup  $\mathcal{I}_k(n)$  becomes equal to a truncated geometric variable  $\min\{G_k, n\}$ , where  $G_k$  is a geometric random variable with parameter  $\frac{k-3}{k}$ ,  $\mathcal{S}_k(n)$  is binomial  $BIN(n-1, \frac{k-3}{k})$ , and  $\mathcal{J}_k(n)$  is distributed as  $1 + BIN(n-1, \frac{3}{k})$ . The challenge in linear order alphabet  $[k]$  is due to the fact that the letters 1 and  $k$  play a special role (in that they both have only two neighbors in the alphabet), and subsequently must be accounted with an extra care. In the proof of the following theorem, even though asymptotic, as  $n \rightarrow \infty$ , results for the moments can still be obtained using short probabilistic arguments, we choose an analytical approach, namely the analysis of generating functions, as a unified method of our analysis in order to compute exact, that is for any given  $n \in \mathbb{N}$ , values of the moments.

**Theorem 1.1.** *For any  $k \geq 3$ ,*

$$(i) \quad E(\mathcal{S}_k(n)) = \frac{(k-2)(k-1)}{k^2}(n-1),$$

$$(ii) \quad \sigma^2(\mathcal{S}_k(n)) = -\frac{4(k-2)}{k^4} + \frac{(k-2)(3k^2-5k+6)}{k^4}(n-1).$$

$$(iii) \quad E(\mathcal{I}_k(n)) \sim \frac{2}{(k-3)^2 U_k(\frac{k-1}{2})} + \frac{2U_{k-1}(\frac{k-1}{2})}{(k-3)^2 U_k(\frac{k-1}{2})} + \frac{(k^2-3k-2)}{(k-3)^2}, \text{ provided that } k \geq 4. \text{ Furthermore, } \\ E(\mathcal{I}_3(n)) = 5 + o(1), \text{ as } n \rightarrow \infty.$$

$$(iv) \quad \text{Provided that } k \geq 4, \text{ we have } \sigma^2(\mathcal{I}_k(n)) = e_1(k) + e_2(k) - e_1^2(k) + o(1), \text{ where the constants } e_1(k) \text{ and } e_2(k) \text{ are explicitly defined below, in (23) and (24), respectively. Furthermore, } \sigma^2(\mathcal{I}_3(n)) = 21 + o(1), \text{ as } n \rightarrow \infty.$$

$$(v) \ E(\mathcal{J}_k(n)) = 1 + \frac{3k-2}{k^2}(n-1).$$

$$(vi) \ \sigma^2(\mathcal{J}_k(n)) = \left( a_k - \frac{(3k-2)(k^2-15k+10)}{k^4} \right) n + O(1), \text{ where } a_k \text{ is defined by (28).}$$

The proof of parts (i) and (ii) of the theorem is given in Section 2.1, the proof of (iii) and (iv) is included in Section 3, and the claims in (v) and (vi) are proved in Section 4.1. It is instructive to compare the results in the above theorem with their analogue in the cyclic setup, when 1 is formally identified with  $k+1$ . For instance, for the expectations we have

$$E(\tilde{\mathcal{S}}_k(n)) = \frac{k-3}{k}(n-1), \quad E(\tilde{\mathcal{I}}_k(n)) = \frac{k}{k-3} \left( 1 - \frac{1}{k^{n-1}} \right), \quad E(\tilde{\mathcal{J}}_k(n)) = 1 + \frac{3}{k}(n-1),$$

where  $\tilde{\mathcal{S}}_k(n)$ ,  $\tilde{\mathcal{I}}_k(n)$ , and  $\tilde{\mathcal{J}}_k(n)$  are counterparts of, respectively,  $\mathcal{S}_k(n)$ ,  $\mathcal{I}_k(n)$ , and  $\mathcal{J}_k(n)$  in the cyclic setup. Curiously enough, even though, in the linear alphabet case, in contrast to the cyclic setup, the random variable  $\mathcal{J}_k(n)$  is not distributed the same as  $n - \mathcal{S}_k(n)$ , we still have that  $E(\mathcal{S}_k(n)) + E(\mathcal{J}_k(n)) = n$ .

Recall (1). The base for our probabilistic analysis is the simple observation that the longest  $L$ -staircase of a random word  $W_n$  is a path of a nearest-neighbor symmetric random walk on  $[k]$ . In particular, the random variable  $\mathcal{S}_k(n)$  can be interpreted in terms of an additive functional of this random walk (see Section 2.2 and also Remark 2.6 below). A similar representation also holds for  $\mathcal{I}_k$  and  $\mathcal{J}_k$  (see Section 4.2 for the latter and the Appendix for the former). Using this observation we obtain the following laws of large numbers and central limit theorems for  $\mathcal{S}_k$  and  $\mathcal{J}_k$ ,

**Theorem 1.2.** *For any  $k \geq 3$  we have:*

$$(i) \text{ (LLN for } \mathcal{S}_k) \ \psi_k := \lim_{n \rightarrow \infty} \frac{\mathcal{S}_k(n)}{n} = \frac{(k-2)(k-1)}{k^2}, \text{ with probability one.}$$

(ii) (CLT for  $\mathcal{S}_k$ ) *There exists a constant  $\sigma_k \in (0, \infty)$  such that*

$$\lim_{n \rightarrow \infty} \mathcal{L} \left( \frac{\mathcal{S}_k(n) - \psi_k n}{\sigma_k \sqrt{n}} \right) = \mathcal{N}(0, 1).$$

$$(iii) \text{ (LLN for } \mathcal{J}_k) \ \varphi_k := \lim_{n \rightarrow \infty} \frac{\mathcal{J}_k(n)}{n} = \frac{3k-2}{k^2}, \text{ with probability one.}$$

(iv) (CLT for  $\mathcal{J}_k$ ) *There exists a constant  $\nu_k \in (0, \infty)$  such that*

$$\lim_{n \rightarrow \infty} \mathcal{L} \left( \frac{\mathcal{J}_k(n) - \varphi_k n}{\nu_k \sqrt{n}} \right) = \mathcal{N}(0, 1).$$

The proof of parts (i) and (ii) of the theorem is given in Section 2.2. The proof of part (iii) and part (iv) is similar, it is outlined in Section 4.2. Some information on  $\sigma_k^2$  and  $\nu_k^2$  is available from the general theory of Markov renewal processes, see Remark 2.4 below for more details. In particular, both the limiting variances can be compactly expressed in terms of a pseudo inverse (the inverse in a  $(k-1)$ -dimensional subspace) of a  $k \times k$  matrix (which is tridiagonal for  $\mathcal{J}_k(n)$ ), see, for instance, Eq. (2) in [19].

In contrast to  $\mathcal{S}_k(n)$  and  $\mathcal{J}_k(n)$ , the distribution of  $\mathcal{I}_k(n)$  converges for all  $k \geq 3$  to a proper limit, as  $n \rightarrow \infty$ , without scaling:

$$\lim_{n \rightarrow \infty} P(\mathcal{I}_k(n) \geq r) = \frac{h_{0,k}(r)}{k^r} \tag{6}$$

for any  $r \in \mathbb{N}$ . Indeed, for any  $n > r$  we have the identity  $P(\mathcal{I}_k(n) \geq r) = \frac{h_{0,k}(r)}{k^r}$ . The limit is a proper distribution since  $\lim_{r \rightarrow \infty} \frac{h_{0,k}(r)}{k^r} = 0$  (for instance, by the result in part (i) of Theorem 1.5 below). Combining (6) with the explicit formula for  $h_{0,k}(r)$  given by Theorem 2.4 in [28], we arrive to the following:

**Theorem 1.3.** *For all  $k \geq 3$  and  $r \geq 1$ ,*

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\mathcal{I}_k(n) = r) &= \frac{h_{0,k}(r)}{k^r} - \frac{h_{0,k}(r+1)}{k^{r+1}} \\ &= \frac{2}{(k+1)k^{r+1}} \sum_{j=1}^k (1 + (-1)^{j+1}) \cot^2 \frac{j\pi}{2(k+1)} \left(1 + 2 \cos \frac{j\pi}{k+1}\right)^{r-1} \left(k - 1 - 2 \cos \frac{j\pi}{k+1}\right). \end{aligned}$$

Various additional aspects of the asymptotic behavior of the sequences  $\mathcal{S}_k(n)$ ,  $\mathcal{I}_k(n)$ , and  $\mathcal{J}_k(n)$  can be analyzed by exploiting the underlying Markovian structure of staircase patterns which is described in Sections 2.2 and 4.2, and Remark 2.6 and the standard machinery of the asymptotic theory of Markov renewal processes (cf. Remark 2.4 in [21]). For instance, one can complement the laws of large numbers in Theorem 1.2 with exponential concentration inequalities and large deviation estimates (cf. Section 3 in [5]) and supplement the central limit theorems in Theorem 1.2 with the associated laws of the iterated logarithm (cf., for instance, Theorem 1-(iv) in [2]).

