



**HAL**  
open science

## Sinai's walk: a statistical aspect

Pierre Andreatti

► **To cite this version:**

| Pierre Andreatti. Sinai's walk: a statistical aspect. 2007. hal-00119187v2

**HAL Id: hal-00119187**

**<https://hal.science/hal-00119187v2>**

Preprint submitted on 4 Sep 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sinai's walk : a statistical aspect

Pierre Andreatti \*

September 4, 2007

**Abstract:** We consider Sinai's random walk in random environment. We prove that the logarithm of the local time is a good estimator of the random potential associated to the random environment. We give a constructive method allowing us to built the random environment from a single trajectory of the random walk.

## 1 Introduction and results

In this paper we are interested in Sinai's walk i.e., a one dimensional random walk in random environment with three conditions on the random environment: two necessary hypothesis to get a recurrent process (see [Solomon(1975)]) which is not a simple random walk and an hypothesis of regularity which allows us to have a good control on the fluctuations of the random environment.

The asymptotic behavior of such walk has been understood by [Sinai(1982)] : this walk is sub-diffusive and at an instant  $n$  it is localized in the neighborhood of a well defined point of the lattice. It is well known, see (Zeitouni [2001] for a survey) that this behavior is strongly dependent of the random environment or, equivalently, by the associated random potential defined Section 1.2.

The question we solve here is the following: given a single trajectory of a random walk  $(X_l, 1 \leq l \leq n)$  where the time  $n$  is fixed, can we estimate the trajectory of the random potential where the walk lives ? Let us remark that the law of this potential is unknown as-well.

In their paper, [Adelman and Enriquez(2004)] are interested in the question of the distribution of the random environment that could be deduced from a single trajectory of the walk, on the other hand, our purpose is to get an approximation of the trajectory of the random potential.

In the paper [V. Baldazzi and Monasson(2006)] the authors are interested in a method to predict the sequence of DNA molecules. They model the unzipping of the molecule as a one-dimensional biased random walk for the fork position (number of open base pair)  $k$  in this landscape. The elementary opening ( $k \rightarrow k + 1$ ) and closing ( $k \rightarrow k - 1$ ) transitions happen with a probability that depends on the unknown sequence. This probability of transition follows an Arrhenius law wich is closed to the one we discuss here. The question they answer is: given an unzipping signal can we predict

---

\*Laboratoire MAPMO - C.N.R.S. UMR 6628 - Fédération Denis-Poisson, Université d'Orléans, (Orléans France).  
MSC 2000 60G50; 60J55; 62G05.

*Key words and phrases : random environment, random walk, Sinai's regime, non-parametric statistics*

the unzipping sequence ? Their approach is based on a Bayesian inference method which gives very good probabilities of prediction for a large amount of data. This means, in term of the walk, several trajectory on the same environment.

Our approach is purely probabilistic, it is based on good properties of the *local time* of the random walk which is the amount of time the walk spends on the points of the lattice. We treat a general case with a very few information on the random environment. We are able to reconstruct the random potential in a significant interval where the walk spends most of its time. Our proof is based on the results of [Andreoletti(2005)], in particular in a weak law of large number for the local time on the point of localization of the walk.

The largest part of this paper is devoted to the proof of a theoretical result (Theorem 1.7), we also present, at the end of the document, numerical simulations to illustrate our result. We give the main steps of the algorithm we use to rebuilt the random potential only by considering a trajectory of the walk. As an introduction we would like to comment one of these simulations:

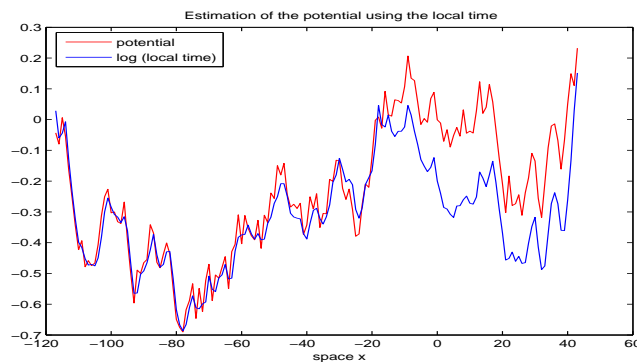


Figure 1: The logarithm of the local time (in blue) and the random potential (in red)

In blue we have represented the logarithm of the local time and in red the potential associated to the random environment. First, remark that we get a good approximation on a large neighborhood of the bottom of the valley around the coordinate -80. Outside this neighborhood and especially after the coordinate -20, the approximation is not precise at all. We will explain this phenomena by the fact that after the walk has reached the bottom of the valley, the walk will not return frequently to the points with coordinate larger than -20, so we lose information for this part of the lattice.

Our method of estimation give us two crucial information: a confidence interval for the differences of potential in sup-norm, on an observable set of sites “sufficiently” visited by the walk and a localization result for the bottom of the valley linked with the hitting time of the maximum of the local times. First we need to define the process:

### 1.1 Definition of Sinai’s walk

Let  $\alpha = (\alpha_i, i \in \mathbb{Z})$  be a sequence of i.i.d. random variables taking values in  $(0, 1)$  defined on the probability space  $(\Omega_1, \mathcal{F}_1, Q)$ , this sequence will be called random environment. A random walk in random environment (denoted R.W.R.E.)  $(X_n, n \in \mathbb{N})$  is a sequence of random variable taking value in  $\mathbb{Z}$ , defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  such that

- for every fixed environment  $\alpha$ ,  $(X_n, n \in \mathbb{N})$  is a Markov chain with the following transition proba-

bilities, for all  $n \geq 1$  and  $i \in \mathbb{Z}$

$$\begin{aligned}\mathbb{P}^\alpha [X_n = i + 1 | X_{n-1} = i] &= \alpha_i, \\ \mathbb{P}^\alpha [X_n = i - 1 | X_{n-1} = i] &= 1 - \alpha_i \equiv \beta_i.\end{aligned}\tag{1.1}$$

We denote  $(\Omega_2, \mathcal{F}_2, \mathbb{P}^\alpha)$  the probability space associated to this Markov chain.

•  $\Omega = \Omega_1 \times \Omega_2$ ,  $\forall A_1 \in \mathcal{F}_1$  and  $\forall A_2 \in \mathcal{F}_2$ ,  $\mathbb{P}[A_1 \times A_2] = \int_{A_1} Q(dw_1) \int_{A_2} \mathbb{P}^{\alpha(w_1)}(dw_2)$ .

The probability measure  $\mathbb{P}^\alpha [\cdot | X_0 = a]$  will be denoted  $\mathbb{P}_a^\alpha [\cdot]$ , the expectation associated to  $\mathbb{P}_a^\alpha$ :  $\mathbb{E}_a^\alpha$ , and the expectation associated to  $Q$ :  $\mathbb{E}_Q$ .

Now we introduce the hypothesis we will use in all this work. The two following hypothesis are the necessary hypothesis

$$\mathbb{E}_Q \left[ \log \frac{1 - \alpha_0}{\alpha_0} \right] = 0,\tag{1.2}$$

$$\text{Var}_Q \left[ \log \frac{1 - \alpha_0}{\alpha_0} \right] \equiv \sigma^2 > 0.\tag{1.3}$$

[Solomon(1975)] shows that under 1.2 the process  $(X_n, n \in \mathbb{N})$  is  $\mathbb{P}$  almost surely recurrent and 1.3 implies that the model is not reduced to the simple random walk. In addition to 1.2 and 1.3 we will consider the following hypothesis of regularity, there exists  $0 < \eta_0 < 1/2$  such that

$$\sup \{x, Q[\alpha_0 \geq x] = 1\} = \sup \{x, Q[\alpha_0 \leq 1 - x] = 1\} \geq \eta_0.\tag{1.4}$$

We call *Sinai's random walk* the random walk in random environment previously defined with the three hypothesis 1.2, 1.3 and 1.4.

