Information recovery from observations by a random walk having jump distribution with exponential tails

ANDREW HART[*] FABIO MACHADO [‡]

HEINRICH MATZINGER[¶]

October 11, 2013

Abstract

A *scenery* is a coloring ξ of the integers. Let $\{S_t\}_{t\geq 0}$ be a recurrent random walk on the integers. Observing the scenery ξ along the path of this random walk, one sees the color $\chi_t := \xi(S_t)$ at time t. The *scenery reconstruction problem* is concerned with recovering the scenery ξ, given only the sequence of observations $\chi := (\chi_t)_{t\geq 0}$. The scenery reconstruction methods presented to date require the random walk to have bounded increments. Here, we present a new approach for random walks with unbounded increments which works when the tail of the increment distribution decays exponentially fast enough and the scenery has five colors.

Keywords: Scenery reconstruction, scenery distinguishing, large deviation, random walk.

2000 MSC: 60K37, 60G50.

[*]Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, UMI 2071 CNRS-UCHILE, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Casilla 170, Correo 3, Santiago, Chile

[†]E-mail: ahart@dim.uchile.cl

[‡]Instituto de Matematica e Estatistica, Universidade de So Paulo, Rua do Mato 1010, So Paulo, SP - Brazil

[§]School of Mathematics, Georgia Institute of Technology, 656 Cherrystreet, Atlanta, 30363 GA, USA

[¶]E-mail: matzi@math.gatech.edu

[‖]Corresponding author.

1 Introduction

Consider a coloring of the integers $\xi : \mathbb{Z} \to \{1, 2, 3, 4, 5\}$ which we shall call a scenery. Let S be a recurrent random walk starting at the origin. We assume that we observe the scenery along the path of the random walk S, that is, we observe the color $\chi_t := \xi(S_t)$ at time t. The scenery reconstruction problem is concerned with determining the scenery ξ using a single realization of the color record

$$\chi := (\chi_0, \chi_1, \chi_2, \ldots).$$

The scenery reconstruction problem can also be formulated as follows: Does one path realization of the process $\{\chi_t\}_{t\geq 0}$ uniquely determine ξ? The answer in such general terms is "no". However, under appropriate restrictions, the answer becomes "yes". Firstly, we can at best hope to be able to reconstruct the scenery up to translation and reflection. Secondly, there are sceneries which can not be reconstructed. Lindenstrauss in [15] exhibited sceneries which cannot be reconstructed. Only 'typical' sceneries can be reconstructed. Sceneries that can be obtained from one another by shift and reflection are called equivalent and we shall use the symbol $\approx$ to denote their equivalence.

The scenery reconstruction problem arose from questions posed by Kesten, Keane, Benjamini, Den Hollander and others. It also falls into the research area concerned with the investigation of the ergodic properties of the color record χ. One of the motivations for studying scenery reconstruction comes from ergodic theory, for example via the T, T^{-1}problem; see Kalikow [8]. The ergodic properties of the observations χ were studied by Heicklen, Hoffman, Rudolph in [4], Kesten and Spitzer in [12], Keane and den Hollander in [9], den Hollander in [2] and den Hollander and Steif in [1].

A related important problem is the distinguishing of sceneries: Benjamini, den Hollander, and Keane independently asked whether all non-equivalent sceneries could be distinguished. We give a brief outline of this problem. Let η_1 and η_2 be two given sceneries. Assume that either η_1 or η_2 is observed along a random walk path, but we do not know which one. Can we tell which of the two sceneries was observed? Kesten and Benjamini proved that one can distinguish almost every pair of sceneries, even in two dimensions and with only two colors. Before that, Howard had proved in [5], [6], and [7] that any two periodic one dimensional non- equivalent sceneries are distinguishable, and that one can almost surely distinguish single defects in periodic sceneries. The problem of distinguishing two sceneries which differ only in one point is called "detecting a single defect in a scenery". Kesten in [10] proved that one can a.s. recognize a single defect in a random scenery with at least five colors. He asked whether one can distinguish a single defect even if there are only two colors in the scenery.

Kesten's question was answered by Matzinger in his Ph.D. thesis [18] Given that the colors in the scenery are taken to be i.i.d. uniformly distributed, he showed that almost every 2-color scenery can be almost surely reconstructed up to equivalence. In [19], Matzinger proved that almost every 3-color scenery can be almost surely reconstructed.

Kesten [11] noticed that the proofs employed in [18] and [19] rely heavily on the skip-free property of the random walk as well as the one-dimensionality of the scenery. He asked whether the result might still hold in more general situations.

In [16], Matzinger and Löwe showed that one can still reconstruct sceneries in two dimensions, provided there are sufficiently many colors. In [23], Rolles and Matzinger adapted the method proposed by Löwe, Merkl and me to the case where random errors occur in the observed color record. They showed that the scenery can be reconstructed provided the probability of the errors is small enough. When the observations are seen with random errors, the reconstruction of sceneries is closely related to some coin tossing problems. These have been investigated by Harris and Keane [3] and Levin, Pemantle and Peres [14].

This paper deals with the problem of whether one can reconstruct a scenery seen along the path of a random walk having unbounded jumps, a question which was asked by den Hollander. Our main result shows that we can a.s. reconstruct a five-color (random) scenery seen along the path of a random walk S with unbounded jumps, provided the probability of making a jump of non-unit size is not too high and the tail of the increment distribution of the random walk decays exponentially fast enough. By unbounded jumps, we mean that the support of the random walk's increment distribution is not bounded.

Methods for carrying out Scenery reconstruction differ greatly depending on the distribution of the scenery and the nature of the random walk. The methods for scenery reconstruction for a simple random walk ([18], [19],[20], [16], [21], [22], [23]), together with the methods appropriate when the support of the increment distribution is bounded ([17] and [13]), fail in the setting of "unbounded jumps". It was not possible to adapt existing methods to the present situation. The approach utilized here is fundamentally different from those which have been fruitfully applied in the setting of random walks having "bounded jumps".

We begin by explaining the setting considered in this article. Let $c > 0$ be an exponential decay rate and $\epsilon > 0$ the probability of jumping a distance that is *not* of unit length. We shall impose the following conditions on the random walk S:

$$P(|S_{t+1} - S_t| \neq 1) = \epsilon \tag{1}$$

and

$$P(|S_{t+1} - S_t| = i \ \mid \ |S_{t+1} - S_t| \neq 1) \leq e^{-ci}. \tag{2}$$

We shall also require that the distribution of the random walk be symmetric:

$$P(S_{t+1} - S_t = i) = P(S_{t+1} - S_t = -i). \tag{3}$$

This condition could be replace by the weaker condition $\mathbb{E}[S_{t+1} - S_t] = 0$, but symmetry simplifies notation. Finally, to ensure that the random walk is non-periodic, we shall stipulate that

$$P(S_{t+1} - S_t = 0) > 0. \tag{4}$$

We can now formulate the main theorem of the paper.

Theorem 1.1 *Assume the scenery $\xi : \mathbb{Z} \to \{1,2,3,4,5\}$ is i.i.d. with the five colors uniformly distributed. If Conditions (1)–(4) are all satisfied, then for every $\epsilon > 0$ small enough and $c > 0$ large enough, we can a.s. reconstruct the scenery ξ. In other words, there exists a measurable map*

$$\mathcal{A} : \{1,2,3,4,5\}^{\mathbb{N}} \to \{1,2,3,4,5\}^{\mathbb{Z}}$$

such that

$$P(\mathcal{A} \circ \chi \approx \xi) = 1.$$

The principal difficulty with proving the theorem is in showing that we can reconstruct the scenery at one point given that we have already managed to reconstruct the scenery on an interval. The next three sections of the paper are dedicated to obtaining an estimate $\hat{\xi}_{n+1}$ of $\xi(n+1)$ given the observations χ and the restriction of the scenery $\xi|_{[-n,n]} := \xi_{-n}\xi_{-n+1}\cdots\xi_{n-1}\xi_n$. Section 2 uses three simplified problems to introduce key concepts that shall be required in the sequel. In Section 3, we describe the algorithm for retrieving ξ_{n+1} from χ given that we know a portion $\xi|_{[-n,n]}$ of the scenery. The following section then proves that the single-point reconstruction algorithm succeeds, that is, $\hat{\xi}_{n+1} = \xi_{n+1}$, with high probability. This is done by showing that the probability of failure is finitely summable, that is,

$$\sum_n P\big(\xi(n+1) \neq \hat{\xi}(n+1)\big) < \infty. \tag{5}$$

Finally, we conclude by showing in Section 5 that finite summability implies that the whole scenery ξ can be reconstructed almost surely up to equivalence.

2 Three simplified problems

To illustrate the key concepts needed for performing scenery reconstruction on 5-color sceneries observed by random walks with unbounded jumps, we begin by presenting three simplified problems and their solutions. Each solution highlights one key idea which we shall use later.

2.1 Reconstructing one point when the scenery is observed at i.i.d. locations

Take any non random scenery $\xi : \mathbb{N} \to \{1,2,3,4,5\}$. Let a and b be two integers with $a < b$ and define $I := [a,b]$. Let $Y_1, Y_2, \ldots$ be i.i.d. random variables such that

$$P(Y_i = b+1) > P(Y_i \notin [a, b+1]). \tag{6}$$

The first simple problem we consider, is to reconstruct the scenery ξ at a single point, namely $b+1$. So we need to determine the value of $\xi(b+1)$. For this, we suppose we are given two things: The restriction of ξ to the interval I, that is, $\xi|_I := (\xi_i \, i \in I)$, and an infinite sequence of observations of the scenery ξ at the random locations Y_i,

$$\xi(Y_1), \xi(Y_2), \ldots. \tag{7}$$

Now, for any $e \in \{1,2,3,4,5\}$, we have

$$P(\xi(Y_i) = e) = P(\xi(Y_i) = e, Y_i \in [a,b]) + \delta_{\xi(b+1)=e} P(Y_i = b+1) + P(Y_i \notin [a,b+1], \xi(Y_i) = e).$$

Let q_e be the quantity

$$q_e := P(\xi(Y_i) = e) - P(\xi(Y_i) = e, Y_i \in [a,b]). \tag{8}$$

When inequality 6 holds, two possibilities manifest. If $\xi(b+1) = e$, then $q_e \geq P(Y_i = b+1)$. On the other hand, if $\xi(b+1) \neq e$, then $q_e < P(Y_i = b+1)$. We can use this dichotomy to reconstruct $\xi(b+1)$ as follows:
Estimate q_e and take the element $e \in \{1,2,3,4,5\}$ which maximizes it for the color at point $b+1$.

