Stabilization time of random strings with three types

Student Research Societies Thesis

Bori Anna Mészáros

Faculty of natural sciences

Mathematics

Supervisor:

Dr. Bálint Vető

Associate professor

Institute of Mathematics

Department of Stochastics

MŰEGYETEM 1782

Budapest University of Technology and Economics (BME)

2024

Contents

Chapter 1

Introduction

In this thesis, we study the stabilization time of evolution on random sequences and its limiting distribution. The aim of the evolution is to arrange the elements in decreasing order over the steps. Evolution for two types of sequences is a well-known and researched topic, with an established and well-regarded interpretation[2] of confused soldiers,which says:

"A large (but finite) number of soldiers are arranged in an east-west line, and all the soldiers are facing north. The commander shouts "Right face!" One second later, all the soldiers ought to be facing east, but they have not completely mastered "right" and "left", so some are facing east and some west. Any soldier who is face-to-face with his neighbor realizes that there was a mistake and turns 180 degrees (disregarding the possibility that the mistake might have been the neighbor's). One second later, when all these 180 degree turns have been completed, any soldier who is now face-to-face with a neighbor turns 180 degrees (even if he had just turned at the previous step). The process repeats in the same manner. Prove that it stops after finitely many steps."

The evolution of two type sequences is a discrete TASEP process mentioned in James Martin and Philipp Schmidt's article[5]. We consider a different approach from the three update rules explained in that article, but our evolution process is also natural and similar to the fully parallel updates in the 5^{th} section of the article.

The distribution of the stabilization time depends on the probability of choosing each type. A natural question arises: what happens during the evolution of sequences with three types, and how does the stabilization time change in this context? The case of three types of sequences is more challenging because the evolution itself is not trivially generalizable from the two-type model. Our goal was to study these three-type sequences in the special case, when there is only one of the third type in the sequence, using the theorem on the limiting distribution of stabilization time provided in a previous paper (see [1]) as a basis. However, for better understanding and applicability to three types, we approached the expression of stabilization time differently, making it more generalizable. We introduced the evolution for three-type sequences and examined the additional time required compared to the stabilization time for their projection into two types. We characterized this additional time and studied its limiting distribution for different probabilities.

This special third type element can be considered a second class particle in some sense, which is defined in Ferrari and Kipnis's paper [6], but they have some differences.

The paper is structured as follows: in Chapter 2 we introduce the rules of the evolution, define the projections, the stabilization time and the evolution steps, in Chapter 3, we state the theorem on the limit distribution of the stabilization times of the two types sequences , which is a known result of the Funk-Nica-Noyes article (see [1]). The results in Chapter 4 are new results, namely the characterization of the excess and Theorem 4.2, in which we state the limit distribution of the excess.

Chapter 2

Model definition

For $n \in \mathbb{N}$ let $\Omega_n = \{0{,}1{,}2\}^n$ be the sample space consisting of the n-length strings with the following three types: 0,1,2.

On this set, we define a discrete time 'evolution' process $S : \Omega_n \mapsto \Omega_n$,a process that, step by step, seeks to reach a stable state in which the elements of the sequence are in descending order. The process works the following way: in a step we replace each occurrence of the length-2 substrings when the smaller number is followed by a bigger one (so $S(01) = 10$, $S(02) = 20$ and $S(12) = 21$). This procedure is well-defined when the string only consists of two types. There is no evolution rule for three types under which both natural projections evolve according to the dynamics for two types. Either of the two projections can have priority over the other. We choose Π_1 below to have the priority.

Definition 2.1. $\Pi_1 : \Omega_n \mapsto \Omega_n$ and $\Pi_2 : \Omega_n \mapsto \Omega_n$ are two projections on Ω_n such that

$$(\Pi_1\omega)_i = \pi_1(\omega_i)$$

and

$$(\Pi_2\omega)_i = \pi_2(\omega_i),$$

where

$\pi_1 : \{0,1,2\} \mapsto \{0,1,2\}$ such that $\pi_1(0) = \pi_1(1) = 0$ and $\pi_1(2) = 2$,

$\pi_2 : \{0,1,2\} \mapsto \{0,1,2\}$ such that $\pi_2(0) = 0$ and $\pi_1(1) = \pi_1(2) = 2$.

So projection Π_1 is the case when we consider 0's and 1's identical in ω, so that the two types are {0,1},{2} and projection Π_2 is when we consider 1's and 2's identical, so that the two types are {0},{1,2}.

However in case of three types it can happen that $\omega \in \Omega_n$ contains a substring for which $\omega_i < \omega_{i+1} < \omega_{i+2}$ (where ω_j is the j^{th} element of ω), which means that $012 \subseteq \omega$, therefore we must choose one of the aforementioned projections. Throughout this paper, in these cases we consider the Π_1 projection, so that $S(012) = 021$.

We repeat this S evolution process until it is stabilized in the sense, the string does not change anymore, so when all the elements are sorted in a decreasing topological order (22..2211..1100..00).

The steps needed for this stabilized state is called the stabilization time of the string, and it is denoted by $T_n : \Omega_n \mapsto \mathbb{N}$.

A concrete example:

$$0122102 \xrightarrow{S} 0212120 \xrightarrow{S} 2021210 \xrightarrow{S} 2202110 \xrightarrow{S} 2220110 \xrightarrow{S} 2221010 \xrightarrow{S} 222110$$

In this case $T(0122102) = 6$, as we needed 6 steps to stabilize the string.

Now let us define the measure $\mathbb{P}_{p,q,r}$, so that the $(\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_{p,q,r})$ represents the probability space consisting of strings in $\{0,1,2\}^n$, where each bit is chosen independently to be 2 with probability p, 0 with probability q or 1 with probability r, when $p,q,r \geq 0$ and $p + q + r = 1$.

In the following sections we will examine the distribution of this stabilization time T_n with respect to the measure $\mathbb{P}_{p,q,r}$ for different values of p,q,r, when $n \to \infty$.

Let us define the two aforementioned operators more precisely:

Definition 2.2. We say that $S : \Omega_n \mapsto \Omega_n$ is the evolution step of the string $\omega \in \Omega_n$, if the followings are satisfied:

If $\exists 012 \subseteq \omega$:

$$S(012) = 021$$

For any other $\omega_i \omega_{i+1}$ substring of ω:

If $\omega_i < \omega_{i+1}$:

$$(S\omega)_i = \omega_{i+1}$$

$$(S\omega)_{i+1} = \omega_i$$

,where $(S\omega)_i$ is the i^{th} element of the string $S\omega \in \Omega_n$.

Definition 2.3. We say that $T_n : \Omega_n \mapsto \mathbb{N}$ is the stabilization time of the string $\omega \in \Omega_n$, if:

$$T_n(\omega) = \min_{k \geq 0}\{k : S^k(\omega) = S^{k+1}(\omega)\}$$

,where $S^k(\omega) = \underbrace{SS...S}_{\text{k-fold}}\omega$ means that we apply the function S k-times on ω.

With other words this means that we apply the evolution step function S on ω until ω reaches the stabilized state, from where it does not change anymore.

Chapter 3

Stabilization time of random strings with two types

In this section we consider the special case when there are only two types in the strings, so when one of p,q,r is 0. Without the loss of generality let us assume that $p = 0$, so for $n \in \mathbb{N}$ and $r = 1 - q \in (0,1)$.

Throughout this chapter let Ω_n denote the probability space consisting of strings in $\{0,1\}^n$.

The aforementioned evolution is a bit simplier in this special case, since there is no 'problematic situation', so we can say that $S : \Omega_n \mapsto \Omega_n$ has no other affect on $\omega \in \Omega_n$ than changing each occurance of 10's to 01's.

It is easy to see that this process must stabilize within a maximum of $n-1$ steps in the following form: 11..10..00, so that all the 1's are on the left to all the 0's.

We are interested in the limit distribution of the random variable T_n with respect to the measure $\mathbb{P}_{0,1-r,r}$, or with more simple term $\mathbb{P}_r$, for different r values, when $n \to \infty$.

The Theorem 3.1 about the limit distributions is known results from the Funk-Nica-Noyes article (see [1]), with a different approach, we gave a new proof, which can be found in the Appendix.