Let us define the local time  $\mathcal{L}$ , at  $k$  ( $k \in \mathbb{Z}$ ) within the interval of time  $[1, T]$  ( $T \in \mathbb{N}^*$ ) of  $(X_n, n \in \mathbb{N})$

$$\mathcal{L}(k, T) \equiv \sum_{i=1}^T \mathbb{I}_{\{X_i=k\}}.\tag{1.5}$$

$\mathbb{I}$  is the indicator function ( $k$  and  $T$  can be deterministic or random variables). Let  $V \subset \mathbb{Z}$ , we denote

$$\mathcal{L}(V, T) \equiv \sum_{j \in V} \mathcal{L}(j, T) = \sum_{i=1}^T \sum_{j \in V} \mathbb{I}_{\{X_i=j\}}.\tag{1.6}$$

To end, we define the following random variables

$$\mathcal{L}^*(n) = \max_{k \in \mathbb{Z}} (\mathcal{L}(k, n)), \quad \mathbb{F}_n = \{k \in \mathbb{Z}, \mathcal{L}(k, n) = \mathcal{L}^*(n)\},\tag{1.7}$$

$$k^* = \inf\{|k|, k \in \mathbb{F}_n\}\tag{1.8}$$

$\mathcal{L}^*(n)$  is the maximum of the local times (for a given instant  $n$ ),  $\mathbb{F}_n$  is the set of all the favourite sites and  $k^*$  the smallest favorite site.

## 1.2 The random potential and the valleys

From the random environment we define what we will call random potential,  
Let

$$\epsilon_i \equiv \log \frac{1 - \alpha_i}{\alpha_i}, \quad i \in \mathbb{Z}, \quad (1.9)$$

define :

**Definition 1.1.** The random potential  $(S_m, m \in \mathbb{Z})$  associated to the random environment  $\alpha$  is defined in the following way:

$$S_k = \begin{cases} \sum_{1 \leq i \leq k} \epsilon_i, & \text{if } k > 0, \\ -\sum_{k+1 \leq i \leq 0} \epsilon_i, & \text{if } k < 0, \\ S_0 = 0. \end{cases}$$

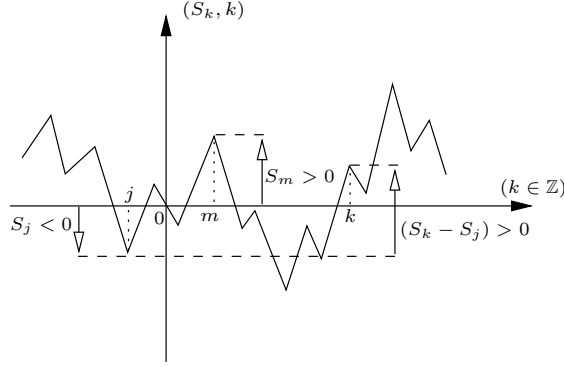


Figure 2: Trajectory of the random potential

**Definition 1.2.** We will say that the triplet  $\{M', m, M''\}$  is a *valley* if

$$S_{M'} = \max_{M' \leq t \leq m} S_t, \quad (1.10)$$

$$S_{M''} = \max_{m \leq t \leq M''} S_t, \quad (1.11)$$

$$S_m = \min_{M' \leq t \leq M''} S_t. \quad (1.12)$$

If  $m$  is not unique we choose the one with the smallest absolute value.

**Definition 1.3.** We will call *depth of the valley*  $\{M', m, M''\}$  and we will denote it  $d([M', M''])$  the quantity

$$\min(S_{M'} - S_m, S_{M''} - S_m). \quad (1.13)$$

Now we define the operation of *refinement*

**Definition 1.4.** Let  $\{M', m, M''\}$  be a valley and let  $M_1$  and  $m_1$  be such that  $m \leq M_1 < m_1 \leq M''$  and

$$S_{M_1} - S_{m_1} = \max_{m \leq t' \leq t'' \leq M''} (S_{t'} - S_{t''}). \quad (1.14)$$

We say that the couple  $(m_1, M_1)$  is obtained by a *right refinement* of  $\{M', m, M''\}$ . If the couple  $(m_1, M_1)$  is not unique, we will take the one such that  $m_1$  and  $M_1$  have the smallest absolute value. In a similar way we define the *left refinement* operation.

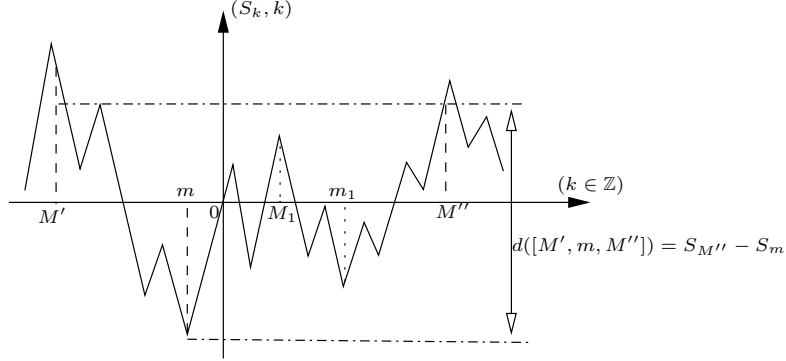


Figure 3: Depth of a valley and refinement operation

We denote  $\log_2 = \log \log$ , in all this section we will suppose that  $n$  is large enough such that  $\log_2 n$  is positive.

**Definition 1.5.** Let  $n > 3$ ,  $\gamma > 0$ , and  $\Gamma_n \equiv \log n + \gamma \log_2 n$ , we say that a valley  $\{M', m, M''\}$  contains 0 and is of depth larger than  $\Gamma_n$  if and only if

1.  $0 \in [M', M'']$ ,
2.  $d([M', M'']) \geq \Gamma_n$ ,
3. if  $m < 0$ ,  $S_{M''} - \max_{m \leq t \leq 0} (S_t) \geq \gamma \log_2 n$ ,  
if  $m > 0$ ,  $S_{M'} - \max_{0 \leq t \leq m} (S_t) \geq \gamma \log_2 n$ .

**The basic valley  $\{M_n', m_n, M_n\}$**

We recall the notion of *basic valley* introduced by Sinai and denoted here  $\{M_n', m_n, M_n\}$ . The definition we give is inspired by the work of [Kesten(1986)]. First let  $\{M', m_n, M''\}$  be the smallest valley that contains 0 and of depth larger than  $\Gamma_n$ . Here smallest means that if we construct, with the operation of refinement, other valleys in  $\{M', m_n, M''\}$  such valleys will not satisfy one of the properties of Definition 1.5.  $M_n'$  and  $M_n$  are defined from  $m_n$  in the following way: if  $m_n > 0$

$$M_n' = \sup \left\{ l \in \mathbb{Z}_-, l < m_n, S_l - S_{m_n} \geq \Gamma_n, S_l - \max_{0 \leq k \leq m_n} S_k \geq \gamma \log_2 n \right\}, \quad (1.15)$$

$$M_n = \inf \{ l \in \mathbb{Z}_+, l > m_n, S_l - S_{m_n} \geq \Gamma_n \}. \quad (1.16)$$

if  $m_n < 0$

$$M_n' = \sup \{l \in \mathbb{Z}_-, l < m_n, S_l - S_{m_n} \geq \Gamma_n\}, \quad (1.17)$$

$$M_n = \inf \left\{ l \in \mathbb{Z}_+, l > m_n, S_l - S_{m_n} \geq \Gamma_n, S_l - \max_{m_n \leq k \leq 0} S_k \geq \gamma \log_2 n \right\}. \quad (1.18)$$

if  $m_n = 0$

$$M_n' = \sup \{l \in \mathbb{Z}_-, l < 0, S_l - S_{m_n} \geq \Gamma_n\}, \quad (1.19)$$

$$M_n = \inf \{l \in \mathbb{Z}_+, l > 0, S_l - S_{m_n} \geq \Gamma_n\}. \quad (1.20)$$

