Random walks and graphs

a few topics

(preliminary version)

N.C.

June 17, 2020

This is the current draft of the lecture notes for the Master 2 course “Marches et graphes aléatoires” given during the years 2016-2019 at université d’Orsay. Thanks are due to the students of this course for having tested this material in particular to Maxime Berger, Jacques De Catelan, Barbara Dembin, Alejandro Fernandez and Corentin Morandeau (promotion 2016) Chenlin Gu, Enzo Miller, Arthur Blanc-Renaudie, Kévin Marchand, Rémy Mahfouf (promotion 2017) Antoine Picard-Weibel (promotion 2018), Pierre-Amaury Monard, Nicolas Zalduendo Vidal, Romain Panis, Lucas d’Alimonte, Baojun Wu (promotion 2019). Additional materials and complements may be found in [13, 17, 25, 27, 30, 33]. Remarks, comments and list of typos are very much welcome at this stage!!!

They are many wonderful books about random walks ([39, 41, 27, 7] just to name a few) and these lecture notes instead of seeking exhaustivity should rather be taken as a pretext to present nice probabilistic techniques and geometric concepts (potential theory, rough isometries, cyclic lemma, second moment method, stable random variables, large deviations estimations, size-biasing, coding of trees via paths, probabilistic enumeration, Fourier transform, coupling, recursive distributional equation, local limit theorems, local convergence...) in a master course. To keep a certain unity in the book the main focus is on ”discrete probability” and we unfortunately do not address Brownian motion limit and the many applications of Donsker’s invariance principle.

We then use one-dimensional random walks to study the geometric properties of random graphs. The prototype example is the coding of random plane trees via skip-free descending random walks via their Lukasiewicz path. Obviously, there are many other shortcomings in these lecture notes. May the reader forgive us.

Contents

Part I:

Random walks on graphs

In this part we study general random walks on (weighted, finite or infinite) graphs. We relate this object to the theory of electrical networks by interpreting the return probabilities for the walk as potential for the associated electrical networks. In particular, we give equivalent characterization of recurrence/transience of a graph in terms of resistance to infinity. This enables us to give robust criteria for recurrence/transience on infinite graphs. We also introduce the discrete Gaussian free field and connect it to potential theory. We refer to [33, Chapter 2] (and its 139 additional exercices!) for more details and to [16] for a smooth introduction to the subject.

Figure 1: ???

Chapter I: Finite electrical networks

In this chapter we develop the well-known connection between potential theory and random walks on finite (weighted) graphs. This enables us to give a probabilistic interpretation of resistance, electric potential and electric current.

1.1 Random walks and Dirichlet problem

1.1.1 Random walk

A graph[1] $\mathfrak{g}$ is described by its set of vertices $V = \mathrm{V}(\mathfrak{g})$ and its set of edges $E = \mathrm{E}(\mathfrak{g})$ which is a multiset of pairs of (non-necessarily distinct) vertices, that is, an edge $e = \{x, y\}$ can appear several times inside E and loops $\{x, x\}$ are allowed. It is useful to also consider the multiset $\vec{E} = \vec{E}(\mathfrak{g})$ of oriented edges obtained by splitting every edge $\{x, y\}$ of E into the two oriented edges (x, y) and (y, x). The starting point of $\vec{e}$ is denoted by $\vec{e}_*$ and its target is $\vec{e}^*$. The degree of a vertex $x \in V$ is the number of oriented edges (counted with multiplicity) starting at x:

$$\deg(x) = \#\{(x, y) \in \vec{E}(\mathfrak{g})\}.$$

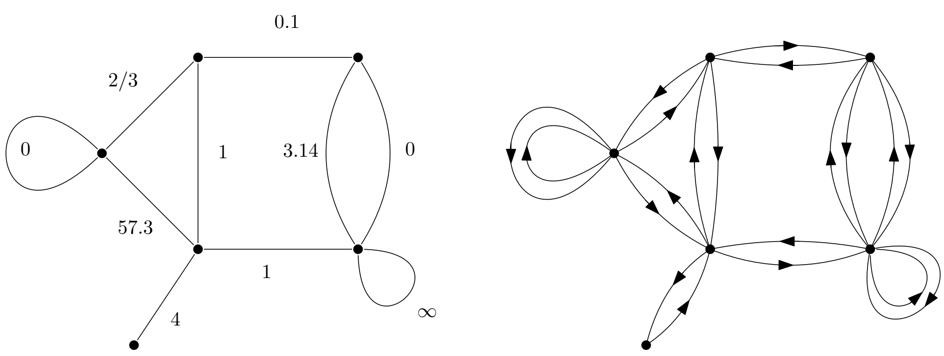


Figure 1.1: On the left a weighted graph. On the right the corresponding graph with oriented edges.

[1] non-oriented multigraph to be precise

Definition 1.1 (Conductance)**.** *A* conductance *is a function $c : E(\mathfrak{g}) \to [0, \infty]$. The conductance of an oriented edge is by definition the conductance of its associated non-oriented edge. The pair $(\mathfrak{g}, c)$ is a weighted graph. The* weight *of a vertex x is the sum of the conductance of oriented edges starting from x:*

$$\pi(x) = \sum_{\vec{e}=(x,y)\in\vec{E}} c(\vec{e}).$$

Remark 1.1. Considering multi-graph may yield to a lack of precision in the notation, for example if the edge $\{x, y\}$ is present with a multiplicity, each of its copies may carry a different conductance. Implicitly when summing over edges we will always sum over the different copies of the edges carrying the possibly different conductances. The confused reader may assume at first reading that all the graphs considered are simple (no loops nor multiple edges) so that this problem disappears.

We write $x \sim y$ if the two vertices $x, y \in V$ share an edge of positive conductance and say that they are in the same connected component. Unless specified:

All the graphs considered in these notes are connected and all degrees are finite.

This notion of conductance is used in the definition of the random walk. The standard notion of random walk is a process evolving on the vertices of the graph, but we will see that sometimes it is more convient to think of it as the trace of a process evolving on the oriented edges of the graph.

Definition 1.2 (Random walk on $(\mathfrak{g}, c)$)**.** *If $(\mathfrak{g}, c)$ is a weighted graph, for any $\vec{e}_0 \in \vec{E}(\mathfrak{g})$, the edge-random walk started from $\vec{e}_0$ is the Markov chain $(\vec{E}_n : n \geqslant 0)$ with values in $\vec{E}$ whose probability transitions are given by*

$$\mathbb{P}(\vec{E}_{n+1} = \vec{e}_{n+1} \mid \vec{E}_n = \vec{e}_n) = \frac{c(\vec{e}_{n+1})}{\pi(\vec{e}_n^{\,*})} \mathbf{1}_{\vec{e}_n^{\,*}=(\vec{e}_{n+1})_*}.$$

The projection of the random walk on the vertices of $\mathfrak{g}$ is thus the Markov chain $(X_n : n \geqslant 0)$ called "random walk on $(\mathfrak{g}, c)$" whose probability transitions $\mathbf{p}$ are

$$\mathbf{p}(x, y) = \mathbb{P}(X_{n+1} = y \mid X_n = x) = \frac{c(x, y)}{\pi(x)},$$

where $c(x, y)$ is the sum of all the conductances of the oriented edges $(x, y) \in \vec{E}$.

Example 1.1. When the conductance function is $c \equiv 1$, the degree of a vertex (for c) coincides with its graph degree and the random walk $(X_n)_{n\geqslant 0}$ is the simple random walk on the graph $\mathfrak{g}$: it chooses, independently of the past, a neighbor uniformly at random (according to the number of connections to that point) and jumps to it.

Since (X_n) is a discrete time Markov [2] chain with a countable state space, it enjoys the

[2]  Andreï Andreïevitch Markov (1856-1922)

Markov property which we now recall: If $\mathcal{F}_n$ is the filtration generated by the first n steps of the walk, a stopping time is a random variable $\theta \in \{0, 1, 2, \dots\} \cup \{\infty\}$ such that $\{\theta = n\}$ is $\mathcal{F}_n$-measurable. The σ-field $\mathcal{F}_\theta$ of the past before θ is made of those events A such that $A \cap \{\theta \leqslant n\}$ is $\mathcal{F}_n$-measurable. The (strong) Markov property says that conditionally on $\mathcal{F}_\theta$ and on $\{\theta < \infty\}$ the law of $(X_{\theta_n+k} : k \geqslant 0)$ is $\mathbb{P}_{X_\theta}$ where we denote as usual

$$\mathbb{P}_x \text{ for the law of the Markov chain started from } x \in V.$$

It is also very important to notice that the random walk $(X_n : n \geqslant 0)$ in fact admits $\pi(\cdot)$ as an invariant and even reversible measure (the easiest way to see this is to notice that the conductance $c(\cdot)$ is in fact an invariant and reversible measure for the edge-random walk chain $(\vec{E}_n)_{n\geqslant 0}$ and the statement follows by projection).

Here is a exercise to remind the reader a few basics in discrete Markov chain theory:

Exercise 1.1. Show that

- a reversible measure is necessarily invariant,
- if there exists a finite ($\neq 0$) invariant measure, then if the chain is irreducible, it is recurrent,
- a recurrent Markov chain admits a unique invariant measure up to multiplicative constant,
- there exist infinite connected weighted graphs (necessarily transient) for which there are two non proportional invariant measures for the random walk.

1.1.2 Harmonic functions and the Dirichlet problem

As we have seen above, Markov chains act on measures and leave invariant measures unchanged. In the dual version, we can consider their actions on functions where the analog of the concept of invariant measure is given by harmonic functions:

Definition 1.3. *A function $h : V(\mathfrak{g}) \to \mathbb{R}$ is harmonic at x for the random walk on $(\mathfrak{g}, c)$ if*

$$h(x) = \sum_{y\sim x} \mathbf{p}(x, y)h(y) = \mathbb{E}_x[h(X_1)].$$

We simply say that h is harmonic if it is harmonic at all vertices.

Exercise 1.2. Prove that $(h(X_n) : n \geqslant 0)$ is a martingale (for the filtration generated by the random walk) for the random walk on $(\mathfrak{g}, c)$ if and only if h is harmonic.

The following problem amounts to find "harmonic extension" to prescribed functions. It is called the Dirichlet[3] problem and pops-up in many areas of physics and mathematics.

[3]  Johann Peter Gustav Lejeune Dirichlet (1805-1859). His paternal grandfather was born in Richelette a small Belgium town, hence the name Lejeune Dirichlet "the young from Richelette"

Dirichlet problem: Let $A \subset V$ be a subset of vertices of the graph $\mathfrak{g}$ and suppose we are given a function $f_0 : V\backslash A \to \mathbb{R}$. The Dirichlet problem consists in finding a harmonic extension inside A, that is, a function f satisfying :

- $f \equiv f_0$ on $V\backslash A$,
- f is harmonic at every point $x \in A$.

Theorem 1.1 (*Dirichlet problem, finite case*)

If $A \subset V$ is finite and $f_0 : V\backslash A \to \mathbb{R}$ then the Dirichlet problem has a unique solution f given by

$$f(x) = \mathbb{E}_x[f_0(X_{\tau_{A^c}})], \qquad \text{where } \tau_{A^c} = \inf\{k \geqslant 0 : X_k \in V\backslash A\}.$$

Since the above theorem is key in this chapter we give two proofs of it:

Proof 1: But our assumptions (A is finite, $\mathfrak{g}$ is connected, all degrees are finite) it is clear that $\tau_{A^c} < \infty$ almost surely and since $f_0(X_{\tau_{A^c}})$ can take only finitely many values, it is bounded so that $f(x) = \mathbb{E}_x[f_0(X_{\tau_{A^c}})]$ is well-defined for every $x \in V$. Clearly, f coincides with f_0 outside of A. The fact that f is harmonic inside A follows from the Markov property applied at time $\theta = 1$ since for $x \in A$ we have

$$f(x) = \mathbb{E}_x[f_0(X_{\tau_{A^c}})] \underset{\tau_{A^c} \geqslant 1}{=} \mathbb{E}_x\left[\mathbb{E}_{X_1}\left[f_0(\tilde{X}_{\tau_{A^c}})\right]\right] = \mathbb{E}_x[f(X_1)].$$

Hence f is a solution to the Dirichlet problem. As for the uniqueness, consider another solution $\hat{f}$ harmonic inside A and coinciding with f_0 on $V\backslash A$. Then by Exercise 1.2 the process $(\hat{f}(X_{n\wedge\tau_{A^c}}) : n \geqslant 0)$ is a bounded martingale and so by the optional sampling theorem we must have for $x \in A$

$$\hat{f}(x) = \hat{f}(X_0) = \mathbb{E}_x\left[\hat{f}(X_{\tau_{A^c}})\right] =: f(x).$$

□

Proof 2: The Dirichlet problem is a *linear* problem with $\#A$ unknowns (the values of the function f inside A) and $\#A$ equations (harmonicity of f at all points in A). It can thus be written under the form $AX = B$ where A is a square matrix, X are the unknowns and B is prescribed by the boundary conditions. Hence the existence and unicity of the Dirichlet problem reduces to show that the matrix A is invertible, or in other words that the equation $AX = 0$ has *at most* a solution. It is easy to see that finding X such that $AX = 0$ is equivalent to finding a harmonic function to the Dirichlet problem where the boundary condition is null i.e. when f_0 equal 0. The uniqueness of a possible solution can then be proved by the maximum principle: Let g be a solution of the above problem and consider $x \in A$ such that g is maximum. If $g \in V\backslash A$ then $g(x) = f_0(x) = 0$, otherwise by harmonicity of g at x we deduce that all neighbors of x share this maximum value (equality case in the triangle inequality). By connectedness the maximal value of g is attained on ∂A and must be zero. Consequently $g \leqslant 0$ and by a similar argument we have $g \geqslant 0$, so $g \equiv 0$ as desired. □

Exercise 1.3. (*) Is there always existence and uniqueness to the Dirichlet problem when A is not necessarily finite? More specifically investigate the cases:

- show that there is always uniqueness the recurrent case,
- the existence may fall down in the case when the function f is required to be positive (and $f_0 \geqslant 0$ as well),
- there is always existence in the general case.

1.2 Electrical networks

A (finite) network is the data formed by a finite weighted graph $(\mathfrak{g}, c)$ together with two (distinct) vertices x_{in}, called the source, and x_{out} called the sink. To help the intuition, it is good to interpret this network as an actual (real) one where the edges have been replaced by resistors whose conductance (the inverse of the resistance) is prescribed by c which is connected to a battery. Beware though: the reader should not confuse the physical intuition with mathematical statements!

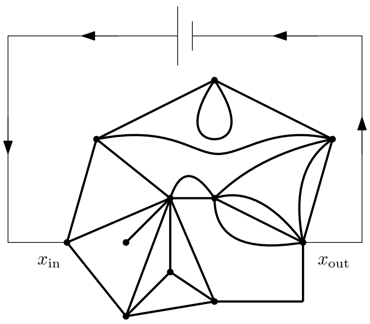
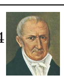


Figure 1.2: Setting of the next two sections: a (finite connected) weighted graph together with two distinguished vertices.

1.2.1 Back to high school : reminder of physics laws

Let us start with the intuition and remind the reader of a few physics laws. In the above network, imagine that we impose potential difference between x_{in} and x_{out} by plugging the network on a battery of 1 Volt[4]. This creates an electrical current, an electric potential and dissipates the energy of the battery. More precisely we can measure the following physical quantities/facts:

- The electrical potential is a function $v : V \to \mathbb{R}$ (defined up to addition but which we can fix to be) equal to 1 at x_{in} and 0 at x_{out} once the battery is plugged,
- The electrical current is a function $i : \vec{E} \to \mathbb{R}$ such that $i(\vec{e}) = -i(\overleftarrow{e})$,

4  Alessandro Giuseppe Antonio Anastasio Volta (1745-1827)

- Ohm's law : for any oriented edge we have $-c(\vec{e}) \cdot \nabla_{\vec{e}} v = i(\vec{e})$, where we used the notation

$$\nabla_{(x,y)} f = f(y) - f(x).$$

- Flow property : for any $x \notin \{x_{\text{in}}, x_{\text{out}}\}$ we have $\sum_{\vec{e}=(x,\cdot)} i(\vec{e}) = 0$.

