

The range of tree-indexed random walk

Jean-François Le Gall, Shen Lin

Institut universitaire de France et Université Paris-Sud Orsay

Erdős Centennial Conference
July 2013

1. Introduction

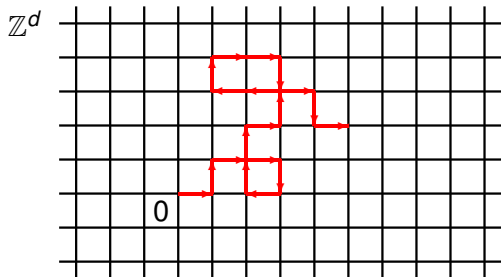
Let $(X_n)_{n \geq 0}$ be a **random walk** in \mathbb{Z}^d :

$$X_n = Y_1 + Y_2 + \cdots + Y_n$$

where Y_1, Y_2, \dots are independent and identically distributed with distribution μ .

The **range** \mathbf{R}_n is the number of distinct sites of the lattice visited by the random walk up to time n :

$$\mathbf{R}_n := \#\{X_0, X_1, \dots, X_n\}.$$



Here $\mathbf{R}_{19} = 17$

The Dvoretzky-Erdős asymptotics

An important result of [Dvoretzky and Erdős](#) in 1951 gives the asymptotics of R_n when $n \rightarrow \infty$, in the case of simple random walk (that is, if μ is uniform over neighbors of 0):

- if $d \geq 3$,

$$\frac{1}{n} \mathbf{R}_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} q_d > 0,$$

- if $d = 2$,

$$\frac{\log n}{n} \mathbf{R}_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \pi,$$

- if $d = 1$,

$$n^{-1/2} \mathbf{R}_n \xrightarrow[n \rightarrow \infty]{(d)} \sup_{0 \leq t \leq 1} B_t - \inf_{0 \leq t \leq 1} B_t,$$

where q_d is the probability that X never returns to its starting point, and $(B_t)_{t \geq 0}$ is a standard linear Brownian motion.

(Dvoretzky and Erdős point out that the method extends to the case when μ is centered with finite second moments)

Applying Kingman's subadditive ergodic theorem

The method of Dvoretzky and Erdős relies on estimating the first and second moment of \mathbf{R}_n .

When $d \geq 3$, a quicker proof follows from **Kingman's subadditive ergodic theorem**. Note that, for every $m, n \geq 0$,

$$\mathbf{R}_{n+m} \leq \mathbf{R}_n + \mathbf{R}_m \circ \theta_n$$

where θ_n is the usual **shift** on trajectories : $X_k \circ \theta_n = X_{n+k} - X_n$
(the number of sites visited between 0 and $n + m$ is smaller than the number visited between 0 and n plus the number visited between n and $n + m$)

Kingman's theorem then gives immediately

$$\frac{1}{n} \mathbf{R}_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} q$$

where $q = \lim_{n \rightarrow \infty} \frac{1}{n} E[\mathbf{R}_n] = P(\text{no return to 0})$

This applies to **any** random walk, and $q > 0$ iff X transient

Tree-indexed random walk

Question (Itai Benjamini): What is the analog of the Dvoretzky-Erdős asymptotics for a tree-indexed random walk?

Consider

- a (random) discrete rooted tree \mathcal{T}_n with n vertices;
- conditionally on \mathcal{T}_n , a collection $(Y_e)_{e \in \mathcal{E}(\mathcal{T})}$ of independent r.v. distributed according to μ , indexed by the set of edges of \mathcal{T}_n .

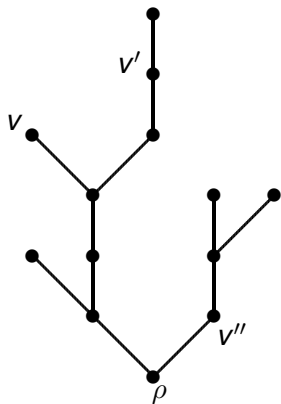
For every vertex v of \mathcal{T}_n , set

