

# A probabilistic approach to the asymptotics of the length of the longest alternating subsequence

Christian Houdré <sup>\*</sup>      Ricardo Restrepo <sup>†‡</sup>

Submitted: May 10, 2010; Accepted: Nov 22, 2010; Published: Dec 10, 2010

Mathematics Subject Classification: 60C05, 60F05 60G15, 60G17, 05A16

## Abstract

Let  $LA_n(\tau)$  be the length of the longest alternating subsequence of a uniform random permutation  $\tau \in [n]$ . Classical probabilistic arguments are used to rederive the asymptotic mean, variance and limiting law of  $LA_n(\tau)$ . Our methodology is robust enough to tackle similar problems for finite alphabet random words or even Markovian sequences in which case our results are mainly original. A sketch of how some cases of pattern restricted permutations can also be tackled with probabilistic methods is finally presented.

*Keywords:* Longest alternating subsequence, random permutations, random words, m-dependence, central limit theorem, law of the iterated logarithm.

## 1 Introduction

Let  $a := (a_1, a_2, \dots, a_n)$  be a sequence of length  $n$  whose elements belong to a totally ordered set  $\Lambda$ . Given an increasing set of indices  $\{\ell_i\}_{i=1}^m$ , we say that the subsequence  $(a_{\ell_1}, a_{\ell_2}, \dots, a_{\ell_m})$  is alternating if  $a_{\ell_1} > a_{\ell_2} < a_{\ell_3} > \dots > a_{\ell_m}$ . The *length of the longest alternating subsequence* is then defined as

$$LA_n(a) := \max \{m : a \text{ has an alternating subsequence of length } m\}.$$

We revisit, here, the problem of finding the asymptotic behavior (in mean, variance and limiting law) of the length of the longest alternating subsequence in the context of random permutations and random words. For random permutations, these problems have seen complete solutions with contributions independently given (in alphabetical order) by

---

<sup>\*</sup>Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia, 30332, USA, houdre@math.gatech.edu. Supported in part by the NSA grant H98230-09-1-0017.

<sup>†</sup>Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia, 30332, USA, restrepo@math.gatech.edu.

<sup>‡</sup>Universidad de Antioquia, Departamento de Matemáticas, Medellin, Colombia.

Pemantle, Stanley and Widom. The reader will find in [18] a comprehensive survey, with precise bibliography and credits, on these and related problems. In the context of random words, Mansour [12] contains very recent contributions where mean and variance are obtained. Let us just say that, to date, the proofs developed to solve these problems are of a combinatorial or analytic nature and that we wish below to provide probabilistic ones. Our approach is developed via iid sequences uniformly distributed on  $[0, 1]$ , counting minima and maxima and the central limit theorem for 2-dependent random variables. Not only does our approach recover the permutation case, but it works as well for random words,  $a \in \mathcal{A}^n$  where  $\mathcal{A}$  is a finite ordered alphabet, recovering known results and providing new ones. Properly modified it also works for several kinds of pattern restricted subsequences. Finally, similar results are also obtained for words generated by a Markov sequence.

## 2 Random permutations

The asymptotic behavior of the length of the longest alternating subsequence has been studied by several authors, including Pemantle [18, page 684], Stanley [17] and Widom [20], who by a mixture of generating function methods and saddle point techniques get the following result:

**Theorem 2.1** *Let  $\tau$ , be a uniform random permutation in the symmetric group  $\mathcal{S}_n$ , and let  $\text{LA}_n(\tau)$  be the length of the longest alternating subsequence of  $\tau$ . Then,*

$$\begin{aligned} \mathbf{E} \text{LA}_n(\tau) &= \frac{2n}{3} + \frac{1}{6}, & n \geq 2 \\ \text{Var} \text{LA}_n(\tau) &= \frac{8n}{45} - \frac{13}{180}, & n \geq 4. \end{aligned}$$

Moreover, as  $n \rightarrow \infty$ ,

$$\frac{\text{LA}_n(\tau) - 2n/3}{\sqrt{8n/45}} \implies \mathcal{Z},$$

where  $\mathcal{Z}$  is a standard normal random variable and where  $\implies$  denotes convergence in distribution.

The present section is devoted to give a simple probabilistic proof of the above result. To provide such a proof we make use of a well known correspondence which transform the problem into that of counting the maxima of a sequence of iid random variables uniformly distributed on  $[0, 1]$ . In order to establish the weak limit result, a central limit theorem for  $m$ -dependent random variables is then briefly recalled.

Let us start by recalling some well known facts (Durrett [4, Chapter 1], Resnick [14, Chapter 4]). For each  $n \geq 1$  (including  $n = \infty$ ), let  $\mu_n$  be the uniform measure on  $[0, 1]^n$  and, for each  $n \geq 1$ , let the function  $T_n : [0, 1]^n \rightarrow \mathcal{S}_n$  be defined

by  $T_n(a_1, a_2, \dots, a_n) = \tau^{-1}$ , where  $\tau$  is the unique permutation  $\tau \in \mathcal{S}_n$  that satisfies  $a_{\tau_1} < a_{\tau_2} < \dots < a_{\tau_n}$ . Note that  $T_n$  is defined for all  $a \in [0, 1]^n$  except for those for which  $a_i = a_j$  for some  $i \neq j$ , and this set has  $\mu_n$ -measure zero. A well known fact, sometimes attributed to Rényi [14], asserts that the pushforward measure  $T_n\mu_n$ , i.e., the image of  $\mu_n$  by  $T_n$ , corresponds to the uniform measure on  $\mathcal{S}_n$ , which we denote by  $\nu_n$ . The importance of this fact relies in the observation that the map  $T_n$  is order preserving, that is,  $a_i < a_j$  if and only if  $(T_n a)_i < (T_n a)_j$ . This implies that any event in  $\mathcal{S}_n$  has a canonical representative in  $[0, 1]^n$  in terms of the order relation of its components. Explicitly, if we consider the language  $L$  of the formulas with no quantifiers, one variable, say  $x$ , and with atoms of the form  $x_i < x_j$ ,  $i, j \in [n]$ , then any event of the form  $\{x : \varphi(x)\}$  where  $\varphi \in L$ , has the same probability in  $[0, 1]^n$  and in  $\mathcal{S}_n$  under the uniform measure. To give some examples, events like  $\{x : x \text{ has an increasing subsequence of length } k\}$ ,  $\{x : x \text{ avoids the permutation } \sigma\}$ ,  $\{x : x \text{ has an alternating subsequence of length } k\}$  have the same probability in  $[0, 1]^n$  and  $\mathcal{S}_n$ . In particular, it should be clear that

$$\text{LA}_n(\boldsymbol{\tau}) \stackrel{d}{=} \text{LA}_n(\mathbf{a}), \tag{1}$$

where  $\boldsymbol{\tau}$  is a uniform random permutation in  $\mathcal{S}_n$ ,  $\mathbf{a}$  is a uniform random sequence in  $[0, 1]^n$  and where  $d$  means equality in distribution.

*Maxima and minima.* Next, we say that the sequence  $a = (a_1, a_2, \dots, a_n)$  has a local maximum at the index  $k$  if (i)  $a_k > a_{k+1}$  or  $k = n$ , and (ii)  $a_k > a_{k-1}$  or  $k = 1$ . Similarly, we say that  $a$  has a local minimum at the index  $k$  if (i)  $a_k < a_{k+1}$  or  $k = n$ , and (ii)  $a_k < a_{k-1}$ . An observation that comes in handy is the fact that counting the length of the longest alternating subsequence is equivalent to counting maxima and minima of the sequence (starting with a local minimum). This is attributed to Bóna in Stanley [18]; for completeness, we prove it next.