Combining Ohm's[5] law for the potential and the flow property of the current we deduce that v should be harmonic on $V \backslash \{x_{\text{in}}, x_{\text{out}}\}$:

$$0 = \sum_{\vec{e}=(x,\cdot)} i(\vec{e}) = \sum_{\vec{e}=(x,\cdot)} \frac{1}{\pi(x)} i(\vec{e}) = \sum_{y \sim x} \mathbf{p}(x,y)(v(x) - v(y)).$$

Similarly, the fact that the current can be seen as the "derivative" of the potential implies the cycle rule: the sum the current times the resistance along an oriented cycle $\mathcal{C} = (\vec{e}_0, \vec{e}_1, \ldots, \vec{e}_k)$ in the graph is equal to 0

$$\sum_{\mathcal{C}} i(\vec{e}_i)/c(\vec{e}_i) = -\sum_{\vec{e} \in \mathcal{C}} \nabla_{\vec{e}} v = 0.$$

In fact the potential and the current point of views are equivalent and it is easy to prove the following:

Proposition 1.2. *Fix a network $((\mathfrak{g}, c); x_{\text{in}}, x_{\text{out}})$. There is a bijection between potential functions $v : V \to \mathbb{R}$ such that $v(x_{\text{out}}) = 0$ which are harmonic on $V \backslash \{x_{\text{in}}, x_{\text{out}}\}$ and current functions $i : \vec{E} \to \mathbb{R}$ which are symmetric, possess the flow property at any $x \notin \{x_{\text{in}}, x_{\text{out}}\}$ and obey the cycle rule. The bijection is simply given by*

$$v \longleftrightarrow i = \big(- c(\vec{e}) \cdot \nabla_{\vec{e}} v \big)_{\vec{e} \in \vec{E}}.$$

Your high-school physics teacher also told you that the energy dissipated in the network by "Joule[6] heating" has a power proportional to $\mathcal{E} = \frac{1}{2} \sum_{\vec{e}} i(\vec{e})^2 / c(\vec{e}) = \frac{1}{2} \sum_{\vec{e}} c(\vec{e}) \big(\nabla_{\vec{e}} v\big)^2$, and this will be proved useful later on.

1.2.2 Probabilistic interpretations

Let us come back to mathematics. We consider the same network $((\mathfrak{g}, c); x_{\text{in}}, x_{\text{out}})$ as above our goal being to give a probabilistic interpretation to the functions $v, i, \ldots$ that are "physically" built by the battery.

[5]  Georg Simon Ohm (1789-1854)

[6]  James Prescott Joule (1818-1889)

Potential

The easiest quantity to interpret is the potential. Indeed, if we impose a unit voltage between the source and sink vertices of the graph and require Ohm's law and the flow property, then by Theorem 1.1 there is indeed a unique function v satisfying $v(x_{\text{in}}) = 1$, $v(x_{\text{out}}) = 0$ and being harmonic inside $V \backslash \{x_{\text{in}}, x_{\text{out}}\}$ it is given by:

Proposition 1.3 (Probabilistic interpretation of the potential)**.** *If τ_{in} and τ_{out} respectively are the hitting times of x_{in} and x_{out} by the random walk on $(\mathfrak{g}, c)$ then we have for any $x \in V$*

$$v(x) = \mathbb{E}_x[v(X_{\tau_{\text{in}} \wedge \tau_{\text{out}}})] = \mathbb{P}_x(\tau_{\text{in}} < \tau_{\text{out}}).$$

Remark 1.2. Notice that if we impose different boundary conditions on the potential (e.g. by plugging a different battery), then the new potential is obtained by an affine transformation.

Remark 1.3. The probabilistic interpretation of the potential suggests that electrons may behave in the networks as random walkers, but unfortunately the physic knowledge of the author does not permit to affirm such an assertion .... :)

Current and effective resistance

In the real-world situation, we could measure the electric current. In our mathematic setup, we *define* it from the potential:

Definition 1.4. *The current i is defined from the potential as follows: for any oriented edge we put*

$$i(\vec{e}) := -c(\vec{e}) \cdot \nabla_{\vec{e}} v = c(\vec{e}) \cdot (v(\vec{e}_*) - v(\vec{e}^*)).$$

Py Proposition 1.2, it is automatic that i is symmetric, obeys the cycle rule and the flow property. The total flux from x_{in} to x_{out} is then

$$\begin{aligned} i_{\text{total}} &= \sum_{\vec{e}=(x_{\text{in}},\cdot)} i(\vec{e}) \\ &= \sum_{\vec{e}=(x_{\text{in}},\cdot)} c(\vec{e}) \cdot (v(\vec{e}_*) - v(\vec{e}^*)) \\ &= \pi(x_{\text{in}}) \sum_{y \sim x_{\text{in}}} \frac{c(x_{\text{in}}, y)}{\pi(x_{\text{in}})} (1 - \mathbb{P}_y(\tau_{\text{in}} < \tau_{\text{out}})) \\ &= \pi(x_{\text{in}}) \mathbb{P}_{x_{\text{in}}} \left(\tau_{\text{out}} < \tau_{\text{in}}^+\right), \end{aligned}$$

where $\tau_{\text{in}}^+ = \inf\{k \geqslant 1 : X_k = x_{\text{in}}\}$. Motivated by the real-world situation we define:

Definition 1.5 (Effective resistance)**.** *The total current i_{total} is proportional to the potential difference $v(x_{\text{in}}) - v(x_{\text{out}})$ and the proportionallity factor is called the* effective conductance C_{eff} *of the graph $(\mathfrak{g}, c)$ between x_{in} and x_{out} (its inverse is the effective resistance R_{eff}). From the above we have*

$$\mathrm{C}_{\text{eff}} = \mathrm{C}_{\text{eff}}\big((\mathfrak{g}, c); x_{\text{in}} \leftrightarrow x_{\text{out}})\big) = {\mathrm{R}_{\text{eff}}}^{-1} = \pi(x_{\text{in}}) \cdot \mathbb{P}_{x_{\text{in}}}(\tau_{\text{out}} < \tau_{\text{in}}^+).$$

Exercise 1.4. It is not clear from Definition 1.5 that the effective conductance is a symmetric expression in x_{in} and x_{out} (i.e. there role can be interchanged). This is intuitively obvious if our interpretations are correct, but prove it only using the reversibility of the random walk.

We can give a quick interpretation of the effective resistance in terms of so-called the Green[7] function of the random walk: For any $x \in V$ let $\tilde{\mathcal{G}}(x)$ be the expected number of visits to x by the random walk strictly before τ_{out}. In particular $\tilde{\mathcal{G}}(x_{\text{out}}) = 0$ and since the number of visits to x_{in} is a geometric random variable with success parameter $\mathbb{P}_{x_{\text{in}}}(\tau_{\text{in}}^+ < \tau_{\text{out}})$ we have

$$\tilde{\mathcal{G}}(x_{\text{in}}) = \frac{1}{1 - \mathbb{P}_{x_{\text{in}}}(\tau_{\text{in}}^+ < \tau_{\text{out}})} = \frac{1}{\mathbb{P}_{x_{\text{in}}}(\tau_{\text{in}}^+ > \tau_{\text{out}})} = \pi(x_{\text{in}})\mathrm{R}_{\text{eff}}.$$

Hence the effective resistance is, up to the weight of $\pi(x_{\text{in}})$ the mean number of visits of x_{in}, before τ_{out}. This interpretation still holds for any $v \in V$ if we add the potential in the game.

Definition 1.6. *The Green function in* $(\mathfrak{g}, c)$ *started at* x_{in} *and killed at* x_{out} *is the function*

$$\mathcal{G}(x) = \frac{\tilde{\mathcal{G}}(x)}{\pi(x)}.$$

Lemma 1.4. *For all* $x \in V$ *we have* $\mathcal{G}(x) = \mathrm{R}_{\text{eff}} \cdot v(x)$ *where* v *is the potential normalized so that* $v(x_{\text{in}}) = 1$ *and* $v(x_{\text{out}}) = 0$.

Proof. Since we know the boundary conditions for $\mathcal{G}(\cdot)$ the lemma follows from the uniqueness in the Dirichlet problem (Theorem 1.1) as soon as we have proved that the function $\mathcal{G}$ is harmonic on $V \backslash \{x_{\text{in}}, x_{\text{out}}\}$. To prove this observe that if $x \notin \{x_{\text{in}}, x_{\text{out}}\}$ we have

$$\begin{aligned}
\mathcal{G}(x) = \frac{1}{\pi(x)} \mathbb{E}_{x_{\text{in}}}\left[\sum_{k=1}^{\infty} \mathbf{1}_{X_k = x} \mathbf{1}_{\tau_{\text{out}} > k}\right] \quad &= \quad \frac{1}{\pi(x)} \sum_{y \sim x} \mathbb{E}_{x_{\text{in}}}\left[\sum_{k=1}^{\infty} \mathbf{1}_{X_{k-1} = y} \mathbf{1}_{X_k = x} \underbrace{\mathbf{1}_{\tau_{\text{out}} > k}}_{\mathbf{1}_{\tau_{\text{out}} > k-1}}\right] \\
&\underset{\text{Markov}}{=} \quad \frac{1}{\pi(x)} \sum_{y \sim x} \tilde{\mathcal{G}}(y) \mathbf{p}(y, x) \\
&\underset{\text{reversibility}}{=} \quad \frac{1}{\pi(x)} \sum_{y \sim x} \tilde{\mathcal{G}}(y) \mathbf{p}(x, y) \frac{\pi(x)}{\pi(y)} \quad \Rightarrow \text{ harmonicity of } \mathcal{G}. \quad \square
\end{aligned}$$

The last lemma enables us to give a wonderful interpretation of the resistance:

Theorem 1.5 (*Commute time captures the resistance [11]*)

We have the following identity

$$\mathrm{R}_{\text{eff}} \cdot \sum_{\vec{e} \in \vec{E}} c(\vec{e}) = \underbrace{\mathbb{E}_{x_{\text{in}}}[\tau_{\text{out}}] + \mathbb{E}_{x_{\text{out}}}[\tau_{\text{in}}]}_{\textit{commute time between } x_{\text{in}} \textit{ and } x_{\text{out}}}.$$

[7]  George Green (1793-1841)

Proof. With the notation of the preceding proof we have

$$\mathbb{E}_{x_{\text{in}}}[\tau_{\text{out}}] = \sum_{x \in V} \tilde{\mathcal{G}}(x) = \mathrm{R}_{\text{eff}} \sum_{x \in V} \pi(x) \cdot v(x),$$

where v is the potential equal to 1 at x_{in} and 0 at x_{out}. Reversing the roles of x_{in} and x_{out} ends up in changing $v(\cdot)$ into $1 - v(\cdot)$. Summing the corresponding equalities, we get

$$\mathbb{E}_{x_{\text{in}}}[\tau_{\text{out}}] + \mathbb{E}_{x_{\text{out}}}[\tau_{\text{in}}] = \mathrm{R}_{\text{eff}} \cdot \sum_{x \in V} \pi(x). \quad \square$$

Remark 1.4. This commute time identity reflects the fact that the effective resistance is in fact symmetric in exchanging the roles of x_{in} and x_{out} (see Exercise 1.4). It gives a practical and a theoretical tool to interpret resistance.

Exercise 1.5. For $n \geqslant 2$, let $\mathbb{K}_n$ be the complete graph on vertices $\{0, 1, \ldots, n-1\}$, i.e. with an edge between each pair of vertices $0 \leqslant i \neq j \leqslant n-1$. All the conductances are set to 1.

1. Compute $\mathrm{R}_{\text{eff}}((\mathbb{K}_n, c \equiv 1); 0 \leftrightarrow 1)$.

2. Using the commute-time identify, deduce the expected hitting time of 1 for the random walk on $\mathbb{K}_n$ started from 0.

We now move on to a probabilistic interpretation of the current. Recall that the edge-random walk starting from x_{in} is a sequence of oriented edges $(\vec{E}_n : n \geqslant 1)$. We denote $S(\vec{e})$ the number of times the random walk has gone through $\vec{e}$ in that particular direction until we first reach the vertex x_{out}. Then we have :

Proposition 1.6 (Probabilistic interpretation of the current)**.** *For any $\vec{e} \in \vec{E}$ we have*

$$\mathbb{E}_{x_{\text{in}}}\left[S(\vec{e}) - S(\overleftarrow{e})\right] \cdot i_{\text{total}} = i(\vec{e}).$$

Proof. Observe that for a given oriented edge $\vec{e}$ we have

$$\begin{aligned} \mathbb{E}[S(\vec{e})] &= \sum_{k=0}^{\infty} \mathbb{P}_{x_{\text{in}}}(X_k = \vec{e}_* \text{ for } k < \tau_{\text{out}} \text{ and } \vec{E}_k = \vec{e}) \\ &= \tilde{\mathcal{G}}(\vec{e}_*) \frac{c(\vec{e})}{\pi(x)} \underset{\text{Lemma 1.4}}{=} c(\vec{e}) \cdot \mathrm{R}_{\text{eff}} \cdot v(\vec{e}_*). \end{aligned}$$

Hence we have $\mathbb{E}_{x_{\text{in}}}\left[S(\vec{e}) - S(\overleftarrow{e})\right] \cdot i_{\text{total}} = c(e) \cdot (v(\vec{e}_*) - v(\vec{e}^{\,*})) = i(\vec{e})$ by definition. $\square$

Exercise 1.6. Let $((\mathfrak{g}, \mathfrak{c}); x_{\text{in}}, x_{\text{out}})$ be a network. Let T be a random spanning tree of $\mathfrak{g}$ chosen with a probability proportional to

$$\mathbb{P}(T = \mathfrak{t}) \propto \prod_{e \in \mathfrak{t}} c(e).$$

We write $\{\vec{e} \in T\}$ for the event where the unique path from x_{in} to x_{out} in T traverses the edge $\vec{e}$ in that particular direction. Prove that

$$i(\vec{e}) = \big(\mathbb{P}(\vec{e} \in T) - \mathbb{P}(\overleftarrow{e} \in T)\big) \cdot i_{\text{total}},$$

is the electric current. Hint: to show that the cycle rule is satisfied consider a random forest i.e. a pair $(T_{\text{in}}, T_{\text{out}})$ of trees attached to x_{in} and x_{out} whose weight is proportional to the product of the conductances of the edges. Show that $\mathbb{P}((a,b) \in T)$ is proportional to

$$\mathbb{P}(a \in T_{\text{in}}, b \in T_{\text{out}} \text{ and } T_{\text{in}} \cup T_{\text{out}} \cup \{a,b\} \text{ is a spanning tree}).$$

Reminder on excursion theory. If $(\mathfrak{g}, c)$ is a weighted graph, the trajectory of a random walk starting from x can be decomposed in a succession of excursions away from x. More precisely, if $0 = \tau_x^{(0)} < \tau_x^{+} = \tau_x^{(1)} < \tau_x^{(2)} < \tau_x^{(3)} < \cdots$ are the successive return times of the walk to the point x then the excursions

$$\mathcal{E}^{(i)} = \left(X_{\tau_x^{(i)}+k}\right)_{0 \leqslant k \leqslant \tau_x^{(i+1)} - \tau_x^{(i)}},$$

are independent identically distributed excursions (piece of a trajectory) starting and ending at x. This is valid more generally as soon as the random walk is recurrent and is key to prove that the invariant measure (unique up to multiplicative constant) is given by

$$\mu(y) = \mathbb{E}_x\left[\sum_{k=1}^{\tau_x^+} \mathbf{1}_{X_k = y}\right]. \tag{1.1}$$

In particular, when the invariant measure is finite we deduce that

$$\frac{\pi(x)}{\pi(V)} = \frac{1}{\mathbb{E}_x[\tau_x^+]}. \tag{1.2}$$

The last formula and the excursion decomposition are used in the following exercises:

Exercise 1.7 (From [34]). Let G be a finite connected graph with n edges having all conductances equal to 1 and $(W_k)_{k \geqslant 0}$ be the associated random walk which starts from $x \in V(G)$ under $\mathbb{P}_x$. We write $\tau_x^+ = \inf\{k \geqslant 1 : W_k = x\}$ for the first return time to x. Since the invariant measure on vertices for the random walk is proportional to the degree, we have by (1.2)

$$\mathbb{E}_x[\tau_x^+] = \frac{2n}{\deg_G(x)}. \tag{1.3}$$

The goal of the exercise is to give an "electrical" proof of this well-known fact. To do this, we consider the graph $\tilde{G}$ obtained from G by adding a new vertex $\tilde{x}$ attached to x via a single edge of conductance 1. We denote by $\tau_{\tilde{x}}$ the hitting time of $\tilde{x}$ by the random walk on $\tilde{G}$. For clarity we denote by $\mathbb{E}$ the expectation under which the random walk moves on G and by $\mathbf{E}$ the expectation under which it moves along $\tilde{G}$.