Another example is the following ‘‘cousin’’ of Theorem 1.1-(i). Let  $\pi^{(1)}\pi^{(2)} \dots \pi^{(\mathcal{S}_k(\pi)+1)}$  be the representation of a random word  $\pi \in [k]^n$  as a concatenation of maximal staircase subwords  $\pi^{(i)}$ ,  $i = 1, \dots, \mathcal{S}_k(\pi) + 1$ . Let  $\ell_i$  be the length (the number of letters) of  $\pi^{(i)}$ . Then, with probability one, the following holds for any  $k \geq 3$ :

$$\lim_{n \rightarrow \infty} \frac{1}{\mathcal{S}_k(n)} \sum_{i=1}^{\mathcal{S}_k(n)} \ell_i = \lim_{n \rightarrow \infty} \frac{1}{\mathcal{S}_k(n)} \sum_{i=1}^{\mathcal{S}_k(n)+1} \ell_i = \frac{k^2}{(k-1)(k-2)}.$$

In words,  $\frac{k^2}{(k-1)(k-2)}$  is the asymptotic average distance between two separations in a random word. The limit result is immediate from (16) and (18) given below, in Section 2.2.

Yet another natural extension of asymptotic results in Theorems 1.1 and 1.2 is the following proposition whose short proof is deferred to the Appendix.

**Proposition 1.4.**

(i) *Let  $\pi_1 \dots \pi_{\mathcal{I}_k(n)}$  be the longest  $L$ -staircase subword of a random word  $\pi \in [k]^n$ . For  $w \in [k]$ , let  $\xi_{w,k}(n) = \#\{i \in [\mathcal{I}_k(n)] : \pi_i = w\}$  be the number of occurrences of the letter  $w$  in the subword. Then, with probability one, the following holds for all  $k \geq 3$  and  $w \in [k]$ :*

$$\lim_{n \rightarrow \infty} E(\xi_{w,k}(n)) = \begin{cases} 6w(4-w) & \text{if } k = 3, \\ \frac{k}{k-3} \cdot \frac{U_k\left(\frac{k-1}{2}\right) - U_{w-1}\left(\frac{k-1}{2}\right) - U_{k-w}\left(\frac{k-1}{2}\right)}{U_k\left(\frac{k-1}{2}\right)} & \text{if } k \geq 4. \end{cases} \quad (7)$$

(ii) *Let  $\pi_1, \dots, \pi_{\mathcal{J}_k(n)}$  be the longest  $L$ -staircase subsequence in a random word  $\pi \in [k]^n$ . For  $w \in [k]$ , let  $\eta_{w,k}(n) = \#\{i \in [\mathcal{J}_k(n)] : \pi_i = w\}$  be the number of occurrences of the letter  $w$  in the subsequence. Then, with probability one, the following holds for any  $k \geq 3$ :*

$$\lim_{n \rightarrow \infty} \frac{\eta_{w,k}(n)}{\mathcal{J}_k(n)} = \begin{cases} \frac{2}{3k-2} & \text{if } w \in \{1, k\}, \\ \frac{3}{3k-2} & \text{if } w \notin \{1, k\}. \end{cases} \quad (8)$$

Next, we study the asymptotic behavior, as  $r \in \mathbb{N}$  remains fixed and  $n$  goes to infinity, of the frequency sequences  $h_{r,k}(n) = k^n P(\mathcal{S}_k(n) = r)$ ,  $f_{r,k}(n) = k^n P(\mathcal{J}_k(n) = r)$ , and  $g_{r,k}(n) = k^n P(\mathcal{I}_k(n) = r)$ . We have:

**Theorem 1.5.** *The following holds for any  $k \geq 3$  :*

- (i)  $\lim_{n \rightarrow \infty} (h_{r,k}(n))^{\frac{1}{n}} = 1 + 2 \cos(\frac{\pi}{k+1})$  for all integers  $r \geq 0$ .
- (ii)  $\lim_{n \rightarrow \infty} (f_{r,k}(n))^{\frac{1}{n}} = k - 2$  for all integers  $r > 0$ .

The proof of part (i) relies on a result of [28], it is given in Section 2.3. A short proof of part (ii) is included in Section 4.3. As regards  $g_{r,k}(n)$ , we note that Theorem 1.3 yields:

$$\lim_{n \rightarrow \infty} (g_{r,k}(n))^{\frac{1}{n}} = \lim_{n \rightarrow \infty} (k^n \mathcal{I}_k(n))^{\frac{1}{n}} = k.$$

for all  $r \in \mathbb{N}$ .

Observe that when  $n$  is fixed and  $k$  is taken to infinity,  $\mathcal{S}_k(n)$  converges to  $n$  while  $\mathcal{J}_k(n)$  converges to 1 in probability. Theorem 1.5 can serve to illustrate the inherent difference between the random mechanism of  $\mathcal{J}_k(\cdot)$  and  $\mathcal{S}_k(\cdot)$  through a sensible quantification of the dependence of their distribution on the size of the alphabet. Specifically, let

$$H_k^{(s)}(n) = - \sum_{r \geq 0} P_{k,n}^{(s)}(r) \log P_{k,n}^{(s)}(r)$$

be the entropy of  $\mathcal{S}_k(n)$ . We similarly denote by  $H_k^{(tw)}(n)$  the entropy of  $\mathcal{I}_k(n)$  and by  $H_k^{(ts)}(n)$  the entropy of  $\mathcal{J}_k(n)$ . Loosely speaking, the entropy measures the amount of uncertainty in the value of a random variable [11]. Therefore, one should expect that the entropy sequence  $H_k^{(s)}(\cdot)$  is subadditive, namely the following holds for all  $n, m \in \mathbb{N}$  :

$$H_k^{(s)}(n+m) \leq H_k^{(s)}(n) + H_k^{(s)}(m). \tag{9}$$

It is not hard to verify that this heuristic inequality indeed holds. The convergence of  $\frac{H_k^{(s)}(n)}{n}$  is thus ensured by Fekete's subadditivity lemma. Our next result specifies the value of the limit. The proof given in Section 2.3 however bypasses (9) entirely and shows the existence of the limit directly, relying on the results in Theorem 1.5.

**Corollary 1.6.** *For  $k \geq 3$ ,*

- (i)  $\lim_{n \rightarrow \infty} \frac{H_k^{(s)}(n)}{n} = \log \frac{k}{1+2 \cos(\frac{\pi}{k+1})}$ .
- (ii)  $\lim_{n \rightarrow \infty} \frac{H_k^{(ts)}(n)}{n} = \log \frac{k}{k-2}$ .

The corollary is a direct implication of Theorem 1.5, and proofs of part (i) and part (ii) are nearly identical. We thus only prove the former in Section 2.3 and omit the proof of the latter. We remark that Theorem 1.3 implies that  $H_k^{(tw)}(n)$  converges, as  $n \rightarrow \infty$ , to the entropy of the limiting distribution of the sequence  $\mathcal{I}_k(n)$  (see, for instance, Proposition 1 in [37] - some care is required because, in general, the entropy is not a continuous function of distributions in infinite spaces, even in a discrete setting [11, 37]).

The rest of the paper is organized as follows. The proof of  $d$ -separation results (namely, Theorem 1.1-(i),(ii), Theorem 1.2-(i),(ii), Theorem 1.5-(i) and Corollary 1.6-(i)) are given in Section 2. The results for  $L$ -staircase subwords (namely, Theorem 1.1-(iii),(iv)) are proven in Section 3 and for  $L$ -staircase subsequences (namely, Theorem 1.1-(v),(vi), Theorem 1.2-(iii),(iv), and Theorem 1.5-(ii)) in Section 4. Finally, the proof of Proposition 1.4 is included in the Appendix.