$$X_v = \sum_{e: \rho \rightarrow v} Y_e$$

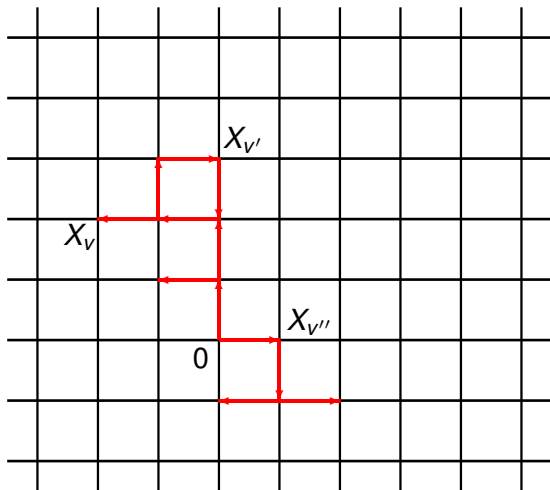
where the sum is over all edges on the path from the root ρ to v .
The range is $\mathcal{R}_n := \#\{X_v : v \text{ vertex of } \mathcal{T}_n\}$.

One expects \mathcal{R}_n to be smaller than \mathbf{R}_n (range of ordinary RW), because there are more self-intersections.

Tree-indexed random walk



Tree T_n



Galton-Watson trees

Let θ be a probability measure on $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ such that:

- $\sum_{k=0}^{\infty} k \theta(k) = 1$ (criticality)
- $\sum_{k=0}^{\infty} k^2 \theta(k) < \infty$ (finite variance)

The Galton-Watson tree \mathcal{T} with offspring distribution θ describes the genealogy of a Galton-Watson branching process with offspring distribution θ :

- the process starts with 1 ancestor at generation 0;
- each individual has k children with probability $\theta(k)$.

→ can be viewed as a rooted ordered tree (put an order on the children of each individual).

θ critical $\Rightarrow \mathcal{T}$ is finite a.s. Notation: $\#\mathcal{T}$ is the number of vertices of \mathcal{T}

Galton-Watson trees

Let θ be a probability measure on $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ such that:

- $\sum_{k=0}^{\infty} k \theta(k) = 1$ (criticality)
- $\sum_{k=0}^{\infty} k^2 \theta(k) < \infty$ (finite variance)

The **Galton-Watson tree** \mathcal{T} with offspring distribution θ describes the genealogy of a Galton-Watson branching process with offspring distribution θ :

- the process starts with 1 ancestor at generation 0;
- each individual has k children with probability $\theta(k)$.

→ can be viewed as a **rooted ordered tree** (put an order on the children of each individual).

θ critical $\Rightarrow \mathcal{T}$ is finite a.s. **Notation:** $\#\mathcal{T}$ is the number of vertices of \mathcal{T}

Galton-Watson trees with a fixed progeny

Let \mathcal{T} be a θ -Galton-Watson tree,

For every $n \geq 1$ such that $P(\#\mathcal{T} = n) > 0$, let

$$\mathcal{T}_n \stackrel{(d)}{=} \mathcal{T} \text{ conditioned on } \#\mathcal{T} = n$$

Then \mathcal{T}_n is a random tree with n vertices.

This setting includes many “combinatorial trees” (meaning that \mathcal{T}_n is then uniformly distributed on a certain class of discrete trees):

- $\theta(k) = 2^{-k-1}$: \mathcal{T}_n is uniform in the class of rooted ordered trees with n vertices;
- $\theta(0) = \theta(2) = \frac{1}{2}$: \mathcal{T}_n is uniform in the class of binary trees with n vertices;
- θ Poisson : \mathcal{T}_n is uniform in the class of Cayley trees with n vertices.

Galton-Watson trees with a fixed progeny

Let \mathcal{T} be a θ -Galton-Watson tree,

For every $n \geq 1$ such that $P(\#\mathcal{T} = n) > 0$, let

$$\mathcal{T}_n \stackrel{(d)}{=} \mathcal{T} \text{ conditioned on } \#\mathcal{T} = n$$

Then \mathcal{T}_n is a random tree with n vertices.

This setting includes many “combinatorial trees” (meaning that \mathcal{T}_n is then uniformly distributed on a certain class of discrete trees):

- $\theta(k) = 2^{-k-1}$: \mathcal{T}_n is uniform in the class of **rooted ordered trees** with n vertices;
- $\theta(0) = \theta(2) = \frac{1}{2}$: \mathcal{T}_n is uniform in the class of **binary trees** with n vertices;
- θ Poisson : \mathcal{T}_n is uniform in the class of **Cayley trees** with n vertices.