1. Show carefully that

$$\mathbf{E}_x[\tau_{\tilde{x}}] = \frac{1}{\deg_G(x)+1} \sum_{k=0}^{\infty} \left(\frac{\deg_G(x)}{\deg_G(x)+1}\right)^k \left(k \cdot \mathbb{E}_x[\tau_x^+] + 1\right).$$

2. Conclude using the commute time identity.

Exercise 1.8. Let $((\mathfrak{g},c);x_{\text{in}},x_{\text{out}})$ be a network. Consider τ_{in}^{+} the first return time at x_{in} and $\sigma_{\text{in,out}}$ the first return time to x_{in} after visiting x_{out}, that is $\mathbb{E}_{x_{\text{in}}}[\sigma_{\text{in,out}}]$ is the commute time between x_{in} and x_{out}. Show that $\tau_{\text{in}}^{+} \leqslant \sigma_{\text{in,out}}$ and that

$$\mathbb{E}_{x_{\text{in}}}[\sigma_{\text{in,out}} - \tau_{\text{in}}^{+}] = \mathbb{P}_{x_{\text{in}}}(\tau_{\text{in}}^{+} < \tau_{\text{out}})\mathbb{E}_{x_{\text{in}}}[\sigma_{\text{in,out}}].$$

Re-deduce Theorem 1.5.

Exercise 1.9. Let $(\mathfrak{g},c)$ be a weighted graph with three distinguished points $x,a,b \in \mathrm{V}(\mathfrak{g})$. By decomposing the random walk into independent and identically distributed excursions from x show that $\mathbb{P}_x(\tau_a < \tau_b) = \mathbb{P}_x(\tau_a < \tau_b \mid \tau_{\{a,b\}} < \tau_x^{+})$ and deduce

$$\mathbb{P}_x(\tau_a < \tau_b) \leqslant \frac{\mathrm{C_{eff}}(x \leftrightarrow a)}{\mathrm{C_{eff}}(x \leftrightarrow \{a,b\})}.$$

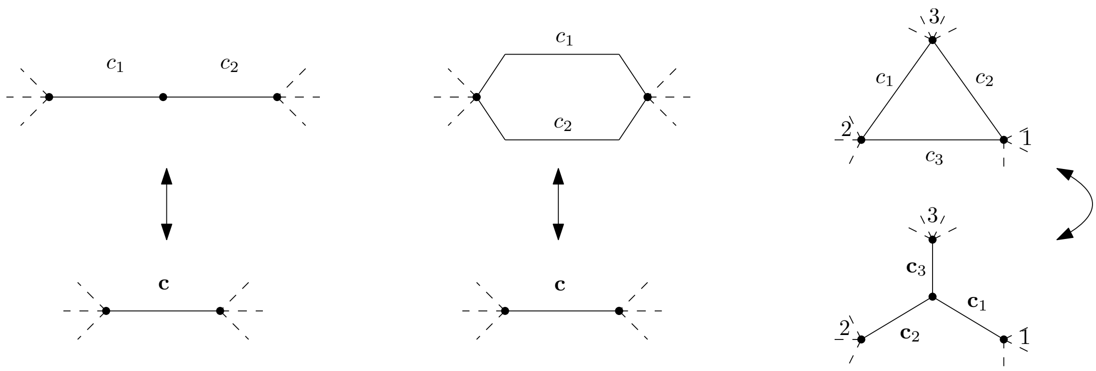
Exercise 1.10. Let $(\mathfrak{g},c)$ be a finite network with n vertices. Using (1.2) show that

$$\sum_{e \in E(\mathfrak{g})} i^{e}(e) = n-1$$

where i^e denotes the unit current flow from the origin to the extremity of e.

1.2.3 Equivalent networks

Proposition 1.7. *In a network we can perform the following operations without affecting the effective resistance (and in fact without changing the potential and the current outside of the current zone of transformation):*



$\mathbf{c} = (c_1^{-1} + c_2^{-1})^{-1}$ $\qquad$ $\mathbf{c} = c_1 + c_2$ $\qquad$ $c_i \mathbf{c}_i = \dfrac{\mathbf{c}_1\mathbf{c}_2\mathbf{c}_3}{\mathbf{c}_1 + \mathbf{c}_2 + \mathbf{c}_3}$

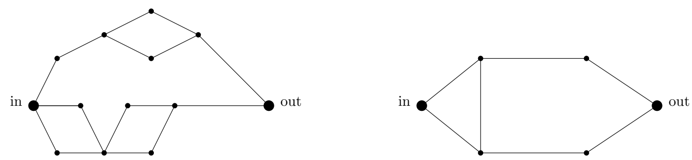
Figure 1.3: The series and parallel transformations. The last one is known as the star-triangle transformation.

Proof. For the series rule, the current flowing through the two edges must be the same hence the resistances add up. For the parallel rule, the potential difference between the two edges are the same hence the conductances add up. For the star-triangle transformation, one has to check that whatever the potentials (v_1, v_2, v_3) the current flowing from the three apexes are the same in the two circuits. If α denotes the value of the potential in the middle of the "star" replacing the triangle we must have

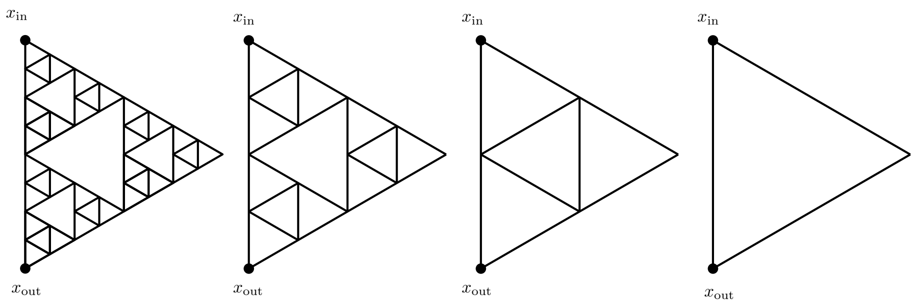
$$\begin{aligned}
(\mathbf{c}_1 + \mathbf{c}_2 + \mathbf{c}_3)\alpha &= \mathbf{c}_1 v_1 + \mathbf{c}_2 v_2 + \mathbf{c}_3 v_3 \\
&\text{and} \\
(v_1 - \alpha)\mathbf{c}_1 &= (v_1 - v_2)c_3 + (v_1 - v_3)c_2, \\
(v_2 - \alpha)\mathbf{c}_1 &= (v_2 - v_1)c_3 + (v_2 - v_3)c_1, \\
(v_3 - \alpha)\mathbf{c}_1 &= (v_3 - v_1)c_2 + (v_3 - v_2)c_1,
\end{aligned}$$

Since these equality must be true for all v_1, v_2, v_3 this indeed imposes $c_i\mathbf{c}_i = \mathbf{c}_1\mathbf{c}_2\mathbf{c}_3/(\mathbf{c}_1+\mathbf{c}_2+\mathbf{c}_3)$ (we leave as an exercise to see that this defines uniquely the $\mathbf{c}_i$'s). □

Exercise 1.11 (From [33]). Find $\mathbb{P}(\tau_{\text{out}} < \tau_{\text{in}}^+)$ in the two networks where all conductances are equal to 1:



Exercise 1.12. Compute the effective resistance in the following sequence of networks (the Sierpinski triangles):



1.3 Energy and variational principles

We finally explore the mathematical interpretation of the "Joule heating". Rather than a probabilistic interpretation, this will motivate the introduction of a functional, the energy – physically this is rather a power–, which the electrical potential minimizes. This is a general concept in physics : we can usually describe the behavior of some process by local laws (here Ohm's law, flow and cycle rules etc) or by the minimization of some potential "the least action principle". The outcome in maths are *variational principles* for the potential/current which are softer to use and in particular enables us to prove *inequalities* instead of computing exactly the quantities of interests.

1.3.1 Energy

Definition 1.7. *Let $((\mathfrak{g}, c); x_{\text{in}}, x_{\text{out}})$ be a finite network, $j : x_{\text{in}} \to x_{\text{out}}$ be a flow in the graph and $f : V(\mathfrak{g}) \to \mathbb{R}$ be a function on its vertices. The energy of the flow j and of the function f are by definition (with a slight abuse of notation)*

$$\mathcal{E}(j) \underset{\text{def.}}{:=} \frac{1}{2}\sum_{\vec{e}\in\vec{E}} j(\vec{e})^2/c(\vec{e}) \qquad \mathcal{E}(f) \underset{\text{def.}}{:=} \frac{1}{2}\sum_{\vec{e}\in\vec{E}} c(\vec{e})\cdot\big(\nabla_{\vec{e}} f\big)^2.$$

Notice that the factors $1/2$ are present in the definition to count each non-oriented edge once in the definition (this is only a matter of convention). In the case of the electric current i and the associated electric potential v, the two energies agree and are

$$\begin{aligned}
\mathcal{E}(i) = \mathcal{E}(v) &= \sum_{e\in E} i(e)^2/c(e) \\
&= \frac{1}{2}\sum_{\vec{e}\in\vec{E}} c(\vec{e})(\nabla_{\vec{e}} v)^2 \\
&= \sum_{x\in V} v(x) \sum_{\vec{e}=(x,y)} c(\vec{e})(v(x)-v(y)) \\
&\underset{\substack{v(x_{\text{out}})=0 \\ \text{harmonicity}}}{=} v(x_{\text{in}}) \sum_{\vec{e}=(x_{\text{in}},\cdot)} c(\vec{e})(v(x_{\text{in}})-v(\vec{e}^*)) \\
&= \Delta v\cdot i_{\text{total}} = \mathrm{R}_{\text{eff}}\cdot(i_{\text{total}})^2 = \mathrm{C}_{\text{eff}}\cdot(\Delta v)^2. \qquad (1.4)
\end{aligned}$$

1.3.2 Variational principles

We now give two variational principles for the energy:

Theorem 1.8 (*Thomson's principle*)

If $j : x_{\text{in}} \to x_{\text{out}}$ is a flow of the same total flux as i then $\mathcal{E}(i) \leqslant \mathcal{E}(j)$ with equality if and only if $i = j$. In words, the electric current minimizes the energy for a given flux and

$$\mathrm{R}_{\text{eff}}((\mathfrak{g}, c); x_{\text{in}} \leftrightarrow x_{\text{out}}) = \min\{\mathcal{E}(j) : j : x_{\text{in}} \to x_{\text{out}} \text{ unit flow}\}.$$

Proof. By assumptions j satisfies the flow rule and has the same flux as i. We consider the flow $i - j$ which has then 0 flux. It is easy to see that we can decompose any 0-flux flow into a sum of currents along oriented cycles. However, it is easy to see that any current along a cycle is orthogonal (for the scalar product whose normed squared is the energy) with respect to i.

Hence we get that

$$\mathcal{E}(j) = \mathcal{E}(i + (j - i)) = \mathcal{E}(i) + \mathcal{E}(j - i) + 2 \underbrace{\sum_{\text{cycles}} (i \mid j - i)}_{=0} \geqslant \mathcal{E}(i),$$

and with equality if and only if $i = j$ as Thomson[8] wanted. The display in the theorem then follows from (1.4). □

Theorem 1.9 (*Dirichlet's principle*)

If $f : \mathrm{V}(\mathfrak{g}) \to \mathbb{R}$ such that $f(x_{\mathrm{in}}) = v(x_{\mathrm{in}})$ and $f(x_{\mathrm{out}}) = v(x_{\mathrm{out}})$ is a function with the same boundary conditions as the electrical potential then $\mathcal{E}(v) < \mathcal{E}(f)$ with equality if and only if $v = f$. In words, the electric potential minimizes the energy for a given boundary condition

$$\mathrm{C}_{\mathrm{eff}}((\mathfrak{g}, c); x_{\mathrm{in}} \leftrightarrow x_{\mathrm{out}}) = \min\{\mathcal{E}(f) : f : \mathrm{V}(\mathfrak{g}) \to \mathbb{R} \text{ such that } f(x_{\mathrm{in}}) = 1 \text{ and } f(x_{\mathrm{out}}) = 0\}.$$

Proof. The proof is almost the same as for Thomson's principle and reduces to showing that for the scalar product whose norm is $\mathcal{E}(\cdot)$ the set of 0-boundary condition functions f is orthogonal to the set of functions which are harmonic except at the boundary. We leave the details as an exercise for the reader. □

The great advantage of the last two results is that if we want upper and lower bounds on the resistance, we just need to produce a flow or a function rather than computing exactly the value of the resistance. This flexibility will be used a lot in the next chapter. Here is some "trivial" corollary whose physical interpretation is clear, but whose proof is far from obvious without the tools developed in this chapter!:

Corollary 1.10 (Rayleigh monotonicity)**.** *The effective conductance is a non-decreasing function of each conductance of the graph.*

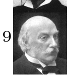
Proof. Let $c \leqslant c'$ two conductances on the same graph $\mathfrak{g}$. We write i and i' respectively for the electrical current carrying a unit flux (in particular, the two potential differences may be different). We write $\mathcal{E}_c$ for the energy relative to the conductances c. Then we have

$$\mathrm{R}_{\mathrm{eff}}((\mathfrak{g}, c); x_{\mathrm{in}} \leftrightarrow x_{\mathrm{out}}) = \mathcal{E}_c(i) \underset{c \leqslant c'}{\geqslant} \mathcal{E}_{c'}(i) \underset{\text{Thm.1.8}}{\geqslant} \mathcal{E}_{c'}(i') = \mathrm{R}_{\mathrm{eff}}((\mathfrak{g}, c'); x_{\mathrm{in}} \leftrightarrow x_{\mathrm{out}}).$$

□

Remark 1.5. The result of Rayleigh[9] is very useful : we can modify the graph in order to estimate the effective resistance. In particular, if the conductance of an edge is put to 0 this boils down to just removing the edge from the graph whereas if its conductance is set to ∞ it is equivalent to identifying its extremities.

[8]  Joseph John Thomson (1856-1940)

[9]  John William Strutt, 3rd Baron Rayleigh (1842–1919)

Exercise 1.13. Suppose that the resistances of the finite graph $\mathfrak{g}$ denoted by $(r(e), e \in \mathrm{E}(\mathfrak{g}))$ are variables. The effective resistance $\mathrm{R_{eff}}(r) = \mathrm{R_{eff}}((\mathfrak{g}, r^{-1}); x_{\mathrm{in}}, x_{\mathrm{out}}))$ is thus a function of $(r(e) : e \in \mathrm{E}(\mathfrak{g}))$.

1. Show that the derivative of $\mathrm{R_{eff}}$ with respect to $r(e)$ is given by $i(e)^2$ where i is the unit-flow electrical current going through e.

2. Deduce that $\mathrm{R_{eff}}$ is a concave function of each resistance.

Exercise 1.14 (From Arvind Singh). (*) Let $(\mathfrak{g}, c; x_{\mathrm{in}}, x_{\mathrm{out}})$ be a finite network and let $i = x_{\mathrm{in}} \to x_{\mathrm{out}}$ the unit-flux electrical current flowing from x_{in} to x_{out}. Show that for every edge $e \in \mathrm{E}(\mathfrak{g})$ we have $|i(e)| \leqslant 1$.