**Proposition 2.2** *For  $\mu_n$ -almost all sequences  $a = (a_1, a_2, \dots, a_n) \in [0, 1]^n$ ,*

$$\text{LA}_n(a) = \# \text{ local maxima of } a + \# \text{ local minima of } a \tag{2}$$

$$= \mathbf{1}(a_n > a_{n-1}) + 2 \mathbf{1}(a_1 > a_2) + 2 \sum_{k=2}^{n-1} \mathbf{1}(a_{k-1} < a_k > a_{k+1}). \tag{3}$$

**Proof.** For  $\mu_n$ -almost all  $a \in [0, 1]^n$ ,  $a_i \neq a_j$  whenever  $i \neq j$ , therefore we can assume that  $a$  has no repeated components. Let  $t_1, \dots, t_r$  be the positions, in increasing order, of the local maxima of the sequence  $a$ , and let  $s_1, \dots, s_{r'}$  be the positions, in increasing order, of the local minima of  $a$ , not including the local minima before the position  $t_1$ . Notice that the maxima and minima are alternating, that is,  $t_i < s_i < t_{i+1}$  for every  $i$ , implying that  $r' = r$  or  $r' = r - 1$ . Also notice, that in case  $r' = r - 1$ , necessarily  $t_r = n$ . Therefore, since  $(a_{t_1}, a_{s_1}, a_{t_2}, a_{s_2}, \dots)$  is an alternating subsequence of  $a$ , we have  $\text{LA}_n(a) \geq r + r' = \# \text{ local maxima} + \# \text{ local minima}$ .

To establish the opposite inequality, take a maximal sequence of indices  $\{\ell_i\}_{i=1}^m$  such that  $(a_{\ell_i})_{i=1}^m$  is alternating. Move every odd index upward, following the gradient of  $a$

(the direction, left or right, in which the sequence  $a$  increases), till it reaches a local maximum of  $a$ . Next, move every even index downward, following the gradient of  $a$  (the direction, left or right, in which the sequence  $a$  decreases), till it reaches a local minimum of  $a$ . Notice, importantly, that this sequence of motions preserves the order relation between the indices, therefore the resulting sequence of indices  $\{\ell'_i\}_{i=1}^m$  is still increasing and, in addition, it is a subsequence of  $(t_1, s_1, t_2, s_2, \dots)$ . Now, since the sequence  $(a_{\ell'_i})_{i=1}^m$  is alternating, it follows that  $LA_n(a) \leq \# \text{ local maxima} + \# \text{ local minima}$ . Finally, associating every local maxima not in the  $n$ -th position with the closest local minima to its right, we obtain a one to one correspondence, which leads to (3).  $\square$

*Mean and variance.* The above correspondence allows us to easily compute the mean and the variance of the length of the longest alternating subsequence by going ‘back and forth’ between  $[0, 1]^n$  and  $\mathcal{S}_n$ . For instance, given a random uniform sequence  $\mathbf{a} = (a_1, \dots, a_n) \in [0, 1]^n$ , let  $M_k := \mathbf{1}(\mathbf{a} \text{ has a local maximum at the index } k)$ ,  $k \in \{2, \dots, n-1\}$ . Then

$$\mathbf{E}M_k = \mu_n(a_{k-1} < a_k > a_{k+1}) = \mu_3(a_1 < a_2 > a_3) = \nu_3(\tau_1 < \tau_2 > \tau_3),$$

where again,  $\nu_n$  is the uniform measure on  $\mathcal{S}_n$ ,  $n \geq 1$ . The event,  $\{\tau_1 < \tau_2 > \tau_3\}$  corresponds to the permutations  $\{132, 231\}$ , which shows that  $\mathbf{E}M_k = 1/3$ .

Similarly,

$$\mathbf{E}M_1 = \nu_2(\tau_1 > \tau_2) = 1/2 \text{ and } \mathbf{E}M_n = \nu_2(\tau_1 < \tau_2) = 1/2.$$

Plugging these values into (3), we get that

$$\mathbf{E}LA_n(\tau) = \frac{2n}{3} + \frac{1}{6}.$$

To compute the variance of  $LA_n(\tau)$ , first note that  $\text{Cov}(M_k, M_{k+r}) = 0$  whenever  $r \geq 3$ , and that  $\mathbf{E}[M_k M_{k+1}] = 0$ . Now, going again back and forth between  $[0, 1]^n$  and  $\mathcal{S}_n$ , we also obtain

$$\mathbf{E}[M_k M_{k+2}] = \nu_5(\tau_1 < \tau_2 > \tau_3 < \tau_4 > \tau_5) = 2/15,$$

$$\mathbf{E}[M_1 M_3] = \nu_4(\tau_1 > \tau_2 < \tau_3 > \tau_4) = 1/6$$

and

$$\mathbf{E}[M_{n-2} M_n] = \nu_4(\tau_1 < \tau_2 > \tau_3 < \tau_4) = 1/6.$$

This implies from Proposition 2.2 and (1), that

$$\text{Var} LA_n(\tau) = \frac{8n}{45} - \frac{13}{180}.$$

*Asymptotic normality.* Recall that collection of random variables  $\{X_i\}_{i=1}^\infty$  is called  $m$ -dependent if  $X_{t+m+1}$  is independent of  $\{X_i\}_{i=1}^t$  for every  $t \geq 1$ . For such sequences

the strong law of large numbers extends in a straightforward manner just partitioning the summand in appropriate sums of independent random variables, but the extension of the central limit theorem to this context is less trivial (although a ‘small block’ - ‘big block’ argument will do the job). For this purpose recall also the following particular case of a theorem due to Hoeffding and Robbins [7] (which can be also found in standard texts such as Durrett [4, Chapter 7] or Resnick [14, Chapter 8]).

**Theorem 2.3** *Let  $(X_i)_{i \geq 1}$  be a sequence of identical distributed  $m$ -dependent bounded random variables. Then*

$$\frac{X_1 + \cdots + X_n - n\mathbf{E}X_1}{\gamma\sqrt{n}} \implies \mathcal{Z},$$

where  $\mathcal{Z}$  is a standard normal random variable, and where the variance term is given by

$$\gamma^2 = \text{Var } X_1 + 2 \sum_{t=2}^{m+1} \text{Cov}(X_1, X_t).$$

Now, let  $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots)$  be a sequence of iid random variables uniformly distributed in  $[0, 1]$ , and let  $\mathbf{a}^{(n)} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$  be the restriction of the sequence  $\mathbf{a}$  to the first  $n$  indices. Recalling (1) and Proposition 2.2, it is clear that if  $\tau$  is a uniform random permutation in  $\mathcal{S}_n$ ,

$$\text{LA}_n(\tau) \stackrel{d}{=} \mathbf{1}[\mathbf{a}_n > \mathbf{a}_{n-1}] + 2\mathbf{1}[\mathbf{a}_1 > \mathbf{a}_2] + 2 \sum_{k=2}^{n-1} \mathbf{1}[\mathbf{a}_{k-1} < \mathbf{a}_k > \mathbf{a}_{k+1}], \quad (4)$$

where  $\stackrel{d}{=}$  denotes equality in distribution. Therefore, since the random variables  $\{\mathbf{1}[\mathbf{a}_{k-1} < \mathbf{a}_k > \mathbf{a}_{k+1}] : k \geq 2\}$  are identically distributed and 2-dependent, we have by the strong law of large numbers that with probability one

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=2}^{n-1} \mathbf{1}[\mathbf{a}_{k-1} < \mathbf{a}_k > \mathbf{a}_{k+1}] = \mu_3(a_1 < a_2 > a_3) = \frac{1}{3}.$$

Therefore, from (4) we get that, in probability,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{LA}_n(\tau) = \frac{2}{3}.$$

Finally, applying the above central limit theorem, we have as  $n \rightarrow \infty$

$$\frac{\text{LA}_n(\tau) - 2n/3}{\sqrt{n}\gamma} \implies N(0, 1), \quad (5)$$

where in our case, the variance term is given by

$$\begin{aligned} \gamma^2 &= \text{Var}(2\mathbf{1}[\mathbf{a}_1 < \mathbf{a}_2 > \mathbf{a}_3]) + 2 \text{Cov}(2\mathbf{1}[\mathbf{a}_1 < \mathbf{a}_2 > \mathbf{a}_3], 2\mathbf{1}[\mathbf{a}_2 < \mathbf{a}_3 > \mathbf{a}_4]) \\ &\quad + 2 \text{Cov}(2\mathbf{1}[\mathbf{a}_1 < \mathbf{a}_2 > \mathbf{a}_3], 2\mathbf{1}[\mathbf{a}_3 < \mathbf{a}_4 > \mathbf{a}_5]) \\ &= \frac{8}{45}, \end{aligned}$$

from the computations carried out in the previous paragraph.