To determine q_e, we first estimate the probability $P(\xi(Y_i) = e)$ from the sequence 7. The probability $P(\xi(Y_i) = e, Y_i \in [a,b])$ can be calculated, since we are given the restriction of ξ to I. For this, we also assume that the distribution of the Y_i's is known to us.

Note that Condition 6 holds automatically when the tail of Y_i decays exponentially with decay rate r strictly less than $1/2$ and

$$a < 0 < b \text{ and } |a| \gg |b|. \tag{9}$$

To see this, let the Y_i's have a symmetric distribution with

$$P(Y_i = y) \cdot r \geq P(Y_i = y+1) \;\; , \;\; \forall y \in \mathbb{N}, \tag{10}$$

where $r < 0.5$. Assuming that $|a| \gg |b|$, $P(Y_i < a)$ is negligible in comparison to $P(Y_i = b+1)$. Furthermore, thanks to the exponential decay, we obtain

$$\frac{P(Y_i > b+1)}{P(Y_i = b+1)} \leq \frac{r}{1-r} < 1.$$

Finally, if $P(Y_i \leq a)$ is small enough then this last inequality implies Condition 6.

2.2 Reconstructing a point when the scenery is seen along an infinite number of random walks

Once again assume that ξ is a non-random scenery. As in the previous setting, the restriction of ξ to $[a,b]$ is known and we try to reconstruct $\xi(b+1)$. We also assume that condition 9 holds. Let

$$\{S_t^1\}_{t \in \mathbb{N}}, \{S_t^2\}_{t \in \mathbb{N}}, \ldots$$

be a sequence of random walks independent of each other and all of which start at the origin. Each of the S_t^i's is identical to the random walk S_t, which we define to have the following increment distribution:

$$P(S_t - S_{t-1} = x) = \begin{cases} \epsilon, \text{if } x = 0, & \\ \gamma, & \text{if } |x| = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where $\gamma = \frac{1-\epsilon}{2}$ and $t \in \mathbb{N}$. Thus, each S_t^i is a simple symmetric random walk modified to allow sojourns of duration greater than unity in any state. This is the simplest random walk satisfying Conditions 1–4. Let χ^i denote the observations of the scenery ξ made by random walk number i:

$$\chi^i := (\chi_i^t,\ t = 0, 1, 2, \ldots),$$

where $\chi_t^i := \xi(S_t^i)$. In addition to knowing the restriction $\xi|_{[a,b]}$, we assume that all the observations $\chi^1, \chi^2, \chi^3, \ldots$ made by the different random walks are given to us.

Take q such that $1 < q < 3$. In order to apply the reconstruction technique from the previous section, we set Y_i to be equal to $Y_i := S_r^i$, where $r = qb$ is an integer. The state distribution of the symmetric random walk S_t is given by

$$P(S_t = x) = \begin{cases} \sum_{s \in [0,t-x]\,:\, s+t \text{ is even}} \binom{t}{s} \binom{t-s}{(t-s+x)/2} \epsilon^s \gamma^{t-s}, & \text{if } |x| \le t, \\ 0, & \text{otherwise.} \end{cases}$$

Provided $|x| \le t$, the decay rate of S_t at the point x is given by

$$\begin{aligned} \rho(t,x) &:= \frac{P(S_t = x+1)}{P(S_t = x)} = \frac{\sum_{s \in [0,t-x-1]\,:\, s+t \text{ is even}} \binom{t}{s} \binom{t-s}{(t-s+x+1)/2} \epsilon^s \gamma^{t-s}}{\sum_{s \in [0,t-x]\,:\, s+t \text{ is even}} \binom{t}{s} \binom{t-s}{(t-s+x)/2} \epsilon^s \gamma^{t-s}} \\ &\le \frac{\sum_{s \in [0,t-x]\,:\, s+t \text{ is even}} \binom{t}{s} \binom{t-s}{(t-s+x+1)/2} \frac{(t-s-x)/2}{(t-s+x+1)/2} \epsilon^s \gamma^{t-s}}{\sum_{s \in [0,t-x]\,:\, s+t \text{ is even}} \binom{t}{s} \binom{t-s}{(t-s+x)/2} \epsilon^s \gamma^{t-s}} \\ &\le \max_{s \in [0,t-x]} \frac{t-s-x}{t-s+x+1} = \frac{t-x}{t+x+1}. \end{aligned}$$

Then,

$$\rho(r,b) = \rho(qb,b) = \frac{q-1}{q+1+1/b} < \frac{q-1}{q+1}. \tag{12}$$

For $q \in (1,3)$, The final expression is always less than $1/2$ and, since q does not depend on b, the tail of S_r decays at a rate less than $1/2$ beyond the point b. Hence, condition 10 is satisfied and we can apply the reconstruction technique described in the previous subsection.

So far, we have described how to perform a one-point reconstruction when presented with an infinite set of realizations of the random walk S. However, in the problem under consideration in this paper, we only have access to the observations made by a single random walk S. We get around this limitation by using stopping times to restart the

random walk in a predetermined way. This technique yields the infinite set of realizations we require. Let τ_i denote the time of the i-th visit by S to the origin. Then due to the strong Markov property, the random walk after time τ_i behaves like a random walk starting at the origin. Hence, if we are given the piece of scenery $\xi|_{[a,b]}$ together with observations χ and the sequence of stopping times $\tau_1, \tau_2, \ldots$, then we can use the reconstruction technique for an infinite number of random walks. For this we simply take the sequence of observations χ^i made by the $i-th$ random walk to be

$$\chi_{\tau_i}, \chi_{\tau_i+1}, \chi_{\tau_i+2}, \ldots, \chi_{\tau_{i+1}-1}.$$

2.3 Finding the way back to the origin

Because we are only given the sequence of observations made by the random walk, we are not able to determine when it is at the origin and hence the stopping times $\tau_1, \tau_2, \ldots$ are not observable. Instead, we must construct stopping times based on the observations which are able to stop the random walk close to some point of reference.

Let $K = [k_1, k_2] \subset [-n, n]$ be an integer interval of length n. The precise position of K within $[-n, n]$ is not important here and will be specified later. Our point of reference will be the location of the finite string $w := \xi_{k_1}\xi_{k_1+1}\ldots\xi_{k_2}$. Let τ_i be the i-th time that we observe the pattern w in the observations χ. Hence

$$\tau_1 := \min\{t \geq n \ : \ \chi_{(t-n)}\chi_{(t-n+1)}\cdots\chi_{(t)} = w\} \tag{13}$$

and

$$\tau_{i+1} := \min\{t > \tau_i \ : \ \chi_{(t-n)}\chi_{(t-n+1)}\cdots\chi_{(t)} = w\}. \tag{14}$$

Let R be a map $R : [0, n] \to \mathbb{Z}$. We call R a *simple random walk path*, if

$$\forall t \in [0, n-1] \ , \ |R(t+1) - R(t)| = 1.$$

If $R(0) = x$, we say the path R starts at x. We assume that the scenery ξ is random, i.i.d. and all five colors appear with equal probability. Let $x \notin [-2n, 2n]$. Let R be a non-random simple random walk path starting at x. What is the probability that ξ seen along the path R gives the string w? In other words, what is the probability that

$$\xi \circ R := \xi(R_0)\xi(R_1)\ldots\xi(R_n) = w?$$

Note that since $x \notin [-2n, 2n]$, the path of R can not reach $[-n, n]$. Since the scenery is i.i.d. and since w only depends on $\xi|_{[-n,n]}$, the pattern w is independent of $\xi \circ R$. The five colors in the scenery having the same probability, we find

$$P(\xi \circ R = w) = \left(\frac{1}{5}\right)^{n+1}.$$

There are 2^n simple random walk paths starting at x of length n. Hence the probability that there exists such a path generating the color record w is bounded:

$$P(\exists R \text{ a simple r.w. path starting at } x \text{ such that } \xi \circ R = w) \leq \left(\frac{1}{5}\right)^{n+1} \cdot 2^n < \left(\frac{2}{5}\right)^n.$$

It follows that with probability close to one, every time we observe the pattern w before the time $T = \rho^n$, the random walk must be in the interval $[-2n, 2n]$:

$$S_{\tau_i} \in [-2n, 2n] \text{ for every } \tau_i \leq T,$$

where $\rho > 0$ is a constant not depending on n such that $\rho < (5/2)^2$.

We shall see that this argument can be refined so that the interval in which S_{τ_i} lies is much narrower. Also, we shall be dealing with non-simple random walks. Hence it will be necessary to adapt the present argument to that situation.

3 Single Point Reconstruction

In this section, we describe the algorithm for reconstructing $\xi(n+1)$ given the observations χ and the restriction $\xi|_{[-n,n]}$. Before beginning, it is useful to consider a small numerical example to illustrate.

3.1 A numerical example

Assume that the random walk is simple and suppose we are given the finite restriction of ξ to the interval $I := [-4, 4]$:

$\xi(z)$	2	4	3	2	4	5	1	5	3	?
z	-4	-3	-2	-1	0	1	2	3	4	5

We wish to reconstruct the value $\xi(5)$. Let $K := [0, 3]$ and let w be the restriction of ξ to K:

$$w := 4515.$$

Occurrences of the pattern w in the observations χ define a sequence of stopping times (τ_i) like that defined by (13) and (14). Note that if we observe the pattern w and this was generated while the random walk was in the interval I, then we must be located at either $z = 1$ or $z = 3$. More precisely, if $S_s \in I$ for all $s \in [\tau_i - 3, \tau_i]$, then $S_{\tau_i} \in \{1, 3\}$ and the random walk will be at one of the points 1 or 3 with equal probability. Let μ denote the two- atom measure which accords probability $1/2$ to $\{1\}$ as well as to $\{3\}$.

Let $P_\mu(.)$ denote the probability distribution for the random walk S starting with initial distribution μ instead of starting at the origin.

Take $r = 4$. As was done in (8), we define

$$q_e := P_\mu\left(\xi(S_r) = e\right) - P_\mu\left(\xi(S_r) = e, \xi(S_r) \in [-4, 4]\right). \tag{15}$$

Then, we look at the empirical frequency of the colors at a fixed offset r following each stopping time τ_i. More precisely, we estimate $P_\mu(\xi(S_r) = e) = P_\mu(\chi_r = e)$ by the empirical distribution of

$$\chi_{\tau_1+r}, \chi_{\tau_2+r}, \ldots, \chi_{\tau_j+r},$$

where j is the largest integer i such that $\tau_i \leq T$. Let us denote this empirical distribution by

$$\hat{P}_\mu(\chi_r = \cdot). \tag{16}$$

We can now describe the reconstruction algorithm.