Theorem 3.1. *We have the following weak limits for the distribution of the random variable T_n with respect to $\mathbb{P}_r$ in the limit $n \to \infty$:*

If $r > \frac{1}{2}$:

$$\frac{T_n - rn}{\sqrt{n}} \Rightarrow \mathcal{N}(0, r(1-r))$$

If $r = \frac{1}{2}$:

$$\frac{T_n - \frac{1}{2}n}{\sqrt{n}} \Rightarrow \frac{\chi_3}{2}$$

In the above limits, $N(0, r(1-r))$ is a mean zero Gaussian variable with variance $r(1-r)$, and $\frac{\chi_3}{2} \sim \frac{1}{2}r\sqrt{Z_1^2 + Z_2^2 + Z_3^2}$ is half of the Euclidean norm of a vector of three independent standard $N(0,1)$ Gaussian variables. By symmetry it is enough to state the theorem for $r \geq \frac{1}{2}$.

Let us introduce some random variables which we use throughout the paper:

Definition 3.2. Let $L : \Omega_n \mapsto \mathbb{N}$ be the number of 1's until the first occurrence of a 0 in the initial string ω, read from the left to the right (so in other words, the number of 1's at the beginning of the string), and similarly, let $R : \Omega_n \mapsto \mathbb{N}$ be the number of 0's until the first occurrence of a 1, read from the right to the left,so the number of 0's at the end of the string.

$$L(\omega) = \{k : \omega_1 = 1 \wedge \omega_2 = 1 \wedge ... \wedge \omega_k = 1 \wedge \omega_{k+1} \neq 1\}$$

$$R(\omega) = \{k : \omega_n = 0 \wedge \omega_{n-1} = 0 \wedge ... \wedge \omega_{n-k+1} = 0 \wedge \omega_{n-k} \neq 0\}$$

Let $\tilde{\omega}$ be the substring of $\omega \in \Omega_n$ in such a way that we chop off the first L and the last R elements of ω, so that we get a string in $\{0,1\}^{n-L-R}$ which starts with a 0 and ends with a 1.

Let $M_1 : \Omega_n \mapsto \mathbb{N}$ denote that position in ω which is the rightmost position in the string $\tilde{\omega} \in \{0{,}1\}^{n-L-R}$ from where any nonempty suffix of the substring to the left of this position contains strictly higher amount of 0's than 1's.

$$M_1(\omega) = \underset{k \geq 1}{max}\{k : L < k < n - R \wedge \forall L < l < k : \sum_{i=l}^{k} \mathbb{1}_{\{\omega_i = 0\}} > \sum_{i=l}^{k} \mathbb{1}_{\{\omega_i = 1\}}\}$$

Remark 3.3. In order to understand M_1 better, we assign a random walk to each $\omega \in \{0{,}1\}^n$ in the following way: it takes a step up for every 0 in ω and a step down for every 1 in ω. So for $1 \leq k \leq n$ let $S_k = \sum\limits_{i=1}^{k}(1 - 2\omega_i)$, where ω_i is the i^{th} element of ω. The function $f : \mathbb{N} \to \mathbb{N}$, where $k \mapsto S_k$ gives the height function of the aforementioned random walk.

With this interpretation, M_1 is the leftmost location of the maximum height in the height function of $\omega \in \{0{,}1\}^n$, or more precisely:

$$M_1 = \min_k\{k | L \leq k \leq n - R \wedge S_k = \max_i\{S_i | L \leq i \leq n - R\}\}$$

, where S_k represents the random walk of ω.

Chapter 4

Stabilization time of random strings with three types in special cases

In this section we consider the special case of the random strings with three types when there is only one 1 in a string of length n, which is constructed in two steps:

Firstly, a string of length n is chosen on $\{0,2\}^n$ with respect to the aforementioned measure $\mathbb{P}_r$, and after that, the position of 1 is chosen uniformly on the n elements of the string with two types, so that a 0 or a 2 is replaced by a 1 on the uniformly chosen place, therefore we can say that the position of 1 in ω is $U(\omega) = \{i : \omega_i = 1\} \stackrel{d}{=} \text{Uni}\{1,2,...,n\}$.

Throughout this chapter, Ω_n stands for that sample space with exactly one 1 among the 0's and 2's.

By Chapter 2, we know the behaviour and the limit distribution of the stabilization time of random strings with two types, therefore for $\forall \omega \in \Omega_n$, $T_n(\Pi_1(\omega))$ (for the definition of Π_1, see 2.1) is clear, but for $T_n(\omega)$, we need to consider an excess time, since it might happen, that after $T_n(\Pi_1(\omega))$ steps, all the 2's in ω are placed in the beginning of the string, but there are still some 0's before the position of 1, so in order to reach the stable state, some more steps are needed, the number of those steps is what we call excess time, and denote it with E_n.

In this Chapter we characterize the existence of the excess and state Theorem 4.2 on the limit distribution of excesses with respect to different r values and these are our new results.

Definition 4.1. $\forall \omega \in \Omega_n$ $E_n : \Omega_n \mapsto \mathbb{N}$ is the excess time, if:

$$T_n(\omega) = T_n(\Pi_1(\omega)) + E_n(\omega)$$

Theorem 4.2. *We have the following weak limits for the distribution of the random variable* E_n *with respect to* $\mathbb{P}_r$ *in the limit* $n \to \infty$*:*

If $r > \frac{1}{2}$*:*

$$E_n \Rightarrow 1,$$

if $r = \frac{1}{2}$*:*

$$E_n \Rightarrow X,$$

if $r < \frac{1}{2}$*:*

$$E_n \Rightarrow 0,$$

where $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$.

To prove the Theorem 4.2, we need to understand how the initial position of 1 is related to the excess and what is the probability of the events when there is an excess and when there is not, with respect to $\mathbb{P}_r$ for different r values.

Let us start by characterising the existence of the excess.

Proposition 4.3. $\forall \omega \in \Omega_n$*:*

if $K(\omega) > U(\omega)$*:*

$$E_n(\omega) = 0$$

if $K(\omega) \leq U(\omega)$ *:*

$$E_n(\omega) \geq 1,$$

where

$$K(\omega) := \begin{cases} \max\{k : \pi_1(\omega_k) = 0,\, \omega_{k+1} = 2,\, \pi_1(\omega_{k-1}) = \pi_1(\omega_{k-2}) = 0, & \\ \quad \textit{such that } \forall l < k-1 : S_l + 1 < S_k\}, & \textit{if } \exists \textit{ such } k, \\ 0, & \textit{otherwise.} \end{cases} \tag{3}$$

Furthermore if $E_n > 1$:

$$U(\omega) > n - R(\pi_1(\omega)).$$

Remark 4.4. The above Theorem states that the existence of the excess depends on the position of this K, in the sense that if the position of 1 is placed before this K, there is no excess, when it is after the K, there will be excess for sure and more than one excess can only happen when the initial position of 1 is such that there is no 2 after it.

So we can say that this K is the switching position of the existence of the excess.

Instead of calculating the distribution of this K random variable, we study its distance from M_1 and will see it is negligible compared to the n-length string.

Proof. Firstly, let us assume that $K(\omega) > U(\omega)$.

In each step 0002 stays together and goes one step to the left, since by the assumption that the height function never reaches the height of S_{k-1}, in any suffix of the substring to the left of 0002 might be at most one more 2's than 0's, so every time a 0 'escapes' on the right as $02 \mapsto 20$, a 0 comes in on the left.

After some steps (it might be the initial form of ω), this incoming element from left will be the 1, and it cannot swap place with the 2 from 0002, since once it reaches ..0102.. position, until this block can move to the left (until there is a 0 before them), $01 \mapsto 10$ and $02 \mapsto 20$ will happen simultaneously, so a 0 stays between them like ..102...

When there are no more 0's before 0102, meaning that the string looks like this at that time: 2...20102.. and the next steps are: 2...21020... $\xrightarrow{S}$ 2...2120... $\xrightarrow{S}$ 2...2210... So there might be 2's after 1's position at this point, but the last 2 will be stabilized by swapping with the 1, therefore there will be no excess.

Now let us assume that $K(\omega) \leq U(\omega)$. Since K is a local maximum to the left so we know that it goes to the left by one in each step until it reaches the front (so when there are only 2's before it and a single 0), from then on it passes a 2 to its left in every step, and since the 1 will never swap with the 2 in the position $K + 1$, the last 2 which has a 0 before its position will stabilize by swapping with a 0, therefore after $T_n(\Pi_1(\omega))$ steps, there will be a 01 substring in the string for sure, so there will indeed be excess.