**Remark 2.4** The above approach via  $m$ -dependence has another advantage, it provides using standard  $m$ -dependent probabilistic statements various types of results on  $LA_n(\tau)$  such as, for example, the exact fluctuation theory via the law of iterated logarithm. In our setting, it gives:

$$\limsup_{n \rightarrow \infty} \frac{LA_n(\tau) - \mathbf{E} LA_n(\tau)}{\sqrt{n \log \log n}} = \frac{4}{3\sqrt{5}},$$

$$\liminf_{n \rightarrow \infty} \frac{LA_n(\tau) - \mathbf{E} LA_n(\tau)}{\sqrt{n \log \log n}} = -\frac{4}{3\sqrt{5}}.$$

Besides the LIL, other types of probabilistic statements on  $LA_n(\tau)$  are possible, e.g., local limit theorems [15], large deviations [8], exponential inequalities [1], etc. This types of statements are also true in the settings of our next sections.

### 3 Finite alphabet random words

Consider a (finite) random sequence  $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  with distribution  $\mu^{(n)}$ , where  $\mu$  is a probability measure supported on a finite set  $[q] = \{1, \dots, q\}$ . Our goal now is to study the length of the longest alternating subsequence of the random sequence  $\mathbf{a}$ . This new situation differs from the previous one mainly in that the sequence can have repeated values. Thus, in order to check if a point is a maximum or a minimum, it is not enough to ‘look at’ its nearest neighbors, losing the advantage of the 2-dependence that we had in the previous case. However, Instead, we can use the stationarity of the property ‘being a local maximum’ with respect to some extended sequence to study the asymptotic behaviour of  $LA_n(\mathbf{a})$ . As a matter of notation, we will use generically, the expression  $LA_n(\mu)$  for the distribution of the length of the longest alternating subsequence of a sequence  $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  having the product distribution  $\mu^{(n)}$ .

In this section we proceed more or less along the lines of the previous section, relating the counting of maxima to the length of the longest alternating subsequence and then, through mixing and ergodicity, obtain results on the asymptotic mean, variance, convergence of averages and asymptotic normality of the longest alternating subsequence. These results are presented in Theorem 3.1 (convergence in probability), and Theorem 3.6 (asymptotic normality).

*Counting maxima and minima.* Given a sequence  $a = (a_1, a_2, \dots, a_n) \in [q]^n$ , we say that  $a$  has a local maximum at the index  $k$ , if (i)  $a_k > a_{k+1}$  or  $k = n$ , and if (ii) for some  $j < k$ ,  $a_j < a_{j+1} = \dots = a_{k-1} = a_k$  or for all  $j < k$ ,  $a_j = a_k$ . Likewise, we say that  $a$  has a local minimum at the index  $k$ , if (i)  $a_k < a_{k+1}$  or  $k = n$ , and if (ii) for some  $j < k$ ,  $a_j > a_{j+1} = \dots = a_{k-1} = a_k$ . The identity (2) can be generalized, in a straightforward

manner to this context, so that

$$\begin{aligned} \text{LA}_n(a) &= \# \text{ local maxima of } a + \# \text{ local minima of } a \\ &= \mathbf{1}(a \text{ has a local maximum at } n) + 2 \sum_{k=1}^{n-1} \mathbf{1}(a \text{ has a local maximum at } k). \end{aligned}$$

Now, the only difficulty in adapting the proof of Theorem 2.2 to our current framework is when moving in the direction of the gradient when trying to modify the alternating subsequence to consist of only maxima and minima. Indeed, we could get stuck at an index of gradient zero that is neither maximum nor minimum. But this difficulty can easily be overcome by just deciding to move to the right whenever we get in such a situation. We then end up with an alternating subsequence consisting of only maxima and minima through order preserving moves.

*Infinite bilateral sequences.* More generally, given an infinite bilateral sequence  $a = (\dots, a_{-1}, a_0, a_1, \dots) \in [q]^{\mathbb{Z}}$ , we say that  $a$  has a local maximum at the index  $k$ , if for some  $j < k$ ,  $a_j < a_{j+1} = \dots = a_k > a_{k+1}$  and that  $a$  has a local minimum at the index  $k$ , if for some  $j < k$ ,  $a_j > a_{j+1} = \dots = a_k < a_{k+1}$ . Also, set  $a^{(n)} = (a_1, \dots, a_n)$  to be the truncation of  $a$  to the first  $n$  positive indices. An important observation is the following: Let

$$\begin{aligned} A_k &= \{a \in [q]^{\mathbb{Z}} : \text{For some } j \leq 0, a_j > a_{j+1} = \dots = a_k > a_{k+1}\}, \\ A'_k &= \{a \in [q]^{\mathbb{Z}} : \text{For some } j \leq 0, a_j \neq a_{j+1} = \dots = a_k \leq a_{k+1}\}, \end{aligned}$$

and

$$A''_k = \{a \in [q]^{\mathbb{Z}} : \text{For some } j \geq 1, a_j < a_{j+1} = \dots = a_k \leq a_{k+1}\}.$$

Then, for any bilateral sequence  $a \in [q]^{\mathbb{Z}}$ , we have

$$\mathbf{1}(a^{(n)} \text{ has a local maximum at } k) = \mathbf{1}(a \text{ has a local maximum at } k) + \mathbf{1}_{A_k}(a), \text{ if } k < n,$$

and

$$\begin{aligned} \mathbf{1}(a^{(n)} \text{ has a local maximum at } n) &= \mathbf{1}(a \text{ has a local maximum at } n) \\ &\quad + \mathbf{1}_{A_n}(a) + \mathbf{1}_{A'_n}(a) + \mathbf{1}_{A''_n}(a). \end{aligned}$$

Hence,

$$\text{LA}_n(a^{(n)}) = 2 \sum_{k=1}^{n-1} \mathbf{1}(a \text{ has a local maximum at } k) + R_n(a), \tag{6}$$

where the remainder term is given by

$$R_n(a) := 2 \sum_{k=1}^{n-1} \mathbf{1}_{A_k}(a) + \mathbf{1}(a^{(n)} \text{ has a local maximum at } n),$$

and is such that  $|R_n(a)| \leq 3$ , since the sets  $\{A_k\}_{k=1}^n$  are pairwise disjoint.

*Stationarity.* Define the function  $f : [q]^{\mathbb{Z}} \rightarrow \mathbb{R}$  via

$f(a) = 2\mathbf{1}(a \text{ has a local maximum at the index } 0)$ .

If  $T : [q]^{\mathbb{Z}} \rightarrow [q]^{\mathbb{Z}}$  is the (shift) transformation such that  $(Ta)_i = a_{i+1}$ , and  $T^{(k)}$  is the  $k$ -th iterate of  $T$ , it is clear that  $f \circ T^{(k)}(a) = 2\mathbf{1}(a \text{ has a local maximum at } k)$ . With these notations, (6) becomes  $\text{LA}_n(a^{(n)}) = \sum_{k=1}^{n-1} f \circ T^{(k)}(a) + R_n(a)$ . In particular, if  $\mathbf{a}$  is a random sequence with distribution  $\mu^{(\mathbb{Z})}$ , and if  $T^{(k)}f$  is short for  $f \circ T^{(k)}(\mathbf{a})$  the following holds true:

$$\text{LA}_n(\mu) \stackrel{d}{=} \sum_{k=1}^{n-1} T^{(k)}f + R_n(\mathbf{a}). \quad (7)$$

The transformation  $T$  is measure preserving with respect to  $\mu^{(\mathbb{Z})}$  and, moreover, ergodic. Thus, by the classical ergodic theorem (see, for example, [16, Chapter V]), as  $n \rightarrow \infty$ ,  $\sum_{k=1}^n T^{(k)}f/n \rightarrow \mathbf{E}f$ , where the convergence occurs *almost surely* and also *in the mean*. The limit can be easily computed:

$$\begin{aligned} \mathbf{E}f &= 2 \sum_{k=0}^{\infty} \mathbb{P}(\mathbf{a}_{-(k+1)} < \mathbf{a}_{-k} = \dots = \mathbf{a}_0 > \mathbf{a}_1) \\ &= 2 \sum_{k=0}^{\infty} \sum_{x \in [q]} L_x^2 p_x^{k+1} \\ &= 2 \sum_{x \in [q]} \frac{p_x}{1 - p_x} L_x^2 \\ &= \sum_{x \in [q]} \left( \frac{L_x^2 + U_x^2}{1 - p_x} \right) p_x, \end{aligned}$$

where for  $x \in [q]$ ,  $p_x := \mu(\{x\})$ ,  $L_x := \sum_{y < x} p_y$  and  $U_x := \sum_{y > x} p_y$ .

