ISRAEL JOURNAL OF MATHEMATICS **142** (2004), 205–220

RANDOM WALKS IN RANDOM ENVIRONMENT: WHAT A SINGLE TRAJECTORY TELLS

BY

OMER ADELMAN AND NATHANAËL ENRIQUEZ

Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 6, 4 place Jussieu, 75252 Paris cedex 05, France
e-mail: adelman@ccr.jussieu.fr, enriquez@ccr.jussieu.fr

ABSTRACT

We present a procedure that determines the law of a random walk in an iid random environment as a function of a single "typical" trajectory. We indicate when the trajectory characterizes the law of the environment, and we say how this law can be determined. We then show how independent trajectories having the distribution of the original walk can be generated as functions of the single observed trajectory.

1. Introduction

Suppose you are given a "typical" trajectory of a random walk in an iid random environment. Can you say what the law of the environment is on the basis of the information supplied by this single trajectory? Can you determine the law of the walk? Such questions may arise if one intends to use the random environment model in applications.

These questions are essentially pointless if the group is finite (in which case the environment at each of the finitely many sites that happen to be visited infinitely many times can of course be determined, but it is hard to say much more). So we assume that the group is infinite, and we go a little further: we assume that the (random) set of sites visited by the walk is almost surely infinite. (See Remark 5.1.)

Questions of this kind have been studied in the context of random walks in random scenery by Benjamini and Kesten [1], Löwe and Matzinger [3], and Matzinger [7].

Received May 1, 2003 and in revised form December 17, 2003

In the case of an iid random environment, the information furnished by a single "typical" trajectory tells us whether the walk is recurrent; indeed, one can show that one of the events {each visited site is visited infinitely many times}, {no site is visited infinitely many times} is an almost sure event. (Cf. Kalikow [4].)

Now, if the walk is recurrent, the problem is quite simple: we can know much more than the *law* of the environment, because we find the environment *itself* at each visited site, which is given by the frequency of each possible jump from this site. In the transient case, the "naïve" approach consisting of doing statistics on sites which have been visited many times cannot be utilized directly, since the assumption of being at a site which has been frequently visited introduces a bias on the environment at that site, which should encourage jumps to sites from which it is easier to come back (loosely speaking, close sites). (See Example 1 in §4.1.)

We present a procedure that eliminates any source of bias, collecting information on sites displaying some specified "histories". Each such "history" which *can* be encountered *is* almost surely encountered infinitely often (Proposition 3). This is combined with an interpretation of the process as a transition reinforced random walk (cf. Enriquez and Sabot [2]), allowing us to find the exact law of the process. Now, there may exist "bad" transitions: if the walk jumps from a site along a "bad" transition, it will never get back to that site. If the set of these "bad" transitions is empty or has just one element, we can find the distribution of the environment (Theorem 1).

Finally, we show how countably many independent trajectories having the distribution of the original walk can be generated by concatenating steps of the observed trajectory. The algorithm is purely "mechanical": it does not imply any computation, and, in particular, the knowledge of the law of the walk (or of the environment) is not needed.

2. Framework and notations

Our "canonical" process $X := (X_n)_{n\geq 0}$ walks on a group G. We denote by $(\mathcal{F}_n)_{n\geq 0}$ the natural filtration of X.

We assume that the group G is Abelian, although this is *never* used in our arguments. Its only utility is in the possibility of writing things like $x - y = e$ or $x = y + e$ indifferently.

We use the additive notation, and the identity element of G is denoted by 0.

We assume, moreover, that the group G is countable. This assumption can be dispensed with — see Remark 5.2 — but we feel that it renders the reading easier and the discussion more tractable. It does not affect the core of the argument.

We denote by $\mathbb{N}$ the set $\{1, 2, \ldots\}$.

2.1. Random walks in random environment. We denote by $\mathcal{P}$ the set of non-negative families $p := (p_e)_{e\in G}$ such that $\sum_{e\in G} p_e = 1$. The environment at the site x, $\nu(x) := (\nu(x,e))_{e\in G}$, is a random element of $\mathcal{P}$. We assume that the environments at sites are iid $\mathcal{P}$-valued random variables with common distribution μ.

We let $\nu_e := \nu(0,e)$.

The random environment $\nu := (\nu(x))_{x\in G}$ is a random element of $\mathcal{P}^G$, and it is governed by the probability measure $\mu^{\otimes G}$.

For all $\pi = ((\pi(x,e))_{e\in G})_{x\in G} \in \mathcal{P}^G$, let Q_π be the probability measure under which X is a G-valued Markov chain started at 0 whose transition probability from x to $x+e$ is $\pi(x,e)$ $(x, e \in G)$.

The law of the random walk in random environment (or the so-called *"annealed"* law) is the probability measure $\mathbf{P}^\mu = \int Q_\pi \mu^{\otimes G}(d\pi)$ $(= \mathbf{E}[Q_\nu]$, Q_ν being what is usually called the "quenched" law).

We recall our "infinitude assumption" according to which the (random) set of sites visited by the walk, $S := \{X_n |\, n \geq 0\}$, is $\mathbf{P}^\mu$-almost surely infinite. (This implies, of course, that the group G itself is infinite.)