Algorithm for single point reconstruction. Given $\xi|_{[-4,4]}$ and χ, our estimate of $\xi(5)$ is the element e of the set $\{1, 2, 3, 4, 5\}$ which maximizes the quantity

$$\hat{q}_e := \hat{P}_\mu(\chi_r = e) - P_\mu\left(\xi(S_r) = e, S_r \in [-4, 4]\right).$$

Note that the second term in the definition of $\hat{q}_e$ can be calculated explicitly because $\xi|_{[-4,4]}$ is given.

To verify that the above algorithm has a high probability of estimating $\xi(5)$ correctly, we first need to check that

$$P_\mu(S_r = 5) > P_\mu(S_t \notin [-4, 5]). \tag{17}$$

This inequality corresponds to Condition 6. Taking $r = 4$ we find

$$P_\mu(S_r = 5) = \frac{5}{32}, \qquad P_\mu(S_r \in (-\infty, -5] \cup [6, \infty)) = \frac{1}{32}$$

and so inequality 17 is satisfied.

The second problem we need to take care of is verifying that the estimate 16 is precise enough with high probability. For the algorithm to work, the estimation error needs to be strictly less than

$$\frac{P_\mu(S_r = 5) - P_\mu(S_r \notin [-4, 5])}{2}. \tag{18}$$

For this we need to show that there exists T such that the following two conditions are satisfied with high probability:

- The time T is large enough so that there are enough visits by the random walk to the origin generating the pattern w up to time T.
- The time T is not too large, since otherwise during the time interval $[0, T]$ the random walk will wander too far away from the origin. Far from the origin there are other places where the pattern w can be generated. Thus, too large a T results in S_{τ_i} being distant from the origin at some of the times $\tau_i \in [0, T]$.

In the numerical example used to illustrate above, n has been taken too small for the required estimate precision to be obtained. One of the main issues in the general context is to show that for n large enough, we can find a T such that with high probability the two conditions above can be satisfied simultaneously.

3.2 The one-point reconstruction algorithm

Now we present an algorithm for estimating $\xi(b+1)$ given the restriction $\xi|_{[-n,n]}$ and the observations χ.

Unlike the numerical example above, our problem is to reconstruct the scenery when the random walk is not simple. for this we shall need the concept of a δ-path. The key feature of a δ-path is that it behaves like a simple random walk most of the time and the proportion δ of time that it doesn't do so, its freedom of movement is linearly restricted pro rata.

Definition 3.1 *Let $\delta > 0$. We call $R : [0, n] \to \mathbb{Z}$ a δ-path (of length n) if the proportion of steps of non-unit length as well as their total variation is less than δ, that is,*

$$|M| \leq \delta n$$

and

$$\sum_{t \in M} |R(t+1) - R(t)| \leq \delta n,$$

where the set M is defined by

$$M := \{t \in [0, n-1] \ : \ |R(t+1) - R(t)| \neq 1\}.$$

For the Reconstruction, we shall use the stopping times τ_i that occur prior to time $T = 2.4^{2n}$. We will subsequently show that by taking the parameter $\epsilon > 0$ small enough, the random walk can be made to follow δ-paths for any time interval of length n before time T.

Our algorithm for reconstructing $\xi(n+1)$ is defined using three intervals. First

$$I := [-n, n]$$

is the interval on which we already know the scenery ξ. Let

$$n^* := n - 61\delta n$$

where $\delta > 0$ is a small constant not depending on n. The second interval

$$K = [k_1, k_2] := [n^* - n, n^*]$$

is the interval from which we take the pattern

$$w := \xi(k_1)\xi(k_1+1)\xi(k_1+2)\dots\xi(k_2).$$

Finally we will show that after reading the pattern w close to the origin we are typically located in the interval

$$J := [n^* - 21\delta n, n^* + \delta n].$$

Next, let ν_i be the i-th time the random walk generates the pattern w on the piece of scenery $\xi|_{[-n,n]}$:

$$\nu_1 := \min\{t \geq n \ : \ \chi_{t-n}\chi_{t-n+1}\dots\chi_t = w \ ; \ \forall s \in [t-n, t], \ S_s \in [-n, n]\} \tag{19}$$

and

$$\nu_{i+1} := \min\{t > \nu_i \ : \ \chi_{t-n}\chi_{t-n+1}\dots\chi_t = w \ , \ S_s \in [-n, n], \forall s \in [t-n, t]\}, \qquad i \geq 1. \tag{20}$$

The difference between the ν_i's and the τ_i's is the additional constraint on the ν_i's that the pattern should be generated while the random walk is in $[-n, n]$. This is apriori not observable. But we will show that the τ_i's and the ν_i's coincide up to time T with high probability.

Note that given ξ, the sequence

$$S_{\nu_1}, S_{\nu_2}, S_{\nu_3}, \dots$$

is a Markov chain whose state space is $[-n, n]$. The transition probability from x to y is given by

$$P_{xy} := P(S_{\nu_{i+1}} = y \ \mid \ S_{\nu_i} = x, \xi).$$

This chain is aperiodic and irreducible. Denote its stationary distribution by μ. Hence, μ depends on ξ and is thus a random measure. As before, $P_\mu(\cdot)$ denotes the measure for which the random walk S has starting distribution μ and for which the distribution of the scenery remains unchanged and is independent of the random walk. We

take $r := 90\delta n$ and look at the frequency of the observed color r time steps after a stopping time. Our estimate for the distribution $\mathcal{L}_\mu(\chi_r)$ is the empirical distribution of

$$\xi(S_{\tau_i+r}), \xi(S_{\tau_2+r}), \ldots, \xi(S_{\tau_j+r}),$$

where j denotes the largest integer for which $\tau_i \leq T$. We shall denote this empirical distribution by

$$\hat{P}_\mu(\chi_r = \cdot).$$

We are now ready to describe the reconstruction algorithm for estimating the value $\xi(n+1)$.

Algorithm: Given $\xi|_{[-n,n]}$ and χ, the estimate $\hat{\xi}(n+1)$ of $\xi(n+1)$ is the element $e \in \{1,2,3,4,5\}$ which maximizes the quantity

$$\hat{q}_e := \hat{P}_\mu(\chi_r = e) - P_\mu\left(\xi(S_r) = e \ , \ S_r \in I\right).$$

Note that the second term on the right-hand side of the definition of $\hat{q}_e$ can be calculated explicitly since we know $\xi|_I$.

3.3 Combinatorial aspects of the one point Reconstruction algorithm

Let A^n denote the event that the above reconstruction algorithm works correctly, that is the event that $\hat{\xi}(n+1) = \xi(n+1)$. We list a few events which are important in making the reconstruction of $\xi(n+1)$ work.

B^n: Performing reconstruction will require the random walk to remain near the origin, in some sense, for a sufficiently long time. To capture this requirement, let B^n be the event that the random walk stays in the interval $[-2.45^n, 2.45^n]$ for all times prior to and including $T = 2.4^{2n}$.

C^n: In order to obtain sufficiently precise estimates, we will need enough observations close to the origin. This is taken care of by C^n which is the event that at least $(1.1)^n$ stopping times ν_i occur before time T:

$$C^n := \{\nu_i \leq T \ : \ i \leq 1.1^n\}.$$

D^n: We take D^n to be The event that, up to time T, all pieces of length n of the path S are δ-paths. More precisely, D^n is the event that for all $s \in [n, T]$, we have

$$S : [s-n, s] \to \mathbb{Z} \quad t \mapsto S_t$$

is a δ-path.

F^n**:** Let F^n be the event that in the interval $[-2.45^n, 2.45^n]$ a δ-path can generate w only if it ends in the interval J and does not leave the interval I. More precisely, F^n is the event that, For all δ-paths

$$R : [0, n] \to [-2.45^n, 2.45^n]$$

for which $\xi \circ R = w$, we have $R(n) \in J$ and $R(l) \in [-n, n]$, for all $l \in [0, n]$.

G^n**:** Let G^n be the event that the precision of the estimate based on the ν_i's is better than

$$\frac{P_\mu(S_r = n + 1) - P_\mu(S_r \notin [-n, n + 1]}{2}. \tag{21}$$

More precisely, denote the empirical distribution of

$$\chi(\nu_1 + r), \chi(\nu_2 + r), \ldots, \chi(\nu_j + r)$$

where j is the largest i for which $\nu_i \leq T$, by

$$\tilde{P}_\mu(\chi_r = \cdot).$$

Recall that r is defined to be $90\delta n$. Then G^n is the event that, for all $e \in \{1, 2, 3, 4, 5\}$, the difference

$$\left|\tilde{P}_\mu(\chi_r = e) - P_\mu(\chi_r = e)\right|$$

is strictly less than the expression given in 21.

The following lemma shows that when all the events B^n through G^n hold, then we can reconstruct $\xi(n + 1)$ correctly.

Lemma 3.1 *Assume that expression 21 is strictly positive. Then,*

$$B^n \cap C^n \cap D^n \cap F^n \cap G^n \subset A^n.$$

Proof. If B^n holds, then the random walk stays within the interval $[-2.45^n, 2.45^n]$ for all times up to and including T. However, thanks to the event F^n, a δ-path within that interval can generate the pattern w only if it stays within the interval $[-n, n]$. This means that when B^n and F^n both hold, $\tau_i = \nu_i$ for all $\tau_i \leq T$. In this situation, the estimates of $P_\mu(\chi_r = e)$ based on the τ_i's and the ν_i's are identical, Since they only make use of stopping times up to time T. The event C^n merely ensures the occurrence of at least 1.1^n stopping times by time T. When G^n holds, we have

$$\left|\tilde{P}(\chi_r = e) - P_\mu(\chi_r = e)\right|$$

is less than 21 and hence for the estimate based on the τ_i's we have that

$$\left|\hat{P}(\chi_r = e) - P_\mu(\chi_r = e)\right| \tag{22}$$

is also less than 21. But this is enough to make the reconstruction algorithm work. To see this, let a be the number

$$a := \frac{P_\mu(S_r = n+1) + P_\mu(S_r \notin [-n, n+1]}{2}.$$

We have proved that when B^n, C^n, D^n, F^n and G^n all hold, then

$$\hat{P}(\chi_r = e) - P_\mu(\chi_r = e, S_r \in [-n, n]) > a$$

if

$$\xi_{n+1} = e$$

and

$$\hat{P}(\chi_r = e) - P_\mu(\chi_r = e, S_r \in [-n, n]) < a$$

if $\xi_{n+1} \neq e$. This means that the reconstruction algorithm works correctly, since it chooses as estimate for the color ξ_{n+1} the value $e \in \{1, 2, 3, 4, 5\}$ which maximizes

$$\hat{P}(\chi_r = e) - P_\mu(\chi_r = e, S_r \in [-n, n]).$$

This completes the proof. ■

4 High probability of the events B^n, C^n, D^n, F^n and G^n

As a consequence of Lemma 3.1, we see that

$$P(A^{nc}) \leq P(B^{nc}) + P(C^{nc}) + P(D^{nc}) + P(F^{nc}) + P(G^{nc}),$$

so in order to show that A^n occurs with probability close to 1, we need only show that each of B^n, C^n, D^n, F^n and G^n occur with probability approaching 1 as n becomes large. We shall do this by showing that the probability that each of these events does not occur is finitely summable over n.