*Oscillation.* Given a probability distribution  $\mu$  supported on  $[q]$ , define the ‘oscillation of  $\mu$  at  $x$ ’, as  $\text{osc}_{\mu}(x) := (L_x^2 + U_x^2)/(L_x + U_x)$  and the total oscillation of the measure  $\mu$  as  $\text{Osc}(\mu) := \sum_{x \in [q]} \text{osc}_{\mu}(x)p_x$ . Interpreting the results of the previous paragraph through

(7), we conclude that

**Theorem 3.1** *Let  $\mathbf{a} = (\mathbf{a}_i)_{i=1}^n$  be a sequence of iid random variables with common distribution  $\mu$  supported on  $[q]$ , and let  $\text{LA}_n(\mu)$  be the length of the longest alternating subsequence of  $\mathbf{a}$ . Then,*

$$\lim_{n \rightarrow \infty} \frac{\text{LA}_n(\mu)}{n} = \text{Osc}(\mu), \text{ in the mean.}$$

In particular, if  $\mu$  a uniform distribution on  $[q]$ ,  $\text{Osc}(\mu) = (2/3 - 1/3q)$ , and thus  $\text{LA}_n(\mu)/n$  is concentrated around  $(2/3 - 1/3q)$  both *in the mean* and *in probability*. We should mention here that Mansour [12], using generating function methods obtained, for  $\mu$

uniform, an explicit formula for  $\mathbf{E}LA_n(\mu)$ , which, of course, is asymptotically equivalent to  $(2/3 - 1/3q)n$ . From (7) it is not difficult to derive also a nonasymptotic expression for  $\mathbf{E}LA_n(\mu)$ :

$$\mathbf{E}LA_n(\mu) = n \operatorname{Osc}(\mu) + \sum_{x \in [q]} R_1(x)p_x + \sum_{x \in [q]} R_2(x)p_x^n, \quad (8)$$

where the terms  $R_1(x)$  and  $R_2(x)$  are given by:

$$R_1(x) = \frac{L_x}{L_x + U_x} + \frac{2L_x U_x}{(L_x + U_x)^2} - \operatorname{osc}_\mu(x) \quad \text{and} \quad R_2(x) = \frac{U_x}{L_x + U_x} - \frac{2L_x U_x}{(L_x + U_x)^2}.$$

Applying (8) in the uniform case recovers computations as given in [12].

As far as the asymptotic limit of  $\operatorname{Osc}(\mu)$  is concerned, we have the following bounds for a general  $\mu$ .

**Proposition 3.2** *Let  $\mu$  be a probability measure supported on the finite set  $[q]$ , then*

$$\frac{1}{2} \left( 1 - \sum_{x \in [q]} p_x^2 \right) \leq \operatorname{Osc}(\mu) \leq \frac{2}{3} \left( 1 - \sum_{x \in [q]} p_x^3 \right). \quad (9)$$

**Proof.** Note that  $\sum_{x \in [q]} L_x p_x = \sum_{i < j} p_i p_j = \sum_{x \in [q]} U_x p_x$  and  $\sum_{x \in [q]} L_x p_x + \sum_{x \in [q]} U_x p_x + \sum_{x \in [q]} p_x^2 = 1$ , which implies that

$$\sum_{x \in [q]} L_x p_x = \sum_{x \in [q]} U_x p_x = \frac{1}{2} \left( 1 - \sum_{x \in [q]} p_x^2 \right). \quad (10)$$

Similarly, for any permutation  $\sigma \in S_3$ , we have that  $\sum_{x \in [q]} L_x U_x p_x = \sum_{i_1 < i_2 < i_3} p_{i_1} p_{i_2} p_{i_3} = \sum_{i_{\sigma(1)} < i_{\sigma(2)} < i_{\sigma(3)}} p_{i_1} p_{i_2} p_{i_3}$ , which implies that  $6 \sum_{x \in [q]} L_x U_x p_x = \sum_{i_1 \neq i_2 \neq i_3} p_{i_1} p_{i_2} p_{i_3}$ . Finally, an inclusion-exclusion argument leads to

$$\sum_{i_1 \neq i_2 \neq i_3} p_{i_1} p_{i_2} p_{i_3} = 1 - 3 \sum_{i_i = i_2} p_{i_1} p_{i_2} + 2 \sum_{i_i = i_2} p_{i_1} p_{i_2} p_{i_3} = 1 - 3 \sum_{x \in [q]} p_x^2 + 2 \sum_{x \in [q]} p_x^3,$$

and therefore

$$\sum_{x \in [q]} L_x U_x p_x = \frac{1}{6} - \frac{1}{2} \sum_{x \in [q]} p_x^2 + \frac{1}{3} \sum_{x \in [q]} p_x^3. \quad (11)$$

Now, to obtain the upper bound in (9), note that

$$\operatorname{Osc}(\mu) = \sum_{x \in [q]} \frac{L_x^2 + U_x^2}{L_x + U_x} p_x = \sum_{x \in [q]} (L_x + U_x) p_x - 2 \sum_{x \in [q]} \frac{L_x U_x}{L_x + U_x} p_x \quad (12)$$

so that in particular,  $\operatorname{Osc}(\mu) \leq \sum_{x \in [q]} (L_x + U_x) p_x - 2 \sum_{x \in [q]} L_x U_x p_x$ . Hence, using (10) and (11),

$$\operatorname{Osc}(\mu) \leq \frac{2}{3} \left( 1 - \sum_{x \in [q]} p_x^3 \right).$$

For the lower bound, note that  $4 \sum_{x \in [q]} \frac{L_x U_x}{L_x + U_x} p_x \leq \sum_{x \in [q]} (L_x + U_x) p_x$ , and from (12) we get

$$\text{Osc}(\mu) \geq \frac{1}{2} \sum_{x \in [q]} (L_x + U_x) p_x = \frac{1}{2} \left( 1 - \sum_{x \in [q]} p_x^2 \right).$$

□

An interesting problem would be to determine the distribution  $\mu$  over  $[q]$  that maximizes the oscillation. It is not hard to prove that such an optimal distribution should be symmetric about  $(q - 1)/2$ , but it is harder to establish its shape (at least asymptotically in  $q$ ).

*Mixing.* The use of ergodic properties to analyze the random variable  $\text{LA}_n(\mu)$  goes beyond the mere application of the ergodic theorem. Indeed, the random variables  $\{T^{(k)}f : k \in \mathbb{Z}\}$  introduced above exhibit mixing, or “long range independence”, meaning that as  $n \rightarrow \infty$

$$\sup_{A \in \mathcal{F}_{\geq 0}, B \in \mathcal{F}_{< -n}} |\mathbb{P}(A|B) - \mathbb{P}(A)| \rightarrow 0,$$

where, for  $n \geq 0$ ,  $\mathcal{F}_{\geq n}$  (respectively  $\mathcal{F}_{< n}$ ) is the  $\sigma$ -field of events generated by  $\{T^{(k)}f : k \geq n\}$  (respectively  $\{T^{(k)}f : k < n\}$ ). This kind of mixing condition is usually called *uniformly strong mixing* or  $\varphi$ -mixing, and the decreasing sequence

$$\varphi(n) := \sup_{A \in \mathcal{F}_{\geq 0}, B \in \mathcal{F}_{< -n}} |\mathbb{P}(A|B) - \mathbb{P}(A)|, \quad (13)$$

is called the *rate* of uniformly strong mixing (see, for example, [11, Chapter 1]). Below, Proposition 3.4 asserts that, in our case, such a rate decreases exponentially. Let us prove the following lemma first.