The range of tree-indexed random walk

- \mathcal{T}_n is a θ -Galton-Watson tree conditioned to have n vertices;
- conditionally on \mathcal{T}_n , $(X_v)_{v \in V(\mathcal{T}_n)}$ is random walk with jump distribution μ indexed by \mathcal{T}_n , and $\mathcal{R}_n = \#\{X_v : v \in V(\mathcal{T}_n)\}$.

Theorem

Suppose μ is centered with sufficiently high moments:

- 1 if $d \geq 5$,
$$\frac{1}{n} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(P)} c_{\mu, \theta} > 0$$
- 2 if $d = 4$, and $\theta(k) = 2^{-k-1}$,
where $\sigma^2 = (\det(\text{cov}(\mu)))^{1/4}$,
$$\frac{\log n}{n} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(P)} 8 \pi^2 \sigma^4,$$
- 3 if $d \leq 3$,
$$n^{-d/4} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(d)} c_{\mu, \theta} \text{Leb}(\text{supp}(\mathcal{I}))$$

where \mathcal{I} is ISE (Integrated Super-Brownian Excursion).

The range of tree-indexed random walk

- \mathcal{T}_n is a θ -Galton-Watson tree conditioned to have n vertices;
- conditionally on \mathcal{T}_n , $(X_v)_{v \in V(\mathcal{T}_n)}$ is random walk with jump distribution μ indexed by \mathcal{T}_n , and $\mathcal{R}_n = \#\{X_v : v \in V(\mathcal{T}_n)\}$.

Theorem

Suppose μ is centered with sufficiently high moments:

- 1 if $d \geq 5$,
$$\frac{1}{n} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(P)} c_{\mu, \theta} > 0$$
- 2 if $d = 4$, and $\theta(k) = 2^{-k-1}$,
where $\sigma^2 = (\det(\text{cov}(\mu)))^{1/4}$,
$$\frac{\log n}{n} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(P)} 8 \pi^2 \sigma^4,$$
- 3 if $d \leq 3$,
$$n^{-d/4} \mathcal{R}_n \xrightarrow[n \rightarrow \infty]{(d)} c_{\mu, \theta} \text{Leb}(\text{supp}(\mathcal{I}))$$

where \mathcal{I} is ISE (Integrated Super-Brownian Excursion).

Remarks on the main theorem

- Results are similar to those for ordinary random walk, BUT

the critical dimension is now $d = 4$

Note that $\max\{|X_v| : v \text{ vertex of } \mathcal{T}_n\} \sim n^{1/4}$ (Janson-Marckert)

- Above the critical dimension, the range grows linearly (can again be viewed as a consequence of Kingman's theorem, but this is less immediate!)
- At the critical dimension $d = 4$, the proof is more involved (our method only works for θ geometric)
- Below the critical dimension, the result is related to the “invariance principles” connecting branching random walk with super-Brownian motion
- The convergence of $\frac{1}{n}\mathcal{R}_n$ to a constant $c_{\mu,\theta} \geq 0$ extends to much more general θ and μ .

Remarks on the main theorem

- Results are similar to those for ordinary random walk, BUT

the critical dimension is now $d = 4$

Note that $\max\{|X_v| : v \text{ vertex of } \mathcal{T}_n\} \sim n^{1/4}$ (Janson-Marckert)

- Above the critical dimension, the range grows linearly (can again be viewed as a consequence of Kingman's theorem, but this is less immediate!)
- At the critical dimension $d = 4$, the proof is more involved (our method only works for θ geometric)
- Below the critical dimension, the result is related to the “invariance principles” connecting branching random walk with super-Brownian motion
- The convergence of $\frac{1}{n}\mathcal{R}_n$ to a constant $c_{\mu,\theta} \geq 0$ extends to much more general θ and μ .

2. The linear growth of \mathcal{R}_n

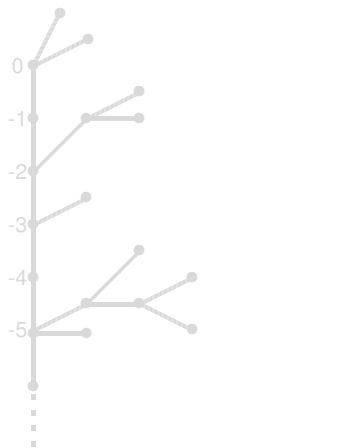
KEY IDEA: Apply Kingman's ergodic theorem

BUT: needs to find a suitable shift transformation on an appropriate space of trees, and a corresponding invariant probability measure.