Since the sections are independent of each other and practically self-contained, with slight abuse of notation, we use the same generic notation  $L_{n,k}(q)$  and  $L_{n,k;l_1l_2\dots l_s}(q)$  for underlying generating functions throughout the rest of the paper. The actual definitions are different in each subsection, they are always specified at the beginning of a subsection and remain unchanged until its end. In addition, if  $f(x) = \sum_{n \geq 0} a_n x^n$ , we frequently use the notation  $[x^n]f(x)$  to denote  $a_n$ , the  $n$ -th coefficient of the power series. For example,  $[x^2] \cos(x) = -\frac{1}{2}$  and  $[x^3] \sin(x) = \frac{1}{3}$ .

## 2 Staircases separation

This goal of this section is to prove the results for the number of separations. The section is divided into three subsections. Section 2.1 contains the proof of part (i) and part (ii) of Theorem 1.1, Section 2.2 includes the proof of part (i) and part (ii) of Theorem 1.2, and Section 4.3 is devoted to the proof of Theorem 1.5-(i) and its implication, Corollary 1.6-(i).

### 2.1 Proof of Theorem 1.1-(i),(ii)

Let  $L_{n,k}(q)$  be the generating function for the  $k$ -ary words of length  $n$ , according to the statistic  $\mathcal{S}_k$ . That is,

$$L_{n,k}(q) = \sum_{\pi \in [k]^n} q^{\mathcal{S}_k(\pi)} = \sum_{r=0}^{n-1} h_{r,k}(n) q^r, \quad q \in \mathbb{C}.$$

In addition, we define  $L_{n,k;l_1l_2\dots l_s}(q)$  to be the generating function for the  $k$ -ary words  $\pi = l_1l_2 \cdots l_s \pi'$  of length  $n$  according to the statistic  $\mathcal{S}_k$ . That is,

$$L_{n,k;l_1l_2\dots l_s}(q) = \sum_{\pi = l_1l_2 \cdots l_s \pi' \in [k]^n} q^{\mathcal{S}_k(\pi)}, \quad q \in \mathbb{C}.$$

Next, for suitable  $q, x \in \mathbb{C}$ , we set

$$L_k(x, q) = \sum_{n \geq 0} L_{n,k}(q) x^n \quad \text{and} \quad L_{k;l_1l_2\dots l_s}(x, q) = \sum_{n \geq 0} L_{n,k;l_1l_2\dots l_s}(q) x^n. \quad (10)$$

Note that the series in (10) converge at least when  $k \cdot |x| \cdot \max\{1, |q|\} < 1$ . It is straightforward to verify that

$$L_1(x, q) = \frac{1}{1-x} \quad \text{and} \quad L_2(x, q) = \frac{1}{1-2x}.$$



Hence, in what follows we will assume that  $k \geq 3$ . By (10), we have

$$\begin{aligned}
L_{k;i}(x, q) &= x + x(L_{k;i-1}(x, q) + L_{k;i}(x, q) + L_{k;i+1}(x, q)) + xq \sum_{j \in [k] \setminus \{i-1, i, i+1\}} L_{k;j}(x, q) \\
&= x + x(L_{k;i-1}(x, q) + L_{k;i}(x, q) + L_{k;i+1}(x, q)) \\
&\quad + xq(L_k(x, q) - 1 - L_{k;i-1}(x, q) + L_{k;i}(x, q) + L_{k;i+1}(x, q)) \\
&= x(1 - q) + x(1 - q)(L_{k;i-1}(x, q) + L_{k;i}(x, q) + L_{k;i+1}(x, q)) + xqL_k(x, q),
\end{aligned}$$

where

$$L_{k;0}(x, q) = L_{k,k+1}(x, q) = 0.$$

Therefore, for all  $i = 1, 2, \dots, k$ , we have

$$L_{k;i}(x, q) = \alpha(L_{k;i-1}(x, q) + L_{k;i+1}(x, q)) + \beta, \quad (11)$$

where

$$\alpha = \frac{x(1-q)}{1-x(1-q)} \quad \text{and} \quad \beta = \alpha + \frac{xq}{1-x(1-q)}L_k(x, q).$$

Define a symmetric tridiagonal matrix  $\mathbf{A}_k = (a_{ij})_{1 \leq i, j \leq k}$  by setting

$$a_{ij} = \begin{cases} 1 & \text{if } i = j, \\ -\alpha & \text{if } |i - j| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In this notation, a matrix form of (11) is

$$\mathbf{A}_k \begin{pmatrix} L_{k;1}(x, q) \\ \vdots \\ L_{k;k}(x, q) \end{pmatrix} = \beta e_k, \quad (12)$$

where  $e_k = (1, \dots, 1) \in \mathbb{R}^k$  is a  $k$ -dimensional vector with all coordinates equal to one. Finally, taking in account that

$$L_k(x, q) = 1 + \sum_{i=1}^k L_{k;i}(x, q),$$

we obtain that

$$L_k(x, q) = 1 + \beta e_k^\top \mathbf{A}_k^{-1} e_k = 1 + \alpha e_k^\top \mathbf{A}_k^{-1} e_k + \frac{xq}{1-x(1-q)} L_k(x, q) e_k^\top \mathbf{A}_k^{-1} e_k. \quad (13)$$

An application of Corollary 4.4 in [17] then yields the following:

**Lemma 2.1.** *For all  $k \geq 1$ ,*

$$L_k(x, q) = 1 - \frac{1 + (k(t-1) - 1)U_k(t) + U_{k-1}(t)}{q + 2(q-1)(t-1)^2 U_k(t) + q(k(t-1) - 1)U_k(t) + qU_{k-1}(t)},$$

where  $t = \frac{1-x(1-q)}{2x(1-q)}$ , as long as  $t \notin \{-1, 1\}$ .

**Example 2.2.** Lemma 2.1 with  $k = 3$  gives

$$\begin{aligned} L_3(x, q) &= \frac{1 + x - qx}{1 - (2 + q)x - (1 - q)x^2} \\ &= \frac{1 + x}{1 - 2x - x^2} + \sum_{r \geq 1} \frac{2x^{r+1}(1 - x)^{r-1}}{(1 - 2x - x^2)^{r+1}} q^r. \end{aligned}$$

Assume now that  $k \geq 4$ . Setting  $\varepsilon = 1 - q$  and expanding the numerator and denominator of the generating function  $L_k(x, 1 - \varepsilon)$  at  $\varepsilon = 0$  with the help of the identity

$$U_n(x) = \sum_{j \geq 0} (-1)^j \binom{n - j}{j} (2x)^{n-2j}, \quad (14)$$

we see that

$$L_k(x, 1 - \varepsilon) = \frac{\sum_{j=0}^2 \alpha_j x^j \varepsilon^j + O(\varepsilon^3)}{\sum_{j=0}^2 \beta_j x^j \varepsilon^j + O(\varepsilon^3)}, \quad (15)$$

where

$$\begin{aligned} \alpha_0 &= k + 2, & \alpha_1 &= -(k + 3)(k + 4), & \alpha_2 &= \frac{1}{2}(k + 3)(k^2 + 6k + 12), \\ \beta_0 &= k + 2 + 6\varepsilon, & \beta_1 &= -(k^2 + 7k + 12) + \frac{1}{2}(k^2 + k - 18)\varepsilon, \\ \beta_2 &= \frac{1}{2}(k + 3)(k^2 + 6k + 12) - \frac{1}{3}k(k^2 + 6k + 11)\varepsilon. \end{aligned}$$

It is not hard to verify that (15) is equivalent to the following result:

**Lemma 2.3.** The power series expansion of the generating function  $L_k(x, q)$  at  $q = 1$  is given by

$$L_k(x, q) = \frac{a_k(x)}{1 - kx} + b_k(x) \frac{x^2(q - 1)}{(1 - kx)^2} + c_k(x) \frac{x^3(q - 1)^2}{(1 - kx)^3} + O((q - 1)^3),$$

where

$$a_k(x) = 1, \quad b_k(x) = (k - 2)(k - 1), \quad c_k(x) = -(k - 2)(2x - k^2 + 4k - 5).$$