4.1 The event B^n

Let σ^2 be the variance of the increment distribution of the random walk S. Conditions (1) and (2) ensure that σ^2 is finite. Then, since S_t is a sum of i.i.d. increments, direct

application of Kolmogorov's inequality yields

$$\begin{aligned} P(B^{nc}) &= P(\max_{0\le t\le T}|S_t| \ge 2.45^n + 1) \\ &\le \frac{2.4^{2n}\sigma^2}{\left(2.45^n+1\right)^2} < \sigma^2\left(\left(2.4/2.45\right)^2\right)^n. \end{aligned}$$

Hence $P(B^{nc})$ is finitely summable and $P(B^n) \to 1$ exponentially fast as $n \to \infty$.

4.2 The event C^n

We define some auxilliary events to assist in proving that C^n occurs with high probability. Let C_0^n be the event that the random walk has exited the interval $[-(2.39)^n, (2.39)^n]$ at time T:

$$C_0^n := \{S_T \notin [-2.39.^n, 2.39^n]\}.$$

Next, let C_1^n be the event that there are at least 2.3^n visits to the point $k_1 = n - n^*$ before the random walk leaves the interval $[-2.39^n, 2.39^n]$ for the first time.

Define C_2^n to be the event that among the first 2.3^n visits to k_1 there are at least 1.1^n for which the random walk subsequently takes n steps to the right. That is, if t_i denotes the time of the i-th visit to k_1, then, C_2^n is the event that

$$|\{i \in \mathbb{N} \,:\, i \le 2.3^n \,,\; S_{t_i+s} = k_1 + s,\; \forall s \in [0,n]\}| \ge 1.1^n.$$

It is easy to see that

$$C_0^n \cap C_1^n \cap C_2^n \subset C^n. \tag{23}$$

Thus, in order to prove that C^n occurs with high probability, we need only prove that each of the three events C_0^n, C_1^n and C_2^n occurs with high probability.

Proof that C_0^n occurs with high probability

Let ρ denote the third absolute moment of the increment distribution of S and note that, as was the case for σ^2, conditions (1) and (2) guarantee $\rho < \infty$. Then, fixing $b = 2.39/2.4$, we have

$$\begin{aligned} P(C_0^n) &= P(S_T \notin [-2.39^n, 2.39^n]) = P\left(\frac{S_{2.4^{2n}}}{2.4^n\sigma} \notin [-b^n, b^n]/\sigma\right) \\ &= 1 - \Phi_{2.4^{2n}}\left(b^n/\sigma\right) + \Phi_{2.4^{2n}}\left(-b^n/\sigma\right), \end{aligned}$$

where $\Phi_t(x) := P\left(\frac{S_t}{\sigma t^{1/2}} \leq x\right)$. Let Φ denote the distribution function of a standard Gaussian random variable. Now, $\Phi_t \xrightarrow{\mathcal{D}} Phi$ as $t \to \infty$ by the central limit theorem. Since $b^n \to 0$, we see that

$$\lim_{n\to\infty} P(C_0^n) = 1.$$

Furthermore,

$$\begin{aligned} 1 - P(C_0^n) &= \Phi_T(B^n/\sigma) - \Phi_T(-b^n/\sigma) \\ &\leq \Phi(B^n/\sigma) - \Phi(-b^n/\sigma) + |\Phi_T(B^n/\sigma) - \Phi(b^n/\sigma)| + \\ &\qquad |\Phi_T(-B^n/\sigma) - \Phi(-b^n/\sigma)|. \end{aligned}$$

By the Berry-Esséen theorem,

$$|\Phi_t(x) - \Phi(x)| \leq \frac{\beta\rho}{\sigma^3 t^{1/2}},$$

for all $x \in \mathbb{R}, t \geq 1$, where $\beta = 0.7655$. Therefore,

$$1 - P(C_0^n) \leq 2\int_0^{b^n/\sigma} (2\pi)^{-1/2} e^{-u^2/2}\, du + \frac{2\beta\rho}{2.4^n\sigma^3}.$$

. Since $\int_0^h (2\pi)^{-1/2} e^{-u^2/2}\, du = \frac{1}{\sigma\sqrt{2\pi}} h + o(h^3)$, we then have

$$1 - P(C_0^n) \leq 2(2\pi\sigma^2)^{-1/2} b^n + o(b^{3n}) + \frac{2\beta\rho}{2.4^n\sigma^3}.$$

When n is large enough, the right-hand side will be bounded by $2(2\pi)^{-1/2}\sigma^{-1}b^{n/2}$ and so $P(C_0^n)$ converges exponentially quickly to 1.

Proof that C_1^n occurs with high probability

Let us imagine for a moment that the random walk S is simple, symmetric and starts at $k_1 + 1$. Let η be the first time that S hits k_1 or 2.39^n:

$$\eta = \min\{t \geq 0 \ : \ S_t \in \{k_1, 2.39^n\}\ \}.$$

The random walk is a Martingale and hence we have that

$$k_1 + 1 = \mathbb{E}[S_0] = \mathbb{E}[S_\eta] = k_1 \cdot P(S_\eta = k_1) + (2.39^n) P(S_\eta = 2.39^n) \geq 2.39^n P(S_\eta = 2.39^n). \tag{24}$$

Hence, $P(S_\eta = 2.39^n) \leq \frac{k_1+1}{2.39^n}$ and the simple random walk starting at $k_1 + 1$ has a probability of less then $(k_1 + 1)(2.39)^{-n}$ of hitting 2.39^n before hitting k_1. The same bound holds for a simple random walk starting at $k_1 - 1$ and its probability of hitting -2.39^n before returning to k_1. We note that it is possible to obtain a slightly better

bound, but this suffices here. Hence the number of visits to the point k_1 by the random walk before hitting either $-2, 39^n$ or 2.39^n is a geometric random variable with parameter $p \leq (k_1+1)2.39^{-n}$. It follows that the probability of making fewer than 2.3^n visits to the point k_1 before hitting the set $\{-2.39^n, 2.39^n\}$ is less than $(k_1+1)\left(\frac{2.3}{2.39}\right)^n$. This imposes a bound on $P(C_1^{nc})$ that is negatively exponentially small in n if the random walk is simple. So we need to explain why the same kind of result holds for our random walk S whose increment distribution has exponentially decaying tails.

This is done by considering a random walk $\bar{S}$, which is generated by S and which makes steps of length at most n^2. Whenever the random walk S jumps further than this, $\bar{S}$ does not move. Hence, $S_0 = \bar{S}_0$ and for all $t \in \mathbb{N}$,

$$\bar{S}_t - \bar{S}_{t-1} := \begin{cases} S_t - S_{t-1}, & \text{if } |S_t - S_{t-1}| \leq n^2, \\ 0, & \text{otherwise.} \end{cases}$$

The key here is that the random walks S and $\bar{S}$ will very likely be identical within a time frame of interest. To capture this, we introduce C_{11}^n which is the event that

$$S_t = \bar{S}_t \ , \ \forall t \leq 7^n.$$

Let C_{12}^n be the event that the random walk $\bar{S}$ leaves the interval $[-2.39^n, 2.39^n]$ no later than 7^n.
We define the stopping times $\bar{t}_i$ inductively. Let $\bar{t}_1 = 0$ and $\bar{t}_{i+1}$ be the first time no earlier than $\bar{t}_i + n^4$ that the random walk $\bar{S}$ visits the interval $-[n^2, n^2]$:

$$\bar{t}_{i+1} := \min\{t \geq \bar{t}_i + n^4 \ : \ \bar{S}_t \in [-n^2, n^2]\}.$$

Let C_{13}^n be the event that at least 2.38^n stopping times $\bar{t}_i$ occur before $\bar{S}$ leaves the interval $[-2.39^n, 2.39^n]$.
Let $\{Y_i\}_1^\infty$ be a sequence of Bbernoulli random variables marking visits to the point k_1 at specific stopping times: $Y_i := 1$ iff $\bar{S}_{\bar{t}_i + n^4} = k_1$.
finally, define C_{14}^n to be the event that

$$\sum_{i=1}^{2.38^n} Y_i \geq 2.3^n.$$

Note that

$$C_{11}^n \cap C_{12}^n \cap C_{13}^n \cap C_{14}^n \subset C_1^n. \tag{25}$$

The following argument explains why this is so. Due to C_{12}^n, the random walk leaves the interval $[-2.39^n, 2.39^n]$ before time 7^n and by C_{13}^n there are at least 2.38^n stopping times $\bar{t}_i$ before $\bar{S}$ leaves the interval $[-2.39^n, 2.39^n]$. Thus, C_{12}^n and C_{13}^n together imply that at least 2.38^n stopping times $\bar{t}_i$ will be seen before time 7^n. The event C_{14}^n forces at least 2.3^n of these 2.38^n stopping times $\bar{t}_i$ to be followed by a visit to the point k_1 at a time n^4 later. Hence, we have that if C_{12}^n, C_{13}^n and C_{14}^n all hold, then prior to time 7^n, the random walk

$\bar{S}$ visits the point k_1 at least 2.3^n times before finally leaving the interval $[-2.39^n, 2.39^n]$. The event C_{11}^n stipulates that S and $\bar{S}$ be identical up to time 7^n. Consequently, S also visits k_1 at least 2.3^n times before leaving the interval $[-2.39^n, 2.39^n]$ and so C_1^n holds.