We let E denote the set of those $g \in G$ such that the probability of the event $\{\nu_g > 0\}$ is strictly positive. (The random set $\{X_{n+1} - X_n |\, n \geq 0\}$ is easily seen to be $\mathbf{P}^\mu$-almost surely exactly E.) We then partition E into two sets, R and T, defined as follows.

- R is the set of elements r of E that can be written as $-(e_1 + \cdots + e_n)$ where $(e_i)_{1\leq i\leq n}$ is a finite nonempty sequence of elements of E. It is easy to see that $r \in R$ if and only if $\mathbf{P}^\mu(X_1 = r$ and, for some $n > 1$, $X_n = 0)$ > 0; and Proposition 3 below implies that if $r \in R$, then the random set $\{n |\, X_{n+1} = X_n + r$ and, for some $k > 0$, $X_{n+k} = X_n\}$ is $\mathbf{P}^\mu$-almost surely infinite. It is therefore quite easy to identify R when observing a single trajectory.
- T is the complement of R in E. It represents the "possible" transitions which do not allow a return to the original site.

2.2. Histories. We start with some definitions.

Definition 1: The **history of the site x at time n**, which we denote by $H(n,x)$, is the random finite sequence of elements of G defined by the successive moves of the process from the site x before time n. More formally, $H(0,x)$ is the empty sequence (); $H(n+1,x) = H(n,x)$ if $X_n \neq x$; and, if $X_n = x$, then $H(n+1,x)$ is the finite sequence obtained by adjoining $X_{n+1} - X_n$ as a new rightmost term to $H(n,x)$.

Let us denote by $(\mathbb{Z}_+^G)_0$ the set of families $(n_g)_{g\in G} \in \mathbb{Z}_+^G$ with a finite number of non null terms.

Definition 2: The **unordered history of the site x at time n**, denoted by $\vec{N}(n,x) := (N_g(n,x))_{g\in G}$, is a random element of $(\mathbb{Z}_+^G)_0$ where, for all $g \in G$, $N_g(n,x)$ is the random number of moves from the site x to $x+g$ before time n. In other words, $N_g(n,x) = \sum_{l=0}^{n-1} 1_{\{X_l = x,\, X_{l+1} - X_l = g\}}$.

Also, the **local unordered history at time n** is the unordered history of the site X_n at time n, $\vec{N}(n) := \vec{N}(n, X_n)$.

2.3. Reinforced random walks. A transition reinforced random walk consists of a discrete random process whose transition probabilities from the currently occupied site are functions of the number of past moves of each type from that site.

Definition 3: A **reinforcement function** is a function

$$V\colon (\mathbb{Z}_+^G)_0 \to \mathcal{P},$$
$$\vec{n} = (n_g)_{g\in G} \mapsto V(\vec{n}) := (V_e(\vec{n}))_{e\in G}.$$

Definition 4: We call a **transition reinforced random walk with reinforcement function** V the random walk defined by the law $\mathbf{P}^V$ on the trajectories starting at 0 given by

$$\mathbf{P}^V(X_{n+1} - X_n = e \mid \mathcal{F}_n) = V_e(\vec{N}(n)).$$

Keane and Rolles [5] considered a special kind of reinforcement (see section 4.3 below). Pemantle [8] considered an essentially equivalent process. (While Keane and Rolles deal with reinforcement of oriented edges of graphs, Pemantle studies reinforcement of non-oriented edges of trees. It can be shown that the replacement of each edge in Pemantle's model by a couple of oriented edges with opposite orientations and with appropriate initial weights results in a process indistiguishable from the original one.)

3. Tools

3.1. RWRE AS A TRANSITION REINFORCED RANDOM WALK. Viewing some specific transition reinforced random walks as random walks in random environment is already present in Pemantle [8] and in Keane and Rolles [5].

In order to get a non-biased procedure of reconstitution of the environment, it is useful to adopt an alternative philosophy, viewing random walks in random environment as transition reinforced random walks. This is the essence of the easy part of the result of Enriquez and Sabot [2] which we now state. (The other part of [2] gives the conditions on a reinforced random walk to correspond to a RWRE. Note that what we call here *transition reinforced random walk* is called *edge-oriented reinforced random walk* in [2].)

PROPOSITION 1: *The annealed law $\mathbf{P}^\mu$ of the RWRE coincides with the law $\mathbf{P}^V$ of the reinforced random walk whose reinforcement function V satisfies, for all $e \in G$,*

$$V_e(\vec{n}) = \frac{\mathbf{E}[\nu_e \prod_{g\in G} \nu_g^{n_g}]}{\mathbf{E}[\prod_{g\in G} \nu_g^{n_g}]}$$

whenever $\vec{n} \in (\mathbb{Z}_+^G)_0$ such that $\mathbf{E}[\prod_{g\in G} \nu_g^{n_g}] > 0$.

In order to be self-contained we recall the proof of this proposition.