Using Lemma 2.3 to evaluate  $\frac{\partial^j}{\partial q^j} L_k(x, q) \Big|_{q=1}$  for  $j = 1, 2$ , we obtain that

$$\begin{aligned} E[\mathcal{S}_k(n)] &= \frac{1}{k^n} [x^n] \frac{\partial}{\partial q} L_k(x, q) \Big|_{q=1} = \frac{1}{k^n} [x^n] \frac{(k - 2)(k - 1)x^2}{(1 - kx)^2} \\ &= \frac{(k - 2)(k - 1)}{k^2} (n - 1), \\ E[\mathcal{S}_k(n)(\mathcal{S}_k(n) - 1)] &= \frac{1}{k^n} [x^n] \frac{\partial^2}{\partial q^2} L_k(x, q) \Big|_{q=1} = \frac{1}{k^n} [x^n] \frac{2(k - 2)x^3(k^2 - 4k + 5 - 2x)}{(1 - kx)^3} \\ &= \frac{(n - 2)(k - 2)}{k^4} ((n - 1)k^3 - 4(n - 1)k^2 + 5(n - 1)k - 2n + 6). \end{aligned}$$

Thus,

$$\begin{aligned}\sigma^2(\mathcal{S}_k(n)) &= E(\mathcal{S}_k(n)(\mathcal{S}_k(n) - 1)) + E(\mathcal{S}_k(n)) - [E(\mathcal{S}_k(n))]^2 \\ &= -\frac{4(k-2)}{k^4} + \frac{(k-2)(3k^2 - 5k + 6)}{k^4}(n-1).\end{aligned}$$

This completes the proof of (i) and (ii) of Theorem 1.1.  $\square$

## 2.2 Proof of Theorem 1.2-(i),(ii)

Recall (1). Without loss of generality we can assume that  $\mathcal{S}_k(n) = \mathcal{S}_k(W_n)$ . Thus, there exists a random sequence of *maximal*  $L$ -staircase words  $(\pi^{(n)})_{n \in \mathbb{N}}$  such that

$$W_n = \pi^{(1)} \dots \pi^{(\mathcal{S}_k(n))} \mu_n,$$

where the remainder  $\mu_n$  is a prefix of  $\pi^{(\mathcal{S}_k(n)+1)}$ , possibly empty. Denote by  $u_n$  the last letter in the word  $\pi^{(n)}$  and by  $\ell_n$  the length of  $\pi^{(n)}$ . For  $n \in \mathbb{N}$ , let

$$\Lambda_n = \sum_{i=1}^n \ell_i.$$

Thus,  $\mathcal{S}_k(n)$  can be characterized as the unique sequence of non-negative integers such that

$$\Lambda_{\mathcal{S}_k(n)} < n \leq \Lambda_{\mathcal{S}_k(n)+1} \quad \forall n \in \mathbb{N}. \quad (16)$$

Equivalently, for any  $n, m \in \mathbb{N}$  we have

$$\{\mathcal{S}_k(n) \geq m\} = \{\Lambda_m < n\}. \quad (17)$$

Clearly, both  $(u_i)_{i \in \mathbb{N}}$  and  $(u_i, \ell_i)_{i \in \mathbb{N}}$  are Markov chains, the former is irreducible and aperiodic, and the latter is formally a *hidden Markov model* (in that the transition kernel of  $(u_i, \ell_i)$  depends only on the current value of the first component, the value of  $(u_{i+1}, \ell_{i+1})$  is independent of  $\ell_i$  when  $u_i$  is given). Furthermore, the distribution tails of  $\ell_i$  are exponential, namely

$$P(\ell_{i+1} \geq x \mid u_i = j) \leq \left(\frac{k-2}{k}\right)^{x-1} \quad \forall i, x \in \mathbb{N}, j \in [k].$$

Therefore, by standard arguments as stated in [8, 23, 32], there exists a strictly positive number  $b > 0$  such that the distribution of the normalized sequence  $\frac{\Lambda_n - an}{b\sqrt{n}}$  converges, as  $n \rightarrow \infty$ , to  $\mathcal{N}(0, 1)$ , where  $a = \lim_{n \rightarrow \infty} \frac{\Lambda_n}{n}$ . Note that by the ergodic theorem,  $a = E_{st}(\Lambda_1)$ , where  $E_{st}$  stands for the expectation in the stationary regime.

The two-component random process  $(u_n, \Lambda_n)_{n \in \mathbb{N}}$  belongs to the class of *Markov renewal processes* [3, 9]. Usual arguments of the renewal theory (see, for instance, (4.2) in [2], Lemma 2.1.17 in [41], or Section 11 in [9]) show that (16) implies that with probability one,

$$a = \lim_{n \rightarrow \infty} \frac{n}{\mathcal{S}_k(n)} = \lim_{n \rightarrow \infty} \frac{n}{E(\mathcal{S}_k(n))} = \frac{k^2}{(k-2)(k-1)}, \quad (18)$$

where the first identity is a direct implication of (16), the second one is an application of the bounded convergence theorem to the sequence  $\mathcal{S}_k(n)/n$ , and the last one is due to the result in Theorem 1.1-(i).

The translation of the CLT for  $\Lambda_n$  into a CLT for  $\mathcal{S}_k(n)$  is a standard duality (locations of the random walk versus first passage times) argument in the renewal theory of random walks on  $\mathbb{Z}$  (which is especially simple in our case because the increments  $\ell_n$  of the random walk  $\Lambda_n$  are strictly positive), see, for instance, [1, p. 344], Theorem 1 in [2], or Section 11 in [9]. Specifically, choose any  $x \in \mathbb{R}$  and, for  $n \in \mathbb{N}$ , let

$$\tau(n) = \lfloor a^{-1}n + a^{-3/2}bx\sqrt{n} \rfloor,$$

where  $\lfloor y \rfloor$  stands for  $\max\{n \in \mathbb{N} \mid n \leq y\}$ , the integer part of  $y$ . Note that  $\tau(n) > 0$  for all  $n$  large enough. It follows then from (17) that, as  $n \rightarrow \infty$ ,

$$\begin{aligned} P\left(\frac{\mathcal{S}_k(n) - a^{-1}n}{a^{-3/2}b\sqrt{n}} \geq x\right) &= P(\mathcal{S}_k(n) \geq \tau(n)) = P(\Lambda_{\tau(n)} \leq n) \\ &= P\left(\frac{\Lambda_{\tau(n)} - (n + a^{-1/2}bx\sqrt{n})}{b\sqrt{a^{-1}n + a^{-3/2}bx\sqrt{n}}} \leq \frac{n - (n + a^{-1/2}bx\sqrt{n})}{b\sqrt{a^{-1}n + a^{-3/2}bx\sqrt{n}}}\right) \\ &\sim \Phi(-x) = 1 - \Phi(x), \end{aligned}$$

where  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$  is the cumulative distribution function of  $\mathcal{N}(0, 1)$ . Thus,

$$\lim_{n \rightarrow \infty} \mathcal{L}\left(\frac{\mathcal{S}_k(n) - a^{-1}n}{a^{-3/2}b\sqrt{n}}\right) = \mathcal{N}(0, 1).$$

The proof of part (i) of the theorem is complete.  $\square$

**Remark 2.4.** *The proof shows that  $\sigma_k = a^{-3/2}b$ . The limiting variance  $b^2$  can be expressed in terms of the underlying Markov chain  $(u_n, \ell_n)_{n \in \mathbb{N}}$  as follows:*

$$b^2 = \text{VAR}_{st}(\ell_1) + 2 \sum_{j=2}^{\infty} \text{COV}_{st}(\ell_1, \ell_j),$$

where the subscript “st” indicates that the covariances are calculated in the stationary regime. For more information on the structure of the limiting variance and its alternative representations see, for instance, Section 17.4.3 in [32] and [19, 39].

## 2.3 Proof of Theorem 1.5-(i) and Corollary 1.6-(i)

We begin with the proof of Theorem 1.5-(i).