The space of trees \mathbb{T} : consists of **infinite trees** \mathcal{T} having

- a “spine” with **infinitely** many vertices labeled $0, -1, -2, \dots$
- attached to each vertex $-k$ of the spine, a **finite** rooted ordered tree \mathcal{T}_k

We assume that there are infinitely many vertices not on the spine.



2. The linear growth of \mathcal{R}_n

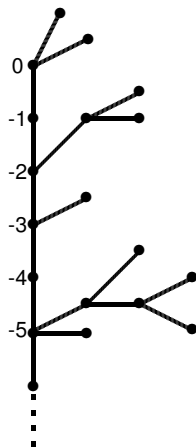
KEY IDEA: Apply Kingman's ergodic theorem

BUT: needs to find a suitable shift transformation on an appropriate space of trees, and a corresponding invariant probability measure.

The space of trees \mathbb{T} : consists of **infinite trees** \mathcal{T} having

- a “spine” with **infinitely** many vertices labeled $0, -1, -2, \dots$
- attached to each vertex $-k$ of the spine, a **finite** rooted ordered tree \mathcal{T}_k

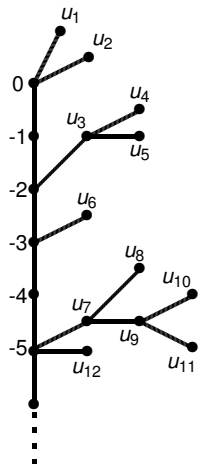
We assume that there are infinitely many vertices not on the spine.



The shift on infinite trees

Let $\mathcal{T} \in \mathbb{T}$ be an infinite tree, let u_1, u_2, \dots be the vertices of \mathcal{T} **not** belonging to the spine enumerated in **lexicographical order** (considering successively the subtrees $\mathcal{T}_0, \mathcal{T}_1, \dots$ in this order)

Define the **shift** $\tau(\mathcal{T})$ by declaring that the top of the spine of $\tau(\mathcal{T})$ is u_1 and removing the vertices of the spine of \mathcal{T} that are not ancestors of u_1 (i.e. the vertices $0, -1, \dots, -k + 1$, if u_1 lies in \mathcal{T}_k)



The shift on infinite trees

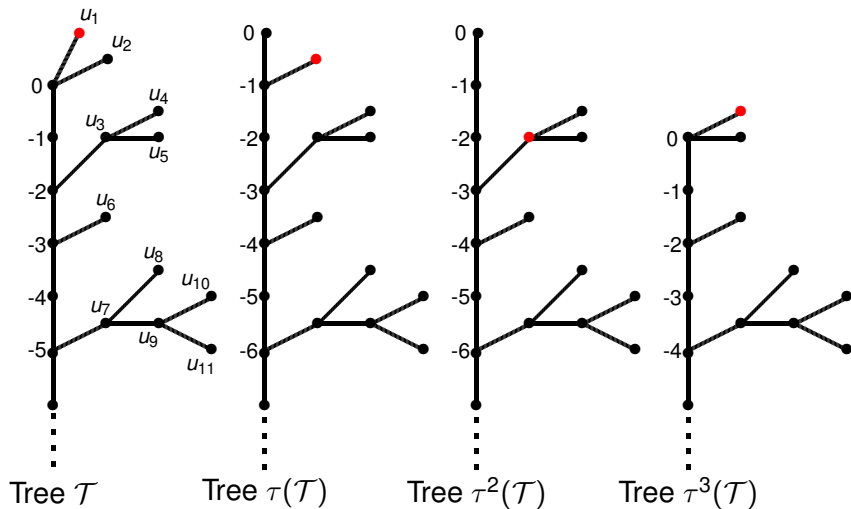


Illustration of the shift: The **red vertex** (the first one not on the spine) becomes the **top of the spine** when applying the shift.

The invariant measure

Define a probability measure \mathbf{P}_θ on \mathbb{T} : Under \mathbf{P}_θ ,

- the subtree attached at the top of the spine is a θ -Galton-Watson tree;
- each vertex of the spine other than 0 has k children (not on the spine) with probability

$$\theta([k + 1, \infty))$$

- these children, and their descendants, then reproduce according to the offspring distribution θ .