Proof: For every x and e in G, for all $n \in \mathbb{N}$, $\mathbf{P}^\mu$-almost everywhere on the event $\{X_n = x\}$,

$$\mathbf{P}^\mu(X_{n+1} = x + e | \mathcal{F}_n) = \frac{\mathbf{E}[\nu(x,e) \prod_{y\in G}\prod_{g\in G} \nu(y,g)^{n_g(n,y)}]}{\mathbf{E}[\prod_{y\in G}\prod_{g\in G} \nu(y,g)^{n_g(n,y)}]}.$$

Now using the independence of the random variables $\nu(y)$ for different sites y, the terms depending on $\nu(y)$ for $y \neq x$ cancel in the previous ratio, and we get the result. ∎

The following result is an analogue of the strong Markov property for reinforced random walks.

PROPOSITION 2: *Let X be a reinforced random walk with reinforcement function V, and let T be a stopping time with respect to the natural filtration of X. Assume T is almost surely finite. Then*

$$\mathbf{P}^V(X_{T+1} - X_T = e | \mathcal{F}_T) = V_e(\vec{N}(T)) \quad \mathbf{P}^V\text{-a.s.}$$

The proof is obtained in an obvious way, by considering the events $\{T = n\}$.

3.2. A ZERO-ONE RESULT. The following zero-one result happens to be quite useful.

PROPOSITION 3: *Let $(r_1, \ldots, r_l)$ be a finite (eventually empty) sequence of elements of R. Let $S_{(r_1,\ldots,r_l)}$ be the random set*

$$\{x \in G \mid \exists n \geq 0,\ H(n,x) = (r_1, \ldots, r_l)\}.$$

Then $S_{(r_1,\ldots,r_l)}$ is either $\mathbf{P}^\mu$-almost surely empty or $\mathbf{P}^\mu$-almost surely infinite.

Proof: Suppose that $S_{(r_1,\ldots,r_l)}$ is not $\mathbf{P}^\mu$-almost surely empty.

This implies that there exists a list of transitions

$$L := (r_1, e_{1,1}, e_{1,2}, \ldots, e_{1,k_1}, r_2, e_{2,1}, \ldots, e_{2,k_2}, r_3, \ldots.., e_{l-1,k_{l-1}}, r_l)$$

such that, for all $m \in \{1, \ldots, l-1\}$,

$$r_m + \sum_{i=1}^{k_m} e_{m,i} = 0 \quad \text{and, for all } k \in \{1, \ldots, k_m - 1\}, \quad r_m + \sum_{i=1}^{k} e_{m,i} \neq 0$$

and

$$\gamma := \mathbf{E}\Bigg[\prod_{k=1}^{l} \nu_{r_k}\Bigg] \mathbf{E}\Bigg[\prod_{m=1}^{l-1} \prod_{i=1}^{k_m - 1} \nu(r_m + e_{m,1} + \cdots + e_{m,i}, e_{m,i+1})\Bigg] > 0.$$

(Note that if $r_m = 0$, then $k_m = 0$.)

Let $q := l + k_1 + \cdots + k_{l-1}$ be the length of the list L.

For all $k \in \{1, \ldots, q\}$, denote by y_k the k-th term of the list L and, for all $k \in \{1, \ldots, q+1\}$, set $x_k := \sum_{i=1}^{k-1} y_i$. ($x_1 = 0$.)

Now consider the list $(g_1, g_2, \ldots)$ of newly visited sites in their order of appearance. So $\mathcal{S} = \{g_1, g_2, \ldots\}$. By the assumption made in the introduction, $\mathcal{S}$ is almost surely infinite.

To any integer $n \geq 1$, we associate a random integer $i(n)$ defined by

$$i(n) := \min\{i \geq 1 \mid \exists k \in \{1, \ldots, q\}, g_i = g_n + x_k\}.$$

We denote by $k(n)$ the random smallest integer $m \geq 1$ such that $g_{i(n)} = g_n + x_m$. (Obviously, $k(n) \leq q$.) The sequence $(g_{i(n)})_{n \geq 1}$ takes infinitely many values (since the infinite set $\mathcal{S}$ is included in $\{g_{i(1)}, g_{i(2)}, \ldots\} - \{x_1, \ldots, x_q\}$). Now, for any $i \geq 1$, we denote by T_i the hitting time of g_i by the walk. By definition of $i(n)$, none of the sites $g_{i(n)} - x_{k(n)} + x_j \quad (1 \leq j \leq q)$ is visited by the trajectory before time $T_{i(n)}$.

As a result, there exist infinitely many sites $g'_1, g'_2, \dots$ visited by the trajectory (enumerated in their order of appearance) such that, for some $k \in \{1, \dots, q\}$, if T'_n denotes the hitting time of g'_n, then none of the sites $g'_n - x_k + x_j \quad (1 \le j \le q)$ is visited by the trajectory before time T'_n. We denote by $k'(n)$ the least such integer k. T'_n are clearly stopping times.

For all $n \ge 1$, let ψ_n be the Bernoulli variable that equals 1 if and only if

$$X_{T'_n+i} = \begin{cases} g'_n - x_{k'(n)} + x_{i+k'(n)} & \text{if } 1 \le i \le q - k'(n), \\ g'_n - x_{k'(n)} + x_{i-k'(n)-q} & \text{if } q - k'(n) < i \le 2q - k'(n) + 1. \end{cases}$$

(Otherwise, ψ_n equals 0.)

Observe that for all n, if $\psi_n = 1$, then $X_{T'_n} - x_{k'(n)} \in S_{(r_1,\dots,r_l)}$.