1.4 Discrete Gaussian Free Field

In this section we introduce the discrete Gaussian free field. Although not used in the rest of these lecture notes, this random object is central in probability theory and especially in planar random geometry. We will see that its definition is intimately connected with the electric notions that we developed so far and this gives us an excuse to shed new light on a few calculations we performed in the preceding pages. In the remaining of this section $(\mathfrak{g}, c)$ is a finite weighted graph and $\partial\mathfrak{g}$ is a set of (at least one) distinguished vertices of $\mathfrak{g}$ (previously we had $\partial\mathfrak{g} = \{x_{\mathrm{in}}, x_{\mathrm{out}}\}$).

1.4.1 The "physics" definition

Given $((\mathfrak{g}, c); \partial\mathfrak{g})$, the discrete Gaussian Free Field (GFF) with zero boundary condition on $\partial\mathfrak{g}$ is a random function $H : \mathsf{V}(\mathfrak{g}) \to \mathbb{R}$ which vanishes on $\partial\mathfrak{g}$. It models the fluctuations of a membrane or a system of springs. As usual in physics, the probability distribution is related to the potential energy of the system through the formalism of Gibbs measure, in our case this boils down to considering the following density

$$\frac{1}{C} \exp\left(-\frac{1}{2} \cdot \frac{1}{2} \sum_{\vec{e} \in \vec{\mathsf{E}}(\mathfrak{g})} c(\vec{e})(\nabla_{\vec{e}} H)^2 \right), \tag{1.5}$$

with respect to the product Lebesgue measure on $\mathbb{R}^{\mathsf{V}(\mathfrak{g}) \setminus \partial\mathfrak{g}}$ where the coordinates along vertices on the boundary are set to 0. The first $1/2$ in the exponential is here to count each non-oriented edge only once, the second one is the usual factor in the density of the standard Gaussian distribution. The normalization constant $C > 0$ is, at this stage, inexplicit (see Exercise 1.15). This definition immediately raises the question of the individual law of $H(x)$ and the correlations between values at different points. To answer these questions, we first need some background and an equivalent definition of H.

1.4.2 The ℓ^2-definition

Another definition of the discrete GFF goes through the L^2 machinery of Gaussian processes. Consider the space ℓ_0^2 of functions $f : \mathsf{V}(\mathfrak{g}) \to \mathbb{R}$ that are vanishing on $\partial\mathfrak{g}$. It is easy to see that this space can be turned into a Hilbert space of dimension $\#\mathsf{V}(\mathfrak{g}) - \#\partial\mathfrak{g}$ endowed with the inner product

$$< f, g >= \frac{1}{2} \sum_{(x,y)\in\vec{\mathsf{E}}(\mathfrak{g})} c(\vec{e})\nabla_{\vec{e}} f \cdot \nabla_{\vec{e}}\, g,$$

where the factor $1/2$ is here to count each non-oriented edge once in the definition. Its associated norm is nothing but $\mathcal{E}(f)$, the energy of f seen in the last section. Now, as soon as we are in possession of a finite dimensional Hilbert space, one can canonically construct a Gaussian random variable on it:

Definition 1.8. *If $(E, < \cdot, \cdot >)$ is a finite dimensional Hilbert space with an orthonormal basis $(e_1, \dots, e_n)$. Then the following random variable is well-defined and its law does not depend on the choice of the orthonormal basis:*

$$H = \sum_{i=1}^{n} \mathcal{N}_i \cdot e_i \in E,$$

where $\mathcal{N}_i$ are independent identically distributed standard real Gaussian variables. The law of H is called the "Gaussian distribution" on $(E, < \cdot, \cdot >)$.

To see that the law of H does not depend on the orthonormal basis, we just have to remark that if $\mathcal{Z}_i$ are the coefficients of H in another orthonormal basis $(f_1, \dots, f_n)$ then these are obtained from $\mathcal{N}_1, \dots, \mathcal{N}_n$ by multiplication by an orthogonal matrix. But by standard properties of Gaussian vectors we then have $(\mathcal{Z}_i)_{1\leqslant i\leqslant n} = (\mathcal{N}_i)_{1\leqslant i\leqslant n}$.

Exercise 1.15. Show that in our case the law of H agrees with (1.5) and prove in particular that the constant $C > 0$ is explicitly given by

$$C = \sqrt{2\pi}^n \det(e_i(x_j))_{1\leqslant i,j\leqslant n},$$

where $x_1, \dots, x_n$ are the n vertices of $\mathfrak{g}\backslash\partial g$ and $e_1, \dots, e_n$ is an orthonormal basis of ℓ_0^2.

1.4.3 Expression of the covariance

By Definition 1.8, the random function H is Gaussian, that is, any linear combination of its entries is Gaussian (and centered). It is then natural to explicit the distribution of H which is encoded in its covariance structure. The starting point is to notice the following integration by part formula (that we already used implicitly in Equation 1.4 of Chapter 1):

$$\begin{aligned} < f, g > &= \frac{1}{2} \sum_{(x,y)\in \vec{\mathsf{E}}(\mathfrak{g})} (f(x) - f(y))(g(x) - g(y))c((x, y)) \\ &= \sum_{x\in \mathsf{V}(\mathfrak{g})\backslash\partial\mathfrak{g}} f(x) \sum_{y\sim x} (g(x) - g(y))c(x, y) \\ &= - \sum_{x\in \mathsf{V}(\mathfrak{g})\backslash\partial\mathfrak{g}} f(x)\pi(x)\Delta g(x), \end{aligned}$$

where we introduced the discrete Laplacian operator Δ acting on functions $g : \mathsf{V}(\mathfrak{g}) \to \mathbb{R}$:

$$\Delta g(x) = \sum_{y\sim x} \mathbf{p}(x, y)(g(y) - g(x)) = \mathbb{E}_x[g(X_1) - g(x)].$$

Hence, if $g_x(\cdot)$ is a function such that $\Delta g_x(\cdot)$ is zero inside $\mathsf{V}(\mathfrak{g})\backslash\partial\mathfrak{g}$ except at x where it is equal to $\frac{-1}{\pi(x)}$ then we have $< f, g_x >= f(x)$. It is easy to see that such a function exist, more generally one can always find a function f satisfying

$$(\star) \qquad \begin{cases} f = f_0 & \text{on } \partial\mathfrak{g}, \\ \Delta f = u_0 & \text{on } \mathsf{V}(\mathfrak{g})\backslash\partial\mathfrak{g}, \end{cases}$$

where u_0 and f_0 are two prescribed functions. Indeed, using our knowledge on the Dirichlet problem, we easily deduce that the above problem has at most one solution (since the difference of two solutions solves the Dirichlet problem with zero boundary condition). The existence can also be deduced using similar lines as those developed in Chapter 1 and the reader can check that a solution can be represented probabilistically as follows (exercise!)

$$f(x) = \mathbb{E}_x \left[f_0(X_{\tau_{\partial\mathfrak{g}}}) - \sum_{k=0}^{\tau_{\partial\mathfrak{g}}-1} u_0(X_k) \right].$$

When $f_0 = 0$ and $u_0(\cdot) = \delta_{x,\cdot}$ we recover the fact that the function $\mathbb{E}_x[\#\text{ visits to } y \text{ before } \tau_{\partial\mathfrak{g}}]$ is harmonic in y except at $y = x$ (this was already observed in Lemma 1.4) and if we recall the definition of the Green function (Definition 1.6) where the dependence in the starting point is now stressed:

$$\mathcal{G}(x, y) = \frac{1}{\pi(y)} \mathbb{E}_x[\#\text{ visits to } y \text{ before } \tau_{\partial\mathfrak{g}}],$$

then we have $\mathcal{G}(x, y) = \mathcal{G}(y, x)$ and $\mathcal{G}$ is harmonic in both coordinates except on the diagonal where its Laplacian equal $\frac{-1}{\pi(\cdot)}$. We can thus come back to our quest and prove

Proposition 1.11. *The covariance structure of the discrete Gaussian free field is given by*

$$\mathsf{Cov}(H(x), H(y)) = \mathcal{G}(x, y).$$

In particular, we have $\mathbb{E}[H(x)^2] = R_{\text{eff}}((\mathfrak{g}, c); x \leftrightarrow \partial\mathfrak{g})$.

Proof: Fix $x, y \in \mathsf{V}(\mathfrak{g})\backslash\partial\mathfrak{g}$. By the above reasoning we deduce that $\mathcal{G}(x,\cdot)$ and $\mathcal{G}(y,\cdot)$ are two functions of ℓ^2_0 which are harmonic in $\mathsf{V}(\mathfrak{g})\backslash\partial\mathfrak{g}$ except respectively at x and y where they satisfy $\Delta\mathcal{G}(x,\cdot)|_x = \frac{-1}{\pi(x)}$ and $\Delta\mathcal{G}(y,\cdot)|_y = \frac{-1}{\pi(y)}$. By the above remark we have

$$H(x) = -\sum_z H(z)\pi(z)\Delta g_x(z) = < H, g_x >,$$

$$H(y) = -\sum_z H(z)\pi(z)\Delta g_y(z) = < H, g_y >,$$

so that

$$\begin{aligned}
\mathsf{Cov}(H(x), H(y)) &= \mathbb{E}[H(x)H(y)] \\
&= \mathbb{E}\left[\left\langle \sum_{i=1}^n \mathcal{N}_i e_i | \mathcal{G}(x,\cdot)\right\rangle \left\langle \sum_{i=1}^n \mathcal{N}_i e_i | \mathcal{G}(y,\cdot)\right\rangle\right] \\
&= < \mathcal{G}(x,\cdot) | \mathcal{G}(y,\cdot) > \\
&= -\sum_z \mathcal{G}(x,z)\pi(z)\, \Delta\mathcal{G}(y,\cdot)|_z \\
&= \mathcal{G}(x,y).
\end{aligned}$$

When $x = y$ we have $\mathcal{G}(x,x) = \mathrm{R_{eff}}((\mathfrak{g}, c); x \leftrightarrow \partial\mathfrak{g})$ as proved in Lemma 1.4. □

1.4.4 The statistical definition

The action (1.5) defining the discrete GFF also pops up in statistics. Assume that on each vertex of your graph $\mathfrak{g}$ is a mountain, a peak, whose altitude you want to measure. Alas, you only know that the peaks on $\partial\mathfrak{g}$ are at height 0 and you are only able to measure altitude differences along edges of your graph (once for each edge)... but with an independent Gaussian error with variance $1/c(e)$. Performing all these measurements you end-up with a collection of noisy height differences

$$\hat{\nabla}(\vec{e}) = \mathcal{N}_{\vec{e}} + \nabla_{\vec{e}} h_0,$$

where the function h_0 is the true height and the errors are independent for each non-oriented edges and so that $\mathcal{N}_{\vec{e}} = -\mathcal{N}_{\overleftarrow{e}}$. What is then the best estimator of the true height of the peaks? The maximum likelihood estimator returns the function h minimizing the Gaussian errors i.e.

$$\mathrm{argmin}_h \exp\left(\frac{1}{4}\sum_{\vec{e}\in\vec{\mathsf{E}}(\mathfrak{g})} c(e)\left(\hat{\nabla}(\vec{e}) - \nabla_{\vec{e}} h\right)^2\right),$$

so that if $h = h_0 + H$ the (random) function H minimizes

$$\sum_{\vec{e}\in\vec{\mathsf{E}}(\mathfrak{g})} c(\vec{e})\left(\nabla_{\vec{e}} H - \mathcal{N}_{\vec{e}}\right)^2. \tag{1.6}$$

Proposition 1.12. *The function H minimizing (1.6) is distributed as the discrete Gaussian Free Field on $((\mathfrak{g}, c); \partial\mathfrak{g})$.*

Proof. The Hilbert space ℓ_0^2 can be embedded in the larger Hilbert space ℓ^2_{sym} of antisymmetric functions $\theta = \vec{\mathsf{E}} \to \mathbb{R}$ on oriented edges. Indeed on ℓ^2_{sym} we can define the inner product

$$< \theta, \psi >= \frac{1}{2} \sum_{\vec{e}} c(\vec{e}) \theta(\vec{e}) \cdot \psi(\vec{e}),$$

and we embed $f \in \ell_0^2 \hookrightarrow (\nabla_{\vec{e}} f)_{\vec{e}} \in \ell^2_{\text{sym}}$ so that the inner products match. In particular, the minimization problem (1.6) is nothing but the projection onto the smaller ℓ_0^2 space made of those functions on oriented edges that actual comes from "derivatives" of usual functions on vertices. But the function $\boldsymbol{\mathcal{N}} = (\mathcal{N}_{\vec{e}}) \in \ell^2_{\text{sym}}$ is standard Gaussian in $\ell^2(\vec{E})$ and thus its coordinates are i.i.d. standard Gaussian in *any* orthonormal basis. We just have to form such a basis by starting with the basis for ℓ_0^2 used in Definition 1.8 and then complete with a basis of the orthogonal. After the projection, we end-up with a function of ℓ_0^2 (hence the derivative of a function) whose coefficients in the basis of Definition 1.8 are just i.i.d. standard normals as wanted. □

Remark 1.6. The discrete Gaussian Free Field is also related to random walks on $(\mathfrak{g}, c)$ via "Dynkin type" isomorphisms relating the occupation (local) times of a random walk to version of the discrete Gaussian free field, see [31]. More generally, the (discrete and continuous) Gaussian free field pops up in many context such as Liouville conformal field theory in dimension 2, uniform spanning trees, loop erased random walks etc. The interested reader should consult the bible [33] and the references therein for much more on this wonderful object.

Bibliographical notes. The connection between random walks on weighted graphs and potential theory is a classic subject covered in e.g. in [33, **?**, **?**, 41, **?**, **?**]. See also [16] for a smooth introduction. Our presentation is mostly inspired from [33] from which we borrowed Exercises .

Chapter II: Infinite graphs and recurrence/transience

In this chapter we use the connection between potential theory, electrical networks and random walks to give robust and practical recurrence/transience criteria for infinite graphs. Here $(\mathfrak{g}, c)$ is a weighted infinite connected graph whose degrees are all finite. In particular, the vertex set V is countable and the random walk on $(\mathfrak{g}, c)$ is irreducible.

2.1 Recurrence and resistance

If $x_{\mathrm{in}} \in V$ we recall that under $\mathbb{P}_{x_{\mathrm{in}}}$ the process $(X_n)_{n\geqslant 0}$ is the random walk (directed by the conductances c) on the graph $\mathfrak{g}$ and started from $X_0 = x_{\mathrm{in}}$. The classical dichotomy for irreducible countable Markov chains then ensures that either $X_n = X_0$ for infinitely many n's in which case $(\mathfrak{g}, c)$ is called *recurrent*, otherwise $X_n = X_0$ finitely many times (and even $\mathbb{E}[\sum_n \mathbf{1}\{X_n = X_0\}] < \infty$) and the graph is *transient*[1]. To relate these concepts with the effective resistance we consider

$$x_{\mathrm{in}} = \mathfrak{g}_0 \subset \mathfrak{g}_1 \subset \cdots \subset \mathfrak{g}_n \subset \cdots \subset \mathfrak{g},$$

an exhaustion of $\mathfrak{g}$ i.e. an increasing sequence of finite connected subgraphs of $\mathfrak{g}$ such that $\cup \mathfrak{g}_n = \mathfrak{g}$. We denote by $\partial \mathfrak{g}_n$ the set of vertices of $\mathfrak{g}_n$ which have a neighbor in $\mathfrak{g}$ which is not in $\mathfrak{g}_n$. We can interpret $\mathfrak{g}_n$ as a finite network where the conductances are inherited from $\mathfrak{g}$ and where $x_{\mathrm{out}} = \partial \mathfrak{g}_n$ where all the vertices are collapsed into a single vertex. By the result of the last chapter we have

$$\pi(x_{\mathrm{in}})\mathbb{P}_{x_{\mathrm{in}}}(\tau_{\mathrm{out}} < \tau_{\mathrm{in}}^+) = \mathrm{C}_{\mathrm{eff}}((\mathfrak{g}_\mathfrak{n}, c); x_{\mathrm{in}} \leftrightarrow \partial \mathfrak{g}_n) \xrightarrow[n\to\infty]{} \pi(x_{\mathrm{in}})\mathbb{P}_\rho(\tau_{\mathrm{in}}^+ = \infty) := \mathrm{C}_{\mathrm{eff}}((\mathfrak{g}, c); x_{\mathrm{in}} \leftrightarrow \infty).$$

Using Rayleigh's monotonicity it is easy to see that the above definition does not depend on the exhaustion of the graph: if $(\mathfrak{g}'_n)_{n\geqslant 0}$ is another exhaustion then for any $n \geqslant 0$ we can find $m, p \geqslant 0$ such that $\mathfrak{g}_n \subset \mathfrak{g}'_m \subset \mathfrak{g}_p$ and by monotonicity of the conductance

$$\mathrm{C}_{\mathrm{eff}}((\mathfrak{g}_n, c); x_{\mathrm{in}} \leftrightarrow \partial \mathfrak{g}_n) \geqslant \mathrm{C}_{\mathrm{eff}}((\mathfrak{g}'_m, c); x_{\mathrm{in}} \leftrightarrow \partial \mathfrak{g}'_m) \geqslant \mathrm{C}_{\mathrm{eff}}((\mathfrak{g}_p, c); x_{\mathrm{in}} \leftrightarrow \partial \mathfrak{g}_p),$$

hence the limits are the same. We have thus proved:

Proposition 2.1. *The graph $(\mathfrak{g}, c)$ is recurrent if and only if there exists (equivalently : for all) $x_{\mathrm{in}} \in V$ such that the effective conductance between x_{in} and ∞ is equal to 0 (the effective resistance is infinite).*

[1] and the notion does not depend upon $x_{\mathrm{in}} \in V$

Example 2.1. The line $\mathbb{Z}$ is recurrent when all conductances are equal to 1 (all trees with at most two ends are recurrent). By monotonicity of the resistance, if $(\mathfrak{g}, c)$ is recurrent then so is any subgraph of it (with the same conductances): this is not trivial even in the case of subgraphs of $\mathbb{Z}^2$!!!