Proposition

The probability measure \mathbf{P}_θ is *invariant* (and ergodic) under the shift τ

The proof is easy by a direct verification.

Remark. This result only requires the fact that θ is critical.

The invariant measure

Define a probability measure \mathbf{P}_θ on \mathbb{T} : Under \mathbf{P}_θ ,

- the subtree attached at the top of the spine is a θ -Galton-Watson tree;
- each vertex of the spine other than 0 has k children (not on the spine) with probability

$$\theta([k + 1, \infty))$$

- these children, and their descendants, then reproduce according to the offspring distribution θ .

Proposition

The probability measure \mathbf{P}_θ is *invariant* (and ergodic) under the shift τ

The proof is easy by a direct verification.

Remark. This result only requires the fact that θ is critical.

Trees with spatial positions

If $\mathcal{T} \in \mathbb{T}$ is an infinite tree, define the μ -random walk $(X_v)_{v \in V(\mathcal{T})}$ indexed by \mathcal{T} by imposing that:

- the increments along edges (oriented from the “bottom” of the spine) are independent and distributed according to μ
- the spatial position X_0 at the top of the spine is 0

Fact

The law of the pair consisting of a tree distributed according to \mathbf{P}_θ and the associated μ -random walk is again invariant under the shift τ .

(To define spatial positions of the shifted tree $\tau(\mathcal{T})$ one translates the spatial positions of the corresponding vertices of \mathcal{T} so that the position of the top of the spine is again 0 in $\tau(\mathcal{T})$)

Linear growth of the range

Assumptions:

- \mathcal{T} is distributed according to \mathbf{P}_θ ;
- conditionally on \mathcal{T} , $(X_v)_{v \in \mathcal{T}}$ is the μ -random walk indexed by \mathcal{T} .

Let u_1, u_2, \dots be the vertices of \mathcal{T} not on the spine, enumerated in lexicographical order, and

$$R_n = \#\{X_{u_1}, X_{u_2}, \dots, X_{u_n}\}$$

Theorem

There exists a constant $c_{\mu, \theta} \in [0, 1]$ such that

$$\frac{1}{n} R_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} c_{\mu, \theta}$$

Proof. Just apply Kingman's subadditive ergodic theorem!

Remark. The theorem requires no assumption on μ or θ , except the fact that the offspring distribution θ is critical.

Positivity of the limiting constant

Assume that the random walk $(S_j)_{j \geq 0}$ with jump distribution μ is **transient** (otherwise $c_{\mu, \theta} = 0$). Let :

- G_μ **Green function** of S
- g_θ **generating function** of θ

Proposition

Assume that

$$\prod_{j=1}^{\infty} \left(\frac{1 - g_\theta((1 - G_\mu(S_j))_+)}{G_\mu(S_j)} \right) > 0 \quad \text{a.s.}$$

Then $c_{\mu, \theta} > 0$.

Corollary

Suppose that μ is centered and has finite moments of order $d - 1$.

- If θ has **finite variance**, $c_{\mu, \theta} > 0$ if $d \geq 5$,
- If θ is in the **domain of attraction of a stable distribution** with index $\alpha \in (1, 2)$, then $c_{\mu, \theta} > 0$ if $d > \frac{2\alpha}{\alpha-1}$.

From the infinite tree to trees with a given size

Need an argument to derive the result for Galton-Watson trees with a **fixed progeny** from the case of the **infinite tree**.

From an infinite tree distributed according to \mathbf{P}_θ can obtain a **sequence** $\mathcal{T}_{(1)}, \mathcal{T}_{(2)}, \dots$ of independent θ -Galton-Watson trees:
→ consider the subtrees branching off the children of the vertices of the spine.

If $k_n = \min\{j : \#(\mathcal{T}_{(j)}) \geq n\}$, then $\mathcal{T}_{(k_n)}$ is a θ -Galton-Watson tree conditioned to have **at least** n vertices.

- Can derive from the theorem an analogous result for a θ -Galton-Watson tree conditioned to have at least n vertices. (this requires limit theorems for the contour of a sequence of independent GW trees, cf Duquesne-LG)
- An absolute continuity argument allows one to deal with a θ -Galton-Watson tree conditioned to have **exactly** n vertices

3. The critical dimension $d = 4$

Assumptions.

- $\theta(k) = 2^{-k-1}$ (geometric).
- μ is symmetric and has small exponential moments.