$\{M_n', m_n, M_n\}$  exists with a  $Q$  probability as close to one as we need. In fact it is not difficult to prove the following lemma

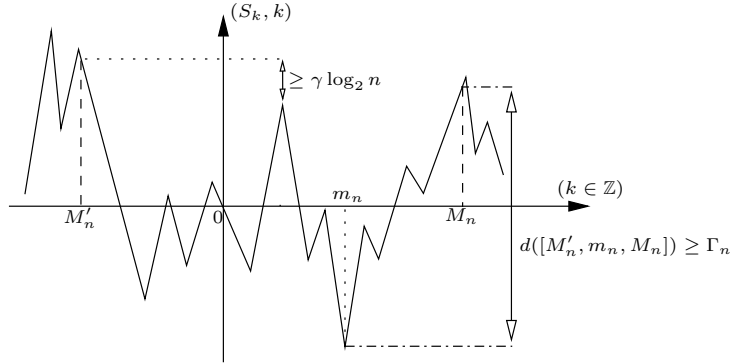


Figure 4: Basic valley, case  $m_n > 0$

**Lemma 1.6.** *There exists  $c > 0$  such that if 1.2, 1.3 and 1.4 hold, for all  $\gamma > 0$  and  $n$  we have*

$$Q [\{M_n', m_n, M_n\} \neq \emptyset] = 1 - \frac{c\gamma \log_2 n}{\log n}. \quad (1.21)$$

**Proof.**

One can find the proof of this Lemma in Section 5.2 of [Andreoletti(2006)]. ■

### 1.3 Main results

We start with some definitions that will be used all along this work. Let  $x \in \mathbb{Z}$ , define

$$T_x = \begin{cases} \inf \{k \in \mathbb{N}^*, X_k = x\} \\ +\infty, \text{ if such } k \text{ does not exist.} \end{cases} \quad (1.22)$$

Let  $n > 1$ ,  $k \in \mathbb{Z}$ , and  $c_0 > 0$ , define:

$$S_{k,m_n}^n = 1 - \frac{1}{\log n} (S_k - S_{m_n}), \quad (1.23)$$

$$\hat{S}_k^n = \frac{\log(\mathcal{L}(k, n))}{\log n}, \quad (1.24)$$

$$u_n = \frac{c_0 \log_3 n}{\log n}. \quad (1.25)$$

$S_{k,m_n}^n$  is the function of the potential we want to estimate,  $\hat{S}_k^n$  is the estimator and  $u_n$  is an error function.

Now let us define the following random sub-set of  $\mathbb{Z}$ :

$$\mathbb{L}_n^\gamma = \left\{ l \in \mathbb{Z}, \sum_{j=T_{k^*}}^n \mathbb{I}_{X_j=l} \geq (\log n)^\gamma \right\}, \quad (1.26)$$

recall that  $\gamma > 0$ . This set  $\mathbb{L}_n^\gamma$  is fundamental for our result, we notice that it depends only on the trajectory of the walk and more especially of its local time:  $\mathbb{L}_n^\gamma$  is the set of points for which we are able to give an estimator of the the random potential. We will see that this set is large and contains a great amount of the points visited by the walk (see Proposition 1.9). We recall that  $T_{k^*}$  is the first time the walk hit a favorite site. In words,  $l \in \mathbb{L}_n^\gamma$ , if and only if: The local time of the random walker in  $l$  after the instant  $T_{k^*}$  is large enough (larger than  $(\log n)^\gamma$ ). Our main result is the following:

**Theorem 1.7.** *Assume 1.2, 1.3 and 1.4 hold, there exists three constants  $c_0, c_1, c_2$  and  $c'_2$  such that for all  $\gamma > 6$ , there exists  $n_0$  such that for all  $n > n_0$  there exists  $G_n \subset \Omega_1$  with  $Q[G_n] \geq 1 - \phi_1(n)$  and*

$$\inf_{\alpha \in G_n} \mathbb{P}^\alpha \left[ \bigcap_{k \in \mathbb{L}_n^\gamma} \left\{ \left| \hat{S}_k^n - S_{k,m_n}^n \right| < u_n \right\} \right] \geq 1 - \phi_2(n). \quad (1.27)$$

where

$$\phi_1(n) = \frac{c_1 \gamma \log_2 n}{\log n}, \quad (1.28)$$

$$\phi_2(n) = \frac{c_2}{(\log n)^{\gamma/2}} + \frac{c'_2}{(\log n)^{\gamma-6}}. \quad (1.29)$$

The fact that our result depends on  $m_n$  seems to be restrictive, we would like to know where is the bottom of the valley only by considering the local time of the walk, we prove the following:

**Proposition 1.8.** *Assume 1.2, 1.3 and 1.4 hold, there exists a constant  $c_3 > 0$  such that for all  $\gamma > 6$ , there exists  $n_0$  such that for all  $n > n_0$  there exists  $G_n \subset \Omega_1$  with  $Q[G_n] \geq 1 - \phi_1(n)$  and*

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha \left[ \max_{x \in \mathbb{F}_n} |m_n - x| \leq (\log_2 n)^2 \right] \geq 1 - \phi_3(n), \quad (1.30)$$

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha \left[ |T_{m_n} - T_{k^*}| \leq (\log n)^3 \right] \geq 1 - \phi_3(n), \quad (1.31)$$

where  $\phi_3(n) = c_3/(\log n)^{\gamma-6}$ .

Notice that the distance between  $m_n$  (coordinate of the point visited by the walk where the minimum of the potential is reached) and a favorite site is negligible comparing to a typical fluctuation of the walk (of order  $(\log n)^2$ ). Thanks to Proposition 1.8 we can replace 1.27 in Theorem 1.7 by

$$\inf_{\alpha \in G_n} \mathbb{P}^\alpha \left[ \bigcap_{k \in \mathbb{L}_n^\gamma} \left\{ \left| \hat{S}_k^n - S_{k,k^*}^n \right| < u_n \right\} \right] \geq 1 - \phi_2(n). \quad (1.32)$$

Now let us give a result giving the main properties of  $\mathbb{L}_n^\gamma$ .



**Proposition 1.9.** *Assume 1.2 and 1.3 and 1.4 hold, there exists a constant  $c_3 > 0$  such that for all  $\gamma > 6$ , there exists  $n_0$  such that for all  $n > n_0$  there exists  $G_n \subset \Omega_1$  with  $Q[G_n] \geq 1 - \phi_1(n)$ ,*

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha [\mathcal{L}(\mathbb{L}_n^\gamma, n) = n(1 - o(1))] \geq 1 - \phi_2(n), \quad (1.33)$$

$$\inf_{\alpha \in G_n} \mathbb{P}_0^\alpha [|\mathbb{L}_n^\gamma| \approx (\log n)^2] \geq 1 - \phi_2(n), \quad (1.34)$$

$$(1.35)$$

where  $\lim_{n \rightarrow +\infty} o(1) = 0$ .