Exercise 2.1. Show that the complete binary tree is transient.

2.2 Criterions for recurrence/transience

We establish a few useful criteria to determine recurrence/transience.

2.2.1 Nash-Williams cutsets

Instead of giving the Nash-Williams[2] criterion in its full generality (see Proposition 2.3) we illustrate the most common application which readily follows from monotonicity of the effective resistance. Fix $(\mathfrak{g}, c)$ a weighted infinite graph and $x_{\text{in}} \in V$. A *cutset* between x_{in} and ∞ in the graph $\mathfrak{g}$ is a subset Γ of edges such that any path $\gamma : x_{\text{in}} \to \infty$ must pass through Γ. Imagine that we can build in the graph $(\mathfrak{g}, c)$ a sequence of disjoints cutsets $\Gamma_1, \Gamma_2, \dots$ which are nested in the sense that we can contract all the edges of the graph except those on the cutsets and identify vertices to obtain a line graph with parallel edges belonging to $\Gamma_1, \Gamma_2, \dots$, see below

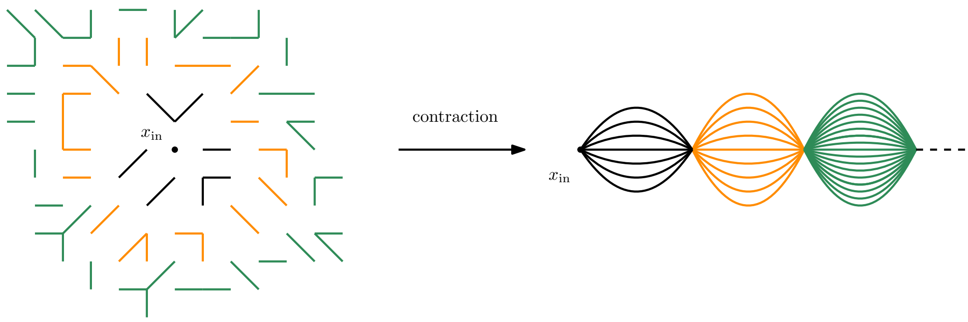


Figure 2.1: Setup of application of the Nash-Williams criterion

It is easy to pass to the limit in Theorem 1.10 and get that the effective resistance $\mathrm{R_{eff}}((\mathfrak{g}, c); x_{\text{in}} \leftrightarrow \infty)$ is monotone in each of the resistance of the graph. However, the previous operation (contraction of edges and identification of vertices) only diminishes the resistance: contracting an edge is the same as setting its resistance to 0 (or its conductance to ∞) and identifying two vertices boils down to adding an edge of resistance 0 between them (we can imagine that before they shared an edge of infinite resistance). Hence we have

$$\mathrm{R_{eff}}((\mathfrak{g}, c); x_{\text{in}} \leftrightarrow \infty) \geqslant \mathrm{R_{eff}}\left(\quad \right) = \sum_{i=1}^{\infty} \left(\sum_{e \in \Gamma_i} c(e) \right)^{-1}.$$

[2]  Crispin St. John Alvah Nash-Williams 1932–2001

Application to $\mathbb{Z}^2$. The last method can be successfully applied in the case of $\mathbb{Z}^2$ (with the obvious edge set and with all conductances equal to 1) : the disjoints cutsets made of the edges in-between $[-n,n]^2$ and $[-(n+1),n+1]^2$ have size $4(2n+1)$ and so the effective resistance between 0 and ∞ in $\mathbb{Z}^2$ is larger than $\sum_{n=0}^{\infty} \frac{1}{4(2n+1)} = \infty$. This proves that $\mathbb{Z}^2$ (as well as any subgraph of it !) is recurrent.

Exercise 2.2. Consider the graph $\mathbb{Z}^3$ made of a stack of horizontal copies of $\mathbb{Z}^2$. We split each vertical edge at height $i \geqslant 1$ in-between two copies of $\mathbb{Z}^2$ into $|i|$ edges in series. Our goal is to prove that the resulting graph G where all the conductances are equal to 1 is recurrent.

1. Prove that we cannot find disjoint cutsets for which we can apply Nash-Williams criterion and deduce recurrence.

We now split the edges in each horizontal $\mathbb{Z}^2$ as follows: for $n \geqslant 1$ each edge connecting $[-n,n]^2$ to $[-(n+1),(n+1)]^2$ is split in n edges in series, each of which has conductance $1/n$. The new weighted graph is denoted by G'.

3. Prove that G and G' are equivalent.

4. Find good cutsets in G' in order to apply Nash-Williams criterion and deduce recurrence.

2.2.2 Variational principles

We now use Theorem 1.8 to give a transience criterion in the infinite setting due to Terry Lyons

Theorem 2.2 (*T. Lyons*)

The graph $(\mathfrak{g},c)$ is transient if and only if there exists (equivalently for all) $x_{\text{in}} \in V$ such that we can create a flow $j : \vec{E} \to \mathbb{R}$ of unit flux whose only source is x_{in} and with finite energy i.e.

$$\mathcal{E}_c(j) = \sum_{e\in E} \frac{j(e)^2}{c(e)} < \infty.$$

Proof. In order to use the result on the finite setting we fix $(\mathfrak{g}_n)_{n\geqslant 0}$ an exhaustion of the graph $\mathfrak{g}$. Suppose first that we possess a flow as in the theorem. By restricting it to $\mathfrak{g}_n$ we obtain a unit flow $x_{\text{in}} \to \partial\mathfrak{g}_n$ whose energy is bounded by the energy of the total flow. Using Theorem 1.8 we deduce that the energy of the unit electric courant $i_n : x_{\text{in}} \to \partial\mathfrak{g}_n$ is also bounded by the same constant, and this proves

$$\forall n, \qquad \mathrm{R}_{\text{eff}}((\mathfrak{g}_n,c); x_{\text{in}} \leftrightarrow \partial\mathfrak{g}_n) \leqslant \mathcal{E}_c(j) < \infty.$$

By passing to the limit $n\to\infty$ we deduce that the effective resistance to ∞ is indeed finite and so the graph is transient.

Conversely, if $(\mathfrak{g},c)$ is transient we know that the unit electric current flow $i_n : x_{\text{in}} \to \partial\mathfrak{g}_n$ has an energy bounded above by the resistance between x_{in} and ∞ in the graph. By taking a diagonal extraction if necessary we can consider a sub sequential limit unit flow $i_n \rightsquigarrow j$. An application of Fatou's lemma then entails that the energy of this flow is again bounded by the resistance between x_{in} and ∞ in the graph. $\square$

Random path. Here is a convenient method to build a unit flux on a graph. Suppose that we dispose of a random infinite oriented path $\vec{\Gamma}$ starting from x_{in}. Erasing loops if necessary, we can suppose that $\vec{\Gamma}$ is a simple path. We then put

$$j(\vec{e}) = \mathbb{P}(\vec{\Gamma} \text{ goes through } \vec{e} \text{ in that direction}).$$

Because $\vec{\Gamma}$ is infinite and starts at x_{in} it is easy to check that j is indeed a unit flow whose only source is x_{in}. Bounding the energy of j reduces to bounding $\sum_{e \in E} \mathbb{P}(e \in \Gamma)^2$.

Application to $\mathbb{Z}^d$ with $d \geqslant 3$. Let $\mathbb{Z}^d$ be the d-dimensional lattice with the usual edge set and conductances equal to 1. We imagine that $\mathbb{Z}^d$ is embedded in $\mathbb{R}^d$ is the natural way. We then consider $\gamma \subset \mathbb{R}^d$ a random semi infinite line starting from 0 such that its intersection with the unit sphere is uniform on the $d-1$ surface. We can then approximate in the discrete γ by a simple oriented path $\vec{\Gamma}$ which stays within a constant distance from γ. It is then easy to see that

$$\sum_{e \in \mathbb{Z}^d} \mathbb{P}(e \in \Gamma)^2 \approx \sum_{e \in \mathbb{Z}^d} (\mathrm{dist}(\mathbf{0}, e))^{2(1-d)} \approx \sum_{r=1}^{\infty} \frac{1}{r^{d-1}} < \infty \text{ whenever } d \geqslant 3,$$

and so by Theorem 2.2 the graphs $\mathbb{Z}^d$ are transient when $d \geqslant 3$.

We can also use the flow criterion in order to prove a strong version of the Nash-Williams criterion without assuming any geometric condition on the cutsets:

Proposition 2.3. *Suppose that $\Gamma_1, \Gamma_2, \dots$ are disjoint cutsets separating x_{in} from ∞ in $\mathfrak{g}$. We then have*

$$\sum_{i \geqslant 1} \left(\sum_{e \in \Gamma_i} c(e) \right)^{-1} = \infty \quad \Rightarrow \quad (\mathfrak{g}, c) \text{ is recurrent.}$$

Proof. If $(\mathfrak{g}, c)$ is transient then by Theorem 2.2 there exists a unit flow j with source x_{in} and finite energy. Clearly the unit flux has to escape through any cutset Γ_i which means that $\sum_{e \in \Gamma_i} |j(e)| \geqslant 1$. On the other hand by Cauchy–Schwarz we have

$$1 \leqslant \left(\sum_{e \in \Gamma_i} |j(e)| \right)^2 \leqslant \left(\sum_{e \in \Gamma_i} j(e)^2 / c(e) \right) \left(\sum_{e \in \Gamma_i} c(e) \right),$$

hence $\left(\sum_{e \in \Gamma_i} j(e)^2/c(e)\right) \geqslant \left(\sum_{e \in \Gamma_i} c(e)\right)^{-1}$ and summing over $i \geqslant 0$ leads a contradiction to the finiteness of the energy. $\square$

Exercise 2.3 (From [33]). If $A \subset \mathbb{Z}^d$ for $d \geqslant 1$ we denote by G_A the subgraph whose vertices are indexed by A and whose edges are inherited from the standard edges in $\mathbb{Z}^d$. All the conductances are set to 1. Let $f : \mathbb{N} \to \mathbb{N}^*$ be a non-decreasing function. We write $A_f = \{(x, y, z) \in \mathbb{Z}^3 : x, y \geqslant 0, 0 \leqslant z \leqslant f(x)\}$.

1. Show that
$$\sum_{n\geqslant 1} \frac{1}{nf(n)} = \infty \quad \Longrightarrow \quad G_{A_f} \quad \text{recurrent.}$$

2. We suppose that $\forall n \geqslant 0, f(n+1) \leqslant f(n)+1$. Show the converse to the last implication (Hint: Consider a random path in G_{A_f} close to $(n, U_1 n, U_2 f(n))$ where U_1, U_2 are i.i.d. uniform over $[0,1]$).

Exercise 2.4 (First passage percolation). Let $\mathfrak{g}$ be an infinite graph. We endow each edge of $\mathfrak{g}$ with an independent exponential edge weight ω_e of parameter 1 which we interpret as a length. The first-passage distance is then defined as

$$\mathrm{d_{fpp}}(x,y) = \inf\left\{ \sum_{e\in\gamma} \omega_e : \gamma \text{ is a path } x \to y \right\}.$$

Show that if $\mathbb{E}[\mathrm{d_{fpp}}(x_{\mathrm{in}}, \infty)] < \infty$ then g is transient (with unit conductances).

As Lyons' criterion follows from a passage to the limit in Thomson criterion in the finite setting (Theorem 1.8), a recurrence criterion can easily be deduced from Dirichlet's formulation of the resistance (Theorem 1.9). We leave the proof as an exercise for the reader.

Theorem 2.4

The graph $(\mathfrak{g}, c)$ is recurrent if and only if there exists (equivalently for all) $x_{\mathrm{in}} \in V$ such that we can create a functions $f_n : V \to \mathbb{R}$ with $f_n(x_{\mathrm{in}}) = 1$ and $f_n = 0$ outside of a finite set of vertices whose energy $\mathcal{E}_c(f_n)$ are tending to 0 as $n \to \infty$.

2.3 Perturbations

In this section we will prove that the concept of recurrence/transience is stable under perturbations of the underlying lattice as long as they are not too severe at large scales. We will restrict ourselves to the case of "bounded geometry" where vertex degrees and conductances are bounded away from 0 and ∞.

2.3.1 Quasi-isometries

Definition 2.1. *Let (E, d) and (F, δ) two metric spaces. A map $\phi : E \to F$ is a* quasi-isometry *if there exist $A, B > 0$ such that*

$$(i) \qquad \forall y \in F, \exists x \in E, \quad \delta(y, \phi(x)) \leqslant B, \qquad \text{“quasi-surjectivity”}$$
$$(ii) \qquad \forall x, x' \in E, \quad \frac{1}{A}\mathrm{d}(x,x') - B \leqslant \delta(\phi(x), \phi(x')) \leqslant A\,\mathrm{d}(x,x') + B \qquad \text{“quasi-isometry”}.$$

If there exists a quasi-isometry between (E, d) and (F, δ) then the two spaces are said to be quasi-isometric. Sometimes this notion is also called rough isometry or coarse isometry. It is an exercise to see that being quasi-isometric is an equivalence relation on all metric spaces.

Example 2.2. The trivial mapping $\mathbb{Z}^d \to \mathbb{R}^d$ shows that they are quasi isometric. All finite metric spaces are trivially quasi isometric to a point. If two graphs $\mathfrak{g}$ and $\mathfrak{g}'$ (endowed with their graph metrics) are quasi isometric then they share the same rough rate of growth of balls. We deduce in particular that $\mathbb{Z}^d$ is not quasi-isometric to $\mathbb{Z}^{d'}$ if $d \neq d'$.

Exercise 2.5 (A few geometric comparisons). (*) We write $\asymp$ for the quasi-isometry relation. All graphs are endowed with the graph distance.

- Prove $\mathbb{T}_3 \asymp \mathbb{T}_4$ where $\mathbb{T}_d$ is the infinite tree where all vertices have degree $d \geqslant 1$.
- Show that $\mathbb{Z}^2 \not\asymp \mathbb{Z} \times \mathbb{N} \asymp \mathbb{N}^2$.
- (Benjamini & Shamov [8]) Show that the bijective quasi-isometries of $\mathbb{Z}$ are within bounded distance for the $\| \cdot \|_\infty$ norm from either identity or -identity.

Exercise 2.6 (Open question of G. Kozma (**)). Does there exist a bounded degree graph which is quasi-isometric to $(\mathbb{R}^2, \| \cdot \|_2)$ where the multiplicative constant A in the quasi-isometry Definition 2.1 is equal to 1?