If $\exists k > U(\omega)$ such that $\omega_k = 2$, then $l := \underset{k>U(\omega)}{max} \{k : \omega_k = 2\}$ will be next to the 1 after some evolution steps for sure, and they will swap places in the next step, so we will get this string: ...210...0, which means that in any evolution step after this, at most one piece of 0 can be placed between them (between the last 2 and the 1), because as soon as 2 swaps with a zero, in the next step 1 swaps with it as well, so the last 0 will act the same, not allowing more than one excess. □

Proposition 4.5. $\forall \omega \in \Omega_n$.

$$\lim_{n \to \infty} \mathbb{P}(E_n(\omega) > 1) = 0$$

Proof. We have seen in the Proposition above, that $E_n(\omega) > 1$ can only happen, when there are only 0's after 1 in ω, which means, that 1 is among the $R(\Pi_1(\omega))$ 0's and since $R(\Pi_1(\omega)) \stackrel{d}{=} \text{Geo}(r)$, which is almost surely finite,

$$\{E_n > 1\} \subseteq \{U \geq n - R\}, \tag{4}$$

therefore

$$\mathbb{P}(\{E_n > 1\}) \leq \mathbb{P}(\{U \geq n - R\}) \to 0 \text{ as } n \to \infty. \tag{5}$$

□

Proposition 4.6. *For* $r > \frac{1}{2}$*:*

$$\underset{n\to\infty}{lim}\,\mathbb{P}(U < M_1) = 0$$

For the proof, let us consider the following lemma first:

Lemma 4.7. *Let us consider a random walk* $S_k = \sum_{i=1}^{k} X_i$ *with a negative drift, where* X_i*'s are iid and* $\mathbb{P}(X_i = -1) = r > \frac{1}{2}$ *and* $\mathbb{P}(X_i = 1) = 1 - r$*,then* $\forall \epsilon > 0$*:*

$$\mathbb{P}(\underset{k\geq\epsilon\sqrt{n}}{max}\,S_k > 0) \to 0 \;\; as \; n \to \infty$$

Proof of Lemma 4.7.

$$\mathbb{P}\left(S_{\epsilon\sqrt{n}} < \frac{1-2r}{2}\epsilon\sqrt{n}\right) = \mathbb{P}\left(\frac{S_{\epsilon\sqrt{n}} - (1-2r)\epsilon\sqrt{n}}{\sqrt{4r(1-r)}n^{\frac{1}{4}}} < \frac{-\frac{1-2r}{2}\epsilon\sqrt{n}}{\sqrt{4r(1-r)}n^{\frac{1}{4}}}\right) \tag{6}$$

By the central limit theorem, we know that

$$\frac{S_{\epsilon\sqrt{n}} - (1-2r)\epsilon\sqrt{n}}{\sqrt{4r(1-r)}n^{\frac{1}{4}}} \to \mathcal{N}(0,1) \text{ as } n \to \infty \tag{7}$$

Since $r > \frac{1}{2}$, we also know that:

$$\frac{-\frac{1-2r}{2}\epsilon\sqrt{n}}{\sqrt{4r(1-r)}n^{\frac{1}{4}}} > 0, \tag{8}$$

therefore

$$\mathbb{P}\left(S_{\epsilon\sqrt{n}} < \frac{1-2r}{2}\epsilon\sqrt{n}\right) \to 1, \text{ as } n \to \infty. \tag{9}$$

$$\mathbb{P}(\max_{k\geq\epsilon\sqrt{n}} S_k > 0) = \tag{10}$$

$$\mathbb{P}(\max_{k\geq\epsilon\sqrt{n}} S_k > 0 | S_{\epsilon\sqrt{n}} < \frac{1-2r}{2}\epsilon\sqrt{n}) \cdot \mathbb{P}(S_{\epsilon\sqrt{n}} < \frac{1-2r}{2}\epsilon\sqrt{n}) + \tag{11}$$

$$\mathbb{P}(\max_{k\geq\epsilon\sqrt{n}} S_k > 0 | S_{\epsilon\sqrt{n}} \geq \frac{1-2r}{2}\epsilon\sqrt{n}) \cdot \mathbb{P}(S_{\epsilon\sqrt{n}} \geq \frac{1-2r}{2}\epsilon\sqrt{n}) \leq \tag{12}$$

$$\mathbb{P}(\max_{k\leq\epsilon\sqrt{n}} S_k > 0 | S_{\epsilon\sqrt{n}} < \frac{1-2r}{2}\epsilon\sqrt{n}) + o(1) \leq a^{-\frac{1-2r}{2}\epsilon\sqrt{n}} + o(1) \to 0, \text{ as } n \to \infty, \tag{13}$$

where $a = \mathbb{P}(\exists k : S_k = 1 | S_0 = 0) < 1$.

So

$$\mathbb{P}(\max_{k\geq\epsilon\sqrt{n}} S_k > 0) \to 0 \text{ as } n \to \infty$$

is indeed the case. □

Proof of Proposition 4.6. We have seen from the Lemma 4.7 that

$$\frac{\underset{1\leq k\leq n}{argmax} S_k}{\sqrt{n}} \to 0, \text{ as } n \to \infty \tag{14}$$

holds for any maximum places and by the properties of M_1 (see, 4.24), for any $\omega \in \Omega_n$ there is a maximum place for which:

$$M_1(\omega) \underset{d}{=} L(\omega) + 1 + \underset{1\leq k\leq n}{argmax} S_k, \tag{15}$$

therefore

$$\frac{M_1}{\sqrt{n}} \overset{d}{=} \frac{L+1}{\sqrt{n}} + \frac{\underset{1\leq k\leq n}{argmax} S_k}{\sqrt{n}} \tag{16}$$

which means that

$$\frac{M_1}{\sqrt{n}} \to 0, \text{ as } n \to \infty, \tag{17}$$

since we have already seen in the proof of 4.31, that:

$$\frac{L+1}{\sqrt{n}} \to 0, \text{ as } n \to \infty \tag{18}$$

The last step of the proof:

$$\mathbb{P}(U < M_1) = \mathbb{P}(\frac{U}{\sqrt{n}} < \frac{M_1}{\sqrt{n}}) \to 0 \tag{19}$$

□

Proposition 4.8. *For* $r = \frac{1}{2}$*:*

$$\underset{n \longrightarrow \infty}{lim} \mathbb{P}(E_n = 0) = \frac{1}{2} \tag{20}$$

Proposition 4.9. *For* $r = \frac{1}{2}$ *:*

$$\underset{n \longrightarrow \infty}{lim} \mathbb{P}(U < M_1) = \frac{1}{2} \tag{21}$$

In order to prove this proposition, we need to prove the first arcsine law (see [3]):

Theorem 4.10. *(First arcsine law for Brownian motion). Let* $\{B(t) : t \geq 0\}$ *be a standard linear Brownian motion.*

Then the random variable $A \in$ [0,1]*, which is uniquely determined by* $B(A) = \underset{s \in [0,1]}{max} B(s)$*, is arcsine distributed.*

Proof of Theorem 4.10. $\{A(t) : 0 \leq t \leq 1\}$ is defined by $A(t) = \max_{0 \leq s \leq t} B(s)$. For $s \in [0, 1]$,

$$\begin{aligned}
\mathbb{P}\{A \leq s\} &= \mathbb{P}\left(\max_{0 \leq u \leq s} B(u) > \max_{s \leq v \leq 1} B(v)\right) &(22)\\
&= \mathbb{P}\left(\max_{0 \leq u \leq s} B(u) - B(s) > \max_{s \leq v \leq 1} B(v) - B(s)\right) &(23)\\
&= \mathbb{P}\{A_1(s) > A_2(1-s)\}, &(24)
\end{aligned}$$

where $\{A_1(t) : 0 \leq t \leq 1\}$ is the maximum process of the Brownian motion $\{B_1(t) : t \geq 0\}$, which is given by $B_1(t) = B(s-t) - B(s)$, and $\{A_2(t) : 0 \leq t \leq 1\}$ is the maximum process of the independent Brownian motion $\{B_2(t) : 0 \leq s \leq 1\}$, which is given by $B_2(t) = B(s+t) - B(s)$. Since, as a consequence of reflection principle, for any fixed t, the random variable $A(t)$ has the same law as $|B(t)|$, we have

$$\mathbb{P}\{A_1(s) > A_2(1-s)\} = \mathbb{P}\{|B_1(s)| > |B_2(1-s)|\} \tag{25}$$

Using the scaling invariance of Brownian motion we can express this in terms of a pair of two independent standard normal random variables N_1 and N_2 , by

$$\begin{aligned}
\mathbb{P}\{A_1(s) > A_2(1-s)\} &= \mathbb{P}\{|B_1(s)| > |B_2(1-s)|\} && (26)\\
&= \mathbb{P}\left(\sqrt{s}|N_1| > \sqrt{1-s}|N_2|\right) && (27)\\
&= \mathbb{P}\left(\frac{|N_2|}{\sqrt{N_1^2+N_2^2}} < \sqrt{s}\right). && (28)
\end{aligned}$$

In polar coordinates, $(N_1, N_2) = (R\cos\theta, R\sin\theta)$ pointwise. The random variable θ is uniformly distributed on $[0, 2\pi]$. So the last quantity becomes

$$\begin{aligned}
\mathbb{P}\left(\frac{|N_2|}{\sqrt{N_1^2+N_2^2}} < \sqrt{s}\right) &= \mathbb{P}\{|\sin(\theta)| < \sqrt{s}\} && (29)\\
&= 4\mathbb{P}\{\theta < \arcsin(\sqrt{s})\} && (30)\\
&= \frac{4}{\pi}\arcsin(\sqrt{s}). && (31)
\end{aligned}$$

It follows by differentiating that A has density $(\pi s(1-s))^{-1}$ for $s \in (0,1)$.