From the implication 25 it follows that

$$P(C_1^{nc}) \leq P(C_{11}^{nc}) + P(C_{12}^{nc}) + P(C_{13}^{nc}) + P(C_{14}^{nc}).$$

To prove finite summability of $P(C_1^{nc})$ it is enough to prove it for each term on the right-hand side of this inequality.

Finite summability of $P(C_{11}^{nc})$. Note that the event C_{11}^n holds as soon as the random walk S does not take any step of size larger than n^2 up to and including time 7^n. Since the tail of the increment distribution of S is exponentially decaying, we have

$$P(|S_t - S_{t-1}| > n^2) \leq c_1 e^{-c_2 n^2},$$

where $c_1, c_2 > 0$ are constants not depending on n. Thus,

$$P(C_{11}^{nc}) \leq c_1 7^n e^{-c_2 n^2}$$

Provided c_2 is large enough, the expression on the right-hand side of this inequality is indeed finitely summable.

Finite summability of $P(C_{12}^{nc})$. To see the finite summability of $P(C_{12}^{nc})$, first realize that

$$C_{12}^n = \{\max_{0 \leq t \leq 7^n} |\bar{S}_t| > 2.39^n\} \supset \{\bar{S}_{7^n} \notin [-2.39^n, 2.39^n]\}$$

and then emulate the proof that $P(C_0^{nc})$ is finitely summable.

Finite summability of $P(C_{13}^n)$. At any stopping time $\bar{t}_i$, the random walk $\bar{S}$ will be in the interval $[-n^2, n^2]$. Since $\bar{S}$ can make steps of size no larger than n^2, it can travel a maximum distance of n^6 from where it started in elapsed time n^4. At time $\bar{t}_i + n^4$, $\bar{S}$ will therefore be in $[-n^2 - n^6, n^2 + n^6]$. If $\bar{S} \in [-n^2, n^2]$ at time $\bar{t}_i + n^4$, then $\bar{t}_{i+1} = \bar{t}_i + n^4$ and so it won't have left $[-2.39^n, 2.39^n]$ by $\bar{t}_{i+1}$. On the other hand, suppose that $\bar{S}$ starts at a point $x_0 \in (n^2, n^2 + n^6]$ and let η be the first time that it enters $[2.39^n, 2.39^n + n^2]$ or $[0, n^2]$. Then

$$x_0 = E_{x_0}[\bar{S}_0] = E_{x_0}[\bar{S}_\eta] = x_2 P_{x_0}(\bar{S}_\eta \in [2.39^n, 2.39^n + n^2]) + x_1 P_{x_0}(\bar{S}_\eta \in [0, n^2]), \quad (26)$$

where x_2 is the conditional expectation of $\bar{S}_\eta$ given that $\bar{S}_\eta \in [2.39^n, 2.39^n + n^2]$ and x_1 denotes the conditional expectation of $\bar{S}_\eta$ given that $\bar{S}_\eta \in [0, n^2]$. From equation 26 it follows that

$$x_0 \geq x_2 P(\bar{S}_\eta \in [2.39^n, 2.39^n + n^2])$$

and, since $x_2 \geq 2.39^n$ and $x_0 \leq n^2 + n^6$, we have

$$P(\bar{S}_\eta \in [2.39^n, 2.39^n + n^2]) \leq \frac{n^2 + n^6}{2.39^n}. \tag{27}$$

An identical argument yields the same bound for the random walk hitting $[-2.39^n - n^2, -2.39^n]$ before $[-n^2, 0]$ given that $\bar{S}$ is in $[-n^2 - n^6, -n^2]$ at time $\bar{t}_i + n^4$. Hence for all stopping times $\bar{t}_i$, the expression on the right-hand side of 27 also serves to bound the probability of the random walk leaving $[-2.39^n, 2.39^n]$ before paying a visit to $[-n^2, n^2]$ after time $\bar{t}_i + n^4$. This implies that the number of stopping times $\bar{t}_i$ appearing before $\bar{S}$ exits $[-2.39, 2.39^n]$ is a geometric random variable with parameter $p \leq (n^2 + n^6)/2.39^n$. Hence, the probability $P(C_{13}^{nc})$ of seeing fewer than 2.38^n stopping times $\bar{t}_i$ before the random walk leaves the interval $[-2.39^n, 2.39^n]$ is less than

$$\left((n^2 + n^6)\frac{2.38}{2.39}\right)^n.$$

This shows that $P(C_{13}^{nc})$ is finitely summable.

Finite summability of $P(C_{14}^n)$. Let $\mathcal{F}_i$ be the σ-algebra generated by $\bar{S}_0, \bar{S}_1, \ldots, \bar{S}_{\bar{t}_i}$. Then by the Local Central Limit Theorem, we find that for n large enough, there exists a constant $c_4 > 0$ (not depending on n) such that

$$P(\bar{S}_{\bar{t}_i + n^4} = k_1 \mid \mathcal{F}_i) \geq \frac{c_4}{n^2}$$

almost surely. (We use the fact that at time $\bar{t}_i$ the random walk $\bar{S}$ is located in the interval $[-n^2, n^2]$.) This then implies that the sum

$$\sum_{i=1}^{2.38^n} Y_i$$

is bounded below by

$$\sum_{i=1}^{2.38^n} Y_i^*$$

where Y_i^* are i.i.d. Bernoulli random variables with parameter c_4/n^2. Thus, we have

$$P(C_{14}^{nc}) \leq P\left(\sum_{i=1}^{2.38^n} Y_i^* < 2.3^n\right) = P\left(\sum_{i=1}^{2.38^n} (1 - Y_i^*) \geq 2.38^n - 2.3^n + 1\right). \tag{28}$$

Note that the $(1 - Y_i^*)$'s remain i.i.d. Bernoulli but have parameter $1 - c_4/n^2$. The second non-central moment of

$$\sum_{i=1}^{2.38^n} (1 - Y_i^*)$$

is equal to $2.38^n(1 - c_4/n^2)$. Applying Chebychev's inequality without centring yields

$$P\left(\sum_{i=1}^{2.38^n}(1 - Y_i^*) \geq 2.38^n - 2.3^n + 1\right) \leq \frac{2.38^n\big(1 - c_4/n^2\big)}{\left(2.38^n - 2.3^n\right)^2} \leq 2.38^{-n}(2.38/0.08)^2,$$

which is finitely summable over n. From Equation 28, this expression also bounds $P(C_{14}^{nc})$.

Proof that C_2^n occurs with high probability

Recall that t_i denotes the i-th visit by the random walk S to the point k_1. We will define a subset $\{\kappa_1, \kappa_2, \ldots\}$ of the t_i's inductively as follows. Fix $\kappa_1 := t_1$ and, for $i \geq 1$, let κ_{i+1} be the first t_j after time $\kappa_i + n$:

$$\kappa_{i+1} := \min\{t_j \geq \kappa_i + n \ : \ j \in \mathbb{N}\}.$$

The κ_i's form a strictly increasing sequence. Note that, among the first 2.3^n stopping times t_i, there are at least $(2.3)^n/n$ stopping times κ_i. Let Y_i be a Bernoulli variable which is equal to one iff the random walk takes n steps to the right immediately following time κ_i:

$$Y_i := \mathbf{1}_{S_{t_i+s}=k_1+s, \forall 0<s\leq n}.$$

The variables $Y_1, Y_2, \ldots$ are i.i.d. and $P(Y_i = 1) = \alpha^n$, where $(1 - \epsilon)/2 \leq \alpha \leq 1/2$. Let C_3^n be the event

$$\sum_{i=1}^{2.3^n/n} Y_i \geq 1.1^n.$$

Then since among the first 2.3^n stopping times t_i there are at least $2.3^n/n$ stopping times κ_i, we see that C_3^n implies C_2^n. Together with Chebychev's inequality, this yields

$$\begin{aligned}
P(C_2^{nc}) &\leq P(C_3^{nc}) = P\left(\sum_{i=1}^{2.3^n/n} Y_i \geq 1.1^n\right) \\
&= P\left(\sum_{i=1}^{2.3^n/n} \big(Y_i - \mathbb{E}[Y_i]\big) \geq 1.1^n - 2.3^n\alpha^n/n\right) \\
&\leq P\left(\sum_{i=1}^{2.3^n/n} \big(Y_i - \mathbb{E}[Y_i]\big) \geq 1.1^n - 1.15^n/n\right) \\
&\leq \frac{2.3^n \cdot 0.5^n/n}{(1.15^n/n - 1.1^n)^2} = 1.15^{-n}\, n\,(1 - n(1.1/1.15)^n)^{-2}.
\end{aligned}$$

Since $n(1.1/1.15)^n \downarrow 0$ as $n \to \infty$, $P(C_2^{nc}) \leq 529n \cdot 1.15^{-n}$ and so $P(_2^{nc})$ is indeed finitely summable.

4.3 The event D^n

Let $R : [0, n] \to \mathbb{Z}$ be a piece of length n taken from the sample path of S. In order to prove D^n occurs with high probability, we must first obtain a bound on the probability of R being a δ-path. towards this end, define

$$D_0^n := \{R : [1, n] \to \mathbb{Z} \text{ is a } \delta\text{-path}\}.$$

Also, set $X_i := |R_{i+1} - R_i| \cdot \mathbf{1}_{|R_{i+1}-R_i|\neq 1}$ and $Y_i := \mathbf{1}_{|R_{i+1}-R_i|\neq 1}$. The X_i's represent the sizes of jumps over distances greater than unity while the Y_i's indicate those times at which a non-unit jump occurred. Note that $X_i = Y_i = 0$ whenever the i-th jump is of unit length. Observe that by definition, $D_0^n = D_1^{nc} \cap D_2^{nc}$, where

$$D_1^n := \left\{ \sum_{i=1}^{n-1} X_i > \delta n \right\}$$

and

$$D_2^n := \left\{ \sum_{i=1}^{n-1} Y_i > \delta n \right\}.$$

Therefore, $P(D_0^{nc}) \leq P(D_1^n) + P(D_2^n)$. Our aim is to show that $P(D_1^n)$ and $P(D_2^n)$ can be bounded exponentially small as n becomes large.