*Proof of Theorem 1.5-(i).* By Corollary 2.5 in [28], we know that the result holds for  $r = 0$ , that is

$$\lim_{n \rightarrow \infty} (h_{0,k}(n))^{1/n} = 1 + 2 \cos\left(\frac{\pi}{k+1}\right). \quad (19)$$

Therefore, in order to complete the proof for general  $r \geq 0$  it suffices to show the following:

**Proposition 2.5.** *For all  $r \geq 0$  and  $k \geq 3$ ,  $\lim_{n \rightarrow \infty} (h_{r,k}(n))^{1/n}$  exists and its value is independent of  $r$ .*

*Proof.* By Example 2.2, the proposition holds for  $k = 3$ . Thus, from now on we will assume that  $k \geq 4$ . We proceed with the proof by induction on  $r$ . Denote

$$\xi_k = 1 + 2 \cos \left( \frac{\pi}{k+1} \right).$$

Assume now that  $\lim_{n \rightarrow \infty} (h_{r-1,k}(n))^{1/n} = \xi_k$  for some  $r \in \mathbb{N}$ . For  $n \in \mathbb{N}$  and integer  $t \geq 0$ , define

$$\Theta_{t,k}(n) = \{\pi \in [k]^n : \mathcal{S}_k(\pi) = t\}.$$

Let  $w_1 w_2 \cdots w_n$  be any word in the set  $\Theta_{r-1,k}(n)$  and choose a letter  $v \in [k]$  such that  $|v - w_n| > 1$ . Then  $h_{r-1,k}(n) \leq h_{r,k}(n+1)$ , which implies

$$\xi_k = \lim_n (h_{r-1,k}(n))^{1/n} \leq \liminf_n (h_{r,k}(n))^{1/n}.$$

Let now  $w = w_1 w_2 \cdots w_n \in \Theta_{r,k}(n)$  and  $i$  be minimal index such that  $|w_i - w_{i+1}| > 1$ . Then  $w_1 w_2 \cdots w_i \in \Theta_{0,k}(i)$  and  $w_{i+1} w_{i+2} \cdots w_n \in \Theta_{r-1,k}(n-i)$ . Thus,

$$h_{r,k}(n) \leq \sum_{i=0}^n h_{0,k}(i) h_{r-1,k}(n-i). \quad (20)$$

By (19) and the induction hypothesis, for every  $\varepsilon > 0$  there exists a constant  $A_\varepsilon > 0$  such that for all  $n \in \mathbb{N}$  we have

$$h_{0,k}(n) \leq A_\varepsilon (\xi_k + \varepsilon)^n \quad \text{and} \quad h_{r-1,k}(n) \leq A_\varepsilon (\xi_k + \varepsilon)^n,$$

which, by virtue of (20), leads to

$$h_{r,k}(n) \leq n A_\varepsilon^2 (\xi_k + \varepsilon)^n.$$

Since  $\varepsilon > 0$  is an arbitrary constant, this implies that  $\limsup_{n \rightarrow \infty} (h_{r,k}(n))^{1/n} \leq \xi_k$ . Combining the liminf and limsup inequalities we get

$$\xi_k \leq \liminf_n (h_{r,k}(n))^{1/n} \leq \limsup_n (h_{r,k}(n))^{1/n} \leq \xi_k,$$

which implies the claim of the proposition.  $\square$

**Remark 2.6.** *The key limit identity (19) is obtained in [28] by analytical methods. We will now outline a short probabilistic argument leading to this result (note however that Corollary 2.5 in [28] gives more information about the asymptotic behavior of  $h_{0,k}(n)$  than the limit in (19) alone). Consider a discrete-time Markov chain  $(Y_\ell)_{\ell \geq 0}$  on the state space  $[k] \cup \{0\}$  with transition kernel*

$$P_{ij} := P(Y_{\ell+1} = j \mid Y_\ell = i) = \begin{cases} \frac{1}{k} & \text{if } |i - j| \leq 1 \text{ and } i, j \neq 0, \\ \frac{k-2}{k} & \text{if } i \in \{1, k\} \text{ and } j = 0, \\ \frac{k-3}{k} & \text{if } i \notin \{1, k\} \text{ and } j = 0, \\ 1 & \text{if } i = j = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Thus 0 is the absorbing state of the Markov chain, which is otherwise irreducible and aperiodic. We will assume that the initial state  $Y_0$  is chosen uniformly from the set  $[k]$ . In terms of this Markov chain,

$$h_{0,k}(n) = k^n P(\mathcal{S}_{n,k} = 0) = k^n P(T_0 \geq n),$$

where  $T_0 = \inf\{\ell \in \mathbb{N} : Y_\ell = 0\}$  is the first hitting time of the absorption state zero. Thus

$$h_{0,k}(n) = k^n \cdot \frac{1}{k} e_k^T Q^{n-1} e_k = k^{n-1} \cdot e_k^T Q^{n-1} e_k,$$

where  $Q$  is  $k \times k$  matrix which is obtained from  $P$  by removing the row and column corresponding to the absorption state zero,  $e_k$  is a  $k$ -vector with all entries equal to one, and  $\frac{1}{k}$  in the formula stands for the initial distribution. Therefore,  $\lim_{n \rightarrow \infty} (h_{0,k}(n))^{1/n} = \lambda_k$ , where  $\lambda_k$  is the Perron-Frobenius eigenvalue of  $kQ$ . It is well-known that  $\lambda_k = 1 + 2 \cos\left(\frac{\pi}{k+1}\right)$  (see, for instance, [34] or [38, p. 239]). We remark in passing that  $kQ = -\frac{1}{\alpha}(\mathbf{A}_k - I)$ , where  $\mathbf{A}_k$  is the matrix that appears in (12). This observation provides a direct link between the analytic (generating function) and probabilistic approaches to the proof of Theorem 1.5-(i).

We now turn to the proof of Corollary 1.6-(i).

*Proof of Corollary 1.6-(i).* We have

$$\begin{aligned} H_k^{(s)}(n) &= -\sum_{r \geq 0} P_{k,n}^{(s)}(r) \log P_{k,n}^{(s)}(r) = -\frac{1}{k^n} \sum_{r \geq 0} h_{r,k}(n) (\log h_{r,k}(n) - n \log k) \\ &= n \log k - \sum_{r \geq 0} P_{k,n}^{(s)}(r) \log h_{r,k}(n). \end{aligned}$$

Thus,

$$\frac{H_k^{(s)}(n)}{n} = \log k - \sum_{r \geq 0} P_{k,n}^{(s)}(r) \log [h_{r,k}(n)]^{\frac{1}{n}}.$$

An application of the bounded convergence theorem (notice that  $1 \leq h_{r,k}(n) \leq k^n$ ) exploiting the result in Theorem 1.5-(i) finishes the proof.  $\square$

### 3 Longest L-staircase subwords

The section is devoted to the proof of parts (iii) and (iv) of Theorem 1.1.

Let  $L_{n,k}(q)$  be the generating function for the  $k$ -ary words of length  $n$  according to the statistic  $\mathcal{I}_k$ . That is,

$$L_{n,k}(q) = \sum_{\pi \in [k]^n} q^{\mathcal{I}_k(\pi)} = \sum_{r=1}^n g_{r,k}(n) q^r, \quad q \in \mathbb{C}.$$

In addition, let  $L_{n,k;i_1 i_2 \dots i_s}(q)$  be the generating function for the  $k$ -ary words  $\pi = i_1 i_2 \dots i_s \pi'$  of length  $n$  according to the statistic  $\mathcal{I}_k$ , that is

$$L_{n,k;i_1 i_2 \dots i_s}(q) = \sum_{\pi=i_1 i_2 \dots i_s \pi' \in [k]^n} q^{\mathcal{I}_k(\pi)}, \quad q \in \mathbb{C}.$$

Similarly to (10), for suitable  $q, x \in \mathbb{C}$ , let

$$L_k(x, q) = \sum_{n \geq 0} L_{n,k}(q) x^n \quad \text{and} \quad L_{k;i_1 i_2 \dots i_s}(x, q) = \sum_{n \geq 0} L_{n,k;i_1 i_2 \dots i_s}(q) x^n.$$

It is easy to verify that  $L_1(x, q) = \frac{1}{1-xq}$  and  $L_2(x, q) = \frac{1}{1-2xq}$ . Hence, from now on we will assume that  $k \geq 3$ . We first observe that

$$\begin{aligned} L_{k;i}(x, q) &= xq + \sum_{j=1}^k L_{k;i_j}(x, q) \\ &= xq + xq(L_{k;i-1}(x, q) + L_{k;i}(x, q) + L_{k;i+1}(x, q)) + xq \sum_{j \in [k] \setminus \{i-1, i, i+1\}} L_{k;j}(x, 1), \end{aligned}$$

where  $L_{k;0}(x, q)$  and  $L_{k,k+1}(x, q)$  are defined as 0. Since  $L_{k;j}(x, 1) = \frac{x}{1-kx}$ , we obtain that the following holds for all  $i = 1, 2, \dots, k$ :

$$\begin{aligned} L_{k;i}(x, q) &= \alpha'(L_{k;i-1}(x, q) + L_{k;i+1}(x, q)) \\ &\quad + \alpha' + (k-2)\beta'\delta_{i=1,k} + (k-3)\beta'\delta_{2 \leq i \leq k-1}, \end{aligned} \quad (21)$$

where

$$\alpha' = \frac{xq}{1-xq}, \quad \beta' = \frac{x^2q}{(1-xq)(1-kx)},$$

and  $\delta_X$  for a statement  $X$  is defined as 1 whenever  $X$  holds true and 0 otherwise.