Due to the fact that the prescribed path the process has to follow during the time period $[T'_n, T'_n + 2q - k'(n) + 1]$ in order to satisfy $\psi_n = 1$ is a path that does not intersect the trajectory before T'_n,

$$\mathbf{P}(\psi_n = 1 | \mathcal{F}_{T'_n}) \ge \mathbf{E}\Bigg[\prod_{k=1}^{q} (\nu(x_k, y_k))^2\Bigg] \ge \mathbf{E}\Bigg[\prod_{k=1}^{q} \nu(x_k, y_k)\Bigg]^2 = \gamma^2.$$

Let $\xi_n := \psi_{(2q+1)n}$ and $\tau_n = T'_{(2q+1)n}$ $(n \ge 1)$.

For all n, $\xi_1, \dots, \xi_n$ are measurable with respect to $\mathcal{F}_{\tau_{n+1}}$.

It is now obvious that for all $n, k \ge 1$

$$\begin{aligned} \mathbf{P}(\xi_{n+1} = \dots = \xi_{n+k} = 0) =& \mathbf{P}(\xi_{n+1} = 0) \times \mathbf{P}(\xi_{n+2} = 0 | \xi_{n+1} = 0) \times \cdots \\ & \cdots \times \mathbf{P}(\xi_{n+k} = 0 | \xi_{n+1} = \dots = \xi_{n+k-1} = 0) \\ \le& (1 - \gamma^2)^k. \end{aligned}$$

Therefore, almost surely, infinitely many of the ψ_n's are equal to 1. This implies that $S_{(r_1,\dots,r_l)}$ is almost surely infinite. ∎

Remark: If G equals $\mathbb{Z}^d$, the notion of convexity can be exploited in a proof slightly different from the one given above.

We deduce that the sets R and T can be "viewed" on the trajectory:

COROLLARY 1: *The sets R and T are such that*

$$R \overset{a.s}{=} \{g \in E | \ S_{(g)} \text{ is infinite}\} \overset{a.s}{=} \{g \in E | \ S_{(g)} \neq \emptyset\} \quad \text{and} \quad T \overset{a.s}{=} \{g \in E | \ S_{(g)} = \emptyset\}.$$

Let $S_{\vec{n}}$ denote the random set $\{x \in G | \ \exists n \ge 0, \vec{N}(n, x) = \vec{n}\}$. An easy corollary of the above proposition is the following analogous result concerning unordered histories.

PROPOSITION 3′: *If $\vec{n} \in (\mathbb{Z}_+^G)_0$, then $S_{\vec{n}}$ is either $\mathbf{P}^\mu$-almost surely empty or $\mathbf{P}^\mu$-almost surely infinite.*

We now distinguish a particular subset of $(\mathbb{Z}_+^G)_0$,

$$\mathcal{N}_{poss} := \{\vec{n} \in (\mathbb{Z}_+^G)_0 |\ S_{\vec{n}} \neq \emptyset \quad \mathbf{P}^\mu\text{-a.s.}\}.$$

Loosely speaking, $\mathcal{N}_{poss}$ is the set of "possible" unordered histories for sites that are presently occupied. An element $\vec{n} = (n_g)_{g\in G}$ of $(\mathbb{Z}_+^G)_0$ belongs to $\mathcal{N}_{poss}$ if $n_g = 0$ whenever $g \notin R$ and if, moreover, it satisfies (any one of) the three following equivalent conditions:

(a) $\mathbf{P}^\mu(\exists n \geq 0 |\ \vec{N}(n,0) = \vec{n}) > 0$;

(b) the random set $S_{\vec{n}}$ is almost surely infinite;

(c) $\mathbf{E}[\prod_{r\in R} \nu_r^{n_r}] > 0$.

Note that, by (c) and Proposition 1, the law of the process is determined by the restriction of the reinforcement function to the set $\mathcal{N}_{poss}$.

4. What a single trajectory tells

In the sequel we assume that the law of the process X is $\mathbf{P}^\mu$ (or, equivalently, $\mathbf{P}^V$).

4.1. DETERMINING THE LAW OF THE WALK. As noted in the introduction, "straightforward" computation based on the frequencies of transitions from sites visited "many" times is not reliable. The following example is an illustration of this fact in the context of a deterministic environment.

Example 1: The process X is a nearest-neighbour random walk on $\mathbb{Z}$: for some fixed $p \in]0,1[$, for all $x \in \mathbb{Z}$, $\nu(x,1) = p$ and $\nu(x,-1) = 1-p$. One might think that if $n \in \mathbb{N}$ is very large and x is some fixed site in $\mathbb{Z}$ then, conditionally on the event $\{x$ is visited at least n times$\}$, the proportion of the transitions from x to $x+1$ among the first n transitions from x is likely to be close to $\nu(x,1)$ (i.e., to p); but this is far from being the case. Indeed, conditionally on the event $\{x$ is visited at least n times$\}$, the first $n-1$ transitions from x are easily seen to be iid and uniformly distributed on $\{-1,1\}$, so the proportion mentioned above is likely to be close to $\frac{1}{2}$, and this is the case *independently of the value of* p.

What we do here, instead, is collecting information on sites displaying some specified histories (or specified *unordered* histories).