From the moment generating function of X_i, we have

$$\begin{aligned}
\mathbb{E}[e^{sX_i}] &= P(X_i = 0) + \sum_{j=2}^{\infty} e^{sj} P(X_i = j) \\
&\leq \epsilon + \epsilon \sum_{j=2}^{\infty} e^{sj} e^{-cj} \\
&= \epsilon + \epsilon \sum_{j=2}^{\infty} \left(e^{s-c}\right)^j \\
&= \epsilon \left(1 + \frac{e^{-2(c-s)}}{1 - e^{-(c-s)}}\right).
\end{aligned}$$

Applying Churnov's inequality, we have

$$\begin{aligned}
P(D_1^n) &= P\left(\sum_{j=0}^{n-1} X_j \geq \delta(n+1)\right) \\
&\leq \mathbb{E}\left[e^{s(\sum_{j=0}^{n-1} X_j - \delta(n+1))}\right] \\
&= e^{-\delta(n+1)s} \left(\epsilon \left(1 + \frac{e^{-2(c-s)}}{1 - e^{-(c-s)}}\right)\right)^n \\
\leq \left(\epsilon \left(1 + \frac{e^{-2(c-s)}}{1 - e^{-(c-s)}}\right)\right)^n,
\end{aligned}$$

for all $s > 0$. Since $\frac{e^{-2(c-s)}}{1-e^{-(c-s)}}$ is strictly increasing on $[0, \infty)$, we see that

$$P(D_1^n) \leq \left(\epsilon\left(1 + \frac{e^{-2c}}{1-e^{-c}}\right)\right)^n = e^{-c^+ n},$$

where $c^+ := -\ln\left(1 + \frac{e^{-2c}}{1-e^{-c}}\right) - \ln\epsilon$. Observe that c^+ can always be made positive and as large as we like by choosing ϵ sufficiently small. In particular, ϵ can be chosen to ensure that $c^+ > c' := c + 2\ln 2.4 + \ln 2$, whence $P(D_1^n) \leq e^{-c'n}$ for all n.

Next,

$$\mathbb{E}[e^{sY_i}] = e^s P(|R_{i+1} - R_i| \neq 1) \leq \epsilon e^s.$$

Applying the same standard large deviation argument used above, we obtain

$$\begin{aligned} P(D_2^n) &= P\left(\sum_{j=0}^{n-1} Y_i \geq \delta(n+1)\right) \\ &\leq \mathbb{E}\left[e^{s\sum_{j=0}^{n-1} Y_i - \delta(n+1)s}\right] \\ &\leq e^{-\delta(n+1)s}\left(\mathbb{E}(e^{sY_1})\right)^n \leq e^{ns}\epsilon^n, \end{aligned}$$

for all $s > 0$. Since $e^n s$ is strictly increasing in s, we have $P(D_2^n) \leq \epsilon^n \leq e^{-c'n}$, provided ϵ is once again chosen small enough.

Now, the probability of seeing a δ-path in a piece of S of length n can be bounded below by

$$P(D_0^n) \geq 1 - P(d_1^n) - P(D_2^n) \geq 1 - 2e^{-c'n} = 1 - e^{-c''n},$$

where $c'' = c + 2\ln 2.4$. Therefore, $P(D_0^n) \nearrow 1$ exponentially fast as $n \to \infty$.

Now, D^n may be expressed as the intersection

$$d^n = \bigcap_{s=n}^{T} D^{n,s},$$

where $D^{n,s}$ is the event that the segment

$$S : [s-n, s] \to \mathbb{Z}$$

of the sample path of S is a δ-path. Then,

$$\begin{aligned} P(D^{nc}) &\leq \sum_{s=n}^{T} P\big((D^{n,s})^c\big) = \sum_{s=n}^{T} P(D_0^{nc}) \\ &\leq (T-n+1)e^{-c''n} \leq 2.4^{2n} e^{-(c+2\ln 2.4)n} = e^{-cn}, \end{aligned}$$

which shows that $P(D^n) \to 1$ exponentially fast.

4.4 The event F^n

Let $R : [0, n] \to \mathbb{Z}$ denote a δ-path of length n. Define F_0^n to be the event that for all $x \in [-2.45^n, 2.45^n] \setminus [-2n - \delta n, 3n]$, there exists no δ- path R such that

$$\xi \circ R = w \text{ and } R(n) = x. \tag{29}$$

Define F_1^n to be the event that for all $x \in [n^* + \delta n, 3n]$ there exist no δ-path R satisfying 29 and let F_2^n denote the event that, for all $x \in [-2n - \delta n, n^* - 21\delta n]$, there exists no δ-path R satisfying 29. Then let F_J^n be the event that for any δ-path $R : [0, n] \to [-2.45^n, 2.45^n]$ satisfying $\xi \circ R = w$, we have $R(n) \in J$. It should be clear that

$$F_0^n \cap F_1^n \cap F_2^n \subset F_J^n. \tag{30}$$

Let F_-^n be the event that for any δ-path $R : [0, n] \to [-2.45^n, 2.45^n]$ satisfying $\xi \circ R = w$, we have $R(0) \in J^-$ where

$$J^- := [n^* - n - \delta n, n^* - n + 21\delta n].$$

Note that for any δ-path $R : [0, n] \to \mathbb{Z}$ with $R(0) \in J^-$ and $R(n) \in J$, we will have

$$n^* - n + (1 - \delta)s - 2\delta n \leq R(s) \leq n^* - n + (1 - \delta)s + 2\delta n \ , \ \forall s \in [0, n]. \tag{31}$$

Fixing δ so that $63\delta < 1$, (31) implies that $R(s) \in [-n, n]$ for all $s \in [0, n]$. Consequently,

$$F_J^n \cap F_-^n \subset F^n$$

and hence

$$P(F^{nc}) \leq P(F_J^{nc}) + P(F_-^{nc}).$$

The proof that F_-^n occurs with high probability is similar to that for F_J^n and we shall leave it to the reader.

To prove the finite summability of $P(F_J^{nc})$ we use 30 which implies that

$$P(F_J^{nc}) \leq P(F_0^{nc}) + P(F_1^{nc}) + P(F_2^{nc}).$$

Hence, we only need to prove high probability of the events F_0^n, F_1^n and F_2^n. The key to proving this is the following lemma:

Lemma 4.1 *Let x be a point. For every n, the total number of δ-paths of length n starting at x is less than*

$$2^{n(1+H(\delta)+2\delta)}$$

Define a (k, δ)-path of length kn to be a path comprising a total of kn steps of which fewer than δn are non-unit and for which the total variation of the non-unit steps is also less than δn. Clearly, all $(1, \delta)$-paths of length n are δ-paths of length n and conversely. Then, the number of $(10\delta, \delta)$-paths of length $10\delta n$ starting at x is less than

$$2^{10\delta n(1+H(0.1)+0.3)}.$$

Proof. First, if all steps are either $+1$ or -1, we have at most 2^n choices, since the path makes at most n jumps of length 1. Then among the n steps, we must consider which of them are not of unit length. There are at most δn such steps and, since δ is taken to be small (in particular $\delta < 1/2$), we shall have fewer than $\binom{n}{\delta n}$ possible arrangements of the non-unit jumps among the unit jumps. An argument involving Stirling's approximation reveals the bound $2^{H(\delta)n}$ on the number of possible arrangements, where the entropy $H(\delta)$ is given by $H(\delta) := -\delta \log_2 \delta - (1-\delta)\log_2(1-\delta)$.

Next we need to take into account the length of each non-unit step. We know that the total variation of the non-unit jumps is no more than δn. Hence, we can partition an interval of length at most δn to determine the lengths of the non-unit steps which are not of length 0. This gives a maximum of $2^{\delta n}$ possibilities. Because we can have steps of length zero, we must also determine which of the non-unit steps have length zero and this gives another $2^{\delta n}$ possibilities at most. Finally, we need to account for the signs of the non-unit steps. Since there are only two possible signs, There are at most $2^{\delta n}$ choices. Multiplying all the preceding bounds yields an upper bound on the maximum number of δ-paths of length n which start at x:

$$2^{n(1+H(\delta)+3\delta)}.$$

To compute a bound on the number of $(10\delta, \delta)$-paths of length $10\delta n$ starting at x, we proceed as before. Determining the steps of size $+1$ and -1, we have at most $2^{10\delta n}$ choices. Then we need to factor in the steps which are not of unit-size. There are at most δn of these, which corresponds to a proportion of no more than 0.1 the length of the path. Hence, we have at most $2^{10\delta n H(0.1)}$ choices. Finally, determining the length and sign of the non-unit steps as before gives a maximum of $2^{3\delta n}$ choices. Combining these bounds in a product, we find that there are less than

$$2^{10\delta n(1+H(0.1)+0.3)}.$$

$(10\delta, \delta)$-paths of length $10\delta n$ starting at x. ■

Proof that F_0^n occurs with high probability

Let x belong to the set

$$X := [-(2.45)^n, (2.45)^n] \setminus [2n - \delta n, 3n].$$

Now, $n + \delta n - 1$ is the maximum distance a δ-path can travel in n steps. So any δ-path of length n which visits x will stay outside $[-n, n]$. Thus, if $R : [0, n] \to \mathbb{Z}$ is a non-random δ-path such that $R(n) = x$, then $\xi \circ R$ is independent of w. To see this recall that The scenery is i.i.d. and so disjoint parts of it are independent of each other. note that the pattern w is a substring of $\xi_{-n}\xi_{-n+1}\ldots\xi_n$ while $\xi \circ R$ only depends on parts of the scenery

outside $[-n, n]$. Furthermore, the string w is i.i.d. with each of the five possible colors appearing with probability $1/5$, that is,

$$P(\xi \circ R = w) = (1/5)^n,$$

where R is a non-random δ-path of length n ending at x. Since there are at most $2^{n(1+H(\delta)+2\delta)}$ δ-paths of length n ending at x, we get

$$P(F_{0,x}^{nc}) \leq \left(\frac{2^{1+H(\delta)+2\delta}}{5}\right)^n \tag{32}$$

where $F_{0,x}^n$ designates the event that there exists no δ-path of length n for which $\xi \circ R = w$ and $R(n) = x$. Finally, F_0^n may be expressed as

$$F_0^n = \cap_{x \in X} F_{0,x}^n.$$

Hence,

$$\begin{aligned} P(F_0^{nc}) &\leq \sum_{x \in X} P(F_{0,x}^{nc}) \\ &\leq |X| \left(\frac{2^{1+H(\delta)+2\delta}}{5}\right)^n = 2\left(\frac{2.45 \cdot 2^{1+H(\delta)+2\delta}}{5}\right)^n, \end{aligned} \tag{33}$$

since X contains fewer than $2 \cdot 2.45^n$ elements. Now, $H(\delta) + 2\delta$ converges to zero as δ tends to zero. Hence, by taking $\delta > 0$ small enough, we see that

$$\frac{2.45 \cdot 2^{1+H(\delta)+2\delta}}{5} \tag{34}$$

can be made as close as we like to $2.45 \cdot 2/5 = 0.98 < 1$. Fixing $\delta > 0$ small enough so that expression 34 is strictly less than 1, we have that inequality 33 constitutes a negative exponential bound for $P(F_0^{nc})$.