**Remark 1.10.** By definition we have  $\mathbb{F}_n \subseteq \mathbb{L}_n^\gamma$ .

Theorem 1.7 is known to be the quenched result that means for a fixed environment  $\alpha$ , a simple consequence (see Remark 2.4) is the following annealed result:

**Corollary 1.11.** *Assume 1.2, 1.3 and 1.4 hold, there exists three constants  $c_0, c_1$  and  $c_2$  such that for all  $\gamma > 6$ , there exists  $n_0$  such that for all  $n > n_0$*

$$\mathbb{P} \left[ \bigcap_{k \in \mathbb{L}_n^\gamma} \left\{ \left| \hat{S}_k^k - S_{k,k^*}^n \right| < u_n \right\} \right] \geq 1 - \phi(n), \quad (1.36)$$

where  $\phi(n) = \phi_1(n) + \phi_2(n)$ .

We would like to notice, that for our purpose the result above is not very interesting, because the aim is to reconstruct one environment whereas the result above give the mean of the probability for the walk over all the possible environments.

This paper is organized as follows. In Section 2 we give the proof of Theorems 1.7 (we easily get the corollary from Remark 2.4), we have split this proof into two parts, the first one deals with the random environment and the other one with the random walk itself. In section 3 we sketch the proofs of Propositions 1.8 and 1.9. In Section 4, as an application of our result, we present an algorithm and some numerical simulations. For completeness, we recall in the appendix, some basic facts on birth and death processes.

## 2 Proof of Theorem 1.7

The proof of a result with a random environment involves both arguments and properties for the random environment and arguments for the random walk itself. I will start to give the properties I need for the random environment. Then we will use it to get the result for the walk.

### 2.1 Properties needed for the random environment

#### 2.1.1 Construction of $(G_n, n \in \mathbb{N})$

Let  $k$  and  $l$  be in  $\mathbb{Z}$ , define

$$E_k^\alpha(l) = \mathbb{E}_k^\alpha [\mathcal{L}(l, T_k)] \quad (2.1)$$

in the same way, let  $A \subset \mathbb{Z}$ , define

$$E_k^\alpha(A) = \sum_{l \in A} \mathbb{E}_k^\alpha [\mathcal{L}(l, T_k)]. \quad (2.2)$$

**Definition 2.1.** Let  $d_0 > 0$ ,  $d_1 > 0$ , and  $\omega \in \Omega_1$ , we will say that  $\alpha \equiv \alpha(\omega)$  is a *good environment* if there exists  $n_0$  such that for all  $n \geq n_0$  the sequence  $(\alpha_i, i \in \mathbb{Z}) = (\alpha_i(\omega), i \in \mathbb{Z})$  satisfies the properties 2.3 to 2.5

$$\bullet \quad \{M_n', m_n, M_n\} \neq \emptyset, \quad (2.3)$$

$$\bullet \quad M_n' \geq -d_0(\sigma^{-1} \log_2 n \log n)^2, \quad M_n \leq d_0(\sigma^{-1} \log_2 n \log n)^2, \quad (2.4)$$

$$\bullet \quad E_{m_n}^\alpha(W_n) \leq d_1(\log_2 n)^2, \quad (2.5)$$

where  $W_n = \{M_n', M_n' + 1, \dots, m_n, \dots, M_n\}$ .

**Remark 2.2.** We will see in Section 2 that we use some results of [Andreoletti(2006)]. Considering this, we need extra properties on the random environment in addition to the three mentioned above, but as we don't need them for our computations we do not make them appear.

Define the *set of good environments*

$$G_n \equiv G_n(d_0, d_1) = \{\omega \in \Omega_1, \alpha(\omega) \text{ is a good environment}\}. \quad (2.6)$$

$G_n$  depends on  $d_0$ ,  $d_1$  and  $n$ , however we only make explicit the  $n$  dependence.

**Proposition 2.3.** *There exists two constants  $d_0 > 0$  and  $d_1 > 0$  such that if 1.2, 1.3 and 1.4 hold, there exists  $n_0$  such that for  $n > n_0$*

$$Q[G_n] \geq 1 - \phi_1(n), \quad (2.7)$$

where  $\phi_1(n)$  is given by 1.28.

**Proof.**

We can find the proof for the first three properties 2.3-2.5 in [Andreoletti(2006)], see Definition 4.1 and Proposition 4.2. ■

To end the section we would like to make the following elementary remark on the decomposition of  $\mathbb{P}$ :

**Remark 2.4.** Let  $\mathcal{C}_n \in \sigma(X_i, i \leq n)$  and  $G_n \subset \Omega_1$ , we have :

$$\mathbb{P}[\mathcal{C}_n] \equiv \int_{\Omega_1} Q(d\omega) \int_{\mathcal{C}_n} d\mathbb{P}^{\alpha(\omega)} \quad (2.8)$$

$$\geq \int_{G_n} Q(d\omega) \int_{\mathcal{C}_n} d\mathbb{P}^{\alpha(\omega)}. \quad (2.9)$$

So assume that  $Q[G_n] \equiv e_1(n) \geq 1 - \phi_1(n)$  and assume that for all  $\omega \in G_n$ ,  $\int_{\mathcal{C}_n} d\mathbb{P}^{\alpha(\omega)} \equiv e_2(\omega, n) \geq 1 - \phi_2(n)$  we get that

$$\mathbb{P}[\mathcal{C}_n] \geq e_1(n) \times \min_{w \in G_n} (e_2(w, n)) \geq 1 - \phi_1(n) - \phi_2(n). \quad (2.10)$$

## 2.2 Arguments for the walk

Let  $(\rho_1(n), n \in \mathbb{N})$  a strictly positive decreasing sequence such that  $\lim_{n \rightarrow \infty} \rho_1(n) = 0$ . First let us show that the Theorem 1.7 is a simple consequence of the following

**Proposition 2.5.** *Assume 1.2, 1.3 and 1.4 hold, there exists  $n_0$  such that for all  $n > n_0$  there exists  $G_n \subset \Omega_1$  with  $Q[G_n] \geq 1 - \phi_1(n)$  and*

$$\sup_{\alpha \in G_n} \left\{ \mathbb{P}_0^\alpha \left[ \bigcup_{k \in \mathbb{L}_n^\gamma} \left\{ \left| \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} \right| \leq w_{k, n} \right\} \right] \right\} \geq 1 - \phi_2(n) \quad (2.11)$$

where  $w_{k, n} = \rho_1(n) \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}$ ,  $\phi_1(n)$  and  $\phi_2(n)$  are given just after 1.27

Taking the logarithm and for  $n$  large enough, using Taylor series expansion, we remark that

$$\frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} (1 - \rho_1(n)) \leq \frac{\mathcal{L}(k, n)}{n} \leq \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} (1 + \rho_1(n)) \quad (2.12)$$

implies

$$-2\rho_1(n) - \log(E_{m_n}^\alpha(W_n)) \leq \log \mathcal{L}(k, n) - \log n - \log(E_{m_n}^\alpha(k)) \leq -\log(E_{m_n}^\alpha(W_n)) + \rho_1(n),$$

rearranging the terms and using A.1 (see the Appendix) we get

$$\frac{1}{\log n} (R_n^\alpha(k) - 2\rho_1(n)) \leq \hat{S}_k^n - S_{k, m_n}^n \leq \frac{1}{\log n} (R_n^\alpha(k) - \rho_1(n)) \quad (2.13)$$

where  $R_n^\alpha(k) = \log \left( \frac{\alpha m_n}{\beta_k} a_{k, m_n} \right) - \log(E_{m_n}^\alpha(W_n))$  and  $a_{k, m_n}$  is given by A.2. Now using A.4 and Property 2.5 we get the Theorem. The proof of Proposition 2.5 is based on the following results (Lemma 2.6) of [Andreoletti(2005)],