2.3.2 Invariance with respect to quasi-isometries

Theorem 2.5 (***recurrence is quasi-isometry invariant***)

Let $\mathfrak{g}$ and $\mathfrak{g}'$ two quasi isometric infinite graphs. Suppose that the vertex degrees and the conductances of $\mathfrak{g}$ and $\mathfrak{g}'$ are bounded away from 0 and ∞. Then $(\mathfrak{g}, c)$ is recurrent if and only if $(\mathfrak{g}', c')$ is recurrent.

Proof. Suppose that $(\mathfrak{g}, c)$ is transient and that $\phi : \mathfrak{g} \to \mathfrak{g}'$ is a quasi-isometry. Up to replacing parallel edges and removing loops we can suppose that $\mathfrak{g}$ is a simple graph. By our assumption and Theorem 2.2 there exists a unit flux flow $j : \vec{E}(\mathfrak{g}) \to \mathbb{R}$ from $x_{\mathrm{in}} \to \infty$ whose energy is finite. We will transform this flow into a flow on $\vec{E}(\mathfrak{g}')$. More precisely for each edge $\vec{e} = (x, y) \in \vec{E}(\mathfrak{g})$ we choose an oriented *geodesic* path in $\mathfrak{g}'$ from $\phi(x)$ to $\phi(y)$. We denote this path $\boldsymbol{\phi}(\vec{e})$. We then change j into a flow j' on $\mathfrak{g}'$ by putting:

$$j'(\vec{e}\,') = \sum_{\vec{e} \in \vec{E}(\mathfrak{g})} j(\vec{e}) \mathbf{1}_{\vec{e}\,' \in \boldsymbol{\phi}(\vec{e})}.$$

It is straightforward to see that j' is a unit flow from $\phi(x_{\mathrm{in}})$ to ∞. Let us now compute its energy:

$$\begin{aligned}
\mathcal{E}_{c'}(j') &\leqslant \mathrm{Cst} \sum_{\vec{e}'} j'(\vec{e}')^2 \\
&\leqslant \mathrm{Cst} \sum_{\vec{e}'} \left(\sum_{\vec{e}} j(\vec{e}) \mathbf{1}_{\vec{e}' \in \boldsymbol{\phi}(\vec{e})} \right)^2 \\
&\underset{\text{Cauchy}-\text{Schwarz}}{\leqslant} \mathrm{Cst} \sum_{\vec{e}'} \sum_{\vec{e}} j(\vec{e})^2 \left(\sum_{\vec{e}} \mathbf{1}_{\vec{e}' \in \boldsymbol{\phi}(\vec{e})} \right) \\
&\underset{\text{see below}}{\leqslant} \mathrm{Cst} \sum_{\vec{e}} j(\vec{e})^2 < \infty.
\end{aligned}$$

We have used the fact that there exists M a number (depending on the constants involved in the quasi-isometry and the maximal degrees in the graphs) such that for any $\vec{e}'$ the total number of oriented edges $\vec{e}$ whose "image" $\boldsymbol{\phi}(\vec{e})$ passes through $\vec{e}'$ is bounded by M. This fact is left as an exercise for the reader. $\square$

Part II:

One dimensional random walks

THIS PART IS DEVOTED TO THE STUDY OF THE FOLLOWING OBJECT:

Definition 2.2 (one-dimensional random walk)**.** *Let μ be a probability distribution on $\mathbb{R}$ and consider $X_1, X_2, \dots$ i.i.d. copies of law μ which we see as the increments of the process (S) on $\mathbb{R}$ defined as follows : $S_0 = 0$ and for $n \geqslant 1$*

$$S_n = X_1 + \cdots + X_n.$$

We say that (S) is a one-dimensional random walk with step distribution μ (or one-dimensional random walk for short).

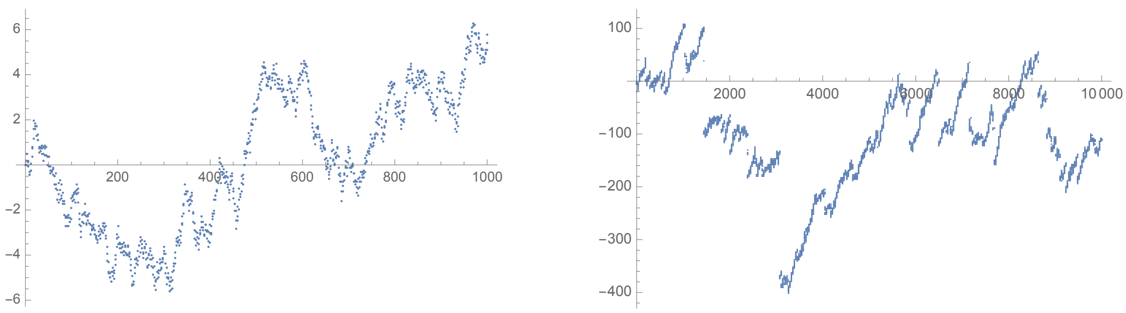


Figure 2.2: Two samples of one-dimensional random walks with different step distributions. The first one seems continuous at large scales whereas the second one displays macroscopic jumps.

The behavior of one-dimensional random walks (recurrence, transience, oscillations...) depends on the step distribution μ in a non-trivial way as we will see. We will develop the fluctuation theory for such walks which studies the return times and entrance heights in half-spaces. This will also enable us to define random walks conditioned to stay positive. This knowledge (especially in the case of skip-free random walks) will later be very useful when random walks appear in the study of random graphs and random trees.

Chapter III: Recurrence and oscillations

In this chapter we fix a law μ on $\mathbb{R}$ whose support is not included in $\mathbb{R}_+$ nor in $\mathbb{R}_-$ and study the one-dimensional random walk (S) with i.i.d. increments (X_i) following this step distribution. This chapter recalls the classic dichotomies recurrence/transience, drifting/oscillating and provides the basic examples for the rest of this part.

3.1 Background and recurrence

Obviously a one-dimensional random walk is a very particular case of Markov chain in discrete time with values in $\mathbb{R}$. We will recall in our context the notion of irreducibility, aperiodicity and (Harris) recurrence.

3.1.1 Lattice walks

Definition 3.1. *We say that the walk is lattice if for some $c > 0$ we have $\mathbb{P}(X_1 \in c\mathbb{Z}) = 1$.*

Remark that when (S) is lattice we have $S_i \in c\mathbb{Z}$ almost surely for every $i \geqslant 0$. We will usually suppose that we have $c = 1$ and that $\gcd(\mathrm{Supp}(\mu)) = 1$ so that (S) induces an irreducible aperiodic Markov chain on $\mathbb{Z}$. The prototype of such walk is the *simple symmetric random walk on* $\mathbb{Z}$ where $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$. When the walk is non lattice (recall that μ is supported neither $\mathbb{R}_+$ nor by $\mathbb{R}_-$) then any real is accessible:

Proposition 3.1. *If the walk is non lattice then*

$$\forall x \in \mathbb{R}, \forall \varepsilon > 0, \exists n \geqslant 0 \quad \mathbb{P}(S_n \in [x \pm \varepsilon]) > 0.$$

Proof. We consider the topological support of μ defined as $\mathrm{Supp}(\mu) = \{x \in \mathbb{R} : \forall \varepsilon > 0, \mu([x - \varepsilon, x + \varepsilon]) > 0\}$. Our goal is to show that $\mathcal{A} = \bigcup_{n \geqslant 0} n \cdot \mathrm{Supp}(\mu)$ is dense in $\mathbb{R}$ where $k \cdot E$ is the kth sum set $E + E + \cdots + E$. Since μ is non lattice the *group* generated by $\mathrm{Supp}(\mu)$ is not discrete, hence it is dense in $\mathbb{R}$. We conclude using the fact that if $\mathcal{A} \not\subset \mathbb{R}_+$ nor $\mathbb{R}_-$ and that the group generated by $\mathcal{A}$ is dense in $\mathbb{R}$ then the semi group generated by $\mathcal{A}$ is also dense in $\mathbb{R}$ (exercise). $\square$

When the walk has a step distribution which has no atoms (it is diffuse) then almost surely the values taken by the random walk are pairwise distinct and in particular $S_i \neq 0$ except for $i = 0$. To see this, fix $0 \leqslant i < j$ and write

$$\mathbb{P}(S_i = S_j) = \mathbb{P}(X_{i+1} + X_{i+2} + \cdots + X_j = 0) = \mathbb{P}(-X_j = X_{i+1} + X_{i+2} + \cdots + X_{j-1}),$$

but since X_j is independent of $X_{i+1} + X_{i+2} + \cdots + X_{j-1}$ and has no atoms this probability is equal to 0. One can then sum over all countable pairs of $i \neq j \geqslant 0$ to get the claim.

3.1.2 Markov property and $0/1$ laws

Before starting with the main course of this chapter, let us recall the very useful Markov property which takes a nice form in our setup: as usual $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is the natural filtration generated by the walk (S) up to time n and a *stopping time* is a random variable $\tau \in \{0, 1, 2, \ldots\} \cup \{\infty\}$ such that for each $n \geqslant 0$ the event $\{\tau = n\}$ is measurable with respect to $\mathcal{F}_n$. In our context the (strong) Markov property can be rephrased as:

Proposition 3.2. *If τ is a stopping time such that $\tau < \infty$ almost surely then the process $(S_n^{(\tau)})_{n \geqslant 0} = (S_{\tau+n} - S_\tau)_{n \geqslant 0}$ is independent of $(S_n)_{0 \leqslant n \leqslant \tau}$ and is distributed as the initial walk $(S_n)_{n \geqslant 0}$.*

Proof. Let f, g be two positive measurable functions and let us compute

$$\begin{aligned}
\mathbb{E}\left[f\left((S_n)_{0 \leqslant n \leqslant \tau}\right) g\left((S_n^{(\tau)})_{n \geqslant 0}\right)\right] &\underset{\tau<\infty}{=} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_{\tau=k} f\left((S_n)_{0 \leqslant n \leqslant k}\right) g\left((S_n^{(k)})_{n \geqslant 0}\right)\right] \\
&\underset{indep.}{=} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_{\tau=k} f\left((S_n)_{0 \leqslant n \leqslant k}\right)\right] \mathbb{E}\left[g\left((S_n^{(k)})_{n \geqslant 0}\right)\right] \\
&\underset{stat.}{=} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_{\tau=k} f\left((S_n)_{0 \leqslant n \leqslant k}\right)\right] \mathbb{E}\left[g\left((S_n)_{n \geqslant 0}\right)\right] \\
&= \mathbb{E}\left[f\left((S_n)_{0 \leqslant n \leqslant \tau}\right)\right] \mathbb{E}\left[g\left((S_n)_{n \geqslant 0}\right)\right],
\end{aligned}$$

and this proves the proposition. $\square$

In the study of random walks, one often uses $0-1$ laws when dealing with asymptotic events such as $\{S_n \to \infty\}$. The most well-known of such laws is Kolmogorov's[1] $0-1$ law which states that if $(X_i)_{i \geqslant 0}$ are independent random variables (not necessarily identically distributed), then any event $\mathcal{A}$ measurable with respect to $\sigma(X_i : i \geqslant 0)$ and which is independent of $(X_1, \ldots, X_{n_0})$ for any n_0 has measure $\mathbb{P}(\mathcal{A}) \in \{0, 1\}$. We give here a stronger version of Kolmogorov $0-1$ law in the case of i.i.d. increments due to Hewitt & Savage[2] which has many applications in the random walk setting:

---

[1]  Andreï Nikolaïevitch Kolmogorov 1903–1987

[2]  Edwin Hewitt 1920–1999,  Leonard Savage 1917–1971

Theorem 3.3 (*Hewitt–Savage exchangeable* $\mathbf{0-1}$ *law*)

Let $(X_i)_{i\geqslant 1}$ be a sequence of independent and identically distributed random variables. Suppose that $\mathcal{A}$ is a measurable event with respect to $\sigma(X_i : i \geqslant 1)$ which is invariant (up to negligible events) by any permutation of the $(X_i : i \geqslant 1)$ with finite support. Then $\mathbb{P}(\mathcal{A}) \in \{0, 1\}$.

Proof. Let $\mathcal{A} \in \sigma(X_i : i \geqslant 1)$ be invariant by any permutations of the X_i with finite support (i.e. only finitely many terms are permuted). By standard measure-theory arguments one can approximate $\mathcal{A}$ by a sequence of events $\mathcal{A}_n \in \sigma(X_1, \dots, X_n)$ in the sense that

$$\mathbb{P}(\mathcal{A}\Delta\mathcal{A}_n) \xrightarrow[n\to\infty]{} 0.$$

Recall that any event $\mathcal{E} \in \sigma(X_i : i \geqslant 1)$ can be written $\mathcal{E} = \mathbf{1}_{(X_i : i\geqslant 1)\in\tilde{\mathcal{E}}}$ where $\tilde{\mathcal{E}}$ is an event of the Borel cylindric σ-field on $\mathbb{R}^{\mathbb{N}}$. We can thus consider the function ψ_n acting on events $\mathcal{E} \in \sigma(X_i : i \geqslant 1)$ by exchanging $X_1, \dots, X_n$ with $X_{n+1}, \dots, X_{2n}$ i.e.

$$\psi_n(\mathcal{E}) = \mathbf{1}_{X_{n+1},\dots,X_{2n},X_1,\dots,X_n,X_{2n+1},\dots\in\tilde{\mathcal{E}}} \in \sigma(X_i : i \geqslant 1).$$

Since the X_i are i.i.d. we have $\mathbb{P}(\psi_n(\mathcal{E})) = \mathbb{P}(\mathcal{E})$ for any event $\mathcal{E}$ and also $\psi_n(\mathcal{A}_n)$ is independent of $\mathcal{A}_n$. Using this we have

$$0 \xleftarrow[n\to\infty]{} \mathbb{P}(\mathcal{A}\Delta\mathcal{A}_n) = \mathbb{P}(\psi_n(\mathcal{A}\Delta\mathcal{A}_n)) = \mathbb{P}(\psi_n(\mathcal{A})\Delta\psi_n(\mathcal{A}_n)) = \mathbb{P}(\mathcal{A}\Delta\psi_n(\mathcal{A}_n)).$$

We deduce that $\mathcal{A}$ is both very well approximated by $\mathcal{A}_n$ but also by $\psi_n(\mathcal{A}_n)$. Since the last two events are independent we deduce that $\mathbb{P}(\mathcal{A}) \in \{0, 1\}$ because

$$\mathbb{P}(\mathcal{A}) = \lim_{n\to\infty} \mathbb{P}(\mathcal{A}_n \cap \psi_n(\mathcal{A}_n)) \underset{\text{indept.}}{=} \lim_{n\to\infty} \mathbb{P}(\mathcal{A}_n)\mathbb{P}(\psi_n(\mathcal{A}_n)) \underset{\text{i.d.}}{=} \lim_{n\to\infty} \mathbb{P}(\mathcal{A}_n)^2 = \mathbb{P}(\mathcal{A})^2. \quad \square$$

Example 3.1. If $A \in \mathbb{R}$ is a measurable subset we write

$$\mathcal{I}_A := \sum_{n=0}^{\infty} \mathbf{1}_{S_n\in A}. \tag{3.1}$$

Then the commutativity of $\mathbb{R}$ (sic) shows that the event $\{\mathcal{I}_A = \infty\}$ is invariant under finite permutations of the X_i's (indeed any finite permutation leaves S_n invariant for large n): hence it has probability 0 or 1. Notice that this cannot be deduced directly from Kolmogorov's $0-1$ law. (This observation is valid in any Abelian group).

3.1.3 Recurrence

In the case of lattice random walks, or random walks on graphs (see Part III), or even irreducible Markov chain on a countable state space, the concept of recurrence is clear: we say that (S) is recurrent if it comes back infinitely often to 0 with probability one. Our first task is to extend this notion in the context of general random walks on $\mathbb{R}$ where the random walk may not even come back exactly at 0 once.