Proof that F_1^n occurs with high probability

Let x be a non-random integer such that

$$x \in [n^* + \delta n, 3n]. \tag{35}$$

Assume also that $R : [0, n] \to \mathbb{N}$ is a non-random δ-path such that $R(n) = x$. Note that because R is a δ-path, $R(n-i)$ is never further from x than $\delta n + i$ and hence because of 35, we have $R(n-i) > n^* - i$. It follows that the $n-i$-th letter of the word w, which we denote by $w(n-i)$ and which is equal to ξ_{n^*-i}, is independent of

$$\xi(R(n-i))\xi(R(n-i+1)) \ldots \xi(R(n)).$$

This holds for all $i = 0, 1, \ldots, n$. Hence we find that if 35 holds with $R(n) = x$ for a non-random δ-path R, then

$$P(\xi \circ R = w) = (1/5)^n.$$

Next, we synthesize the proof that F_0^n occurs with high probability. First let $F_{1,x}^n$ be the event that there exists no δ-path of length n ending in x and generating w. Since the number of such paths, according to Lemma 4.1, is no greater than $2^{n(1+H(\delta)+2\delta)}$, we have

$$P(F_{1,x}^{nc}) \leq \left(\frac{2^{(1+H(\delta)+2\delta)}}{5}\right)^n.$$

Finally, there are fewer than $3n$ points x in the set $[n^* + \delta n, 3n]$ and hence the last inequality above yields

$$P(F_1^{nc}) \leq 3n \left(\frac{2^{(1+H(\delta)+2\delta)}}{5}\right)^n. \tag{36}$$

Once again choosing $\delta > 0$ small enough, we have

$$\frac{2^{(1+H(\delta)+2\delta)}}{5} < 1,$$

whence inequality 36 gives the desired negative exponential bound on $P(F_1^{nc})$.

Proof that F_2^n occurs with high probability.

Let $x \in [-2n - \delta n, n^* - 21\delta n]$. Assume that $R : [0, n] \to \mathbb{Z}$ is a δ-path such that $R(n) = x$. Note then that for all $i \leq 10\delta n$ we have

$$R(n - i) \leq x + i + \delta n \leq x + 10\delta n + \delta n < n^* - 10\delta n.$$

Hence

$$\xi(R(n - 10\delta n))\xi(R(n - 10\delta n + 1))\xi(R(n - 10\delta n + 2)) \ldots \xi(R(n))$$

is independent of

$$\xi_{n^* - 10\delta n}\xi_{n^* - 10\delta n + 1} \cdots \xi_{n^*}$$

which corresponds to the last $10\delta n$ letters of the word w. Hence, we get

$$P(\xi \circ R = w) \leq \left(\frac{1}{5}\right)^{10\delta n}.$$

By Lemma 4.1, there are no more than $2^{10\delta n(1+H(0.1)+0.3)}$ δ-paths of length 10δ ending in a given point x. Let $F_{2,x}^n$ be the event that there is no δ-path of length n such that $R(n) = x$ and $\xi \circ R = w$. Then

$$P(F_{2,x}^{nc}) \leq \left(\frac{2^{10\delta(1+H(0.1)+0.3)}}{5}\right)^n$$

for all $x \in [-2n - \delta n, n^* - 21\delta n]$. Since there are no more than $3n$ points in $[-2n - \delta n, n^* - 21\delta n]$, we find

$$P(F_2^{nc}) \leq 3n \left(\frac{2^{1+H(0.1)+0.3}}{5} \right)^{10\delta n}. \tag{37}$$

Now, $\frac{2^{1+H(0.1)+0.3)}}{5}$ is strictly less than 1 so expression 37 provides the desired exponential bound on $P(F_2^{nc})$.

4.5 The event G^n.

We shall use $\tilde{\mu}$ to denote the empirical distribution of

$$S_{\nu_1}, S_{\nu_2}, \ldots, S_{\nu_j},$$

where the random variable j is the largest i such that $\nu_i \leq T = 2.4^n$; that is,

$$\tilde{\mu}(x) := \tilde{\mu}(\{x\}) = \frac{1}{j} \sum_{i=0}^{j} \mathbf{1}_{S_{\nu_i} = x}.$$

Let G_1^n denote the event that the difference between $\tilde{\mu}$ and μ is less than or equal to $n^{3/2}/1.1^{n/2}$ in total variation:

$$G_1^n := \left\{ \sum_{x \in [-n,n]} |\tilde{\mu}(x) - \mu(x)| < \frac{n^{3/2}}{\sqrt{1.1}^n} \right\}.$$

Next, set $G_2^n := B^n \cap C^n \cap D^n \cap F^n$. Clearly, we can write $P(G_1^{nc}) \leq P(G_1^{nc} \mid G_2^n) + P(G_2^{nc})$. Since $P(G_2^{nc}) \leq P(B^{nc}) + P(C^{nc}) + P(D^{nc}) + P(F^{nc})$ and the 4 quantities on the right-hand side are finitely summable, $P(G_2^{nc})$ is itself finitely summable. Therefore, to show that $P(G_1^{nc})$ is finitely summable, it is enough to prove finite summability of $P(G_1^{nc} \mid G_2^n)$. Towards this end, observe that the event G_2^n implies that at least 1.1^n stopping times ν_i manifest by time T and that each of these times marks the end of a δ-path of length n in I whose final position lies in J. As a consequence, $\tilde{\mu}$ has support on $J \subset I$ where $|J| = 22\delta n + 1 < n < 2n + 1 = |I|$ for $\delta > 0$ sufficiently small. Now, $\sum_{x \in J} |\tilde{\mu}(x) - \mu(x)| \leq n \sup_{x \in J} |\tilde{\mu}(x) - \mu(x)|$. Conditioning on $j = k$, we have

$$\begin{aligned}
& P\left(\sum_{x \in J} |\tilde{\mu}(x) - \mu(x)| \geq \frac{n^{3/2}}{\sqrt{1.1}^n} \,\middle|\, \{j = k\} \cap G_2^n \right) \\
\leq\ & P\left(\sup_{x \in J} |\tilde{\mu}(x) - \mu(x)| \geq \frac{\sqrt{n}}{\sqrt{1.1}^n} \,\middle|\, \{j = k\} \cap G_2^n \right) \\
=\ & P\left(\cup_{x \in J} \{ |\tilde{\mu}(x) - \mu(x)| \geq \frac{\sqrt{n}}{\sqrt{1.1}^n} \} \,\middle|\, \{j = k\} \cap G_2^n \right) \\
\leq\ & \sum_{x \in J} P\left(abs\tilde{\mu}(x) - \mu(x) \geq \frac{\sqrt{n}}{\sqrt{1.05}^n} \,\middle|\, \{j = k\} \cap G_2^n \right).
\end{aligned} \tag{38}$$

Next, define $Y_0(x) = 0$ and $Y_i(x) = \sum_{k=1}^{i} X_i(x)$, for $i \in \mathbb{N}$, where $X_i(x) = \mathbf{1}_{S_{\nu_i}=x} - \mu(x)$. Now $Y_i(x)$ is a Martingale with increment size bounded by 1 and $Y_k(x) - Y_0(x) = k(\tilde{\mu}(x) - \mu(x)$. An application of the Azuma-Hoeffding inequality shows that

$$\begin{aligned} P\left(\tilde{\mu}(x) - \mu(x) \geq \frac{\sqrt{n}}{\sqrt{1.1}^n} \,\middle|\, \{j = k\} \cap G_2^n\right) &= P\left(Y_k(x) - Y_0(x) \geq \frac{k\sqrt{n}}{\sqrt{1.1}^n} \,\middle|\, \{j = k\} \cap G_2^n\right) \\ &\leq e^{-\frac{k^2 n}{2k \cdot 1.1^n}} \leq e^{-0.5nk/1.1^n}. \end{aligned}$$

Similarly, $P\left(\tilde{\mu}(x) - \mu(x) \leq -\frac{\sqrt{n}}{\sqrt{1.1}^n} \,\middle|\, \{j = k\} \cap G_2^n\right) \leq e^{-0.5nk/1.1^n}$ so that

$$P\left(|\tilde{\mu}(x) - \mu(x)| \geq \frac{\sqrt{n}}{\sqrt{1.1}^n} \,\middle|\, \{j = k\} \cap G_2^n\right) \leq 2e^{-0.5nk/1.1^n}.$$

Combining this bound with 38 then yields

$$P\left(\sum_{x \in J} |\tilde{\mu}(x) - \mu(x)| \geq \frac{n^{3/2}}{\sqrt{1.1}^n} \,\middle|\, \{j = k\} \cap G_2^n\right) \leq 2\,|J|\, e^{-0.5nk/1.1^n} < 2ne^{-0.5nk/1.1^n}.$$

Since C^n guarantees $j \geq 1.1^n$, k will belong to $[1.1^n, T]$ and we find that

$$\begin{aligned} P(G_1^{nc} \mid G_2^n) &= \sum_{k=1.1^n}^{T} P(G_1^{nc} \mid \{j = k\} \cap G_2^n) P(j = k \mid G_2^n) \\ &\leq \sum_{k=1.1^n}^{T} 2ne^{-0.5nk/1.1^n} P(j = k \mid G_2^n) \leq 2ne^{-n/2}, \end{aligned}$$

which is finitely summable.

Now, if it could be shown that $G_1^n \cap G_2^n \subset G^n$, then we would have $P(G^{nc}) \leq P(G_1^{nc}) + P(G_2^{nc})$. However, we have already shown that the two terms on the right-hand side are finitely summable, whence $P(G^{nc})$ would be finitely summable.