### 2.2.1 Known facts

Let  $(\rho(n), n \in \mathbb{N})$  be a positive decreasing sequence such that  $\lim_{n \rightarrow \infty} \rho(n) = 0$ , we define

$$\mathcal{A}_1 = \left\{ \left| \frac{\mathcal{L}(m_n, n)}{n} - \frac{1}{E_{m_n}^\alpha(W_n)} \right| > \frac{\rho(n)}{E_{m_n}^\alpha(W_n)} \right\}, \quad (2.14)$$

$$\mathcal{A}_2 = \{T_{m_n} \leq n/(\log n)^4, \mathcal{L}(W_n, n) = 1\}. \quad (2.15)$$

**Lemma 2.6.** *Assume 1.2, 1.3 and 1.4 hold, there exists a constant  $b_1 > 0$  such that for all  $\gamma > 6$ , there exists  $n_0$  such that for all  $n > n_0$  there exists  $G_n \subset \Omega_1$  with  $Q[G_n] \geq 1 - \phi_1(n)$  and*

$$\sup_{\alpha \in G'_n} \{\mathbb{P}_0^\alpha [\mathcal{A}_1]\} \leq r_1(n), \quad (2.16)$$

where  $r_1(n) = b_1/(\log n)^{\gamma-6}$ .

**Proof.**

We do not give the details of the computations because the reader can find it in the referenced paper (Theorem 3.8 of [Andreoletti(2006)]), just notice that comparing to the Theorem 3.8 we have a better rate of convergence for the probability obtained just by using a weaker result for the concentration of the walk. ■

We will also need the following elementary fact :

**Lemma 2.7.** *Assume 1.2, 1.3 and 1.4 hold, there exists a constant  $b_2 > 0$  such that for all  $\gamma > 2$ , there exists  $n_0$  such that for all  $n > n_0$  there exists  $G_n \subset \Omega_1$  with  $Q[G_n] \geq 1 - \phi_1(n)$  and*

$$\sup_{\alpha \in G'_n} \{\mathbb{P}_0^\alpha [\mathcal{A}_2]\} \leq r_2(n), \quad (2.17)$$

where  $r_2(n) = b_2/(\log n)^{\gamma-2}$ .

**Proof.**

Once again this can be find in [Andreoletti(2006)]: Proposition 4.7 and Lemma 4.8. ■

Using these results we can give the proof of Proposition 2.5 into two steps :

**2.2.2 Step 1**

Let us define the following subsets :

$$\bar{v}_1^n \equiv \{M'_n \leq k \leq m_n - 1, (\max_{k \leq j \leq m_n} S_j - S_{m_n}) < \log n - \frac{\gamma}{2} \log_2 n\}, \quad (2.18)$$

$$\bar{v}_2^n \equiv \{m_n + 1 \leq k \leq M_n, (\max_{m_n \leq j \leq k} S_j - S_{m_n}) < \log n - \frac{\gamma}{2} \log_2 n\}, \quad (2.19)$$

and

$$V_n^\gamma = \bar{v}_1^n \cap \bar{v}_2^n. \quad (2.20)$$

In words  $V_n^\gamma$  is a subset of points included in  $W_n$ , such that for all  $k \in V_n^\gamma$  the largest difference of potential between  $m_n$  and  $k$  is smaller than  $\log n - \gamma/2 \log_2 n$ . For the walk we will see (Lemma below) that if  $k \in V_n^\gamma$  then the walk will hit  $k$  after it has reached  $m_n$  and it will hit this point  $k$  a number of time large enough (see figure 5).

First let us prove the following Lemma :

**Lemma 2.8.** *Assume 1.2, 1.3 and 1.4 hold, there exists a constant  $b_3 > 0$  such that for all  $\gamma > 6$ , there exists  $n_0$  such that for all  $n > n_0$  there exists  $G_n \subset \Omega_1$  with  $Q[G_n] \geq 1 - \phi_1(n)$*

$$\sup_{\alpha \in G'_n} \{\mathbb{P}_0^\alpha [\mathbb{L}_n^\gamma \subseteq V_n^\gamma]\} \geq 1 - r_3(n), \quad (2.21)$$

where  $r_3(n) = b_3/(\log n)^{\gamma/2}$ .

Notice that  $\mathbb{L}_n^\gamma$  is a  $\mathbb{P}$  random variable (with two levels of randomness) whereas  $V_n^\gamma$  is only a  $Q$  random variable (with one level of randomness), this Lemma makes the link between a trajectory of the walk and the random environment.

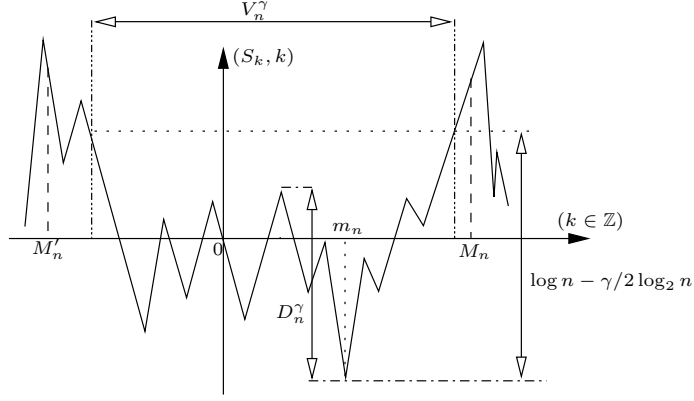


Figure 5:  $V_n^\gamma$  with  $m_n > 0$ , case 1:  $D_n^\gamma \equiv \max_{k \in \bar{v}_1} \max_{k \leq j \leq m_n} (S_j - S_{m_n}) \leq \log n - \gamma \log_2 n$

**Proof.**

To prove this Lemma we use Proposition 1.8. First notice that

$$\mathbb{P}_0^\alpha [\mathbb{L}_n^\gamma \subseteq V_n^\gamma] = 1 - \mathbb{P}_0^\alpha \left[ \bigcup_{k \in (v_n^1 \cup v_n^2)} \{k \in \mathbb{L}_n^\gamma\} \right] \quad (2.22)$$

where

$$v_1^n \equiv \{M'_n \leq k \leq m_n - 1, (\max_{k \leq j \leq m_n} S_j - S_{m_n}) \geq \log n - \frac{\gamma}{2} \log_2 n\}, \quad (2.23)$$

$$v_2^n \equiv \{m_n + 1 \leq k \leq M_n, (\max_{m_n \leq j \leq k} S_j - S_{m_n}) \geq \log n - \frac{\gamma}{2} \log_2 n\}. \quad (2.24)$$