**Lemma 3.3** *Let  $\mathbf{a} = (\mathbf{a}_i)_{i \in \mathbb{Z}}$  be a bilateral sequence of iid random variables with common distribution  $\mu$  supported on  $[q]$ . Let  $C_{n,t} = \{\mathbf{a}_{-n} = \cdots = \mathbf{a}_{-n+t-1} \neq \mathbf{a}_{-n+t}\}$ ,  $n \geq 1$ ,  $0 \leq t \leq n$ , then:*

(i) *For any  $A \in \mathcal{F}_{\geq 0}$  and any  $t \leq n$ , the event  $C_{n,t} \cap A$  is independent of the  $\sigma$ -field  $\mathcal{G}_{< -n}$  of events generated by  $\{\mathbf{a}_i : i < -n\}$ .*

(ii) *Restricted to the event  $C_{n,t}$ , the  $\sigma$ -fields  $\mathcal{F}_{\geq 0}$  and  $\mathcal{G}_{< -n}$  are independent.*

**Proof.** Let the event  $B_{r,s} := \{\mathbf{a}_r < \mathbf{a}_{r+1} = \cdots = \mathbf{a}_s > \mathbf{a}_{s+1}\}$ . Then, for  $s_1 < s_2 < \cdots < s_m$ ,  $\prod_{i=1}^m T^{(s_i)}f = \sum \prod_{i=1}^n \mathbf{1}_{B_{r_i, s_i}}$  holds true, where the sum runs over the  $r_1, \dots, r_n$  such that  $s_{i-1} < r_i < s_i$  (letting  $s_0 = -\infty$ ) and where

$$f(\mathbf{a}) = 2 \mathbf{1}(\mathbf{a} \text{ has a local maximum at the index } 0).$$

Now, since the random variables  $\{T^{(i)}f, i \in \mathbb{Z}\}$  are binary, then for any  $A \in \mathcal{F}_{\geq 0}$  the random variable  $\mathbf{1}_A$  can be expressed as a linear combination of terms of the form  $\prod_{i=1}^m T^{(s_i)}f$ , where  $0 \leq s_1 < \cdots < s_m$ .

Next,  $\mathbf{1}_{C_{n,t}} \prod_{i=1}^m T^{(s_i)} f = \mathbf{1}_{C_{n,t}} \left( \sum_{i=1}^n \prod_{i=1}^n \mathbf{1}_{B_{r_i, s_i}} \right) = \mathbf{1}_{C_{n,t}} \left( \sum_{r_1 \geq -n+t-1} \prod_{i=1}^n \mathbf{1}_{B_{r_i, s_i}} \right)$ , which implies that  $\mathbf{1}_{C_{n,t}} \prod_{i=1}^m T^{(s_i)} f$  and  $\mathcal{G}_{<-n}$  are independent. This implies, in particular, the independence of the events  $C_{n,t} \cap A$  and  $B$ , for any  $A \in \mathcal{F}_{\geq 0}$  and  $B \in \mathcal{G}_{<-n}$ , proving (i). The statement (ii) follows directly from (i).  $\square$

**Proposition 3.4** *Let  $\mathbf{a} = (\mathbf{a}_i)_{i \in \mathbb{Z}}$  be a bilateral sequence of iid random variables with  $\mu$  supported on  $[q]$ . If the event  $A$  belongs to the  $\sigma$ -field  $\mathcal{F}_{\geq 0}$ , then for any  $n \geq 1$ ,*

$$\|\mathbb{P}(A | \mathcal{G}_{<-n}) - \mathbb{P}(A)\|_{\infty} := \sup_{B \in \mathcal{G}_{<-n}} |\mathbb{P}(A | B) - \mathbb{P}(A)| \leq 2q\kappa^n,$$

where  $\kappa := \max_{x \in [q]} \mu(\{x\})$ . In particular, the rate of uniform strong mixing of the sequence  $\{T^{(k)} f : k \in \mathbb{Z}\}$  (see (13)), satisfies  $\varphi(n) \leq 2q\kappa^{n-1}$ .

**Proof.** Let  $A \in \mathcal{F}_{\geq 0}$ . By Lemma 3.3,  $\mathbb{P}(A \cap C_{n,r} | \mathcal{G}_{<-n}) = \mathbb{P}(A \cap C_{n,r})$ , whenever  $r \leq n$ . Therefore,

$$\begin{aligned} \mathbb{P}(A | \mathcal{G}_{<-n}) &= \sum_{r=1}^n \mathbb{P}(A \cap C_{n,r} | \mathcal{G}_{<-n}) + \mathbb{P}(A \cap \{\mathbf{a}_{-n} = \dots = \mathbf{a}_0\} | \mathcal{G}_{<-n}) \\ &= \sum_{r=1}^n \mathbb{P}(A \cap C_{n,r}) + \mathbb{P}(A \cap \{\mathbf{a}_{-n} = \dots = \mathbf{a}_0\} | \mathcal{G}_{<-n}) \\ &= \mathbb{P}(A) + (\mathbb{P}(A \cap \{\mathbf{a}_{-n} = \dots = \mathbf{a}_0\} | \mathcal{G}_{<-n}) - \mathbb{P}(A \cap \{\mathbf{a}_{-n} = \dots = \mathbf{a}_0\})). \end{aligned}$$

Then, it follows:

$$\begin{aligned} \|\mathbb{P}(A | \mathcal{G}_{<-n}) - \mathbb{P}(A)\|_{\infty} &\leq \mathbb{P}(A \cap \{\mathbf{a}_{-n} = \dots = \mathbf{a}_0\}) \\ &\quad + \|\mathbb{P}(A \cap \{\mathbf{a}_{-n} = \dots = \mathbf{a}_0\} | \mathcal{G}_{<-n})\|_{\infty} \\ &\leq 2 \|\mathbb{P}(\mathbf{a}_{-n} = \dots = \mathbf{a}_0 | \mathcal{G}_{<-n})\|_{\infty} \\ &\leq 2q\kappa^n \end{aligned}$$

where the last conclusion follows trivially from  $\mathcal{G}_{<-n} \supseteq \mathcal{F}_{\leq -(n+1)}$ .  $\square$

Taking advantage of the mixing property we can now infer without much effort the behaviour of the asymptotic variance and also deduce the asymptotic normality of the statistic  $\text{LA}_n(\mu)$ . This is done in the next two paragraphs.

*Variance.* The computation of the variance of the sequence  $S_n = \sum_{k=1}^n T^{(k)} f$  is straightforward. Indeed

$$\text{Var } S_n = n \left[ \text{Cov}(f, f) + 2 \sum_{k=1}^{n-1} \text{Cov}(f, T^{(k)} f) \right] - 2 \sum_{k=1}^{n-1} k \text{Cov}(f, T^{(k)} f), \quad (14)$$

and the mixing property from Proposition 3.4 implies that  $|\text{Cov}(f, T^{(k)}f)|$  decreases geometrically in  $k$ , so that all the series involved in (14) converge. Therefore,

$$\text{Var } S_n = n\gamma^2 + O(1) \text{ , where } \gamma^2 = \text{Cov}(f, f) + 2\sum_{k=1}^{n-1} \text{Cov}(f, T^{(k)}f) . \quad (15)$$

Moreover, for  $k \leq l$ ,  $|\text{Cov}(\mathbf{1}_A(\mathbf{a}), T^{(k)}f)| \leq \mathbf{E}\mathbf{1}_A(\mathbf{a}) \leq \kappa^l$ , and for  $k \geq l$ , and making use of Proposition 3.4,  $|\text{Cov}(\mathbf{1}_A(\mathbf{a}), T^{(k)}f)| \leq 4q\kappa^{k-l-2}\mathbf{E}\mathbf{1}_A(\mathbf{a}) \leq 4q\kappa^{k-2}$ . This implies that, as  $n \rightarrow \infty$ ,

$$\left| \text{Cov} \left( \sum_{k=1}^{n-1} T^{(k)}f, \sum_{k=1}^{n-1} \mathbf{1}_{A_k}(\mathbf{a}) \right) \right| \leq 4q^3 \left\{ \sum_{k \leq l} \kappa^l + \sum_{l \leq k} \kappa^k \right\} = O(1) .$$