For any $\vec{n} \in \mathcal{N}_{poss}$, let $T_i^{\vec{n}}$ $(i \geq 1)$ be the successive times where the unordered history of the currently occupied site is $\vec{n}$:

- $T_0^{\vec{n}} := \inf\{k \geq 0 | \vec{N}(k) = \vec{n}\}$,
- $\forall i \geq 0, T_{i+1}^{\vec{n}} := \inf\{k > T_i^{\vec{n}} | \vec{N}(k) = \vec{n}\}$.

(We ignore the case (happening on the negligible event $\{S_{\vec{n}}$ is finite$\}$) where some $T_i^{\vec{n}}$ is infinite.)

PROPOSITION 4: *The E-valued random variables $\Delta_i^{\vec{n}} := X_{T_i^{\vec{n}}+1} - X_{T_i^{\vec{n}}}$ ($i \geq 1$, $\vec{n} \in \mathcal{N}_{poss}$) are independent. Also, for all $\vec{n} \in \mathcal{N}_{poss}$, the random variables $\Delta_i^{\vec{n}}$ have the same law:*

$$\forall e \in E, \quad \mathbf{P}(\Delta_i^{\vec{n}} = e) = V_e(\vec{n}).$$

Proof: Let $\Theta_i^{\vec{n}}$ ($i \geq 1$, $\vec{n} \in (\mathbb{Z}_+^G)_0$) be independent random variables such that for all $e \in E$,

$$\mathbf{P}(\Theta_i^{\vec{n}} = e) = V_e(\vec{n}).$$

Now we consider the process $(Y_n)_{n\geq 0}$:

$$Y_0 = 0, \quad Y_{n+1} = Y_n + \Theta_{\tau(n)}^{\vec{N}_y(n)},$$

where $\vec{N}_y$ is to the process Y what $\vec{N}$ is to the process X, and

$$\tau(n) := \mathrm{card}\{j \in \{0, \ldots, n\} | \vec{N}_y(j) = \vec{N}_y(n)\}.$$

We have

$$\begin{aligned}
\mathbf{P}(Y_{n+1} &= Y_n + e | \sigma(Y_k, k \leq n)) \\
&= \mathbf{P}(\Theta_{\tau(n)}^{\vec{N}_y(n)} = e | \sigma(Y_k, k \leq n)) \\
&= \mathbf{E}\Bigg[\sum_{l \geq 0, \vec{m} \in (\mathbb{Z}_+^G)_0} 1_{\tau(n)=l, \vec{N}_y(n)=\vec{m}} 1_{\Theta_l^{\vec{m}}=e} | \sigma(Y_k, k \leq n)\Bigg].
\end{aligned}$$

But on the event $A_{l,\vec{m},n} := \{\tau(n) = l, \vec{N}_y(n) = \vec{m}\}$, $\Theta_l^{\vec{m}}$ is independent of $\sigma(Y_k, k \leq n)$. Thus,

$$\begin{aligned}
\mathbf{P}(Y_{n+1} = Y_n + e | \sigma(Y_k, k \leq n)) &= \sum_{l \geq 0, \vec{m} \in (\mathbb{Z}_+^G)_0} 1_{\tau(n)=l, \vec{N}_y(n)=\vec{m}} \mathbf{E}[1_{\Theta_l^{\vec{m}}=e}] \\
&= \sum_{l \geq 0, \vec{m} \in (\mathbb{Z}_+^G)_0} 1_{\tau(n)=l, \vec{N}_y(n)=\vec{m}} V_e(\vec{m}) \\
&= V_e(\vec{N}_y(n)).
\end{aligned}$$

Consequently, the two processes X and Y have the same law. But $\Delta_i^{\vec{n}}$ is to the process $(X_n)_{n\geq 0}$ exactly what $\Theta_i^{\vec{n}}$ is to the process Y. The result follows. ∎

We deduce from this proposition the following corollary, which describes the construction of the reinforcement function on $\mathcal{N}_{poss}$ or, equivalently (by Proposition 1), of the annealed law:

COROLLARY 2: *If $\vec{n} \in \mathcal{N}_{poss}$, then almost surely, for all $e \in E$,*

$$\frac{1_{\Delta_1^{\vec{n}}=e} + \cdots + 1_{\Delta_m^{\vec{n}}=e}}{m} \to V_e(\vec{n}) \quad \text{as } m \to \infty.$$

4.2. THE LAW OF THE ENVIRONMENT. The next result follows easily.

THEOREM 1: (a) *A single trajectory determines almost surely the moments of the form $\mathbf{E}_\mu[\nu_{r_1}^{n_1} \cdots \nu_{r_k}^{n_k} \nu_t^\varepsilon]$ for all $r_1, \ldots, r_k \in R$, $t \in T$, $n_1, \ldots, n_k \in \mathbb{Z}_+$, $\varepsilon = 0$ or 1. Moreover, if these moments coincide for two distinct environment distributions, the induced RWRE have the same annealed law (and, consequently, two such environment distributions cannot be distinguished).*

(b) *If $\operatorname{card} T = 0$ or 1, a single trajectory determines almost surely the distribution of the environment.*

Proof: (a) By Corollary 2, the restriction of V to $\mathcal{N}_{poss}$ is almost surely determined by a single trajectory. So, for all $\vec{n} = (n_g)_{g \in G} \in \mathcal{N}_{poss}$ and for all $e \in G$, a single trajectory determines the moments $\mathbf{E}[(\prod_{g \in G} \nu_g^{n_g}) \cdot \nu_e]$ almost surely. Moreover, all the other moments of the type $\mathbf{E}_\mu[\nu_{r_1}^{n_1} \cdots \nu_{r_k}^{n_k} \nu_t^\varepsilon]$ ($r_j \in R$, $t \in T$) are zero. Finally, the restriction of V to $\mathcal{N}_{poss}$ determines the law of the process.