□

Lemma 4.11. *For any X random variable with symmetric distribution and Y random variable with uniform distribution, when the density function of X is $f : [0,1] \to$*

[0,1]*:*

$$\mathbb{P}(X > Y) = \frac{1}{2}, \tag{32}$$

Proof of Lemma 4.11.

$$\mathbb{P}(X > Y) = \mathbb{E}[X], \tag{33}$$

since

$$\mathbb{P}(X > Y) = \int_0^1 \mathbb{P}(X > Y | X = x) \cdot f(x) dx \tag{34}$$

$$= \int_0^1 x \cdot f(x) dx \tag{35}$$

$$= \mathbb{E}[X] \tag{36}$$

As X is symmetric on [0,1], $\mathbb{E}[X] = \frac{1}{2}$, so we are done with the proof.

□

Now let us consider the proof of Proposition 4.9:

Proof of Proposition 4.9. As $r = \frac{1}{2}$, for $\omega \in \Omega_n$ with respect to $\mathbb{P}_r$, $S_k(\omega)$ is a simple symmetric random walk, therefore by Donsker's theorem, $\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}$ converges to $B(t)$, $t \in [0,1]$ in distribution, so by the first arcsine law (see 4.9), $\frac{M_1}{n}$ converges to the arcsine distribution, which is a symmetric distribution, so for its density function $f : [0,1] \to [0,1]$: $f(x) = f(1 - x)$.

It is trivial that $\frac{U}{n} \stackrel{d}{=} \text{Uni}\{\frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n}\}$.

Therefore

$$\mathbb{P}(M_1 > U) = \mathbb{P}(\frac{M_1}{n} > \frac{U}{n}) \to \frac{1}{2}, \text{ as } n \to \infty. \tag{37}$$

□

Remark 4.12. However M_1 is not necessarily the switching position of the existence of the excess (by switching position we mean the special position from which, if the 1 is to the left in the string, there is no excess, and if it is to the right, there is), the switching position might have some shift from M_1, which is the distance of M_1 and

K, but we will see that this distance will be 'small enough' in the string of length $n \to \infty$.

Proposition 4.13. $\exists c \in \mathbb{R}$ *such that* $\forall \omega \in \Omega_n$:

$$\mathbb{E}[M_1(\omega) - K(\omega)] \leq c\sqrt{n} \tag{38}$$

Remark 4.14. For the proof of Proposition 4.13, we need to understand the behaviour of the string to the left of the maximum position M_1.

It is essential to see that in the $r = \frac{1}{2}$ case, starting from M_1 and going backwards, the height function is a simple symmetric random walk conditioned on the fact that it never returns to the value of the maximum.

In order to estimate the distance of M_1 and K we split the string between K and M_1 into parts that we call excursions and one or two single steps between the excursions and consider the height function of the string. One excursion starts at some height and continues until it reaches one height below. There are some excursions to this lower level and there is a last one. After that we never reach that level again. Then we start the excursion from there conditioning on that we never reaches that level again. In this case after an excursion ended, the next one or two steps are downward steps for sure in order to avoid the 'banned' height. Two steps when it's the last excursion to a certain height and one when it's not the last one. That's why there are single steps between the excursions. Therefore we can interpret these excursions (after applying a reflection to the horizontal axis) as simple symmetric random walks' first visit to one step higher level than the starting point.

$$V_1 = min\{n|S_n = 1\}$$

is the visiting time of one,

$$V_0 = min\{n \geq 1|S_n = 0\}$$

is the returning time to 0, and S_n is a simple symmetric random walk.

So we have to study some properties of this conditioned random walk, such as the excursion probabilities and expected excursion length.

Proposition 4.15. *We consider the* $\tilde{X_n} := (X_n|T_1 = \infty)$ *random walk, where* X_n *is the simple symmetric random walk.*

Then the distribution of the number of excursions is the following:

$$\mathbb{P}(X = k) = \begin{cases} \frac{1}{2}, & \text{if } k = 0 \\ \frac{1}{8}(\frac{3}{4})^{k-1}, & \text{if } k \geq 1. \end{cases} \tag{39}$$

Proof. Let us introduce the following random variables:

$$X = \{\#\text{excursions}\} \tag{40}$$

$$G_i = \{\#\text{returns to } i \text{ from } i+1 \text{ in } \tilde{X_n}\} \tag{41}$$

It's easy to see that G_i's are iids and $G_i \stackrel{d}{=} Geo(\frac{1}{2})$ (pessimistic geometric) for $i = 1,2,...$

Let

$$N = min\{k|G_k = 0\} - 1. \tag{42}$$

$N \stackrel{d}{=} Geo(\frac{1}{2})$ (pessimistic), since $\mathbb{P}(G_i = 0) = \mathbb{P}(G_i > 0) = \frac{1}{2}\ \forall i$.

By the definition of the random variables

$$X = \sum_{i=1}^{N} G_i,$$

which is a random sum. N depends on the G_i's, but this dependence is only that in case of $N = n$, $G_i > 0\ \forall i < n$, therefore $(G_i|N = n, i < n) \stackrel{d}{=} Geo(\frac{1}{2})$ (optimistic geometric). So we can use the probability generating function of them in the following way:

$$g_X(z) = g_N(g_{G_1}(z)) = \frac{1}{2 - \frac{z}{2-z}} = \frac{2-z}{4-3z}. \tag{43}$$

From the PGF of X, we get the distribution we wanted:

$$\mathbb{P}(X = k) = \begin{cases} \frac{1}{2}, & \text{if } k = 0 \\ \frac{1}{8}(\frac{3}{4})^{k-1}, & \text{if } k \geq 1. \end{cases} \tag{44}$$

. □

Remark 4.16. The distribution of the number of excursions is exponentially decaying, from which we can conclude that there are not expected to be many excursions so we can guess that the shift from M_1 will not be large.

However this is not a formal argument, moreover this is not exactly the case we are interested in.

In our case it is not necessary that the random walk never returns to 0 (the starting point), it is only sufficient that for fixed $M_1 = m$ the string does not return to 0 on any finite trajectory to the left of M_1.

So from now on we study the excursions of the random walk conditioned on not returning to 0 in $M_1 = m$ steps.

Proposition 4.17.

$$\mathbb{P}_2(V_1 < m | V_0 > m) = \frac{1}{2}(1 + \mathcal{O}(\sqrt{m^{1-\epsilon}})) \tag{45}$$

This means that the excursion probability is near to $\frac{1}{2}$ if $m > \sqrt{n}$, where n is the size of the string, therefore the number of excursions exponentially decays.

Proof. We know that

$$\begin{aligned}\mathbb{P}_2(V_1 = 2k+1) &= \mathbb{P}_1(V_0 = 2k+1) \\ &= \mathbb{P}_1(S_1 > 0,...,S_{2k} > 0, S_{2k+1} = 0) \\ &\sim \frac{1}{2\sqrt{\pi}k^{\frac{3}{2}}}\end{aligned} \tag{46}$$

and

$$\mathbb{P}_1(\{V_0 > m\}) = \mathbb{P}_2(\{V_1 > m\}) \sim \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m}}, \tag{47}$$

these asymptotics follow from the computations done in the Feller book [7] by reflection principle.