Let  $k \in v_1^n$  let us give an upper bound for

$$\begin{aligned} \mathbb{P}_0^\alpha [k \in \mathbb{L}_n^\gamma, |T_{k^*} - T_{m_n}| \leq (\log n)^3] &\leq \mathbb{P}_0^\alpha \left[ \sum_{j=T_{k^*}}^n \mathbb{I}_{X_j=k} \geq (\log n)^\gamma, |T_{k^*} - T_{m_n}| \leq (\log n)^3 \right] \\ &\leq \mathbb{P}_0^\alpha \left[ \sum_{j=T_{m_n}}^n \mathbb{I}_{X_j=k} \geq (\log n)^\gamma - (\log n)^3 \right] \\ &\leq \mathbb{P}_{m_n}^\alpha \left[ \sum_{j=1}^{T_{m_n},n} \mathbb{I}_{X_j=k} \geq (\log n)^\gamma - (\log n)^3 \right], \end{aligned} \quad (2.25)$$

for the third inequality we have used the strong Markov property, where

$$T_{m_n,j} \equiv \begin{cases} \inf\{k > T_{m_n,j-1}, X_k = m_n\}, & j \geq 2 \\ +\infty, & \text{if such } k \text{ does not exist.} \end{cases}$$

$$T_{m_n,1} \equiv T_{m_n} \text{ (see 1.22).}$$

Now using the Markov inequality and Lemma A.1 we get

$$\mathbb{P}_0^\alpha [k \in \mathbb{L}_n^\gamma, |T_{k^*} - T_{m_n}| \leq (\log n)^3] \leq \frac{n \mathbb{E}_{m_n}^\alpha [\mathcal{L}(k, T_{m_n})]}{(\log n)^\gamma - (\log n)^3} \quad (2.26)$$

$$\leq \frac{n}{\eta_0 \exp(S_k - S_{m_n}) ((\log n)^\gamma - (\log n)^3)} \quad (2.27)$$

$$\leq \frac{1}{\eta_0 (\log n)^{\gamma/2} (1 - (\log n)^3 / (\log n)^\gamma)}, \quad (2.28)$$

notice that in the last inequality we have used the fact that  $k \in v_1^n$ . A similar computation give the same inequality when  $k \in v_2^n$ . Collecting what we did above, and using the Property 2.4 together with 1.31 yields the Lemma. ■

### 2.2.3 Step 2

This second step is devoted to the proof of the following Lemma.

**Lemma 2.9.** *For all  $\alpha$  and  $n$  we have*

$$\mathbb{P}_0^\alpha \left[ \left| \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} \right| > w_{k,n}, \mathcal{A}_1, \mathcal{A}_2 \right] \leq 2 \exp(-n/2\psi_2^\alpha(n)) \quad (2.29)$$

recall that  $w_{k,n} = \rho_1(n) \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)}$  and  $\psi_2^\alpha(n) = 2 \frac{(\rho_1(n) - \rho(n))^2 (\alpha_{m_n} \wedge \beta_{m_n}) \exp(-(S_{M_k} - S_{m_n}))}{1 + \rho(n) |k - m_n| E_{m_n}^\alpha(W_n)}$ .  $M_k$  is such that  $S_{M_k} = \max_{m_n+1 \leq j \leq k} S_j$  if  $k > m_n$  and conversly if  $k < m_n$   $S_{M_k} = \max_{k \leq j \leq m_n-1} S_j$ .

**Proof.**

We essentially use an inequality of concentration (see [Ledoux(2001)]), for simplicity we only give the proof for  $k > m_n$ , the other case ( $k \leq m_n$ ) is very similar. Using the Markov property and the fact that  $\mathcal{L}(k, T_{m_n}) = 0$ , we get

$$\mathbb{P}_0^\alpha \left[ \left| \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} \right| > w_{k,n}, \mathcal{A}_1, \mathcal{A}_2 \right] \leq \mathbb{P}_{m_n}^\alpha \left[ \left| \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} \right| > w_{k,n}, \mathcal{A}_1 \right] \quad (2.30)$$

We have

$$\begin{aligned} & \mathbb{P}_{m_n}^\alpha \left[ \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} > w_{k,n}, \mathcal{A}_1 \right] \\ & \leq \mathbb{P}_{m_n}^\alpha \left[ \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} > w_{k,n}, \frac{\mathcal{L}(m_n, n)}{n} - \frac{1}{E_{m_n}^\alpha(W_n)} \leq \frac{\rho(n)}{E_{m_n}^\alpha(W_n)} \right] \end{aligned} \quad (2.31)$$

$$\leq \mathbb{P}_{m_n}^\alpha \left[ \frac{\mathcal{L}(k, T_{m_n, n_1})}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} > w_{k,n} \right] \quad (2.32)$$

$$\equiv \mathbb{P}_{m_n}^\alpha \left[ \frac{\mathcal{L}(k, T_{m_n, n_1})}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} (1 + \rho(n)) > w'_{k,n} \right] \quad (2.33)$$

where  $n_1 = \frac{n}{E_{m_n}^\alpha(W_n)} (1 + \rho(n))$ , notice that  $n_1$  is not necessarily an integer but for simplicity we disregard that, and  $w'_{k,n} = \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} (\rho_1(n) - \rho(n))$ . The strong Markov property implies that  $\mathcal{L}(k, T_{m_n, n_1})$  is a sum of  $n_1$  i.i.d. random variables, the inequality of concentration gives

$$\mathbb{P}_{m_n}^\alpha \left[ \frac{\mathcal{L}(k, T_{m_n, n_1})}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} > w'_{k,n}, \mathcal{A}_1 \right] \leq \exp \left[ -\frac{n}{2} \frac{E_{m_n}^\alpha(W_n)}{\text{Var}_{m_n}(\mathcal{L}(k, T_{m_n}))} \frac{(w'_{k,n})^2}{1 + \rho(n)} \right]. \quad (2.34)$$

With the same method we also get

$$\mathbb{P}_{m_n}^\alpha \left[ \frac{\mathcal{L}(k, T_{m_n, n_1})}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} < -w'_{k, n}, \mathcal{A}_1 \right] \leq \exp \left[ -\frac{n}{2} \frac{E_{m_n}^\alpha(W_n)}{\text{Var}_{m_n}(\mathcal{L}(k, T_{m_n}))} \frac{(w'_{k, n})^2}{1 + \rho(n)} \right] \quad (2.35)$$

Using A.3 we get Lemma 2.9. ■

### 2.2.4 End of the proof of the Theorem

Using Lemmata 2.6, 2.7 and 2.8 we have:

$$\begin{aligned} & \mathbb{P}_0^\alpha \left[ \bigcup_{k \in \mathbb{L}_n^a} \left\{ \left| \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} \right| > w_{k, n} \right\} \right] \\ & \leq |V_n^\gamma| \sup_{k \in V_n^\gamma} \mathbb{P}_0^\alpha \left[ \left| \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} \right| > w_{k, n}, \mathcal{A}_1, \mathcal{A}_2 \right] + 3 \max_{1 \leq i \leq 3} \{r_i(n)\}, \end{aligned}$$

then using Lemma 2.9 we get

$$\begin{aligned} \sup_{k \in V_n^\gamma} \mathbb{P}_0^\alpha \left[ \left| \frac{\mathcal{L}(k, n)}{n} - \frac{E_{m_n}^\alpha(k)}{E_{m_n}^\alpha(W_n)} \right| > w_{k, n}, \mathcal{A}_1, \mathcal{A}_2 \right] & \leq 2 \sup_{k \in V_n^\gamma} \exp(-n/2\psi_2^\alpha(k, n)) \\ & \leq 2 \exp(-(\log n)^{\gamma/2-2}/(\rho_1(n) \log_2 n)), \end{aligned} \quad (2.36)$$

where the last inequality comes from the definition of  $V_n^\gamma$  (see 2.20) and the Properties 2.4 and 2.5. To end we use again the Property 2.4 together with the definition of  $V_n^\gamma$ .