Using argument similar to those we employed in the previous section (comapre (11) with (21)), we can write (21) in terms of a symmetric tridiagonal matrix, and then, inverting the matrix with the help of Corollary 4.4 in [17], obtain the following analogue of Lemma 2.1:

$$\begin{aligned} L_k(x, q) &= 1 + \frac{\alpha' + (k-3)\beta'}{\alpha'U_k(t)} \left( \sum_{i=0}^{k-1} U_i(t)U_{k-1-i}(t) + 2 \sum_{i=0}^{k-2} \sum_{j=0}^{k-2-i} U_i(t)U_j(t) \right) + \frac{2\beta'}{\alpha'U_k(t)} \sum_{i=0}^{k-1} U_i(t) \\ &= 1 + \frac{\alpha' + (k-3)\beta'}{2(t-1)^2\alpha'U_k(t)} ((k(t-1)-1)U_k(t) + U_{k-1}(t) + 1) \\ &\quad + \frac{\beta'}{(t-1)\alpha'U_k(t)} (U_k(t) - U_{k-1}(t) - 1). \end{aligned} \quad (22)$$

It follows from (22) that for any  $k \geq 3$ ,

$$\begin{aligned} &\frac{\partial}{\partial q} L_k(x, q) \Big|_{q=1} \\ &= \frac{2x^2}{(1-kx)(1-3x)^2 U_k(\frac{1-x}{2x})} + \frac{2x^2 U_{k-1}(\frac{1-x}{2x})}{(1-kx)(1-3x)^2 U_k(\frac{1-x}{2x})} - \frac{x(3kx-k+2x)}{(1-kx)(1-3x)^2}. \end{aligned}$$

By using the formula  $\frac{d}{dx}U_n(x) = \frac{(k+1)T_{k+1}(x) - xU_k(x)}{x^2 - 1}$ , where  $T_k(x)$  is the  $k$ -th Chebyshev polynomial of the first kind, we obtain:

$$\begin{aligned} \frac{\partial^2}{\partial q^2} L_k(x, q) \Big|_{q=1} &= \frac{-2(9kx - 3k + 6x + 2)x^2}{(1 - kx)(1 - 3x)^3} + \frac{4x^3(3x + 5)}{(1 - 3x)^3(1 - kx)(1 + x)U_k(\frac{1-x}{2x})} \\ &+ \frac{4(1 + 3x)x^2U_{k-1}(\frac{1-x}{2x})}{(1 - kx)(1 - 3x)^2U_k(\frac{1-x}{2x})} + \frac{8(k + 1)x^3(U_{k-1}(\frac{1-x}{2x}) + 1)T_{k+1}(\frac{1-x}{2x})}{(1 - kx)(1 - 3x)^3(1 + x)U_k^2(\frac{1-x}{2x})} \\ &- \frac{8kx^3T_k(\frac{1-x}{2x})}{(1 - kx)(1 - 3x)^3(1 + x)U_k(\frac{1-x}{2x})}. \end{aligned}$$

For  $k \geq 4$ , this implies that  $E(\mathcal{I}_k(n)) \sim e_1$  and  $E(\mathcal{I}_k(n)(\mathcal{I}_k(n) - 1)) \sim e_2$ , where

$$e_1(k) = \frac{2}{(k - 3)^2U_k(\frac{k-1}{2})} + \frac{2U_{k-1}(\frac{k-1}{2})}{(k - 3)^2U_k(\frac{k-1}{2})} + \frac{(k^2 - 3k - 2)}{(k - 3)^2} \quad (23)$$

and

$$\begin{aligned} e_2(k) &= \frac{-2(6 + 11k - 3k^2)}{(k - 3)^3} + \frac{4(3 + 5k)}{(k - 3)^3(k + 1)U_k(\frac{k-1}{2})} + \frac{4(k + 3)U_{k-1}(\frac{k-1}{2})}{(k - 3)^2U_k(\frac{k-1}{2})} \\ &+ \frac{8k(U_{k-1}(\frac{k-1}{2}) + 1)T_{k+1}(\frac{k-1}{2})}{(k - 3)^3U_k^2(\frac{k-1}{2})} - \frac{8k^2T_k(\frac{k-1}{2})}{(k - 3)^3(k + 1)U_k(\frac{k-1}{2})}. \end{aligned} \quad (24)$$

In particular,  $\sigma^2(\mathcal{I}_k(n)) \sim e_2 + e_1 - e_1^2$ . The proof is complete.  $\square$

## 4 Longest L-staircase subsequences

This section is devoted to the proof of the results for longest  $L$ -staircase subsequences. Section 4.1 contains the proof of parts (v) and (vi) of Theorem 1.1, Section 4.2 includes the proof of parts (iii) and (iv) of Theorem 1.2, and the proof part (ii) of Theorem 1.5 is given in Section 4.3.

### 4.1 Proof of Theorem 1.1-(v), (vi)

Let  $L_{n,k}(q)$  be the generating function for the  $k$ -ary words of length  $n$ , according to the statistic  $\mathcal{J}_k$ . That is,

$$L_{n,k}(q) = \sum_{\pi \in [k]^n} q^{\mathcal{J}_k(\pi)} = \sum_{r=1}^n f_{r,k}(n)q^r, \quad q \in \mathbb{C}.$$

In addition, we define  $L_{n,k;l_1l_2 \dots l_s}(q)$  to be the generating function for the  $k$ -ary words  $\pi = l_1l_2 \dots l_s\pi'$  of length  $n$  according to the statistic  $\mathcal{J}_k$ . That is,

$$L_{n,k;l_1l_2 \dots l_s}(q) = \sum_{\pi = l_1l_2 \dots l_s\pi' \in [k]^n} q^{\mathcal{J}_k(\pi)}, \quad q \in \mathbb{C}.$$



Next, similarly to (10), for suitable  $q, x \in \mathbb{C}$ , we set

$$L_k(x, q) = \sum_{n \geq 0} L_{n,k}(q)x^n \quad \text{and} \quad L_{k;l_1 l_2 \dots l_s}(x, q) = \sum_{n \geq 0} L_{n,k;l_1 l_2 \dots l_s}(q)x^n. \quad (25)$$

It is straightforward to verify that  $L_1(x, q) = \frac{1}{1-xq}$  and  $L_2(x, q) = \frac{1}{1-2xq}$ . Hence, in what follows we will assume that  $k \geq 3$ . By (25), we have

$$\begin{aligned} L_{k;i}(x, q) &= xq + xq(L_{k;i-1}(x, q) + L_{k;i}(x, q) + L_{k;i+1}(x, q)) + x \sum_{j \in [k] \setminus \{i-1, i, i+1\}} L_{k;i}(x, q) \\ &= xq + xq(L_{k;i-1}(x, q) + L_{k;i}(x, q) + L_{k;i+1}(x, q)) + (k-3)xL_{k;i}(x, q), \end{aligned}$$

for  $i = 2, \dots, k-1$ , and

$$\begin{aligned} L_{k;1}(x, q) &= xq + xq(L_{k;1}(x, q) + L_{k;2}(x, q)) + x \sum_{j \in [k] \setminus \{1, 2\}} L_{k;1}(x, q) \\ &= xq + xq(L_{k;1}(x, q) + L_{k;2}(x, q)) + (k-2)xL_{k;1}(x, q), \\ L_{k;k}(x, q) &= xq + xq(L_{k;k}(x, q) + L_{k;k-1}(x, q)) + x \sum_{j \in [k] \setminus \{k-1, k\}} L_{k;k}(x, q) \\ &= xq + xq(L_{k;k}(x, q) + L_{k;k-1}(x, q)) + (k-2)xL_{k;k}(x, q). \end{aligned}$$