It is trivial that from 2 to reach 1 we need odd many steps for sure, and

$$\mathbb{P}_2(\{V_1 < m < V_0 | V_1 = 2i+1\}) = \mathbb{P}_2(\{V_1 = 2i+1\})\mathbb{P}_1(\{V_0 > m-2i\}). \tag{48}$$

Therefore

$$\begin{aligned}\mathbb{P}_2(V_1 < m \mid V_0 > m) &= \frac{\mathbb{P}_2(V_1 < m < V_0)}{\mathbb{P}_2(V_0 > m)} \\ &\sim \frac{\sum_{i=1}^{\frac{m}{2}} \frac{1}{2\sqrt{\pi}i^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m-2i}}}{\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m}} + \sum_{i=1}^{\frac{m}{2}} \frac{1}{2\sqrt{\pi}i^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m-2i}}},\end{aligned} \tag{49}$$

We can estimate the numerator of the above expression in the following way: we sum it up until k instead of $\frac{m}{2}$, So

$$\begin{aligned}\sum_{i=1}^{k} \frac{1}{2\sqrt{\pi}i^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m-2i}} &\sim (1 - \frac{1}{\sqrt{\pi k}})\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m}} \\ &= \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m}}(1 + \mathcal{O}(\frac{1}{\sqrt{\pi k}}))\end{aligned} \tag{50}$$

Let us choose $k := \frac{n^{1-\epsilon}}{2}$ $(\epsilon > 0)$, so that the estimation is

$$\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m}}(1+\mathcal{O}(\frac{1}{\sqrt{m^{1-\epsilon}}})) \tag{51}$$

So

$$\mathbb{P}_2(V_1 < m|V_0 > m) = \frac{1}{2}(1+\mathcal{O}(\sqrt{m^{1-\epsilon}}) \tag{52}$$

is indeed the case. □

Proposition 4.18.

$$\mathbb{E}_2[V_1 \mathbb{1}_{\{V_1 \leq m\}}|V_0 > m, V_1 \leq m] \sim \sqrt{\frac{\pi}{2}}\sqrt{m} \tag{53}$$

Remark 4.19. According to Proposition 4.18, if there is an excursion in the remaining time m (remaining steps until the end of the string, which is basically the beginning of the string as we go backwards from M_1 now), the expected length of an excursion is in the order of magnitude $\sqrt{m}$.

Proof.

$$\begin{aligned}\mathbb{P}_2(V_1 = 2i+1|V_0 > m, V_1 < m) &\sim \frac{\frac{1}{2\sqrt{\pi}i^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m-2i}}}{\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{m}}\frac{1}{2}(1+\mathcal{O}(\frac{1}{\sqrt{m^{1-\epsilon}}}))} \\ &\sim \frac{1}{\sqrt{\pi}}\frac{1}{i^{\frac{3}{2}}}\sqrt{\frac{1}{1-\frac{2i}{m}}}\end{aligned} \tag{54}$$

$$\begin{aligned}\mathbb{E}_2[V_1 \mathbb{1}_{\{V_1 \leq m\}}|V_0 > m, V_1 \leq m] &\sim \sum_{i=1}^{\frac{m}{2}}(2i)\frac{1}{\sqrt{\pi}}\frac{1}{i^{\frac{3}{2}}}\sqrt{\frac{1}{1-\frac{2i}{m}}} \\ &\sim \sqrt{m}\frac{2}{\sqrt{\pi}}\sum_{i=1}^{\frac{m}{2}}\frac{1}{\sqrt{i}}\frac{1}{\sqrt{m-2i}} \\ &\sim \sqrt{2\pi}\sqrt{m},\end{aligned} \tag{55}$$

since

$$\frac{1}{m}\sum_{i=1}^{\frac{m}{2}}\frac{1}{\sqrt{i}}\frac{1}{\sqrt{m-2i}}\to\int_0^{\frac{1}{2}}\frac{1}{\sqrt{x}}\frac{1}{\sqrt{1-2x}}dx=\frac{\pi}{\sqrt{2}},\text{ as } n\to\infty. \tag{56}$$

□

Now we have the tools to prove Proposition 4.13.

Proof of Proposition 4.13. Let

$$W_i=\{\text{the length of the } i^{th} \text{ excursion}\} \tag{57}$$

and

$$N=\{\text{number of excursions in } m \text{ steps}\} \tag{58}$$

The shift can be interpreted as excursions of simple symmetric random walks from the current level to the level one level below and one or two deterministic downward steps between them (and starting the next random walk from there) so the following estimation holds:

$$M_1-K\le\sum_{i=1}^{N}W_i+2N \tag{59}$$

We have already seen that the number of excursions exponentially decays, therefore $\mathbb{E}[N]=\mathcal{O}(1)$.

Now we need to estimate the expected value of the sum.

N and the W_i's are not independent, hence a trick is needed. The conditioning to $N\ge k$ means that the k^{th} excursion finishes before m hence the previous upper bound can be applied to its length.

$$\begin{aligned}\mathbb{E}[\sum_{k=1}^{N} W_k] &= \sum_{k=1}^{\infty} \mathbb{E}[W_k \mathbb{1}_{\{N\geq k\}}] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[\mathbb{E}[W_k \mathbb{1}_{\{N\geq k\}} | \mathbb{1}_{\{N\geq k\}}]] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[W_k | N \geq k]\mathbb{P}(N \geq k) \leq \tilde{c}\sqrt{n}\end{aligned} \tag{60}$$

So

$$\mathbb{E}[M_1(\omega) - K(\omega)] \leq c\sqrt{n} \tag{61}$$

is indeed the case. □

Proof of Proposition 4.8.

$$\begin{aligned}\underset{n\longrightarrow\infty}{lim} \mathbb{P}(E_n = 0) &= \underset{n\longrightarrow\infty}{lim} \mathbb{P}(U < K) \\ &= \underset{n\longrightarrow\infty}{lim} (\mathbb{P}(U < M_1) + \underset{n\longrightarrow\infty}{lim} \mathbb{P}(U \in (K,M_1))) \\ &= \frac{1}{2},\end{aligned} \tag{62}$$

since

$$0 \leq \mathbb{E}[\mathbb{P}(U \in (K,M_1))|K,M_1] = \mathbb{E}[\frac{M_1 - K}{n}] \leq \frac{c}{\sqrt{n}}, \tag{63}$$

so it goes to 0 as n goes to infinity.

□

Proposition 4.20. *For $r < \frac{1}{2}$:*

$$\underset{n\longrightarrow\infty}{lim} \mathbb{P}(U < M_1) = 1$$

Proof of Proposition 4.20. The proof comes directly from Proposition 4.6. □

Proposition 4.21. *For $\omega \in \Omega_n$ with respect to $\mathbb{P}_r$,where $r < \frac{1}{2}$:*

$$\mathbb{P}(U(\omega) < K(\omega)) \to 1 \text{ as } n \to \infty$$

.

Proof of Proposition 4.21. Let us read the string ω and the height function from the end, so that the first element of ω is now considered to be the last and so on.

Let us split the string of length-n into $n^{\frac{1}{4}}$ parts, and on these sections, separately check whether it starts with such a K, i.e. whether it starts with a substring of 000 such that to the left of this position the height function never returns to the height after the first 0 of 000 plus two.

$$\mathbb{P}(\{\exists 000 \subseteq \omega\}) = (1-r)^3 > 0 \tag{64}$$

$$\mathbb{P}(\{\text{the height function never reaches the height after the first 0 of 000 plus two}\}) \tag{65}$$

$$= 1 - (\frac{r}{1-r})^2 > 0 \tag{66}$$

This equality comes from [4].

$$\{\text{the height function never reaches the height after the first 0 of 000 plus two,} \tag{67}$$

$$\text{when there is a 000}\} \tag{68}$$

$$\subseteq \{\text{the height function do not reach the height of 000 plus two} \tag{69}$$

$$\text{on a section of length } n^{\frac{1}{4}}, \text{when there is a 000 in the beginning of the section}\} \tag{70}$$

So for

$$b = \mathbb{P}(\{\text{the height function do not reach the height after the first 0 of 000 plus two} \tag{71}$$

$$\text{on a section of length } n^{\frac{1}{4}}, \text{ when there is a 000 in the beginning of the section}\}) > 0 \tag{72}$$

The sections are independent of each other, therefore

$$W \stackrel{d}{=} \mathrm{Geo}(b \cdot (1-r)^3)$$

, where W is the number of the section of length $n^{\frac{1}{4}}$ in which there is the 000 in the beginning from where the height do not return to the height after the first 0 of 000 (counted from the end of the string) within the section, but we will see that the probability of that the return time is greater than $n^{\frac{1}{4}}$ goes to zero:

Let V denote the return time.By the Hoeffding's inequality:

$$\mathbb{P}(S_n \geq 0) \leq e^{-cn} \text{ for some } c > 0. \tag{73}$$

Therefore

$$\mathbb{P}(V > n) \le \mathbb{P}(\underset{k\ge n}{max} \ge 0) \le e^{-\frac{cn}{2}} \text{ for n big enough,} \tag{74}$$

so

$$\mathbb{P}(V > n^{\frac{1}{4}}) \le e^{-\frac{cn^{\frac{1}{4}}}{2}} \tag{75}$$

So the probability that the return time is greater than $n^{\frac{1}{4}}$, anywhere over the string ω, is:

$$\mathbb{P}(\text{the return time is greater than } n^{\frac{1}{4}}\text{, anywhere over the string}) \tag{76}$$

$$\le n \cdot e^{-\frac{cn^{\frac{1}{4}}}{2}} \to 0 \text{ as } n \to \infty. \tag{77}$$

So the order of magnitude of the distance of the position K from n will be

$$\text{Geo}(b \cdot (1 - r^3)) \cdot n^{\frac{1}{4}}$$

, and

$$\frac{\text{Geo}(b \cdot (1 - r^3)) \cdot n^{\frac{1}{4}}}{n^{\frac{1}{2}}} = \frac{\text{Geo}(b \cdot (1 - r)^3)}{n^{\frac{1}{4}}} \to 0 \text{ as } n \to \infty, \tag{78}$$

meaning that the distance of this K position from the end of the string is less than $n^{\frac{1}{4}}$, hence

$$\mathbb{P}(U < K) = \mathbb{P}(\frac{n-U}{\sqrt{n}} > \frac{n-K}{\sqrt{n}}) \to 1 \text{ as } n \to \infty \tag{79}$$

□

Proof of Theorem 4.2. For $r > \frac{1}{2}$, $E_n \Rightarrow 1$ comes directly from Proposition 4.5, Proposition 4.3 and Proposition 4.6, for $r < \frac{1}{2}$, $E_n \Rightarrow 0$ comes from Proposition 4.3, and Proposition 4.21, and in case of $r = \frac{1}{2}$, $E_n \Rightarrow X$, where $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$, it comes from Proposition 4.5 and Proposition 4.8. □

Let us consider the critical case of r:

Theorem 4.22. *If* $r = \frac{1}{2} + \frac{\lambda}{\sqrt{n}}$*:*

$$\mathbb{P}(E_n = 1) \to C_\lambda, \text{ as } n \to \infty, \tag{80}$$

where

$$C_\lambda = \mathbb{E}[\underset{s\in[0,1]}{argmax} B_s^\lambda], \tag{81}$$

where B_s^λ *is a scaled Brownian motion with* λ *drift.*

The above theorem means that for a critically scaled r, the excess probability can be expressed exactly by the argmax location of the brownian motion with λ drift scaled to [0,1]. We have already seen that $C_0 = \frac{1}{2}$.

Appendix

A Proof of Theorem 3.1

This was already proven in a previous paper [1], but this is a slightly different approach of the problem, which is helpful for the understanding of the three type string case as well.

Remark 4.23. In order to prove the Theorem we use the following random variables on the probability space Ω_n:L,R,M_1,N,Z.

Let $L : \Omega_n \mapsto \mathbb{N}$ be the number of 1's until the first occurrence of a 0 in the initial string ω, read from the left to the right (so in other words, the number of 1's at the beginning of the string), and similarly, let $R : \Omega_n \mapsto \mathbb{N}$ be the number of 0's until the first occurrence of a 1, read from the right to the left,so the number of 0's at the end of the string.

$$L(\omega) = \{k : \omega_1 = 1 \wedge \omega_2 = 1 \wedge ... \wedge \omega_k = 1 \wedge \omega_{k+1} \neq 1\}$$

$$R(\omega) = \{k : \omega_n = 0 \wedge \omega_{n-1} = 0 \wedge ... \wedge \omega_{n-k+1} = 0 \wedge \omega_{n-k} \neq 0\}$$

Let $\tilde{\omega}$ be the substring of $\omega \in \Omega_n$ in such a way that we chop off the first L and the last R elements of ω, so that we get a string in $\{0,1\}^{n-L-R}$ which starts with a 0 and ends with a 1.

Let $M_1 : \Omega_n \mapsto \mathbb{N}$ denote that position in ω which is the rightmost position in the string $\tilde{\omega} \in \{0,1\}^{n-L-R}$ from where any nonempty suffix of the substring to the left of this position contains strictly higher amount of 0's than 1's.

$$M_1(\omega) = \underset{k \geq 1}{max}\{k : L < k < n - R \wedge \forall L < l < k : \sum_{i=l}^{k} \mathbb{1}_{\{\omega_i = 0\}} > \sum_{i=l}^{k} \mathbb{1}_{\{\omega_i = 1\}}\}$$

Let $Z : \Omega_n \mapsto \mathbb{N}$ be the number of 0's before the position of M_1 and let $N : \Omega_n \mapsto \mathbb{N}$ be the number of 1's after the position of M_1.

$$Z(\omega) = \{ \sum_{i=1}^{M_1(\omega)} \mathbb{1}_{\{\omega_i = 0\}}\}$$

$$N(\omega) = \{ \sum_{i=M_1(\omega)+1}^{n} \mathbb{1}_{\{\omega_i = 1\}}\}$$

A concrete example:

$$\omega = \underbrace{11}_{L(\omega)}0100 \overset{M_1(\omega)}{|} 1\underbrace{000}_{R(\omega)}$$

In this case $L(\omega) = 2$,$R(\omega) = 3$, $M_1(\omega) = 6$,$Z(\omega) = 3$, and $N(\omega) = 1$.

Remark 4.24. In order to understand M_1 better, we assign a random walk to each $\omega \in \{0,1\}^n$ in the following way: it takes a step up for every 0 in ω and a step down for every 1 in ω. So for $1 \leq k \leq n$ let $S_k = \sum_{i=1}^{k}(1 - 2\omega_i)$, where ω_i is the i^{th} element of ω. The function $f : \mathbb{N} \to \mathbb{N}$, where $k \mapsto S_k$ gives the height function of the aforementioned random walk.

With this interpretation, M_1 is the leftmost location of the maximum height in the height function of $\omega \in \{0,1\}^n$, or more precisely:

$$M_1 = \min_k\{k | L \leq k \leq n - R \wedge S_k = \max_i\{S_i | L \leq i \leq n - R\}\}$$

, where S_k represents the random walk of ω.

Remark 4.25. The two interpretations of M_1 are indeed the same, so being the rightmost position from where any suffix of the substring to the left of this position contains strictly more 0's than 1's means exactly that the height function never reaches the height of the M_1 position before (to the left of) the position M_1 and it never surpasses it. It is because if the substring to the left of this position has no suffix such that the number of 1's in them is greater or equal to the number of 0's in them means that the height function restricted to $[0,M_1]$ has a unique maximum in M_1 and as it is the rightmost position like this, it must be a maximum of the height function over the entire domain as well so it is indeed the leftmost maximum.

Now let us express the stabilization time of the strings with the help of these random variables, which is a different approach to the one in the [1] article, in order to understand the development of the evolution process better.

Proposition 4.26.

$$T_n(\omega) = Z(\omega) - 1 + N(\omega) \tag{83}$$

With the Proposition above, we can express the stabilization time in a slightly different way to the one on the previous paper [1].

For the proof of this proposition, we need to examine the changes of the variables throughout the evolution.

Remark 4.27. Firstly, let us check that M_1 is in the following position for $\forall\omega \in \Omega_n$ for which $T_n(\omega) > 0$:

$$...0 \overset{M_1}{|} 1...$$

(so its position is always between a 0 and a 1 unless ω is stabilized already).

It is true, because if it would be positioned like this: ...0|0..., than ...00|... would be a righter position for which it is true that every suffix of the substring to the left contains more 0's than 1', so it would be contradiction.

And of course M_1 cannot be followed by a 1 as it would not satisfy the condition for the suffixes.

Lemma 4.28. *For any $\omega \in \{0{,}1\}^n$ such that $\exists$ 01 $\in \omega$:*

$$T(S\omega) = T(\omega) - 1$$

It is trivial, since with each evolution step S, we get one step closer to the stabilized state of ω, and we need exactly $T(\omega)$ steps to reach it.