Definition-Proposition 3.1. *The random walk* (S) *is said to be* recurrent *if one of the following equivalent conditions holds:*

(i) *For every* $\varepsilon > 0$ *we have* $\mathbb{P}(\exists n \geqslant 1 : |S_n| < \varepsilon) = 1$,

(ii) $\mathbb{P}(|S_n| < 1$ *for infinitely many* $n) = 1$,

(iii) $\mathbb{P}(|S_n| \to \infty) = 0$,

(iv) $\mathbb{E}\left[\sum_{n=0}^{\infty} \mathbf{1}_{|S_n|<1}\right] = \infty$.

Otherwise the walk is said to be transient *and the complementary events hold.*

Notice that in the case of random walks on $\mathbb{Z}$, or more generally the lattice case (even in the case of any irreducible Markov chain on $\mathbb{Z}$), the above conditions reduce to the well-know equivalences:

$$\mathbb{P}(|S_n| \to \infty) = 0 \iff \mathbb{P}(\exists n \geqslant 1 : S_n = 0) = 1 \iff \mathbb{E}[\#\text{returns to } 0] = \infty.$$

Proof. We consider only the non-lattice case. Let us suppose (i) and prove (ii). Fix $\varepsilon > 0$ small and let $\tau_\varepsilon = \inf\{k \geqslant 1 : |S_0 - S_k| \leqslant \varepsilon\}$ be the first return of the walk inside $[-\varepsilon, \varepsilon]$. By (i) we know that the stopping time τ_ε is almost surely finite. If we define by induction

$$\tau_\varepsilon^{(i)} = \inf\{k > \tau_\varepsilon^{(i-1)} : |S_k - S_{\tau_\varepsilon^{(i-1)}}| \leqslant \varepsilon\}$$

then an application of the strong Markov property shows that $\tau_\varepsilon^{(i)}$ is finite almost surely for every $i \geqslant 1$. Now if $k < \frac{1}{\varepsilon}$ it is clear that $|S_{\tau_\varepsilon^{(k)}}| \leqslant k\varepsilon < 1$ and hence $\#\{n \geqslant 0 : |S_n| < 1\}$ is almost surely larger than $\lfloor \varepsilon^{-1} \rfloor$. Since this holds for any $\varepsilon > 0$ we deduce that $\#\{n \geqslant 0 : |S_n| < 1\} = \infty$ almost surely as desired in (ii).

The implication $(ii) \Rightarrow (iii)$ is clear.

$(iii) \Rightarrow (i)$. We will prove first that for any non trivial interval (a, b) the event $\{\mathcal{I}_{(a,b)} = \infty\}$ is equal to $\{\mathcal{I}_{(-1,1)} = \infty\}$ up to negligible events (recall the notation (3.1)). To see this denote the successive returns times of S in $(-1, 1)$ by $\tau(0) = 0$ and $\tau(k) = \inf\{n > \tau(k-1) : S_n \in (-1, 1)\}$. We then claim (Exercise!) that using Proposition 3.1 one can find $n_0 \geqslant 1$ and $\varepsilon > 0$ such that irrespectively of $x \in (-1, 1)$ we have $\mathbb{P}(x + S_{n_0} \in (a, b)) > \varepsilon$. Using this and the strong Markov property, it is easy to see by standard Markov chain arguments that the events $\{\mathcal{I}_{(-1,1)} = \infty\}$, $\{\tau(k) < \infty, \forall k \geqslant 0\}$ and $\{\mathcal{I}_{(a,b)} = \infty\}$ coincide up to a null event. By the $0-1$ law established in Example 3.1 we deduce that either

$$\text{a.s. } \forall a \neq b \in \mathbb{Q} \text{ we have } \mathcal{I}_{(a,b)} < \infty \quad \text{ or } \quad \forall a \neq b \in \mathbb{Q} \text{ we have } \mathcal{I}_{(a,b)} = \infty. \tag{3.2}$$

Now suppose (iii). Since $|S_n|$ does not diverge with probability 1, this means that there exists a (random) value A such that $S_n \in [-A, A]$ for infinitely many n. This clearly implies that we are in the second option in the last display and in particular $\mathcal{I}_{(-\varepsilon,\varepsilon)} = \infty$ almost surely for any $\varepsilon > 0$.

Since clearly $(ii) \Rightarrow (iv)$ it remains to prove $(iv) \Rightarrow (i)$. Suppose (iv) and assume non (i) by contradiction. This means that for some $\varepsilon > 0$ we have $\mathbb{P}(\forall n \geqslant 1 : |S_n| \geqslant \varepsilon) > \varepsilon$. By considering the successive return times to $(-\varepsilon/2, \varepsilon/2)$, the strong Markov property shows that

$$\mathbb{P}\left(\sum_{i \geqslant 0} \mathbf{1}_{|S_i| < \varepsilon/2} \geqslant k\right) = \mathbb{P}(\mathcal{I}_{(-\varepsilon/2,\varepsilon/2)} \geqslant k) \leqslant (1-\varepsilon)^k.$$

We deduce that $\mathbb{E}[\mathcal{I}_{(-\varepsilon/2,\varepsilon/2)}] < \infty$. Now if if (a,b) is any interval of length $\varepsilon/2$, by applying the Markov property at the hitting time $\tau_{(a,b)}$ of (a,b) we get that

$$\mathbb{E}[\mathcal{I}_{(a,b)}] \leqslant \mathbb{P}(\tau_{(a,b)} < \infty)\mathbb{E}[\mathcal{I}_{(-\varepsilon/2,\varepsilon/2)}].$$

Since $\mathcal{I}_{(-1,1)}$ is less than the sum of roughly $2/\varepsilon$ terms $\mathcal{I}_{(a_i,b_i)}$ where $(-1,1) \subset \cup_i (a_i, b_i)$ and $|a_i - b_i| \leqslant \varepsilon/2$ we reach a contradiction since this implies that $\mathcal{I}_{(-1,1)}$ is of finite expectation. $\square$

The above proof shows in fact that (S) is recurrent if and only

$$\exists a \neq b \in \mathbb{R}, \quad \mathbb{E}[\mathcal{I}_{(a,b)}] = \infty \qquad \Longleftrightarrow \qquad a.s. \forall a \neq b \in \mathbb{R}, \quad \mathcal{I}_{(a,b)} = \infty.$$

Remark 3.1. We will furnish later a criterion for recurrence based on the Fourier transform of the step distribution, see the Chung–Fuchs Theorem 7.5.

Exercise 3.1. Show that $(S_n)_{n \geqslant 0}$ is recurrent if and only if $(S_{2n})_{n \geqslant 0}$ is recurrent.

3.2 Oscillation and drift

In the last section we focused on the dichotomy between recurrence/transience. We now further split transient walks into two finer categories: oscillating random walks and drifting random walks.

3.2.1 Dichotomy

Definition-Proposition 3.2. *A (non-trivial) one-dimensional random walk (S) falls into exactly one of the three categories:*

(i) Either $S_n \to \infty$ a.s. as $n \to \infty$ in which case (S) is said to drift towards ∞*,*

(ii) Or $S_n \to -\infty$ a.s. as $n \to \infty$ in which case (S) is said to drift towards $-\infty$*,*

(iii) Or (S) oscillates *i.e. $\limsup_{n\to\infty} S_n = +\infty$ and $\liminf_{n\to\infty} S_n = -\infty$ almost surely.*

Proof. Note that our background assumption on μ forces $\mu \neq \delta_0$ for which none of the above cases apply. In the lattice case, this proposition is well known. Let us suppose that we are in the non-lattice case. Each of the events defining points $(i) - (ii) - (iii)$ are independent of the values of the first few values of the increments. By Kolmogorov's $0-1$ law they thus appear with probability 0 or 1. Let us suppose that we are in none of the above cases. With the notation of the proof Definition-Proposition 3.1 this means that $\cup_{k\geqslant 1}\{\mathcal{I}_{(-k,k)}=\infty\}$ is of full probability and so we are in the second alternative of (3.2). This is clearly a contradiction because then the range would be dense in $\mathbb{R}$. □

Example 3.2.
- A recurrent walk automatically oscillates.
- However, a transient walk *does not* necessarily drifts towards $+\infty$ or $-\infty$ (see below).
- A random walk whose increments are symmetric necessarily oscillates.

3.2.2 Ladder variables

Definition 3.2 (Ladder heights and epochs)**.** *We define by induction* $T_0^{>}=T_0^{<}=T_0^{\geqslant}=T_0^{\leqslant}=0$ *as well as* $H_0^{>}=H_0^{<}=H_0^{\geqslant}=H_0^{\leqslant}=0$ *and for* $i\geqslant 1$ *we put*

$$\begin{aligned}
T_i^{>} &= \inf\left\{k>T_{i-1}^{>}: S_k > H_{i-1}^{>}\right\} \quad \text{and} \quad H_i^{>}=S_{T_i^{>}},\\
T_i^{\geqslant} &= \inf\left\{k>T_{i-1}^{\geqslant}: S_k \geqslant H_{i-1}^{\geqslant}\right\} \quad \text{and} \quad H_i^{\geqslant}=S_{T_i^{\geqslant}},\\
T_i^{<} &= \inf\left\{k>T_{i-1}^{<}: S_k < H_{i-1}^{<}\right\} \quad \text{and} \quad H_i^{<}=S_{T_i^{<}},\\
T_i^{\leqslant} &= \inf\left\{k>T_{i-1}^{\leqslant}: S_k \leqslant H_{i-1}^{\leqslant}\right\} \quad \text{and} \quad H_i^{\leqslant}=S_{T_i^{\leqslant}}.
\end{aligned}$$

If T_i^* *is not defined (i.e. we take the infimum over the empty set) then we put* $T_j^*=H_j^*=\pm\infty$ *for all* $j\geqslant i$. *The variables* $(T^{>}/T^{\geqslant})$ *(resp.* $(T^{<}/T^{\leqslant})$*) are called the strict/weak ascending (resp. descending) ladder epochs. The associated* H *process are called the (strict/weak ascending/descending) ladder heights.*

Remark 3.2. When the walk (S) has continuous increments, since almost surely S does not take twice the same value the weak and strict ladder variables are the same.

In the following we write H and T generically for one of the four couples

$$(T^{\geqslant},H^{\geqslant}),(T^{>},H^{>}),(T^{<},H^{<}) \text{ or } (T^{\leqslant},H^{\leqslant}).$$

The ladder epochs are clearly stopping times for the natural filtration generated by the walk. The strong Markov property then shows that $N=\inf\{i\geqslant 0: T_i=\infty\}$ is a geometric random variable with distribution

$$\mathbb{P}(N=k)=\mathbb{P}(T_1=\infty)\mathbb{P}(T_1<\infty)^{k-1},$$

and that conditionally on N the random variables $((T_i-T_{i-1}),(H_i-H_{i-1}))_{1\leqslant i\leqslant N-1}$ are i.i.d. with law (T_1,H_1) conditioned on $T_1<\infty$. Combining these observations with Definition-Proposition 3.2, we deduce a characterization of drift/oscillation in terms of the ladder epochs:

Proposition 3.4. *The walk* (S) *drifts towards* $+\infty$ *if and only if* $\mathbb{P}(T_1^{<}=\infty)>0$ *and* $\mathbb{P}(T_1^{>}=\infty)=0$. *The walk* (S) *oscillates if and only if* $\mathbb{P}(T_1^{<}=\infty)=\mathbb{P}(T_1^{>}=\infty)=0$.

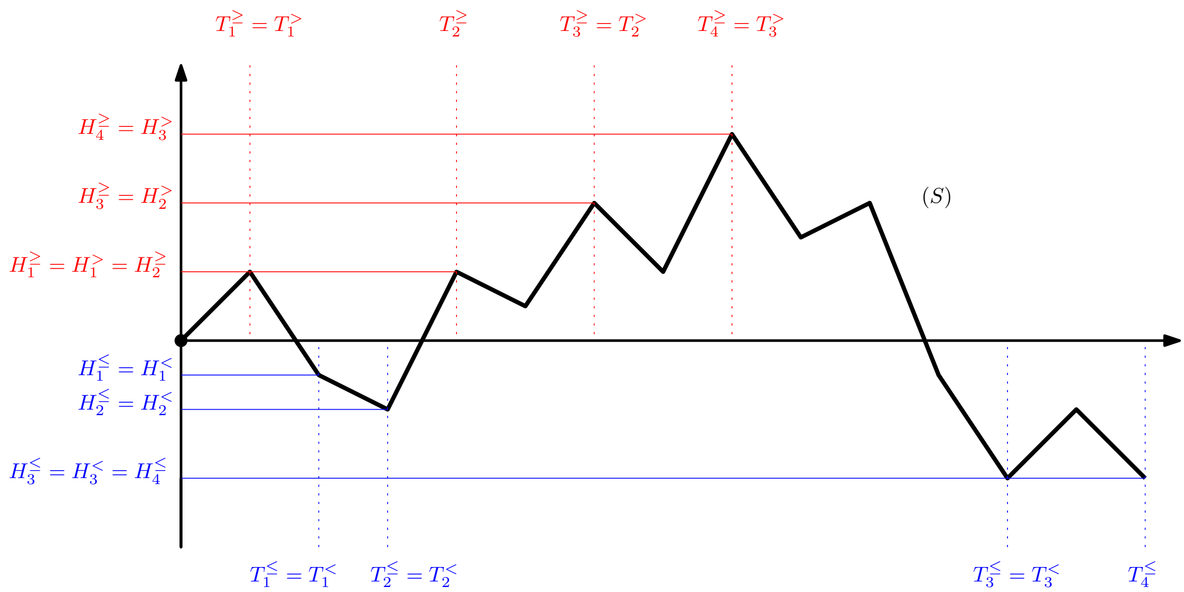


Figure 3.1: Illustration of the definition of the ladder heights and epochs.

3.3 Walks with finite mean

In this section we examine the particular case when $\mathbb{E}[|X|] < \infty$ and show in particular that the walk is recurrent whenever it is centered, otherwise it is transient and drifts. It will be a good opportunity to wiggle around the (strong/weak) law of large numbers, see Exercises 3.2 and 3.3 (see also Exercise 4.9 for an enhanced version of the strong law of large numbers).

3.3.1 Recurrence/transience

Theorem 3.5

Suppose $\mathbb{E}[|X_1|] < \infty$ then

(i) If $\mathbb{E}[X_1] \neq 0$ then (S) is transient and drifts,

(ii) otherwise if $\mathbb{E}[X_1] = 0$ then (S) is recurrent.

Proof. The first point (i) is easy since by the strong law of large numbers we have $n^{-1}S_n \to \mathbb{E}[X_1]$ almost surely: when $\mathbb{E}[X_1] \neq 0$ this automatically implies that $|S_n| \to \infty$ and so (S) is transient by Definition-Proposition 3.1 and drifts towards $\pm\infty$ depending on the sign of $\mathbb{E}[X]$.

In the second case we still use the law of large numbers to deduce that $S_n/n \to 0$ almost surely as $n \to \infty$. This implies that for any $\varepsilon > 0$ we have $|S_n| \leqslant \varepsilon n$ eventually and so for n

large enough

$$\sum_{i=0}^{\infty} \mathbf{1}_{|S_i| \leqslant \varepsilon n} \geqslant n \quad \text{so that} \quad \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{|S_i| \leqslant \varepsilon n}\right] \geqslant n/2. \tag{3.3}$$

We claim that this inequality is not compatible with transience. Indeed, according to Definition-Proposition 3.1, if the walk (S) is transient then for some constant $C > 0$ we have

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{|S_i|<1}\right] \leqslant C.$$

If $x \in \mathbb{R}$, applying the strong Markov property at the stopping time $\tau = \inf\{i \geqslant 0 : S_i \in [x, x+1]\}$ we deduce that

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{S_i \in [x,x+1]}\right] \leqslant \mathbb{P}(\tau < \infty)\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{|S_i|<1}\right] \leqslant C.$$

Dividing the interval $[-\varepsilon n, \varepsilon n]$ into at most $2\varepsilon n + 2$ interval of length at most 1 and applying the above inequality we would deduce that $\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{|S_i| \leqslant \varepsilon n}\right] \leqslant (2\varepsilon n + 2)C$ which contradicts (3.3) for $\varepsilon > 0$ small enough. Hence the walk cannot be transient. □

Exercise 3.2. Let $(S_n)_{n \geqslant 0}$ be a one-dimensional random walk with step distribution μ on $\mathbb{R}$.