Let

$$\alpha = \frac{xq}{1 - (k-2+q)x} \quad \text{and} \quad \beta = \frac{xq}{1 - (k-3+q)x}.$$

Suppose that  $xq \neq 0$ , so that  $\alpha\beta \neq 0$ . We define a tridiagonal matrix  $\mathbf{B}_k = (a_{ij})_{1 \leq i, j \leq k}$  by setting

$$b_{ij} = \begin{cases} 1/\alpha & \text{if } i = j = 1 \text{ or } i = j = k, \\ 1/\beta & \text{if } i = j \text{ and } i \notin \{1, k\}, \\ -1 & \text{if } |i - j| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In the matrix notation, we have

$$\mathbf{B}_k \begin{pmatrix} L_{k;1}(x, q) \\ \vdots \\ L_{k;k}(x, q) \end{pmatrix} = e_k, \quad (26)$$

where  $e_k \in \mathbb{R}^k$  is a  $k$ -vector with all entries equal to zero. Similarly to (13), we obtain that

$$L_k(x, q) = 1 + \sum_{i=1}^k L_{k,i}(x, q) = 1 + e_k^T \mathbf{B}_k^{-1} e_k.$$

It follows from Corollaries 4.2-4.4 in [17] (apply first Corollary 4.4 in [17] to the  $(k-2) \times (k-2)$  matrix obtained by deleting the first and the last rows and columns in  $\mathbf{B}$ , alternatively we could use Theorem 2 in [40]) that

$$L_k(x, q) = \frac{\beta \left( k - 2 - \frac{2\beta \left( 1 - \frac{1+U_{k-3}(1/2\beta)}{U_{k-2}(1/2\beta)} \right)}{1-2\beta} \right)}{1-2\beta} + 2\alpha \left( 1 - \frac{U_{k-3}(1/2\beta)}{U_{k-2}(1/2\beta)} \right). \quad (27)$$

Differentiating  $L_k(x, q)$  with respect to  $q$  at  $q = 1$  we obtain that

$$\left. \frac{\partial}{\partial q} L_k(x, q) \right|_{q=1} = \frac{x(k - (k-2)(k-1)x)}{(1-kx)^2},$$

which leads to  $E(\mathcal{J}_k(n)) = 1 + \frac{3k-2}{k^2}(n-1)$ . Also, by differentiating  $L_k(x, q)$  twice with respect to  $q$  and then substituting  $q = 1$ , we obtain that there exist two constants  $a_k$  and  $b_k$  and a function  $\tilde{L}_k(x)$  analytic at  $x = 1/k$  such that

$$\left. \frac{\partial^2}{\partial q^2} L_k(x, q) \right|_{q=1} = \frac{2(3k-2)^2}{k^4(1-kx)^3} + \frac{a_k}{(1-kx)^2} + \frac{b_k}{1-kx} + \tilde{L}_k(x),$$

where

$$\begin{aligned} a_k &= \lim_{x \rightarrow 1/k} \left( \left. \frac{\partial^2}{\partial q^2} L_k(x, q) \right|_{q=1} - \frac{2(3k-2)^2}{k^4(1-kx)^3} \right) (1-kx)^2, \\ b_k &= \lim_{x \rightarrow 1/k} \left( \left. \frac{\partial^2}{\partial q^2} L_k(x, q) \right|_{q=1} - \frac{2(3k-2)^2}{k^4(1-kx)^3} - \frac{a_k}{(1-kx)^2} \right) (1-kx). \end{aligned} \quad (28)$$

Thus, by using the method of residuals [16] to calculate the coefficient in front of  $x^2$ , we obtain that

$$E(\mathcal{J}_k(n)(\mathcal{J}_k(n) - 1)) = \frac{(3k-2)^2(n+1)(n+2)}{k^4} + a_k(n+1) + O(1),$$

and

$$\begin{aligned} \sigma^2(\mathcal{J}_k(n)) &= E(\mathcal{J}_k(n)(\mathcal{J}_k(n) - 1)) + E(\mathcal{J}_k(n)) - E[(\mathcal{J}_k(n))]^2 \\ &= \left( a_k - \frac{(3k-2)(k^2 - 15k + 10)}{k^4} \right) n + O(1). \end{aligned}$$

The proof is complete.  $\square$

## 4.2 Proof of Theorem 1.2-(iii),(iv)

Recall (1). Without loss of generality we can assume that  $\mathcal{J}_k(n) = \mathcal{J}_k(W_n)$ . Let  $T(1) = 1$  and, recursively, for  $\ell \in \mathbb{N}$ ,

$$T(\ell + 1) = \inf \{ t \in \mathbb{N} : |w_t - w_{T(\ell)}| \leq 1 \}.$$

Denote  $X_\ell = w_{T(\ell)}$ . Then, for all  $n \in \mathbb{N}$ ,  $X_1 \cdots X_{\mathcal{J}_k(n)}$  is the longest  $L$ -staircase subsequence of  $W_n$ , and  $\mathcal{J}_k(n)$  is the unique integer such that

$$T(\mathcal{J}_k(n)) \leq n < T(\mathcal{J}_k(n) + 1). \quad (29)$$

Let  $T(0) = 0$  and  $\tau_n = T_n - T_{n-1}$  for  $n \in \mathbb{N}$ . Then  $(X_n)_{n \in \mathbb{N}}$  is a reflected nearest-neighbor symmetric random walk on  $[k]$ , and  $(X_n, \tau_n)_{n \in \mathbb{N}}$  is a Markov chain with transition kernel

$$P(X_{n+1} = y, \tau_n = t | X_n = x) = \begin{cases} \frac{1}{k} \left( \frac{k-3}{k} \right)^{t-1} & \text{if } t \in \mathbb{N}, x \in \{2, \dots, k-1\}, |y-x| \leq 1, \\ \frac{1}{k} \left( \frac{k-2}{k} \right)^{t-1} & \text{if } t \in \mathbb{N}, x \in \{1, k\}, |y-x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The rest of the proof is similar to the one given in Section 2.2: the Markov renewal process  $(X_n, T(n))_{n \in \mathbb{N}}$  plays the role similar to that of  $(u_n, \Lambda_n)_{n \in \mathbb{N}}$ , we first proof a CLT for the additive process  $T(n)$ , and then transform it into a CLT for  $\mathcal{J}_k(n)$  using (29). In particular, similarly to (18),

$$\varphi_k = \lim_{n \rightarrow \infty} \frac{\mathcal{J}_k(n)}{n} = \lim_{n \rightarrow \infty} \frac{E(\mathcal{J}_k(n))}{n} = \frac{3k-2}{k^2},$$

where in the last step we used the result in Theorem 1.1-(iv).  $\square$

### 4.3 Proof of Theorem 1.5-(ii)

Using (27), it is not hard to check that the function  $L_k(x, 0) = \sum_{n=1}^{\infty} f_{r,k}(n)x^n$  is meromorphic and that  $1/(k-2)$  is its smallest by absolute value pole. The result in Theorem 1.5-(ii) follows from this observation [16].

Alternatively, one can use the following simple counting argument to prove the result. On one hand,

$$f_{r,k}(n) \leq \binom{n+r-1}{r-1} \cdot h_{0,k}(r) \cdot (k-2)^{n-r}.$$

The first term on the right-hand side of the inequality is the number of ways to locate a given  $L$ -staircase within a word of length  $n$ , the second is for the number of staircase words of length  $r$ , and the third one is an upper bound for the number of ways to fill the remaining  $n-r$  positions. On the other hand, considering exclusively  $L$ -staircase subsequences that contain letter 1 only, we obtain that

$$f_{r,k}(n) \geq \binom{n+r-1}{r-1} \cdot (k-2)^{n-r}.$$

The comparison of the lower and upper bounds concludes the proof.  $\square$

## Appendix. Proof of Proposition 1.4

(i) Recall the random walk  $(Y_n)_{n \in \mathbb{N}}$ , the absorption time  $T_0$ , and the sub-stochastic transition matrix  $Q$  introduced in Remark 2.6. By the monotone convergence theorem we have:

$$\lim_{n \rightarrow \infty} E(\xi_{w,k}(n)) = \sum_{m=1}^{\infty} P(Y_m = w, T_0 > m) = \sum_{m=0}^{\infty} \sum_{j=1}^k \frac{1}{k} Q^m(j, w) = \frac{1}{k} e(I - Q)^{-1}(w),$$

where  $e$  is a  $k$ -vector with all entries equal to one. After a little algebra, the result then follows from Corollary 4.3 in [17] (see also Theorem 2 in [40]).  $\square$