## 3 Proof of Proposition 1.8 and 1.9

*Sketch of the proof of Proposition 1.8* 1.30 is an improvement of the proof of Corollary 3.17 of [Andreoletti(2006)] in order to get a better rate of convergence for the probability. To get 1.31, we have used the same idea of the proof of Corollary 3.17 of [Andreoletti(2006)], so once again we will not repeat the computations here. We recall just the intuitive idea: once the walk has reached  $k^*$ , we know from 1.30 that  $m_n$  is at most at a distance  $(\log_2 n)^2$ , therefore the walk need at most the amount of time  $\exp(\sqrt{((\log_2 n)^2)}) = (\log n)$  to reach  $m_n$ . We take  $(\log n)^3$  to get a better rate of convergence for the probability.

*Sketch of the proof of Proposition 1.9* The first two properties can be deduced from the following inequality, let  $\epsilon > 1$ , for all  $n$  large enough and all  $\alpha \in G_n$ :

$$\mathbb{P}_0^\alpha \left[ V_n^{2(\gamma+\epsilon)} \subseteq \mathbb{L}_n^\gamma \right] \geq \phi_3(n) + r_1(n) + cte(\log t)^2 \exp(-(\log n)^{\gamma+\epsilon-2}(1 - (\log n)^{1-\epsilon})). \quad (3.1)$$

Indeed, thanks to 3.1 we have

$$\mathbb{P}^\alpha [\mathcal{L}(\mathbb{L}_n^\gamma, n) \geq n(1 - o(1))] \geq \mathbb{P}^\alpha [\mathcal{L}(V_n^{2(\gamma+\epsilon)}, n) \geq n(1 - o(1))] \quad (3.2)$$

we get 1.33 by using the same method [Andreoletti(2006)] uses to get Theorem 3.1. To get 1.34, we only need to show that  $|V_n^{\gamma+\epsilon}| \approx (\log t)^2$ , which is a basic fact for a simple random walk. Now, to get

3.1, first we notice that, by using a similar method of the proof of Theorem 1.7 we can get

$$\mathbb{P}_0^\alpha \left[ V_n^{2(\gamma+\epsilon)} \notin \mathbb{L}_n^\gamma \right] \leq |V_n^{\gamma+\epsilon}| \max_{k \in V_n^{2(\gamma+\epsilon)}} \mathbb{P}_{m_n}^\alpha \left[ \sum_{j=1}^{n_1} \eta_j^k < (\log n)^\gamma \right] + \phi_3(n) + r_1(n). \quad (3.3)$$

where  $(\eta_j^k, j)$  is a i.i.d. sequence with the law of  $\mathcal{L}(k, T_{m_n})$ . Then using an inequality of concentration, we get 3.1.

## 4 Algorithm and Numerical simulations

### 4.1 General and recall of the main definitions

First notice that we have no criteria to determine whether or not we can apply this method to an unknown series of data. All we know is that it works for Sinai's walk, however we can apply the following algorithm to every process. Let us recall the basic random variables that will be used for our simulations, let  $x \in \mathbb{Z}$ ,  $n \in \mathbb{N}$ ,

$$T_x = \begin{cases} \inf\{k \in \mathbb{N}^*, X_k = x\} \\ +\infty, \text{ if such } k \text{ does not exist.} \end{cases}, \quad (4.1)$$

$$\mathcal{L}(x, n) \equiv \sum_{i=1}^n \mathbb{I}_{\{X_i=x\}}, \quad (4.2)$$

$$\mathcal{L}^*(n) = \max_{k \in \mathbb{Z}} (\mathcal{L}(k, n)), \quad \mathbb{F}_n = \{k \in \mathbb{Z}, \mathcal{L}(k, n) = \mathcal{L}^*(n)\}, \quad (4.3)$$

$$k^* = \inf\{|k|, k \in \mathbb{F}_n\}. \quad (4.4)$$

$$(4.5)$$

We recall also the set  $\mathbb{L}_n^\gamma$ , the function of the potential we want to estimate and its estimator:

$$\mathbb{L}_n^\gamma = \left\{ k \in \mathbb{Z}, \sum_{j=T_k^*}^n \mathbb{I}_{X_j=k} \geq (\log n)^\gamma \right\}, \quad (4.6)$$

$$S_{k, m_n}^n = 1 - \frac{1}{\log n} (S_k - S_{m_n}), \quad (4.7)$$

$$\hat{S}_k^n = \frac{\log(\mathcal{L}(k, n))}{\log n}, \quad (4.8)$$

We also recall that thanks to Proposition 1.8, in probability we have  $|m_n - k^*| \leq \text{cte}(\log_2 n)^2$ .

### 4.2 Main steps of the algorithm

Step 1: We have to determine  $\mathbb{L}_n^\gamma$  and to get it we have to compute  $T_{k^*}$  and therefore the local time of the process. First we compute  $\mathcal{L}(k, n)$  for every  $k$ , notice that  $\mathcal{L}(k, n)$  is not equal to zero only if  $k$  has been visited by the walk within the interval of time  $[1, n]$ . Then we can compute  $\mathcal{L}^*(n)$  and determine  $k^*$  and  $T_{k^*}$ . Notice that  $T_{k^*}$  is not a stopping time, therefore we need two passes to compute what we



need. We are now able to determine  $\mathbb{L}_n^\gamma$  computing  $\sum_{j=T_{k^*}}^n \mathbb{I}_{X_j=k}$ .

Step 2: We can check that  $\mathbb{L}_n^\gamma$  is connex, contains  $k^*$  and that its size is of the order of a typical fluctuation of the walk. Now, keeping only the  $k$  that belongs to  $\mathbb{L}_n^\gamma$  we compute for those  $k$ :  $\hat{S}_k^n = \frac{\log(\mathcal{L}(k,n))}{\log n}$  the estimator of the potential. We localize the bottom of the valley  $m_n$  using  $k^*$ .

### 4.3 Simulations

For the first simulation (Figure 6) we show a case where  $\mathbb{L}_n^\gamma$  is large i.e.  $\mathbb{L}_n^\gamma$  contains most of the points visited by the walk. The trajectory of the random potential is in red the interval of confidence in

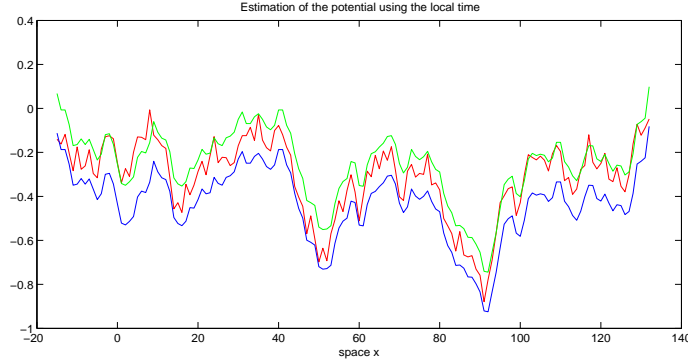


Figure 6: in red  $S_{x,m_n}^n$ , in blue  $\hat{S}_x^n - u_n$ , in green  $\hat{S}_x^n + u_n$

blue and green. We took  $n = 500000$  and  $\gamma = 7$ , notice that the larger is  $\gamma$ , the smaller is  $\mathbb{L}_n^\gamma$  but better is the rate of convergence of the probability. We get that  $\mathbb{L}_n^\gamma = [10, 94]$ . In Figure 7 we plot the difference  $S_{x,m_n}^n - \hat{S}_x^n$  and its the linear regression. We notice that the slope of the linear regression is

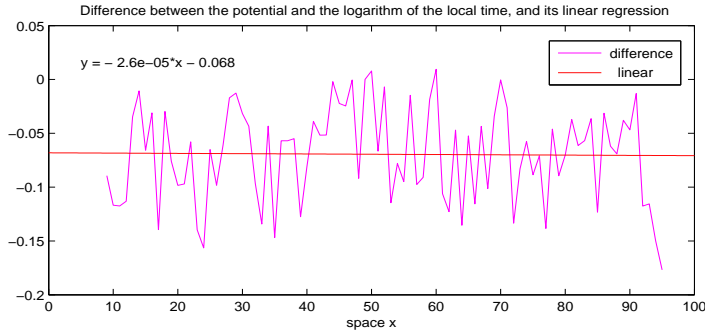


Figure 7: in magenta  $S_{x,m_n}^n - \hat{S}_x^n$ , in red the linear regression

of order  $10^{-5}$ . We also notice that we have taken  $n = 500000$ , so the error function  $u_n \approx \frac{\log_3 n}{\log n} \approx 0,7$  this match with the  $\max_x(S_{x,m_n}^n - \hat{S}_x^n) \approx 0.8$  for this simulation.