Set $\sigma^2 = (\det(\text{cov}(\mu)))^{1/4}$.

IDEA: Use the path-valued Markov chain called the **discrete snake** to generate the spatial positions of an infinite tree distributed according to \mathbf{P}_θ .

→ Then exploit the **Markovian properties** of the discrete snake to derive the needed estimates.

(If θ is not geometric, the discrete snake approach does not work and things become more complicated!)

The discrete snake

The discrete snake is a Markov chain with values in the set \mathcal{W} of all **semi-infinite** discrete paths

$$w : (-\infty, \zeta] \cap \mathbb{Z} \longrightarrow \mathbb{Z}^4$$

where $\zeta = \zeta_{(w)} \in \mathbb{Z}$ is called the lifetime of w .

Transition kernel. Suppose that $W_0 = w$:

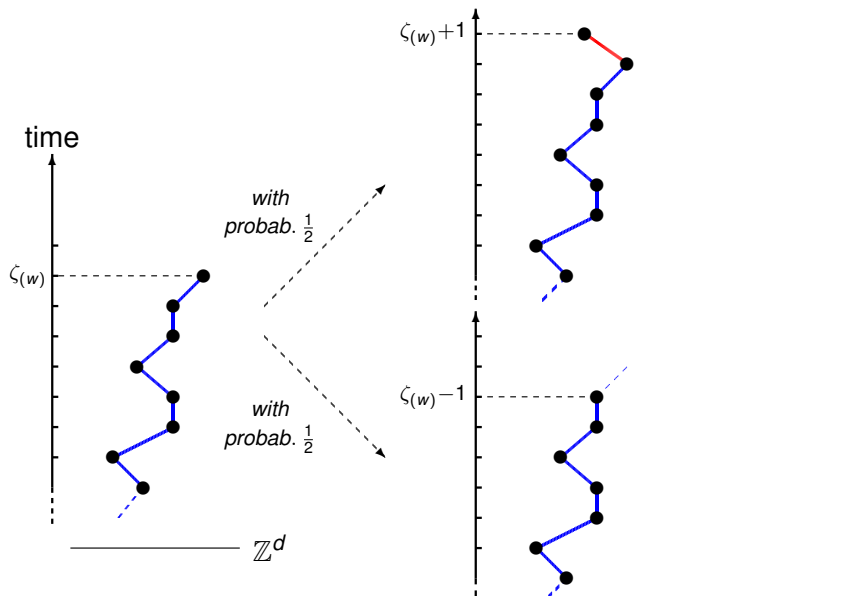
- With probability $\frac{1}{2}$,
 - ▶ $\zeta_{(W_1)} = \zeta_{(w)} - 1$,
 - ▶ $W_1(k) = w(k)$ for all $k \leq \zeta_{(w)} - 1$.

(the last step of w is removed)

- With probability $\frac{1}{2}$,
 - ▶ $\zeta_{(W_1)} = \zeta_{(w)} + 1$
 - ▶ $W_1(k) = w(k)$ for all $k \leq \zeta_{(w)}$
 - ▶ $W_1(\zeta_{(w)} + 1) - W_1(\zeta_{(w)})$ has law μ

(one step is added to w using the jump distribution μ)

Transition kernel of the discrete snake



A key estimate

Suppose that $\zeta_{(W_0)} = 0$ and $(W_0(-k))_{k \geq 0}$ is distributed as a random walk with jump distribution μ started from 0.

Write $\widehat{W}_k := W_k(\zeta_{(W_k)})$ for the terminal point of W_k (the “head of the discrete snake”)

Lemma

We have

$$\lim_{n \rightarrow \infty} (\log n) P(\widehat{W}_k \neq \widehat{W}_0 \text{ for all } k = 1, 2, \dots, n) = 4\pi^2 \sigma^4.$$

Remark. Analogous result for a (centered, finite variance) random walk S on \mathbb{Z}^2 started from 0,

$$\lim_{n \rightarrow \infty} (\log n) P(S_k \neq 0 \text{ for all } k = 1, 2, \dots, n) = c > 0$$

This is much easier to prove than the lemma.

A key estimate

Suppose that $\zeta_{(W_0)} = 0$ and $(W_0(-k))_{k \geq 0}$ is distributed as a random walk with jump distribution μ started from 0.