So if $T(\omega) = n$,

$$\omega_0 = \omega \text{ (steps needed until stable state: n)}$$

$$\omega_1 = S\omega \text{ (steps needed until stable state: n-1)}$$

$$\omega_2 = S^2\omega \text{ (steps needed until stable state: n-2)}$$

$$\vdots$$

$$\omega_{n-1} = S^{n-1}\omega = 1...1010...0 \text{ (steps needed until stable state: 1)}$$

$$\omega_n = S^n\omega = 1...10...0 \text{ (steps needed until stable state: 0)}$$

Lemma 4.29. *For any $\omega \in \{0{,}1\}^n$ such that $\exists$ 01 $\in \omega$:*

$$Z(S\omega) = \begin{cases} Z(\omega) - 1 & \textit{,if } Z(\omega) > 1 \\ 1 & \textit{,if } Z(\omega) = 1 \textit{ and } N(\omega) > 1 \\ 0 & \textit{,if } Z(\omega) = 1 \textit{ and } N(\omega) = 1 \end{cases}$$

$$N(S\omega) = \begin{cases} N(\omega) - 1 & \textit{,if } Z(\omega) = 1 \textit{ and } N(\omega) > 1 \\ N(\omega) & \textit{otherwise} \end{cases}$$

If this is true, then we are done with the proof of Proposition 4.26's first half (the left equality), since the above mentioned behaviour of Z and N during the evolution process means the following:

After $Z(\omega)-1$ steps (while the value of Z decreases from $Z(\omega)$ to 1), the value of N starts to decrease from $N(\omega)$ to 1 in $N(\omega)-1$ steps, and then we have $\omega' = S^{(Z(\omega)-1)+(N(\omega)-1)}\omega$, and as $N(\omega')=1$ and $Z(\omega')=1$, ω' has the following form: $\omega' = 11...110|100...00$, so there is only one 01 in the string so it is stabilized in one step, which means that ω needs indeed $Z(\omega)-1+N(\omega)-1+1 = Z(\omega)+N(\omega)-1$ evolution steps until it reaches the stable state, so $T(\omega)=Z(\omega)+N(\omega)-1$.

To prove the aforementioned changes of Z and N, we need to understand the behaviour of M_1 during the evolution.

Lemma 4.30.

$$M_1(S\omega) = \begin{cases} M_1(\omega)-1 & ,\textit{if } Z(\omega)>1 \\ M_1(\omega)+1 & ,\textit{if } Z(\omega)=1 \end{cases}$$

Proof. In order to prove that if $Z(\omega)>1$, then $M_1(S\omega)=M_1(\omega)-1$ (so that with an evolution step the leftmost maximum position moves one step to the left), it is enough to prove that to the left of the position $M(\omega)-1$ in $S\omega$, there is no such position where the height function reaches the value at $M(\omega)-1$, and to the right of it, there is no such position where the height function exceeds it, because in that case, it is indeed the position of the leftmost maximum.

More precisely, we need to prove the following:

$$\nexists L(S\omega)<k<M_1(S\omega), \text{ such that } S_k(S\omega)\geq S_{M_1(\omega)-1}(S\omega)$$

and

$$\nexists M_1(S\omega)<k<n-R(S\omega), \text{ such that } S_k(S\omega)>S_{M_1(\omega)-1}(S\omega)$$

We know that the height of the position $M_1(\omega) - 1$ in ω is one less than the height of the position $M_1(\omega)$ in ω, and it does not change during the evolution step, so in other words:

$$S_{M_1(\omega)-1}(S\omega) = S_{M_1(\omega)-1}(\omega) = S_{M_1(\omega)}(\omega) - 1,$$

since:

$$\omega = ...00 \overset{M_1(\omega)}{|} 1... \xrightarrow{S} ...0 \overset{M_1(\omega)-1}{|} 10... = S\omega$$

And we also know the following by the definition of M_1:

$$\nexists L(\omega) < k < M_1(\omega), \text{ such that } S_k(\omega) \geq S_{M_1(\omega)}(\omega)$$

and

$$\nexists M_1(\omega) < k < n - R(\omega), \text{ such that } S_k(S\omega) > S_{M_1(\omega)}(\omega)$$

Therefore if there is a k before the position $M_1 - 1$ in ω for which the height is one less than at the maximum (it cannot be the same or higher), it must be in a ...0|1... position (because it has to reach the $\underset{k}{max}\{S_k(\omega)|L(\omega) < k < n - R(\omega)\} - 1 = S_{M_1(\omega)} - 1$ with a 0, which has to be followed by a 1, otherwise it would reach the maximum to the left of M_1), which turns into ...1|0... with the evolution step, so it decreases by two, so there is no position to the left of $M_1(\omega) - 1$ in $S\omega$ where the height function reaches the height at $M_1(\omega) - 1$.

Now we need to check the other part: No position k to the right of the position $M_1(\omega)$ in ω can exceed the maximum height, but there can be other maximum values, which are ...0|1... positions for sure, so with the evolution step the height at that position decreases by two, so there is no position to the right of $M_1(\omega) - 1$ in $S\omega$ where the height function exceeds the height at $M_1(\omega) - 1$. So for $Z(\omega) > 1$, $M_1(S\omega) = M_1(\omega) - 1$ is indeed satisfied.

Now let us consider the case when $Z(\omega) = 1$. This means that an evolution step of the string ω looks like this:

$$\underbrace{11...1}_{L(\omega)}0\overset{M_1(\omega)}{|}1... \xrightarrow{S} \underbrace{11...1}_{L(\omega)+1}0\overset{M_1(S\omega)}{|}1...,$$

because the positions different from $M_1(\omega)$ where the height function has maximum, also decreases by the evolution step, so $M_1(S\omega)$ cannot be in any other position, and this also means that once M_1 reaches the edge (so when its positioned after the first 0 in the string so when $M_1 = L + 1$), it stays there and L increases by one at each step since in every step one 1 from the right of M_1 moves to the other side, so becomes one of the 1's from the beginning, so part of L.

So we have seen that in case of $Z(\omega) = 1$, $M_1(\omega) = L(\omega) + 1$, and $L(S\omega) = L(\omega) + 1$, therefore $M_1(S\omega) = M_1(\omega) + 1$ is indeed satisfied, and with that we proved the behaviour of the random variables M_1, N and Z during an evolution step and throughout the evolution,so now we are done with the proof of Proposition 4.26 which says: $T_n(\omega) = Z(\omega) + N(\omega) - 1$. □

Lemma 4.31. *In the case* $r > \frac{1}{2}$*:*

$$\frac{T_n - rn}{\sqrt{n}} \Rightarrow \mathcal{N}(0, r(1-r))$$

Proof. We can express Z and N with the aforementioned random walk associated with ω, $S_k = \sum_{i=1}^{k}(1 - 2\omega_i)$, where $X_i = 1 - 2\omega_i$ are independent and identically distributed (iid) random variables for $i = 1,....,n$ with distribution $\mathbb{P}(X_i = -1) = r > \frac{1}{2}$ and $\mathbb{P}(X_i = 1) = 1 - r < \frac{1}{2}$.

$$S_{M_1} = 2Z - M_1, \text{ so } Z = \frac{S_{M_1} + M_1}{2}$$

And similarly

$$S_n - S_{M_1} = n - M_1 - 2N, \text{ so } N = \frac{n - M_1 - S_n + S_{M_1}}{2}$$

Therefore we get the following formula for the stabilization time of ω:

$$Z + N - 1 = \frac{1}{2}(n - S_n) + S_{M_1} - 1$$

S_n is the sum of iid random variables with mean $1 - 2r$ and variance $4r(1 - r)$, or more precisely: $\mathbb{E}[X_i] = r \cdot (-1) + (1-r) \cdot 1 = 1 - 2r$ and $Var(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = 1 - (1 - 2r)^2 = 4r(1 - r)$.

So by the central limit theorem, we can estimate S_n with normal distribution:

$$S_n = (1 - 2r)n + 2\sqrt{r(1 - r)}\sqrt{n}\eta_n \text{ ,where } \eta_n \xrightarrow{d} \mathcal{N}(0,1)$$

So

$$\frac{n - S_n}{2} = rn + \sqrt{r(1 - r)}\sqrt{n}\eta_n$$

Hence,

$$\frac{T_n - rn}{\sqrt{n}} =$$
$$\frac{Z + N - 1 - rn}{\sqrt{n}} =$$
$$\frac{\frac{1}{2}(n - S_n) - rn}{\sqrt{n}} + \frac{S_{M_1} - 1}{\sqrt{n}}$$

And by the above mentioned equality, we can easily see that:

$$\frac{\frac{1}{2}(n - S_n) - rn}{\sqrt{n}} \Rightarrow \mathcal{N}(0, r(1 - r))$$

so we only need to check that

$$\frac{S_{M_1} - 1}{\sqrt{n}} \xrightarrow{\mathbb{P}} 0$$

is indeed the case, and then we are done with the proof of the lemma.