(b) If $\operatorname{card} T = 0$, we get all the moments of the ν_e's. Since these variables are compactly supported variables, this determines all the finite dimensional marginals of the distribution of ν.*

If $\operatorname{card} T = 1$, we get all the moments involving the ν_r's. And if t is the unique element of T, then $\nu_t = 1 - \sum_{r \in R} \nu_r$, and we get all the moments of ν. ∎

When $\operatorname{card} T \geq 2$ the law of the environment can be determined in some cases, but not in general. (Accordingly, Corollary 1 of [2] should be amended; it holds in fact if $\operatorname{card} T \leq 1$, but not in complete generality.) Here are two examples:

---

* We recall a standard fact: if $U_1, \ldots, U_l$ are positive random variables such that, almost surely, $U_1 + \cdots + U_l \leq 1$, then, for Lebesgue-almost all $(a_1, \ldots, a_l) \in \mathbb{R}^l$,

$$\mathbf{P}(U_1 < a_1, \ldots, U_l < a_l) = \lim_{n \to \infty} \sum_{\substack{k_0, \ldots, k_l \geq 0 \\ k_0 + \cdots + k_l = n \\ k_1/n < a_1, \ldots, k_l/n < a_l}} \frac{n!}{k_0! \cdots k_l!} \mathbf{E}[(1 - U_1 - \cdots - U_l)^{k_0} \cdot U_1^{k_1} \cdots U_l^{k_l}].$$

Example 2: We consider the two following walks on $\mathbb{Z}$:

- The first one has a deterministic environment, and moves from $x \in \mathbb{Z}$ with probability $\frac{1}{2}$ to $x+1$, with probability $\frac{1}{2}$ to $x+2$.
- The environment in the second walk is coin-tossed independently at each site $x \in \mathbb{Z}$: with probability $\frac{1}{2}$, the transition probability to $x+1$ is 1, and with probability $\frac{1}{2}$, the transition probability to $x+2$ is 1.

Here $T = \{1, 2\}$, and, obviously, the two walks have the same law.

Example 3: Again, $G = \mathbb{Z}$; for any $x \in \mathbb{Z}$, with probability $\frac{1}{2}$, the transition probability from x to x is equal to $\frac{1}{2}$, and the transition probability from x to $x+1$ is also equal to $\frac{1}{2}$; and, with probability $\frac{1}{2}$, the transition probability from x to $x+2$ is equal to 1. In this case, $T = \{1, 2\}$, but the distribution of the environment is almost surely completely determined by the single trajectory we observe.

We can only see 0-transitions (from a site x to itself), 1-transitions ($x \to x+1$) and 2-transitions ($x \to x+2$). Hence μ, which is the law of $\nu(0)$, is such that $\mu(\nu_0 + \nu_1 + \nu_2 = 1) = 1$. The fact that a 0-transition is never followed by a 2-transition tells us that if $\nu_2 > 0$, then $\nu_0 = 0$. Statistics on sites from which there are 0-transitions or 1-transitions reveals that the (conditional) distribution of the number of 0-transitions from such a site is geometric. But a geometric distribution cannot be a nondegenerate convex combination of geometric distributions, and the conditional number of 0-transitions from a visited site x for which $\nu(x, 0)$ is given (and is in $]0, 1[$) has a geometric distribution. We deduce that there is exactly one value a $(= \frac{1}{2})$ such that almost surely, for all $x \in \mathbb{Z}$, if $\nu(x, 0) \neq 0$, then $\nu(x, 0) = a$. And a simple computation shows that the proportion of visited sites from which a 0-transition is possible but does not take place fits exactly with the proportion of visited sites from which the first (and unique) transition is a 1-transition; so $\mu(\nu_1 > 0 \text{ and } \nu_2 > 0) = 0$. We deduce that μ is a convex combination of μ^1 and μ^2, where $\mu^1(\nu_0 = \nu_1 = \frac{1}{2}) = 1$ and $\mu^2(\nu_2 = 1) = 1$. And the coefficients in this convex combination are easily determined (and, of course, both coefficients equal $\frac{1}{2}$).

4.3. Efficiency. Generally speaking, the efficiency of the approach described in 4.2 depends on the asymptotic behaviour of the number of sites visited by the walk up to time n: even a perfect knowledge of the environments at these sites will only give a vague idea of the law, if the visited sites are very few.