Write $\widehat{W}_k := W_k(\zeta_{(W_k)})$ for the terminal point of W_k (the “head of the discrete snake”)

Lemma

We have

$$\lim_{n \rightarrow \infty} (\log n) P(\widehat{W}_k \neq \widehat{W}_0 \text{ for all } k = 1, 2, \dots, n) = 4\pi^2 \sigma^4.$$

Remark. Analogous result for a (centered, finite variance) random walk S on \mathbb{Z}^2 started from 0,

$$\lim_{n \rightarrow \infty} (\log n) P(S_k \neq 0 \text{ for all } k = 1, 2, \dots, n) = c > 0$$

This is much easier to prove than the lemma.

Application of the key estimate 1

Set $R_n = \#\{\widehat{W}_0, \widehat{W}_1, \dots, \widehat{W}_n\}$.

Then,

$$\begin{aligned} E[R_n] &= E \left[\sum_{j=0}^n \mathbf{1}_{\{\widehat{W}_k \neq \widehat{W}_j, \text{ for all } k=j+1, \dots, n\}} \right] \\ &= \sum_{j=0}^n P \left[\widehat{W}_k \neq \widehat{W}_j, \text{ for all } k = j+1, \dots, n \right] \\ &= \sum_{j=0}^n P \left[\widehat{W}_k \neq \widehat{W}_0, \text{ for all } k = 1, \dots, n-j-1 \right] \end{aligned}$$

by stationarity. The lemma now gives

$$\lim_{n \rightarrow \infty} \frac{\log n}{n} E[R_n] = 4\pi^2 \sigma^4.$$

Application of the key estimate 2

By similar arguments,

$$\lim_{n \rightarrow \infty} \left(\frac{\log n}{n} \right)^2 E[(R_n)^2] = (4\pi^2 \sigma^4)^2.$$

Suffices to get

$$\frac{\log n}{n} R_n \xrightarrow[n \rightarrow \infty]{L^2} 4\pi^2 \sigma^4.$$

→ This gives the desired asymptotics for the range of random walk indexed by the [infinite geometric tree](#).

→ Technical work (more difficult than in the supercritical case) is needed to get the asymptotics for a tree with [fixed size](#) n .

Ideas of the proof of the main estimate 1

We aim at proving that

$$P(\widehat{W}_k \neq \widehat{W}_0 \text{ for all } k = 1, 2, \dots, n) \underset{n \rightarrow \infty}{\sim} \frac{4\pi^2 \sigma^4}{\log n}.$$

Start from the identity

$$\begin{aligned} 1 &= \sum_{k=0}^n P(\widehat{W}_k = 0; \widehat{W}_\ell \neq 0, \forall \ell \in \{k+1, \dots, n\}) \\ &= \sum_{k=0}^n E \left[\mathbf{1}_{\{\widehat{W}_k=0\}} P_{W_k}(\widehat{W}_\ell \neq 0, \forall \ell \in \{1, \dots, n-k\}) \right] \quad (\text{Markov}) \\ &= \sum_{k=0}^n E \left[\mathbf{1}_{\{\widehat{W}_k=0\}} P_{W_0}(\widehat{W}_\ell \neq 0, \forall \ell \in \{1, \dots, n-k\}) \right]. \end{aligned}$$

(symmetry argument: (W_0, W_k) and (W_k, W_0) have the same distribution under $P(\cdot \mid \widehat{W}_k = 0)$)

Ideas of the proof of the main estimate 2

It follows that

$$1 = E \left[E_{W_0} \left[\sum_{k=0}^n \mathbf{1}_{\{\widehat{W}_k=0\}} \right] P_{W_0}(\widehat{W}_\ell \neq 0, \forall \ell \in \{1, \dots, n-k\}) \right].$$

Direct calculations show that

$$E \left[\sum_{k=0}^n \mathbf{1}_{\{\widehat{W}_k=0\}} \right] \underset{n \rightarrow \infty}{\sim} \frac{\log n}{4\pi^2 \sigma^4}.$$

Needs to verify that

$$E_{W_0} \left[\sum_{k=0}^n \mathbf{1}_{\{\widehat{W}_k=0\}} \right]$$

is **very concentrated** near its mean:

→ First get a continuous version of this concentration property involving Brownian motion

→ Then use a strong invariance principle (**Komlós-Major-Tusnády** and **Zaitsev**) to complete the proof.