Similarly, using the Cauchy-Schwarz inequality, we have that  $\text{Cov} \left( \sum_{k=1}^n T^{(k)}f, \mathbf{1}_{\tilde{A}_n}(\mathbf{a}) \right) \rightarrow 0$  where  $\tilde{A}_n$  is either one of the events  $A_n, A'_n$  or  $A''_n$ . Finally using the fact that  $\text{Cov} \left( \sum_{k=1}^{n-1} T^{(k)}f, T^{(n)}f \right) = \sum_{k=1}^{n-1} \text{Cov}(f, T^{(k)}f)$  is bounded as  $n \rightarrow \infty$ , we conclude that  $\text{Cov} \left( \sum_{k=1}^{n-1} T^{(k)}f, R(n) \right) = O(1)$ , as  $n \rightarrow \infty$ . This implies the corresponding extension of (15) to  $\text{LA}_n(\mu)$ :

$$\text{Var LA}_n(\mu) = n\gamma^2 + O(1) \text{ as } n \rightarrow \infty .$$

Note that the bound just established is not meaningless since the boundedness of  $R_n(\mathbf{a})$  only guarantees the weaker estimate  $\text{Var LA}_n(\mu) = n\gamma^2 + O(n^{1/2})$ .

Let us proceed to compute  $\gamma^2$ . Let  $f_l : [q]^{\mathbb{Z}} \rightarrow \mathbb{R}$  via

$$f_l(a) = 2\mathbf{1}(a_{-l} < a_{-l+1} = \dots = a_0 > a_1) ,$$

so that  $f(a) = \sum_{l=1}^{\infty} f_l(a)$ . Note that

$$\text{Cov}(f, T^{(k)}f_l) = \begin{cases} 0 & \text{if } k \geq l+2 \\ 4 \sum_{x,y \in [q]} \left( \frac{L_x}{1-p_x} \right) (L_y p_y^l) L_{x \wedge y} p_x - 2 \text{Osc}(\mu) \sum_{y \in [q]} L_y^2 p_y^l & \text{if } k = l+1 \\ -2 \text{Osc}(\mu) \sum_{y \in [q]} L_y^2 p_y^l & \text{if } 1 \leq k \leq l \\ 4 \sum_{y \in [q]} L_y^2 p_y^l - 2 \text{Osc}(\mu) \sum_{y \in [q]} L_y^2 p_y^l & \text{if } 0 = k \leq l, \end{cases}$$

and thus

$$\begin{aligned} \gamma^2 &= \text{Var } f + 2 \sum_{k=1}^{\infty} \sum_{l=k-1}^{\infty} \text{Cov}(f, T^{(k)}f_l) \\ &= \text{Osc}(\mu) \left( 2 - 3 \text{Osc}(\mu) - 4 \sum_{x \in [q]} \left( \frac{L_x}{1-p_x} \right)^2 p_x \right) + 8 \sum_{x,y \in [q]} \frac{L_x L_y L_{x \wedge y}}{(1-p_x)(1-p_y)} p_x p_y . \end{aligned}$$

We further mention at this point that Mansour [12] already obtained, with generating function methods, an exact expression for the variance when  $\mu$  is the uniform distribution on  $[q]$ . It is given (as it can also be checked from (15)) by

$$\gamma^2 = \frac{8}{45} \left[ \frac{(1 + 1/q)(1 - 3/4q)(1 - 1/2q)}{(1 - 1/2q)} \right].$$

*Asymptotic normality.* Under appropriate conditions (say, asymptotic positive variance and fast enough mixing), it is natural to expect for the sequence of partial sums to be asymptotically normal. In our model, this is indeed the case. Let us recall the following central limit theorem which goes back to Volkonskii and Rozanov [19, Theorem 1.2] and which can be found, greatly generalized, in texts such as Bradley [2, Theorem 10.3].

**Theorem 3.5** *Let  $\mathbf{x} = (\mathbf{x}_i)_{i \in \mathbb{Z}}$  be a strictly stationary sequence of bounded random variables such that the sequence*

$$\alpha(n) := \sup_{A \in \mathcal{F}_{\geq 0}, B \in \mathcal{F}_{< -n}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

*is summable (i.e.  $\sum_{n \geq 1} \alpha(n) < \infty$ ), where  $\mathcal{F}_{\geq 0}$  is the  $\sigma$ -field generated by the random variables  $\{\mathbf{x}_i : i \geq 0\}$  and  $\mathcal{F}_{< -n}$ ,  $n \geq 1$ , is the  $\sigma$ -field generated by the random variables  $\{\mathbf{x}_i : i < -n\}$ . Then,*

*i.  $\gamma^2 := \text{Var } \mathbf{x}_0 + 2 \sum_{t=1}^{\infty} \text{Cov}(\mathbf{x}_0, \mathbf{x}_t)$  exists in  $[0, \infty)$ , the sum being absolutely convergent.*

*ii. If  $\gamma^2 > 0$ , then as  $n \rightarrow \infty$ ,*

$$\frac{\sum_{t=1}^n \mathbf{x}_t - n \mathbf{E} \mathbf{x}_0}{\sqrt{n\gamma}} \Longrightarrow \mathcal{Z},$$

*where  $\mathcal{Z}$  is a standard normal random variable.*

Now, the asymptotic normality of  $\text{LA}_n(\mu)$ , namely, the fact that as  $n \rightarrow \infty$ ,

$$\frac{\text{LA}_n(\mu) - n \text{Osc}(\mu)}{\sqrt{n\gamma}} \Longrightarrow \mathcal{Z},$$

is clear: By Proposition 3.4, the mixing coefficients  $\alpha(n)$  decrease geometrically, implying the summability of  $\sum \alpha(n)$ . Summarizing, we get:

**Theorem 3.6** *Let  $\mathbf{a} = (\mathbf{a}_i)_{i=1}^n$  be a sequence of iid random variables, with common distribution  $\mu$  supported on  $[q]$ , and let  $\text{LA}_n(\mu)$  be the length of the longest alternating subsequence of  $\mathbf{a}$ . Then, as  $n \rightarrow \infty$ ,*

$$\frac{\text{LA}_n(\mu) - n \text{Osc}(\mu)}{\sqrt{n\gamma}} \Longrightarrow \mathcal{Z},$$

*where  $\mathcal{Z}$  is a standard normal random variable and  $\gamma$  is given by (15).*

**Remark 3.7** *It is clear that the above proofs extend to countable infinite alphabets, without major modification. A parallel situation for the longest increasing subsequence is given in [9], though in that context a more delicate “sandwich” argument is required.*

## 4 Markovian words

Our probabilistic methodologies also provide results beyond the iid framework. Let now  $(\mathbf{x}_k)_{k \geq 0}$  be an ergodic Markov chain started at stationarity and whose state space is a finite linearly ordered set  $\mathcal{A}$ , so that without loss of generality,  $\mathcal{A} = [q]$ . Our objective (as before), is to study the behavior of the statistics  $\text{LA}_n(\mathbf{x}_0, \dots, \mathbf{x}_n)$ .

*Adding gradient information to the chain.* Let us consider the related process  $(\mathbf{y}_k)_{k \geq 0}$  defined recursively as follows:

- $\mathbf{y}_0 = 1$ .
- $\mathbf{y}_{k+1} = 1$  if  $\mathbf{x}_{k+1} > \mathbf{x}_k$  or if  $\mathbf{x}_{k+1} = \mathbf{x}_k$  and  $\mathbf{y}_k = 1$ .
- $\mathbf{y}_{k+1} = -1$  if  $\mathbf{x}_{k+1} < \mathbf{x}_k$  or if  $\mathbf{x}_{k+1} = \mathbf{x}_k$  and  $\mathbf{y}_k = -1$ .