(ii) Recall the Markov chain  $(X_n)_{n \in \mathbb{N}}$  from Section 4.2. Let  $(P_{ij})_{i,j \in [k]}$  denote transition kernel and  $\mu = (\mu_1, \dots, \mu_k)$  be the unique stationary distribution of the Markov chain. That is,  $\mu_i \geq 0$  for all  $i \in [k]$ ,  $\sum_{i \in [k]} \mu_i = 1$ , and  $\mu_j = \sum_{i \in [k]} \mu_i P_{ij}$  for all  $j \in [k]$ . It is easy to verify that

$$\mu_1 = \mu_k = \frac{2}{3k-2} \quad \text{and} \quad \mu_2 = \dots = \mu_{k-1} = \frac{3}{3k-2}.$$

Since the Markov chain  $X_n$  is irreducible and aperiodic, the law of large numbers (ergodic theorem) implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=w\}} = \mu_w \quad \forall w \in [k], \quad (30)$$

with probability one, regardless of the initial distribution of the Markov chain. Here  $\mathbf{1}_A$  denotes the indicator of the event  $A$ , that is its value is one if the event occurs and is zero otherwise. To complete the proof, we note that  $\eta_{w,k}(n) = \sum_{i=1}^{\mathcal{J}_k(n)} \mathbf{1}_{\{X_i=w\}}$  and pass in (30) from the regular limit to the limit along the (random, but diverging with probability one) subsequence  $\mathcal{J}_k(n)$ ,  $n \in \mathbb{N}$ .  $\square$

## References

- [1] S. Alili, *Asymptotic behaviour for random walks in random environments*, J. Appl. Prob. **36** (1999), 334–349.
- [2] G. Alsmeyer and A. Gut, *Limit theorems for stopped functionals of Markov renewal processes*, Ann. Inst. Statist. Math. **51** (1999), 369–382.
- [3] V. S. Barbu and N. Limnios, *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications. Their Use in Reliability and DNA Analysis*, Lecture Notes in Statistics, Vol. 191, Springer, 2008.
- [4] M. Bóna, *Combinatorics of Permutations*, Chapman & Hall/CRC, Boca Raton, 2004.

- [5] J. Bucklew, *The blind simulation problem and regenerative processes*, IEEE Trans. Inform. Theory **44** (1998), 2877–2891.
- [6] R. G. Buschman, *Fibonacci numbers, Chebyshev polynomials, generalizations and difference equations*, Fibonacci Quart. **1** (1963), 1–7.
- [7] A. Carmona, A. M. Encinas, and M. Mitjana, *Discrete elliptic operators and their Green operators*, Linear Algebra Appl. **442** (2014), 115–134.
- [8] X. Chen, *Limit Theorems for Functionals of Ergodic Markov Chains with General State Space*, Mem. Amer. Math. Soc., Vol. 139, 1999.
- [9] E. Çinlar, *Markov renewal theory: a survey*, Management Sci. **21** (1974/75), 727–752.
- [10] P. Coolen-Schrijner and E. A. van Doorn, *Analysis of random walks using orthogonal polynomials*, J. Comput. Appl. Math. **99** (1998), 387–399.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (2nd ed.), Wiley-Interscience, 2006.
- [12] H. Dette, *First return probabilities of birth and death chains and associated orthogonal polynomials*, Proc. Amer. Math. Soc. **129** (2001), 1805–1815.
- [13] A. M. Encinas and M. J. Jiménez, *Explicit inverse of a tridiagonal  $(p, r)$ -Toeplitz matrix*, Linear Algebra Appl. **542** (2018), 402–421.
- [14] A. M. Encinas and M. J. Jiménez, *Explicit inverse of nonsingular Jacobi matrices*, Discrete Appl. Math. **263** (2019), 130–139.
- [15] A. M. Encinas and M. J. Jiménez, *Boundary value problems for second order linear difference equations: application to the computation of the inverse of generalized Jacobi matrices*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM, in press.
- [16] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2008.
- [17] C. M. da Fonseca and J. Petronilho, *Explicit inverses of some tridiagonal matrices*, Linear Algebra Appl. **325** (2001), 7–21.
- [18] F. A. Grünbaum, *Random walks and orthogonal polynomials: some challenges*, In M. Pinsky and B. Birnir (Eds.), *Probability, Geometry and Integrable systems*, Math. Sci. Res. Inst. Publ., Vol. 55, Cambridge Univ. Press, 2008, 241–260.
- [19] O. Häggström and J. S. Rosenthal, *On variance conditions for Markov chain CLTs*, Electron. Comm. Probab. **12** (2007), 454–464.
- [20] S. Heubach and T. Mansour, *Combinatorics of Compositions and Words*, Chapman and Hall/CRC, 2009.

- [21] C. Houdré and R. Restrepo, *A probabilistic approach to the asymptotics of the length of the longest alternating subsequence*, Electron. J. Combin. **17** (2010), paper no. 168.
- [22] G. Y. Hu and R. F. O’Connell, *Analytical inversion of symmetric tridiagonal matrices*, J. Phys. A **29** (1996), 1511–1513.
- [23] G. L. Jones, *On the Markov chain central limit theorem*, Probab. Surveys **1** (2004), 299–320.
- [24] S. Karlin and J. McGregor, *The classification of birth and death processes*, Trans. Amer. Math. Soc. **86** (1957), 366–400.
- [25] S. Kitaev, *Patterns in Permutations and Words*, Springer, 2011.
- [26] S. Kitaev and J. Remmel, *(a, b)-rectangle patterns in permutations and words*, Discrete Appl. Math. **186** (2015), 128–146.
- [27] A. Knopfmacher, T. Mansour, and A. Munagi, *Smooth compositions and smooth words*, Pure Math. Appl. (P.U.M.A.) **11** (2011), 209–226.
- [28] A. Knopfmacher, T. Mansour, A. Munagi, and H. Prodinger, *Staircase words and Chebyshev polynomials*, Appl. Anal. Discrete Math. **4** (2010), 81–95.
- [29] W. Ledermann and G. E. H. Reuter, *Spectral theory for the differential equations of simple birth and death processes*, Philos. Trans. Roy. Soc. London. Ser. A. **246** (1954), 321–369.
- [30] T. Mansour, *Smooth partitions and Chebyshev polynomials*, Bull. Lond. Math. Soc. **41** (2010), 961–970.
- [31] T. Mansour, R. Rastegar, and A. Roitershtein, *On ballistic deposition process on a strip*, Journal of Statistical Physics **177(4)** (2019), 626–650.
- [32] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd. ed., Cambridge University Press, 2009.
- [33] A. Nalli and H. Cıvcıv, *A generalization of tridiagonal matrix determinants, Fibonacci and Lucas numbers*, Chaos Solitons Fractals **40** (2009), 355–361.
- [34] P. J. Nash, *Chebyshev polynomials and quadratic path integrals*, J. Math. Phys. **27** (1986), 2963.
- [35] F. Qi, V. Čerňanová, and Y. S. Semenov, *Some tridiagonal determinants related to central Delannoy numbers, the Chebyshev polynomials, and the Fibonacci polynomials*, Politehn. Univ. Bucharest Sci. Bull. Ser. A Appl. Math. Phys. **81** (2019), 123–136.
- [36] T. J. Rivlin, *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*, Wiley-Interscience, 1990.

- [37] J. F. Silva, *Shannon entropy estimation in  $\infty$ -alphabets from convergence results: studying plug-in estimators*, Entropy **20** (2018), paper no. 397.
- [38] F. Spitzer, *Principles of Random Walk*, 2nd. ed., Graduate Texts in Mathematics, Vol. 34, Springer, 2001.
- [39] D. J. Spitzner and T. R. Boucher, *Asymptotic variance of functionals of discrete time Markov chains via the Drazin inverse*, Electron. Comm. Probab. **12** (2007), 120–133.
- [40] L. S. L. Tan, *Explicit inverse of tridiagonal matrix with applications in autoregressive modeling*, IMA J. Appl. Math., in press.
- [41] O. Zeitouni, *Random Walks in Random Environment*, XXXI Summer School in Probability, (St. Flour, 2001). Lecture Notes in Math., Vol. 1837, Springer, 2004, pp. 193–312.