The study of nearest-neighbour random walks on $\mathbb{Z}$ sheds some light on the variety of possible situations. We restrict our study to the case where ν_1 is known to follow some beta distribution whose parameters are unknown, i.e., for some unknown real numbers $\alpha, \beta > 0$

$$\mathbf{P}(\nu_1 \in [0,t]) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^t x^{\alpha-1}(1-x)^{\beta-1}dx \quad (0 < t < 1).$$

This corresponds to the reinforcement

$$V_1(n_-, n_+) = \frac{n_+ + \alpha}{n_- + n_+ + \alpha + \beta}$$

where n_- (resp. n_+) denote the number of past moves to the left (resp. to the right) from the current site. This can be checked by a straightforward computation based on Proposition 1. See also Zabell [12], Keane and Rolles [5] and Rolles [9] for a Polya urn approach.

The reinforcement we are considering is present in the following situation: from a newly visited site $x \in \mathbb{Z}$, the transition $x \to x+1$ has weight α and the transition $x \to x-1$ has weight β. Now, any time a given transition occurs, its weight is augmented by one.

We distinguish three cases:

CASE 1: $\alpha = \beta$. Our process is a Sinai's walk [6]; it is reccurent and visits, during its n first steps, a number of sites of the order of $(\log n)^2$. Among these sites, an asymptotically strictly positive proportion is visited at least twice before time n. (The asymptotic proportion is in fact 1.) We base our observation on these sites. Among them, the proportion of sites from where the two first moves of the walk are the same gives an estimator for $(\alpha+1)/(2\alpha+1)$. By the central limit theorem, we can obtain, with a high probability, a confidence interval (at a given level) whose length is of the order of $1/\log n$.

CASE 2: $\beta < \alpha \leq \beta + 1$. This was studied by Kesten, Kozlov and Spitzer [6]. The process is transient (it walks towards $+\infty$ almost surely) and visits, during its first n steps, a number of sites of the order of n^γ where $0 < \gamma < 1$. The exponent γ is an explicit function of the unique positive number κ satisfying $\mathbf{E}[(\frac{1-\nu_1}{\nu_1})^\kappa] = 1$, i.e., $\Gamma(\alpha-\kappa)\Gamma(\beta+\kappa) = \Gamma(\alpha)\Gamma(\beta)$ (see [6]).

In order to estimate α and β, we make observations on sites of two categories. Among the visited sites, the proportion of those from which the first move is to the left gives an estimator for $\beta/(\alpha+\beta)$. Among the sites visited at least twice from which the first move is to the left, the proportion of those from which the second move is also to the left gives an estimator for $(\beta+1)/(\alpha+\beta+1)$. Using these two observations, we can estimate α and β. These two categories of sites have both a cardinal of the order of n^γ. Again, by the central limit theorem, we can obtain, with a high probability, a confidence interval (at a given level) whose length is of the order of $1/n^{\gamma/2}$.

CASE 3: $\alpha > \beta + 1$. This is the so-called ballistic case considered by Solomon [11]: for some real constant $u > 0$, $\mathbf{P}(X_n/n \to u) = 1$. We proceed as in Case 2 to get, with a high probability, a confidence interval (at a given level) whose length is of the order of $1/n$.

Remark: The remaining case ($\beta > \alpha$) splits into obvious analogues of the above Case 2 and Case 3.

Of course, the above discussion is succinct, and it deals with a special family of environment. Our aim is just to examplify things; and one cannot hope any reasonably fast algorithm if one does not restrict the search of the law of the environment to some "parametrized" family of probability measures.

4.4. SAMPLING IID TRAJECTORIES. Here we show that infinitely many independent trajectories drawn under $\mathbf{P}^{\mu}$ can be generated by concatenating steps of the single trajectory at our disposal.

For simplicity, we describe the construction of just two trajectories, X^1 and X^2. This is enough, since we can do the following: once X^1 and X^2 are constructed, leave X^1 as it is and extract two trajectories, say X^3 and X^4, out of X^2 exactly the way we extracted X^1 and X^2 out of X, then, out of X^4, get X^5 and X^6, and so on; and the family $(X^1, X^3, X^5, \ldots)$ is exactly what we want.

All we do is construct X^1 (resp. X^2) using in the "natural" way the transitions $\Delta_i^{\vec{n}}$ with i odd (resp. even). More formally, X^1 and X^2 are defined as follows:

$$X_0^1 = 0, \quad X_{n+1}^1 = X_n^1 + \Delta_{2\tau^1(n)-1}^{\vec{N}^1(n)},$$

where $\vec{N}^1$ is to the process X^1 what $\vec{N}$ is to the process X, and

$$\tau^1(n) := \operatorname{card}\{j \in \{0, \ldots, n\} | \ \vec{N}^1(j) = \vec{N}^1(n)\};$$

and, similarly,

$$X_0^2 = 0, \quad X_{n+1}^2 = X_n^2 + \Delta_{2\tau^2(n)}^{\vec{N}^2(n)},$$

where $\vec{N}^2$ is to the process X^2 what $\vec{N}$ is to the process X, and

$$\tau^2(n) := \operatorname{card}\{j \in \{0, \ldots, n\} | \ \vec{N}^2(j) = \vec{N}^2(n)\}.$$

The validity of this construction is an immediate consequence of Proposition 4.