This new sequence basically carries the information indicating that the sequence is increasing or decreasing at  $k$  (we define the sequence  $\mathbf{x}_1, \mathbf{x}_2, \dots$  to be increasing at  $k$  if  $\mathbf{x}_k > \mathbf{x}_{k-1}$  or if it is increasing at  $k-1$  and  $\mathbf{x}_k = \mathbf{x}_{k-1}$ , likewise, the sequence is decreasing at  $k$  if  $\mathbf{x}_k < \mathbf{x}_{k-1}$  or if it is decreasing at  $k-1$  and  $\mathbf{x}_k = \mathbf{x}_{k-1}$ ).

The following holds true for the process  $(\mathbf{x}_k, \mathbf{y}_k)_{k \geq 0}$ :

**Proposition 4.1** *The process  $(\mathbf{x}_k, \mathbf{y}_k)_{k \geq 0}$  is Markov, with transition probabilities given by*

$$\begin{aligned} p_{(r, \pm 1) \rightarrow (s, 1)} &= p_{r,s} \mathbf{1}(s > r), & p_{(r, 1) \rightarrow (r, 1)} &= p_{r,r} \\ p_{(r, \pm 1) \rightarrow (s, -1)} &= p_{r,s} \mathbf{1}(s < r), & p_{(r, -1) \rightarrow (r, -1)} &= p_{r,r} \end{aligned}$$

and stationary measure given by

$$\pi_{(r, 1)} = (1 - p_{r,r})^{-1} \sum_{s < r} \pi_s p_{s,r}, \quad \pi_{(r, -1)} = (1 - p_{r,r})^{-1} \sum_{s > r} \pi_s p_{s,r}.$$

Moreover, the Markov process  $(\mathbf{x}_k, \mathbf{y}_{k-1}, \mathbf{y}_k)_{k \geq 0}$  has a stationary measure given by

$$\begin{aligned} \pi_{(r, 1, 1)} &= \sum_{t < s \leq r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}}, & \pi_{(r, -1, -1)} &= \sum_{t > s \geq r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}} \\ \pi_{(r, 1, -1)} &= \sum_{t < s > r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}}, & \pi_{(r, -1, 1)} &= \sum_{t > s < r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}} \end{aligned}$$

**Proof.** The process is Markov since by definition  $\mathbf{y}_{k+1} \in \sigma(\mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{y}_k)$  and since  $(\mathbf{x}_k)_{k \geq 0}$  is Markov. The transition probabilities are easily obtained from the definition, and moreover,

$$\begin{aligned}
 & \sum_r \pi_{(r,1)} p_{(r,1) \rightarrow (u,1)} + \sum_r \pi_{(r,-1)} p_{(r,-1) \rightarrow (u,1)} \\
 &= \sum_{r \leq u} (1 - p_{r,r})^{-1} \sum_{t < r} \pi_t p_{t,r} p_{r,u} + \sum_{r < u} (1 - p_{r,r})^{-1} \sum_{t > r} \pi_t p_{t,r} p_{r,u} \\
 &= \sum_{r < u} (1 - p_{r,r})^{-1} \sum_{t \neq r} \pi_t p_{t,r} p_{r,u} + (1 - p_{u,u})^{-1} \sum_{t < u} \pi_t p_{t,u} p_{u,u} \\
 &= \sum_{t < u} \pi_t p_{t,u} + (1 - p_{u,u})^{-1} \sum_{t < u} \pi_t p_{t,u} p_{u,u} \\
 &= \pi_{(u,1)}.
 \end{aligned}$$

Similar computations show that

$$\sum_r \pi_{(r,1)} p_{(r,1) \rightarrow (u,-1)} + \sum_r \pi_{(r,-1)} p_{(r,-1) \rightarrow (u,-1)} = \pi_{(u,-1)},$$

thus proving that  $\pi_{(u,\pm 1)}$  is the stationary measure of  $(\mathbf{x}_k, \mathbf{y}_k)_{k \geq 0}$ .

For the chain  $(\mathbf{x}_k, \mathbf{y}_{k-1}, \mathbf{y}_k)_{k \geq 1}$  let us only verify one case since the others are similar:

$$\begin{aligned}
 & \sum_r \pi_{(r,1,1)} p_{(r,1,1) \rightarrow (u,1,1)} + \sum_r \pi_{(r,-1,1)} p_{(r,-1,1) \rightarrow (u,1,1)} \\
 &= \sum_{r \leq u} \sum_{t < s \leq r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}} p_{r,u} + \sum_{r \leq u} \sum_{t > s < r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}} p_{r,u} \\
 &= \sum_{s < r \leq u} \frac{p_{s,r}}{1 - p_{s,s}} p_{r,u} \sum_{t < s} \pi_t p_{t,s} + \sum_{s < r \leq u} \frac{p_{s,r}}{1 - p_{s,s}} p_{r,u} \sum_{t > s} \pi_t p_{t,s} + \sum_{s=r \leq u} \frac{p_{s,r}}{1 - p_{s,s}} p_{r,u} \sum_{t < s} \pi_t p_{t,s} \\
 &= \sum_{s < r \leq u} \pi_s p_{s,r} p_{r,u} + \sum_{s < r \leq u} \frac{\pi_s p_{s,r} p_{r,u} p_{r,r}}{1 - p_{r,r}} \\
 &= \pi_{(u,1,1)}.
 \end{aligned}$$

□

*Oscillations of a Markov chain.* Given an ergodic Markov chain  $\mathbf{x} := (\mathbf{x}_k)_{k \geq 1}$  whose state space is a finite linearly ordered set, define

$$\begin{aligned}
 \text{Osc}^+(\mathbf{x}) &:= \sum_{t < s > r} (\pi_t p_{t,s} p_{s,r}) / (1 - p_{s,s}) \\
 \text{Osc}^-(\mathbf{x}) &:= \sum_{t > s < r} (\pi_t p_{t,s} p_{s,r}) / (1 - p_{s,s})
 \end{aligned}$$

and  $\text{Osc}(\mathbf{x}) := \text{Osc}^+(\mathbf{x}) + \text{Osc}^-(\mathbf{x})$  ( $= 2 \text{Osc}^+(\mathbf{x}) = 2 \text{Osc}^-(\mathbf{x})$ ). With these notations, we have:

**Theorem 4.2** Let  $LA_n(\mathbf{x}_0, \dots, \mathbf{x}_n)$  be the length of the longest alternating subsequence of the first  $n + 1$  elements of the Markov chain  $(\mathbf{x}_k)_{k \geq 0}$ . Then, as  $n \rightarrow \infty$ ,

$$\frac{LA_n(\mathbf{x}_0, \dots, \mathbf{x}_n)}{n} \rightarrow \text{Osc}(\mathbf{x}),$$

in the mean and almost surely.

**Proof.** From the very definition of  $\mathbf{y}_k$ ,

$$LA_n(\mathbf{x}_0, \dots, \mathbf{x}_n) = \sum_{k=0}^{n-1} \mathbf{1}(\mathbf{y}_k \mathbf{y}_{k+1} = -1),$$

therefore, by the ergodic theorem,

$$\frac{LA_n(\mathbf{x}_0, \dots, \mathbf{x}_n)}{n} \rightarrow \Pi(\mathbf{y}_0 \mathbf{y}_1 = -1),$$

in the mean and almost surely and where  $\Pi$  is the stationary measure of the chain. Now, from Proposition 4.1,

$$\Pi(\mathbf{y}_0 \mathbf{y}_1 = -1) = \sum_{t < s > r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}} + \sum_{t > s < r} \frac{\pi_t p_{t,s} p_{s,r}}{1 - p_{s,s}},$$

from which the result follows. □

**Remark 4.3** Above, the case  $p_{t,s} = p_s$  (and therefore  $\pi_t = p_t$ ), corresponds to iid letters thus recovering Theorem 3.1.

*Central limit theorem:* In case the asymptotic variance term of  $LA_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is nonzero, then since

$LA_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$  is an additive functional of the finite Markov chain  $(\mathbf{x}_k, \mathbf{y}_{k-1}, \mathbf{y}_k)_{k \geq 0}$ , and since the mixing rate of an ergodic Markov chain with finite state space is exponentially decreasing, Theorem 3.5 imply that, for some  $\gamma > 0$ ,

$$\frac{LA_n(\mathbf{x}_0, \dots, \mathbf{x}_n) - n \text{Osc}(\mathbf{x})}{\sqrt{n\gamma}} \implies \mathcal{Z},$$

where  $\mathcal{Z}$  is a standard normal random variable. The reader should contrast this last fact with the increasing subsequence results where the iid and Markov limiting laws differ when the alphabet has a size of four or more ([10]).