Now let us choose another example where  $\mathbb{L}_n^\gamma$  is much more smaller. For the following simulation (Figure 8) we have only changed the sequence of random number. We get that  $\mathbb{L}_n^\gamma = [-150, -85]$ . We

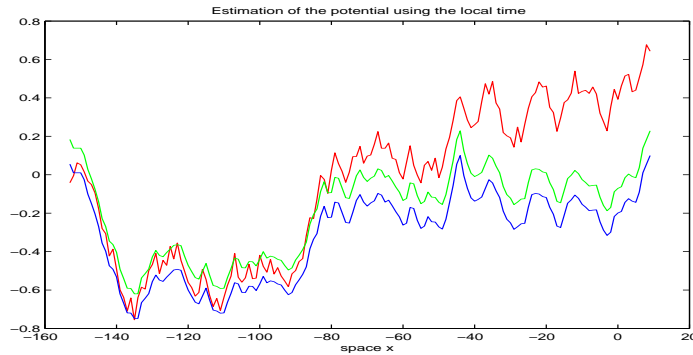


Figure 8: in red  $S_{x,m_n}^n$ , in blue  $\hat{S}_x^n - u_n$ , in green  $\hat{S}_x^n + u_n$

notice that for the coordinates larger than -85 and especially after -40, our estimator is not good at all. In fact once the walk has reached the minimum of the valley (coordinate -111) it will never reach again one of the points of coordinate larger than -40 before  $n = 500000$ , so our estimator can not say anything about the difference  $S_{x,m_n}^n - \hat{S}_x^n$ . However if we look in the past of the walk and especially at a the time  $T_{k^*}$  which is the first time it has reached the coordinate -111, the favorite point for this time is localized around the point -2, so a good estimator between the coordinate -40 and 10 may be given by  $(\frac{\log(\mathcal{L}(k,T_{k^*}))}{\log T_{k^*}}, k)$ . The difference  $S_{x,m_n}^n - \hat{S}_x^n$  and the linear regression in the interval  $\mathbb{L}_n^\gamma = [-150, -85]$  is presented Figure 9.

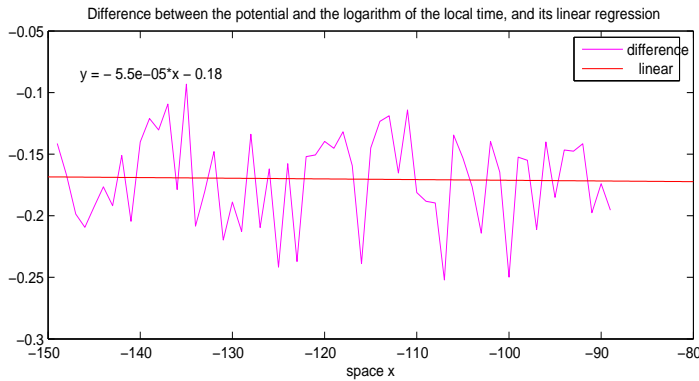


Figure 9: in magenta  $S_{x,m_n}^n - \hat{S}_x^n$ , in red the linear regression

## A Basic results for birth and death processes

For completeness we recall an explicit expression for the mean and an upper bound for the variance of the local times at a certain stopping time, we can be found a proof of these elementary facts in [Révész(1989)] (page 279)

**Lemma A.1.** *For all  $\alpha$ , Let  $k > m_n$*

$$\mathbb{E}_{m_n}^\alpha [\mathcal{L}(k, T_{m_n})] = \frac{\alpha_{m_n}}{\beta_k} \frac{1}{e^{S_k - S_{m_n}}} a_{k, m_n}, \text{ where} \quad (\text{A.1})$$

$$a_{k, m_n} = \frac{\sum_{i=m_n+1}^{k-1} e^{S_i} + e^{S_k}}{\sum_{i=m_n+1}^{k-1} e^{S_i} + e^{S_{m_n}}}. \quad (\text{A.2})$$

$$\text{Var}_{m_n} [\mathcal{L}(k, T_{m_n})] \leq 2(\mathbb{E}_{m_n}^\alpha [\mathcal{L}(k, T_{m_n})])^2 \frac{e^{S_{M_k} - S_{m_n}}}{\beta_k} |k - m_n|. \quad (\text{A.3})$$

$M_k$  is such that  $S_{M_k} = \max_{m_n+1 \leq j \leq k-1} S_j$ . For  $Q$ -a.a. environment  $\alpha$

$$\frac{\eta_0}{1 - \eta_0} \leq \frac{\alpha_{m_n}}{\beta_k} a_{k, m_n} \leq \frac{1}{\eta_0}. \quad (\text{A.4})$$

A similar result is true for  $k < m_n$  and  $\mathbb{E}_{m_n}^\alpha [\mathcal{L}(m_n, T_{m_n})] = 1$ .

**Acknowledgements** I would like to thank Nicolas Klutchnikoff for stimulating discussions.

## REFERENCES

- [Solomon(1975)] F. Solomon. Random walks in random environment. *Ann. Probab.*, **3**(1): 1–31, 1975.
- [Sinai(1982)] Ya. G. Sinai. The limit behaviour of a one-dimensional random walk in a random medium. *Theory Probab. Appl.*, **27**(2): 256–268, 1982.
- [Adelman and Enriquez(2004)] O. Adelman and N. Enriquez. Random walks in random environment: What a single trajectory tells. *Israel J. Math.*, 142:205–220, 2004.
- [V. Baldazzi and Monasson(2006)] E. Marinari V. Baldazzi, S. Cocco and R. Monasson. Inferring dna sequences from mechanical unzipping: an ideal-case study. *Physical Review Letters E*, **96**: 128102–1–4, 2006.
- [Andreoletti(2005)] P. Andreoletti. Alternative proof for the localisation of Sinai’s walk. *Journal of Statistical Physics*, 118:883–933, 2005.
- [Kesten(1986)] H. Kesten. The limit distribution of Sinai’s random walk in random environment. *Physica*, **138A**: 299–309, 1986.
- [Andreoletti(2006)] P. Andreoletti. On the concentration of Sinai’s walk. *Stoch. Proc. Appl.*, 116: 1377–1408, 2006.

[Ledoux(2001)] M. Ledoux. *The concentration of measure phenomenon*. Amer. Math. Soc., Providence, RI, 2001.

[Révész(1989)] P. Révész. *Random walk in random and non-random environments*. World Scientific, 1989.

Laboratoire MAPMO - C.N.R.S. UMR 6628 - Fédération Denis-Poisson  
Université d'Orléans, UFR Sciences  
Bâtiment de mathématiques - Route de Chartres  
B.P. 6759 - 45067 Orléans cedex 2  
FRANCE