So, it remains to show that $G_1^n \cap G_2^n \subset G^n$ for n large enough. In order to do this, we need to show that

$$\frac{P_\mu(S_r = n+1) - P_\mu(S_r \notin [-n, n+1])}{2} \geq \left(\frac{1-\epsilon}{2}\right)^{90\delta n + 2}. \tag{39}$$

Recall that r is defined to be $90\delta n$. We shall take $\delta > 0$ small enough so that $(2/(1-\epsilon))^{90\delta} < \sqrt{1.1}$. Then for all n large enough, the bound on the right side of 39 is strictly larger than $n^{3/2}/1.1^{n/2}$. However, the event G_1^n guarantees $n^{3/2}/1.1^{n/2}$ as an upper bound on the difference in total variation between the measure $\tilde{\mu}$ and μ. In turn, the difference in total variation between $\tilde{\mu}$ and μ bounds the difference between $P_{\tilde{\mu}}(\chi_r = e)$ and $P_\mu(\chi_r = e)$ for every $e \in \{1, 2, 3, 4, 5\}$. Thus, we have:

$$|P_{\tilde{\mu}}(\chi_r = e) - P_\mu(\chi_r = e)| \leq \sum_x |\tilde{\mu}(x) - \mu(x)| \leq \frac{n^{3/2}}{\sqrt{1.1}^n}. \tag{40}$$

Note that $\tilde{P}_\mu(\chi_r = e) = P_{\tilde{\mu}}(\chi_r = e)$. Therefore, if (39) holds, $\left|\tilde{P}_\mu(\chi_r = e) - P_\mu(\chi_r = e)\right|$ will be bounded by the quantity in (21) for n large enough and hence G^n will hold.

To finish, we return to the question of (39) holding. Since G_2^n holds, the support of μ is in the interval J. Let us first assume that the random walk S is aperiodic and symmetric with the increment distribution given in (11). We will deal with the more general case later. Now, S is able to reach the point $n+1$ from any point of J in r steps with probability greater than or equal to $\left(\frac{1-\epsilon}{2}\right)^r = \left(\frac{1-\epsilon}{2}\right)^{90\delta n}$. The minimum distance from any point $x \in J$ to the point $n+1$ is $60\delta n$. Using (12) with $b = 60\delta n$ and $q = 1.5$, we calculate the decay in the tail of the state probability distribution beyond a distance of $60\delta n$ from x after $90\delta n$ steps to be

$$\rho(90\delta n, 60\delta n) = \frac{P(S_r = n+2)}{P(S_r = n+1)} < \frac{1.5-1}{1.5+1} = \frac{1}{5}.$$

Hence, for any point $x \in J$, we have

$$\begin{aligned} P_x(S_r = n+1) - P_x(S_r > n+1) - P_x(S_r < -n) &\geq (1 - 0.25 - 0.25)P_x(S_r = n+1) \\ &= 0.5P_x(S_r = n+1), \end{aligned}$$

but as mentioned above, $P_x(S_r = n+1) \geq ((1-\epsilon)/2)^r = ((1-\epsilon)/2)^{90\delta n}$ and hence we obtain

$$P_x(S_r = n+1) - P_x(S_r \notin [-n, n+1]) \geq \left(\frac{1-\epsilon}{2}\right)^{90\delta n+1}.$$

Here $P_x(\cdot)$ refers to the distribution when S is started at the point x. As the event G_2^n holds, the measure μhas support on the interval J and so

$$P_\mu(S_r = n+1 \mid F^n) - P_\mu(S_r \notin [-n, n+1] \mid F^n) \geq \left(\frac{1-\epsilon}{2}\right)^{90\delta n+1}.$$

Dividing both sides by two then yields (39).

While the veracity of (39) has been discussed assuming the aperiodic, symmetric random walk of Section 2.2, it is not too difficult to show that it holds for any random walk satisfying Conditions (1)– (4) given G_2^n, provided we select the parameters of the random walk appropriately. In other words, for large n, $G_1^n \cap G_2^n \subset G^n$ when $\epsilon > 0$ is sufficiently small and $c > 0$ is sufficiently large.

5 Reconstruction of the whole scenery

To conclude, we prove that we can reconstruct the whole scenery ξ a.s. up to equivalence. for this we use the following lemma which was proved in [17].

Lemma 5.1 *Assume that there exists an algorithm which reconstructs ξ with probability strictly greater than $1/2$. Then we can also reconstruct ξ with probability one. More precisely, if there exists a measurable map*

$$\mathcal{A}^* : \{1,2,3,4,5\}^{\mathbb{N}} \to \{1,2,3,4,5\}^{\mathbb{Z}}$$

such that

$$P(\mathcal{A}^*(\xi) \approx \xi) > 1/2,$$

then there exists a measurable map

$$\mathcal{A} : \{1,2,3,4,5\}^{\mathbb{N}} \to \{1,2,3,4,5\}^{\mathbb{Z}}$$

such that

$$P(\mathcal{A}(\xi) \approx \xi) = 1.$$

So, we need to build a map which reconstructs ξ with probability strictly greater than $1/2$. Due to inequality 5, there exists a non-random n_0 such that

$$P\left(\hat{\xi}(n+1) = \xi(n+1), \hat{\xi}(-n-1) = \xi(-n-1), \forall n \geq n_0\right) \geq \frac{4}{5}. \tag{41}$$

We tune the parameter $\epsilon > 0$ small enough so that the random walk S only makes ± 1 steps for a long time. It is known that we can reconstruct a finite piece of ξ close to the origin (see [17]) with probability as close to one as we like when we are dealing with a simple random walk on a five-color scenery provided we have enough observations. So by taking $\epsilon > 0$ small enough we can ensure that the random walk follows a simple random walk path long enough to reconstruct the finite piece $\xi|_{[-n_0,n_0]}$ with probability larger than $\frac{4}{5}$. So, we start by reconstructing the finite piece $\xi|_{[-n_0,n_0]}$. Then we proceed inductively in n for $n \geq n_0$. Once we have reconstructed $\xi|_{[-n,n]}$ we estimate $\xi(n+1)$ and $\xi(-n-1)$ using the algorithm described in Section 3.2. In this way we end up reconstructing ξ correctly with probability at least $3/5$ which is strictly larger than $1/2$. This establishes that we can reconstruct ξ with probability one and thus completes the proof of Theorem 1.1.

One more remark: Since in general one can not reconstruct the scenery but only reconstruct it up to equivalence, it follows that we can not reconstruct $\xi_{|[-n,n]}$ exactly. Instead, from the result of matzinger [17], we have that we can only perform reconstruction successfully with high probability on an approximately centered interval $\xi_{|[-n+l,n+l]}$ where l is small compared to n, but not known to us. Note however that our one-point reconstruction algorithm works if we are given the string $\xi(-n+l)\xi(-n+l+1)\ldots\xi(n+l)$ without knowing its exact position (i.e. not knowing l), but only that l is of order smaller than n.

Acknowledgements

The first and third author would like to thank the Millennium Nucleus in Information and Randomness P04-069-F, MSI, as well as SFB701 in Bielefeld for supporting this work. The third author would like to also thank FAPESP for supporting his visit to the IME-USP during which time an important part of this article was written.

References

[1] F. den Hollander and J. E. Steif. Mixing properties of the generalized T, T^{-1}-process. *J. Anal. Math.*, 72:165–202, 1997.

[2] W. Th. F. den Hollander. Mixing properties for random walk in random scenery. *Ann. Probab.*, 16(4):1788–1802, 1988.

[3] M. Harris and M. Keane. Random coin tossing. *Probab. Theory Related Fields*, 109(1):27–37, 1997.

[4] D. Heicklen, C. Hoffman, and D. J. Rudolph. Entropy and dyadic equivalence of random walks on a random scenery. *Adv. Math.*, 156(2):157–179, 2000.

[5] C. D. Howard. Detecting defects in periodic scenery by random walks on $\mathbb{Z}$. *Random Structures Algorithms*, 8(1):59–74, 1996.

[6] C. D. Howard. Orthogonality of measures induced by random walks with scenery. *Combin. Probab. Comput.*, 5(3):247–256, 1996.

[7] C. D. Howard. Distinguishing certain random sceneries on $\mathbb{Z}$ via random walks. *Statist. Probab. Lett.*, 34(2):123–132, 1997.

[8] S. A. Kalikow. T, T^{-1} transformation is not loosely Bernoulli. *Ann. of Math. (2)*, 115(2):393–409, 1982.

[9] M. Keane and W. Th. F. den Hollander. Ergodic properties of color records. *Phys. A*, 138(1-2):183–193, 1986.

[10] H. Kesten. Detecting a single defect in a scenery by observing the scenery along a random walk path. In *Itô's stochastic calculus and probability theory*, pages 171–183. Springer, Tokyo, 1996.

[11] H. Kesten. Distinguishing and reconstructing sceneries from observations along random walk paths. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, pages 75–83. Amer. Math. Soc., Providence, RI, 1998.

[12] H. Kesten and F. Spitzer. A limit theorem related to a new class of self-similar processes. *Z. Wahrsch. Verw. Gebiete*, 50(1):5–25, 1979.

[13] J. Lember and H. Matzinger. Information recovery from a randomly mixed up message-text. *Electronic Journal of Probability*, 13:396 –466, 2008.

[14] D. A. Levin, R. Pemantle, and Y. Peres. A phase transition in random coin tossing. *Ann. Probab.*, 29(4):1637–1669, 2001.

[15] E. Lindenstrauss. Indistinguishable sceneries. *Random Structures Algorithms*, 14(1):71–86, 1999.

[16] M. Löwe and H. Matzinger. Scenery reconstruction in two dimensions with many colors. *Ann. Appl. Probab.*, 12(4):1322–1347, 2002.

[17] M. Löwe, H. Matzinger, and F. Merkl. Reconstructing a multicolor random scenery seen along a random walk path with bounded jumps. *Electronic Journal of Probability*, 15:436 – 507, 2004.

[18] H. Matzinger. *Reconstructing a 2-color scenery by observing it along a simple random walk path with holding.* PhD thesis, Cornell University, 1999.

[19] H. Matzinger. Reconstructing a three-color scenery by observing it along a simple random walk path. *Random Structures Algorithms*, 15(2):196–207, 1999.

[20] H. Matzinger. Reconstructing a 2-color scenery by observing it along a simple random walk path. *Ann. Appl. Probab.*, 1B:778 – 819, 2005.

[21] H. Matzinger and S. W. W. Rolles. Retrieving random media. *Probab. Theory Related Fields*, 136(6):469 – 507, 2006.

[22] H. Matzinger and S.W.W. Rolles. Reconstructing a piece of scenery with polynomially many observations. *Stochastic Processes and their Applications*, 107(2):289–300, 2003.

[23] H. Matzinger and S.W.W. Rolles. Reconstructing a random scenery observed with random errors along a random walk path. *Probab. Theory Related Fields*, 125(4):539 – 577, 2003.