## 5 Concluding remarks

Determining the length of the longest alternating subsequence of a random *pattern-avoiding* permutation or word, has been recently studied by Firro, Mansour and Wilson [5, 6, 13] inspired by the work of Deutsch, Hildebrand and Wilf [3] on the longest increasing subsequence of pattern-avoiding permutations. In such a case, a probabilistic (i.e. measure theoretic) approach is also possible once an appropriate recursive description of the pattern-avoiding permutations is given. Such recursive description is the subject of an extensive list of works, originating from an old standing conjecture of Zeilberger [21] claiming in particular, that the set of pattern avoiding permutations is  $P$ -recursive. In the case of avoiding patterns of length 3 a concise work is found in [5]. A canonical example of this situation is the case of permutations avoiding the pattern (123), or equivalently, sequences in  $[0, 1]^n$  avoiding the pattern (123) (recall the observation at the beginning of Section 2). In this context, if we let  $\mathcal{G}_n$  to be the set of sequences in  $[0, 1]^n$  that avoid the pattern (123), and for  $n \geq 1$  let

$$v_n(x_n, \dots, x_1) = dx_n \dots dx_1 \mathbf{1}((x_n, \dots, x_1) \in \mathcal{G}_n),$$

then, the recursive construction  $v_1 = dx_1$  and

$$\begin{aligned} v_{n+1}(x_{n+1}, \dots, x_1) &= dx_{n+1} v_n(x_n, \dots, x_1) \mathbf{1}(x_{n+1} > x_n) \\ &+ dx_n v_n(x_{n+1}, x_{n-1}, \dots, x_1) \mathbf{1}(x_n > \max\{x_1, \dots, x_{n-1}, x_n\}). \end{aligned}$$

for  $n \geq 1$ , holds. This recursive formulation for the restricted measure translates to a recursive formula for the distribution of the number of local maxima of the sequence  $(x_n, \dots, x_1)$  on  $\mathcal{G}_n$ : Let  $M_n = \max\{x_1, \dots, x_n\}$ , let  $L_n = \#\{i : x_i < x_{i+1} > x_{i+2}, i = 1, \dots, n-2\}$  and let  $\chi_n = \mathbf{1}(M_n = x_n)$ ,  $\varrho_n = \mathbf{1}(x_n < x_{n-1} > x_{n-2})$ , then

$$\begin{aligned} &v_{n+1}(M_{n+1} = m, x_{n+1} = x, L_n = k, \chi_{n+1} = 0, \varrho_{n+1} = 1) \\ &= v_n(M_n < m, x_n = x, L_n = k, \chi_n = 0, \varrho_n = 1) dm \\ &+ v_n(M_n < m, x_n = x, L_n = k-1, \chi_n = 0, \varrho_n = 0) dm \\ &+ v_n(M_n = x, x_n = x, L_n = k-1, \chi_n = 1, \varrho_n = 0) dm \\ &v_{n+1}(M_{n+1} = m, x_{n+1} = x, L_n = k, \chi_{n+1} = 0, \varrho_{n+1} = 0) \\ &= v_n(M_n = m, x_n < x, L_n = k, \chi_n = 0) dx \\ &v_{n+1}(M_{n+1} = x, x_{n+1} = x, L_n = k, \chi_{n+1} = 1, \varrho_{n+1} = 0) \\ &= v_n(M_n < x, x_n < x, L_n = k, \chi_n = 0) dx \\ &+ v_n(M_n < x, x_n < x, L_n = k, \chi_n = 1, \varrho_n = 0) dx. \end{aligned}$$

These formulas can be interpreted as Markovian formulations of the process of counting local maxima (therefore, the length of the longest alternating subsequence), in the restricted space of permutations avoiding the pattern (123). Therefore the appropriate extension of the methods of Section 4 lead to the corresponding results in this context. Notice however, that such Markovian formulation is not measure preserving, and the corresponding modification of the ergodic theorem, central limit theorem, etc., should be

introduced. It is our goal in subsequent research, to study these methods for tractable (in the above sense), sets of pattern avoiding permutations or words, following this alternative probabilistic path just presented.

## References

- [1] AVERKAMP, R. and HOUDRÉ, C., “Wavelet thresholding for nonnecessarily Gaussian noise: functionality,” *Annals of Statistics*, vol. 33, no. 5, pp. 2164–2193, 2005.
- [2] BRADLEY, R., *Introduction to strong mixing conditions*. Kendrick Press, Heber City, Utah, 2007.
- [3] DEUTSCH, E., HILDEBRAND, A.J. and WILF, H.S., “Longest increasing subsequences in pattern-restricted permutations,” *The Electronic Journal of Combinatorics*, vol. 9(2), no. R12, 2003.
- [4] DURRETT, R., *Probability: Theory and Examples*. Thomson, 2005.
- [5] FIRRO, G., MANSOUR, T. and WILSON, M.C., “Three-letter-pattern-avoiding permutations and functional equations,” *Elect. J. Combin.*, vol. 13, no. R51, 2006.
- [6] FIRRO, G., MANSOUR, T. and WILSON, M.C., “Longest alternating subsequences in pattern-restricted permutation,” *The Electronic Journal of Combinatorics*, vol. 14, no. R34, 2007.
- [7] HOEFFDING, W. and ROBBINS, H., “The central limit theorem for dependent random variables,” *Duke Math. J.*, vol. 15, no. 3, pp. 773–780, 1948.
- [8] HEINRICH, L., “Non-uniform estimates, moderate and large deviations in the central limit theorem for  $m$ -dependent random variables,” *Mathematische Nachrichten*, vol. 121, no. 1, pp. 107–121, 1985.
- [9] HOUDRÉ, C. and LITHERLAND, T.L., “On the longest increasing subsequence for finite and countable alphabets,” *High Dimensional Probability V: The Luminy Volume IMS Collections 5*, pp. 185–212, 2009.
- [10] HOUDRÉ, C. and LITHERLAND, T.L., “On the Limiting Shape of Random Young Tableaux For Markovian Words,” *arXiv:0810.2982*, 2009.
- [11] LIN, Z., ZHENGYAN, L., LU, C., and CHUANRONG, L., *Limit theory for mixing dependent random variables*. Kluwer Academic Pub, 1996.
- [12] MANSOUR, T., “Longest alternating subsequences of  $k$ -ary words,” *Discrete Applied Mathematics*, vol. 156, no. 1, pp. 119–124, 2008.
- [13] MANSOUR, T., “Longest alternating subsequences in pattern-restricted  $k$ -ary words,” *Online J. Analytic Combin.*, vol. 3, 2008.
- [14] RESNICK, S., *A probability path*. Birkhauser, 1999.
- [15] RIAUBA, B., “A local limit theorem for dependent random variables,” *Lithuanian Mathematical Journal*, vol. 17, no. 1, pp. 119–129, 1977.
- [16] SHIRYAEV, A., “Probability. Number 95 in Graduate Texts in Mathematics,” 1996.

- [17] STANLEY, R., “Longest alternating subsequences of permutations,” *Michigan Mathematical Journal*, vol. 57, pp. 675–687, 2008.
- [18] STANLEY, R., “Increasing and decreasing subsequences and their variants,” *Proc. Internat. Cong. Math (Madrid 2006)*, American Mathematical Society, pp. 549–579, 2007.
- [19] VOLKONSKII, V. and ROZANOV, Y., “Some limit theorems for random functions. I,” *Theory of Probability and its Applications*, vol. 4, p. 178, 1959.
- [20] WIDOM, H., “On the Limiting Distribution for the Length of the Longest Alternating Sequence in a Random Permutation,” *The Electronic Journal of Combinatorics*, vol. 13, no. R25, 2006.
- [21] ZEILBERGER, D., “A holonomic systems approach to special functions identities\* 1,” *Journal of Computational and Applied Mathematics*, vol. 32, no. 3, p. 321-368, 1990.