Let us consider $S_{M_1} + L$, which is like we have shifted the height function with L, thus we get the maximum value of the height function related to the substring $\tilde{\omega} \in \{0,1\}^{n-L-R}$, which starts with a 0 deterministically by definition of L and $\tilde{\omega}$, so we need to subtract one in order to get to the weighted random walk S_k with a negative drift, where it takes one step upwards with probability $1 - r < \frac{1}{2}$ and one step downwards with probability $r > \frac{1}{2}$.

Therefore $S_{M_1} + L - 1 \stackrel{d}{=} \underset{1 \leq k \leq n}{max} \{S_k\}$ for any n-length string, and we can overestimate this maximum for any n with the maximum over the whole trajectory, which has a geometric distribution with parameter $a = 1 - \frac{1-\sqrt{1-4r(1-r)}}{2r} < 1$, so we can say that $\underset{1 \leq k \leq n}{max} \{S_k\}$ is stochastically dominated by $\underset{1 \leq k < \infty}{max} \{S_k\}$ for any $n \in \mathbb{N}$, or more precisely:

$$\underset{1 \leq k \leq n}{max} \{S_k\} \preceq \underset{1 \leq k < \infty}{max} \{S_k\},$$

and since stochastic dominance is the property of distributions:

$$S_{M_1} + L - 1 \preceq \text{Geo}(1-a)$$

It is easy to see that

$$L \sim \text{Geo}(1-r),$$

so the following is also true:

$$S_{M_1} - 1 \preceq S_{M_1} + L - 1 \preceq \text{Geo}(1-a)$$

Geo$(1-a)$ is almost surely finite, therefore, by the Markov's inequality, we get that:

$$\frac{\text{Geo}(1-a)}{\sqrt{n}} \to 0,$$

so by the stochastic dominance

$$\frac{S_{M_1} - 1}{\sqrt{n}} \to 0$$

as well, and that is exactly what we wanted.

The parameter a comes from one step reasoning method, where

$$a = \mathbb{P}(\exists k : S_k = 1 | S_0 = 0)$$

so the probability that we ever reach one from zero during the random walk, which we can achieve in one step if the first step is up (which happens with probability $1 - r$) or move one step further away from it in the first step if it is a step down (with probability r), so from

$$a = (1 - r) + ra^2$$

we can get the aforementioned $a = 1 - \frac{1-\sqrt{1-4r(1-r)}}{2r} < 1$.

□

Lemma 4.32. *If* $r = \frac{1}{2}$*:*

$$\frac{T_n^- \frac{1}{2}n}{\sqrt{n}} \Rightarrow \frac{\chi_3}{2}$$

Lemma 4.33. $\frac{\chi_3}{2} \stackrel{d}{=} A_1 - \frac{1}{2}B_1$, *where* B_t *is a Brownian motion,* $t \in [0,1]$ *and* $A_t = \max_{s \leq t} B_s$

Proof. We know the probability density function of $\frac{\chi_3}{2}$:

$$d\frac{\chi_3}{2} = \frac{8\sqrt{2}}{\sqrt{\pi}} x^2 exp(-2x^2)dx \text{ for } x > 0$$

In order to prove that they are identically distributed, it is enough to check the density function of $A_1 - \frac{1}{2}B_1$, and for that, first let us consider the joint distribution function of A_1 and B_1.

By the reflection principle, we know that:

$$\mathbb{P}(A_1 > a) = 2\mathbb{P}(B_1 > a) = 2(1 - \Phi(a))$$

So the joint density function is the following:

$$F(a,b) = \mathbb{P}(A_1 \leq a, B_1 \leq b) \tag{84}$$
$$= \mathbb{P}(B_1 \leq b) - \mathbb{P}(A_1 \geq a, B_1 \leq b) \tag{85}$$
$$= \Phi(b) - (1 - \Phi(2a - b)) \tag{86}$$

We can get the joint density function by deriving the joint distribution function:

$$f(a,b) = \frac{\partial^2 F(a,b)}{\partial a \partial b} \tag{87}$$
$$= \frac{\partial}{\partial b}\phi(2a - b) \cdot 2 \tag{88}$$
$$= \sqrt{\frac{2}{\pi}}(2a - b)e^{-\frac{1}{2}(2a-b)^2} \tag{89}$$

This is true for $b > a$, $b > 0$, and 0 otherwise.

Now we are only one last step from the density function of $A_1 - \frac{1}{2}B_1$, $\rho : \mathbb{R} \mapsto [0,1]$, which comes from an elementary convolution of the two random variables:

$$\rho(x) = \int_{-2x}^{2x} f(\frac{y}{2} + x,y)dy \tag{90}$$
$$= \sqrt{\frac{2}{\pi}} \int_{-2x}^{2x} ((y + 2x) - y) \exp\left(-\frac{((y + 2x) - y)^2}{2}\right) dy \tag{91}$$
$$= 8\sqrt{\frac{2}{\pi}}x^2 \exp(-2x^2) \tag{92}$$

So we have seen that $\frac{\chi_3}{2} \stackrel{d}{=} A_1 - \frac{1}{2}B_1$ is indeed the case. □

Therefore it is enough to prove the following proposition and we are done with the proof of the 4.32.

Proposition 4.34.

$$\frac{T_n - \frac{1}{2}n}{\sqrt{n}} \Rightarrow A_1 - \frac{1}{2}B_1$$

Proof. Similarly to the previous proof:

$$\frac{T-\frac{n}{2}}{\sqrt{n}} = \frac{Z+N-1-\frac{n}{2}}{\sqrt{n}} \tag{93}$$

$$= \frac{S_{M_1}-\frac{1}{2}S_n-1}{\sqrt{n}} \tag{94}$$

$$= \frac{S_{M_1}+L-L-\frac{1}{2}S_n-1}{\sqrt{n}} \tag{95}$$

$$= \frac{S_{M_1}+L-1}{\sqrt{n}} - \frac{L}{\sqrt{n}} - \frac{\frac{1}{2}S_n}{\sqrt{n}} \tag{96}$$

We have seen before, that $\frac{L}{\sqrt{n}} \xrightarrow{\mathbb{P}} 0$, and for the rest we can apply the Donsker's theorem (also known as functional central limit theorem), which states that an appropriately centered and scaled version of the empirical distribution function converges weakly to a Gaussian process, so in our case this means the following:

If we consider the aforementioned random walk as a piecewise linear function under the correct scaling, it converges to a Brownian motion. More precisely, $L_n : [0,1] \mapsto \mathbb{R}$ by $L_n(t) = \sqrt{\frac{1}{n}}\left(\sum_{k=1}^{\lfloor nt \rfloor} X_k + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor+1}\right)$, which means that we have rescaled the random walk, where $\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = \frac{1}{2}$ and in some sense made it continuous.

Therefore by Donsker's theroem, $L_n(t) \Rightarrow B(t)$, for any $t \in C[0,1]$ as $n \to \infty$.

Since $L_n\left(\frac{k}{n}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{k} X_i = \frac{1}{\sqrt{n}} S_k$, by applying the following function on L_n: $h(f(t)) = \underset{t\in[0,1]}{sup} f(t) - \frac{1}{2}f(1)$ for any $t \in [0,1]$, which is continuous on $C[0,1]$ with the sup norm, thus it respects weak limit(so it does not ruin it), we get exactly what we wanted, namely: $\frac{S_{M_1}+L-1-\frac{1}{2}S_n}{\sqrt{n}} = h(L_n(t)) \Rightarrow h(B(t) = A_1 - \frac{1}{2}B_1$ for any $t \in [0,1]$ as n goes to ∞, because $S_{M_1} + L - 1$ has the same distribution as the simple maximum of the random walk, only shifted. □

Now we are done with the whole proof of the Theorem 3.1.

Bibliography

[1] Jacob Funk, Mihai Nica and Michael Noyes, Stabilization time for a type of evolution on binary strings, 2013

[2] University of Michigan Undergraduate Mathematics Competition 13, 1996, Problem 1.

[3] Peter Mörters and Yuval Peres Brownian Motion, 2008

[4] Karoly Simon, Stochastic processes, 2023 File A, Corollary 10.3

[5] James Martin and Philipp Schmidt, Multi-type TASEP in discrete time

[6] A. Ferrari and C. Kipnis. Second class particles in the rarefaction fan, 1995

[7] Feller W.,An introduction to probability theory and its applications,1968