5. Remarks

5.1. INFINITUDE OF S. The "infinitude assumption" (according to which the random set S of sites visited by X is almost surely an infinite set) is made in order to avoid discussing rather trivial cases. (If S is finite, then precise knowledge of the environment at *some* sites is almost surely available; but, unless some specific conditions are imposed on μ, complete knowledge of μ is out of reach.)

Proceeding along the general lines of the proof of Proposition 3, one can show that the random set S is either almost surely infinite or almost surely finite.

5.2. COUNTABILITY OF G. If we abandon the countability assumption on G, the set $E = \{x \in G | \ P(X_1 = x) > 0\}$ remains countable, and our sampling procedure works just as well. Consequently, $\mathbf{P}^\mu$ can be determined in non-pathological situations (and, in particular, if G is the real line).* (This can also be seen by introducing a new kind of reinforcement function which, to a given unordered history at a site, associates the probability that the next transition falls into some measurable set.) If there is no countable set $D \subset G$ such that $X_1 \in D$ almost surely, then μ cannot be determined (as one can see after studying the first example of section 4). (If there is some countable set $D \subset G$ such that $X_1 \in D$ almost surely then, almost surely, *each* X_n is in the subgroup generated by D; and since this subgroup is countable, all we did is adaptable in an obvious way.)

5.3. STRUCTURE OF G. The choice of dealing with RWRE on a *group* captures, we think, the essence of the matter. We could have restricted ourselves to groups like $\mathbb{Z}^d$ (or some other subgroups of $\mathbb{R}^d$) without a substantial gain in simplicity. A group structure is suitable (though not absolutely indispensable) if the notion of iid random environment is to make sense. We could have dealt with RWRE on homogenous spaces, or on some trees, without gaining new insight.

---

* Of course, if G is countable, the σ-field we use (without explicitly saying so) is the set of all subsets of G; and if G is the real line, we take the Borel σ-field on the line. But problems may arise if a σ-field on G is not specified in advance and there is no "natural" σ-field on G: the very notion of the law of X is problematic (and, in fact, even the notion of the law of X_1 does not make much sense). But even if a σ-field on G is "given", we aren't through. What we want is to be able to determine, on the basis of the observation of one realization of a random sequence $(U_1, U_2, \ldots)$ of iid random variables taking values in G, the probability distribution of U_1. Now if the σ-field is generated by some countable π-system of subsets of G, things are all right. Otherwise, there is no general positive result.

5.4. Assumptions on the environment. The requirement that environment at sites are iid can be loosened in various ways.

Example: $G = \mathbb{Z}$; there are two laws for the environment at sites, say μ_0 and μ_1; environments at sites are independent; and $\nu(n)$ is governed by μ_0 if n is even, by μ_1 if n is odd.

Example: $G = \mathbb{Z}$; there are laws $\mu_0, \mu_1, \ldots$ for the environment at a site; K is a random variable taking values in the set $\{2, 3, \ldots\}$; and, conditioned on K, the $\nu(n)$ are independent and, for all n, $\nu(n)$ is governed by $\mu_{n(\text{mod}\, K)}$.

Example: $G = \mathbb{Z}$; the couples $(\nu(2n), \nu(2n+1))$ $(n \in \mathbb{Z})$ are iid, but $\nu(0)$ and $\nu(1)$ are *not* independent.

5.5. Other reinforcements. Our results on the determination of the law of the process and on sampling iid trajectories can be extended to various other transition reinforced random walks that do not correspond to a random environment. Whenever an appropriate analogue of Proposition 3 is valid, things work quite well. (A sufficient condition is strict positivity of the restriction of V_r to $(\mathbb{Z}_+^R)_0$ for all real r.)

References

[1] I. Benjamini and H. Kesten, *Distinguishing sceneries by observing the scenery along a random walk path*, Journal d'Analyse Mathématique **69** (1996), 97–135.

[2] N. Enriquez and C. Sabot, *Edge-oriented reinforced random walks and RWRE*, Comptes Rendus de l'Académie des Sciences, Paris **335** (2002), 941–946.

[3] M. Löwe and H. Matzinger, *Scenery reconstruction in two dimensions with many colors*, The Annals of Applied Probability **12** (2002), 1322–1347.

[4] S. A. Kalikow, *Generalized random walk in a random environment*, The Annals of Probability **9** (1981), 753–768.

[5] M. Keane and S. Rolles, *Tubular recurrence*, Acta Mathematica Hungarica **97** (2002), 207–221.

[6] H. Kesten, M. Kozlov and F. Spitzer, *A limit law for random walk in a random environment*, Compositio Mathematica **30** (1975), 145–168.

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

[8] R. Pemantle, *Phase transition in reinforced random walk and RWRE on trees*, The Annals of Probability **16** (1988), 1229–1241.

[9] S. Rolles, *How edge-reinforced random walk arises naturally*, Probability Theory and Related Fields **126** (2003), 243–260.

[10] Y. Sinai, *The limit behavior of a one-dimensional random walk in a random environment*, (Russian) Teoriya Veroyatnostei i Primeneniya **27** (1982), 247–258.

[11] F. Solomon, *Random walks in a random environment*, The Annals of Probability **3** (1975), 1–31.

[12] S. Zabell, *Characterizing Markov exchangeable sequences*, Journal of Theoretical Probability **8** (1995), 175–178.