1. Show that if $n^{-1} \cdot S_n \xrightarrow[n\to\infty]{(\mathbb{P})} 0$ then (S) is recurrent.
2. Using the Cauchy law (Section 3.4) show that the converse is false.

3.3.2 Wald's equality

Theorem 3.6 (***Wald equality***)

Suppose $\mathbb{E}[|X_1|] < \infty$. *Let* τ *be a stopping time with finite expectation. Then we have*

$$\mathbb{E}[\tau] \cdot \mathbb{E}[X_1] = \mathbb{E}[S_\tau].$$

Proof with martingales. We present a first proof based on martingale techniques. If we denote by m the mean of X_1 then clearly the process $(S_n - nm)_{n \geqslant 0}$ is a martingale for the canonical filtration $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. By the optional sampling theorem we deduce that

$$\mathbb{E}[S_{n\wedge\tau}] = m\mathbb{E}[n \wedge \tau]. \tag{3.4}$$

Since τ is almost surely finite by Corollary 4.2, we can let $n \to \infty$ and get by monotone convergence that the right hand side tends to $\mathbb{E}[\tau]$. However, to deduce that the left hand side also converges towards $\mathbb{E}[S_\tau]$ one would need a domination... To get this just remark that for all n we have

$$|S_{n\wedge\tau}| \leqslant \sum_{i=1}^{\tau} |X_i|,$$

and the last variable has a finite expectation equal to $\mathbb{E}[\tau]\mathbb{E}[|X|] < \infty$ by a standard exercise. On can then use this domination to prove convergence of the left-hand side in (3.4). $\square$

Proof with law of large numbers. We give a second proof based on the law of large numbers. The idea is to iterate the stopping rule. Let $\tau = \tau_1 \leqslant \tau_2 \leqslant \tau_3 \leqslant \dots$ be the successive stopping times obtained formally as

$$\tau_{i+1} = \tau_{i+1}\left((S)_{n\geqslant 0}\right) = \tau_i((S_n)_{n\geqslant 0}) + \tau((S_{n+\tau_i((S_n)_{n\geqslant 0})})_{n\geqslant 0}),$$

for $i \geqslant 1$. In particular $(\tau_{i+1} - \tau_i; S_{\tau_{i+1}} - S_{\tau_i})_{i\geqslant 0}$ are i.i.d. of law (τ, X_τ). By the law of large numbers (we suppose here that $\tau \neq 0$ otherwise the result is trivial) we get that

$$\frac{S_{\tau_i}}{\tau_i} \xrightarrow[i\to\infty]{a.s.} \mathbb{E}[X_1].$$

On the other hand since τ has finite expectation by assumption, applying once more the law of large numbers we deduce that

$$\frac{S_{\tau_i}}{i} = \frac{S_{\tau_i}}{\tau_i} \cdot \frac{\tau_i}{i} \xrightarrow[i\to\infty]{a.s.} \mathbb{E}[X_1] \cdot \mathbb{E}[\tau].$$

We then use the reciprocal of the law of large numbers (see Exercise 3.3) to deduce that S_τ has finite expectation and equal to $\mathbb{E}[\tau] \cdot \mathbb{E}[X_1]$ as claimed by Wald[3]. $\square$

Exercise 3.3 (Converse to the strong law of large numbers). Let $(X_i)_{i\geqslant 0}$ be i.i.d. real variables and suppose that for some constant $c \in \mathbb{R}$ we have

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n\to\infty]{a.s.} c.$$

The goal is to show that X_i have a finite first moment and $\mathbb{E}[X_1] = c$. For this we argue by contradiction and suppose that $\mathbb{E}[|X|] = \infty$.

(i) Show that $\sum_{n\geqslant 1} \mathbb{P}(|X| > n) = \infty$.

(ii) Deduce that $|X_n| > n$ for infinitely many n's.

(iii) Conclude.

(iv) By considering increments having law $\mathbb{P}(X = k) = \mathbb{P}(X = -k) \sim c/(k^2 \log k)$ for some $c > 0$, show that the converse of the *weak* law of large numbers does not hold.

3.4 Examples

We illustrate in this section the concept of recurrence/transience and oscillation/drift in the case when the step distribution has a stable law (either the Gaussian, Lévy or Cauchy distribution) or when it is heavy-tailed (and regular varying).

[3]  Abraham Wald 1902–1950

3.4.1 Stable laws

Gaussian law. First, let us suppose that $\mu(\mathrm{d}x) = \frac{\mathrm{d}x}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ is the standard Gaussian[4] distribution on $\mathbb{R}$. In this case, since μ has a first moment and is centered, one can apply Theorem 3.5 and deduce that the associated walk is recurrent (and thus oscillates). Let us deduce this via another route. A well-know property of Gaussian distribution shows that

$$S_n = X_1 + \cdots + X_n \overset{(d)}{=} \sqrt{n} X_1.$$

This is called the so-called stability property of the Gaussian law of index 2 (appearing in $\sqrt{n} = n^{1/2}$). In particular the last display shows that

$$\sum_{n \geqslant 1} \mathbb{P}(|S_n| < 1) = \sum_{n \geqslant 1} \underbrace{\mathbb{P}(|X_1| \leqslant \frac{1}{\sqrt{n}})}_{\sim \frac{1}{\sqrt{2\pi n}}}.$$

Hence the last series is infinite and so the walk is recurrent by Definition-Proposition 3.1. One could of course have deduced the same result using the criterion 7.5 together with the well-known fact that the Fourier transform of μ is equal to $\hat{\mu}(t) = \mathrm{e}^{-t^2/2}$ for all $t \in \mathbb{R}$. Let us generalize these approaches to other interesting laws.

Lévy law. We now consider $\mu(\mathrm{d}x) = \frac{1}{\sqrt{2\pi x^3}} \mathrm{e}^{-\frac{1}{2x}} \mathrm{d}x \mathbf{1}_{x>0}$ called the standard Lévy[5]. This distribution appears as the first hitting time of -1 by a standard Brownian real motion $(B_t)_{t \geqslant 0}$ starting from 0 (reflexion principle). Using the strong Markov property of the Brownian motion, if $X_1, \ldots, X_n$ are i.i.d. of law μ then $X_1 + \cdots + X_n$ is equal in law to the first hitting time of $-n$ by (B). Combining this observation with the scaling property of B we deduce that

$$S_n = X_1 + \cdots + X_n \overset{(d)}{=} n^2 X_1,$$

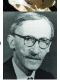
and we say that μ is a stable law with index $1/2$ (the last display can also be seen by a direct calculation). In particular the walk (S) is transient since

$$\sum_{n \geqslant 1} \mathbb{P}(|S_n| < 1) = \sum_{n \geqslant 1} \underbrace{\mathbb{P}(|X_1| \leqslant \frac{1}{n^2})}_{=O(n^{-2})} < \infty.$$

This is not surprising since μ is supported on $\mathbb{R}_+$ and so summing positive increments yields (S) to drift to infinity. But the last calculation is more surprising once we realized it is also valid if we consider for μ the law of $X + cX'$ where X and X' are independent copies of standard Lévy law! In particular if c is negative then $\mathbb{P}(S_n > 0) = \mathbb{P}(X_1 > 0) \in (0,1)$ and so the walk cannot drift. This gives an example of a transient yet oscillating random walk.

Exercise 3.4. Compute the Laplace transform $\mathcal{L}(\mu)(t) = \int \mu(\mathrm{d}x) \mathrm{e}^{-tx}$ for $t > 0$.

[4]  Johann Carl Friedrich Gauß 1777–1855

[5]  Paul Pierre Lévy 1886–1971

Cauchy walk. Our last example is when $\mu(\mathrm{d}x) = \frac{\mathrm{d}x}{\pi(1+x^2)}$ is the standard Cauchy[6] distribution on $\mathbb{R}$. This is again an instance of a stable random variable (here of index 1) since it is well known that if $X_1, \ldots, X_n$ are i.i.d. copies of law μ then

$$S_n = X_1 + \cdots + X_n \overset{(d)}{=} nX_1.$$

One way to see this is to realize X_1 as the x-value of a standard two dimensional Brownian motion started from $(0,0)$ and stopped at the first hitting time of the line $y = -1$. Performing the same calculation as above we realize that

$$\sum_{n \geqslant 1} \mathbb{P}(|S_n| < 1) = \sum_{n \geqslant 1} \underbrace{\mathbb{P}(|X_1| \leqslant \frac{1}{n})}_{\sim \frac{1}{n\pi}} = \infty,$$

and so (S) is an example of a recurrent random walk even though its increment does not admit a first moment! It may also be surprising to the reader that the walk $(S_n + cn)_{n \geqslant 0}$ is also recurrent for any value of $c \in \mathbb{R}$ by the same argument! Another way to prove recurrence for the last walk is to apply the Fourier criterion of Theorem 7.5 provided the next exercise is solved:

Exercise 3.5. Show that the Fourier transform $\hat{\mu}$ of the standard Cauchy distribution is given by $\hat{\mu}(t) = \mathrm{e}^{-|t|}$ for $t \in \mathbb{R}$.

As we said, the Gaussian, Lévy and Cauchy distributions are particular instances of stable distributions. We just give the definition since their detailed study would need a full course following the steps of Paul Lévy.

Definition 3.3. *A stable distribution μ is a law on $\mathbb{R}$ such that for all $n \geqslant 1$ if $X_1, X_2, \ldots, X_n$ are n independent copies of law μ then for some $A_n, B_n \in \mathbb{R}$ we have*

$$X_1 + \cdots + X_n \overset{(d)}{=} A_n \cdot X + B_n.$$

It turns out that necessarily $A_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$ which is called the index of stability of the law. The only stable laws with "explicit" densities are the Gaussian laws ($\alpha = 2$), the Lévy laws ($\alpha = \frac{1}{2}$) and the Cauchy laws ($\alpha = 1$).

3.4.2 Heavy-tailed random walks

Here is an example of a random walk with "heavy tails" which is transient and yet oscillates:

Proposition 3.7. *Let μ be a step distribution on $\mathbb{Z}$ such that $\mu_k = \mu_{-k} \sim k^{-\alpha}$ as $k \to \infty$ for $\alpha \in (1, 2)$. In particular μ has no first moment. Then the associated random walk (S) is transient and yet oscillates.*

[6]  Augustin Louis, baron Cauchy 1789–1857

Remark 3.3. If $\alpha > 2$, the walk (S) is centered and so by Theorem 3.5. The critical case $\alpha = 2$ where the step distribution has Cauchy-type tail is more delicate: the walk is recurrent (see Exercise 3.2).

Proof. Since (S) is a symmetric random walk, it cannot drift and must oscillate. We just need to show that (S) is transient e.g.

$$\sum_{n \geqslant 0} \mathbb{P}(S_n = 0) < \infty.$$

The idea is to use the randomness produced by a single big jump of the walk. More precisely, let us introduce the event $A = \{\exists 1 \leqslant i \leqslant n : |X_i| > n^{1+\varepsilon}\}$ where $\varepsilon > 0$ will be chosen small enough later on. We can write

$$\mathbb{P}(S_n = 0) \leqslant \mathbb{P}(A^c) + \mathbb{P}(S_n = 0 \mid A).$$

The first term of the right-hand side is easy to evaluate:

$$\mathbb{P}(A^c) = \big(1 - \mathbb{P}(|X| \geqslant n^{1+\varepsilon})\big)^n \approx \exp(-c \cdot n \cdot n^{(1+\varepsilon)(1-\alpha)}) \leqslant \exp(-n^\delta),$$

for some $\delta > 0$ provided that $\varepsilon > 0$ is small enough (we used here the fact that $1 < \alpha < 2$). On the other hand, conditionally on A, one can consider the first increment $|X_{i_0}|$ of S of absolute value larger than $n^{1+\epsilon}$. Clearly, the law μ of X_{i_0} is that of X conditioned on being of absolute value larger than $n^{1+\varepsilon}$ and in particular

$$\forall k \in \mathbb{Z}, \qquad \mu(k) = \mathbf{1}_{|k| \geqslant n^{1+\varepsilon}} \frac{\mathbb{P}(X = k)}{\mathbb{P}(|X| > n^{1+\varepsilon})} \leqslant C n^{-1-\varepsilon},$$

for some constant $C > 0$ and X_{i_0} is furthermore independent of all the other increments. Hence we can write

$$\mathbb{P}(S_n = 0 \mid A) = \mathbb{P}(X_{i_0} = -(X_1 + \cdots + \widehat{X_{i_0}} + \cdots + X_n) \mid A) \leqslant \sup_{k \in \mathbb{Z}} \mu(k) \leqslant C n^{-1-\varepsilon}.$$

Gathering-up the pieces, we deduced that $\mathbb{P}(S_n = 0) \leqslant \exp(-n^\delta) + C n^{-1-\varepsilon}$ for $\delta > 0$ provided that $\varepsilon > 0$ is small enough. The implies summability of the series and ensures transience of the walk. □

Let us also state without proof a theorem of Shepp[7] which shows the disturbing fact that there exists recurrent one-dimensional random walk with arbitrary fat tails (but not regular varying):

Theorem 3.8 (***Shepp*** *[38]*)

For any position function $\epsilon(x)$ tending to 0 as $x \to \infty$, there exists a symmetric step distribution μ such that $\mu(\mathbb{R}\backslash[-x, x]) \geqslant \epsilon(x)$ for any $x \geqslant 0$ and such that the associated random walk (S) is recurrent.

---

[7]  Lawrence Alan Shepp 1936–2013

More exercises.

Exercise 3.6 (Sums of random walks). Let $(S_n)_{n\geqslant 0}$ and $(S'_n)_{n\geqslant 0}$ be two independent one-dimensional random walks with independent increments of law μ and μ' on $\mathbb{R}$.

1. Give an example where (S) and (S') are transient and yet $(S+S')$ is recurrent.
2. We suppose that μ and μ' are both symmetric. Show that as soon as (S) or (S') is transient then so is $(S+S')$.
3. Give an example where (S) is recurrent, (S') transient and yet $(S+S')$ is recurrent. (Hint: Use the Cauchy law of Section 3.4).
4. (*) Can we have both (S) and (S') recurrent and $(S+S')$ transient ?

Exercise 3.7 (Subordinated random walks). Let $(S_n)_{n\geqslant 0}$ be a one-dimensional random walk with step distribution μ. Let also $(Y_n)_{n\geqslant 0}$ be another independent one-dimensional random walk whose step distribution ν is supported on $\{1,2,3,\dots\}$ and is aperiodic. We form the process

$$Z_n = S_{Y_n}.$$

1. Show that (Z) is again a one-dimensional random walk with independent increments and characterize its step distribution.
2. Show that if $\int \nu(\mathrm{d}x)x < \infty$ then Z and S have the same type (recurrent, transient, oscillating, drifting towards $\pm\infty$).
3. Let ν be the return time to $(0,0)$ of the simple random walk on $\mathbb{Z}^2$. Using Theorem 7.7 show that for any μ the walk (Z) is transient.
4. (*) Can we find (μ,ν) so that (S) oscillates but (Z) drifts?

Bibliographical notes. The material in this chapter is standard and can be found in many textbooks, see e.g. [39, Chapter II], [13, Chapter 8] or [19, 26]. Shepp's Theorem 3.8 is based on the Chung–Fuchs Fourier criterion for recurrence that we will see in Section 7.3.1. A solution due to Edouard Maurel-Segala of the last question of Exercice 3.6 can be found at *https://mathoverflow.net/questions/314312/sum-of-independent-random-walks*. The reader eager to learn more about one-dimensional stable distributions is refereed to [42].

Chapter IV: Fluctuations theory

In this chapter we still consider a one-dimensional random walk (S) based on i.i.d. increments of law μ (whose support is not contained in $\mathbb{R}_+$ nor $\mathbb{R}_-$). The goal is to get information on the distribution of the ladder processes and reciprocally get information on the walk from the ladder processes.

4.1 Duality and applications

We begin with a simple but surprisingly important observation called *duality*.

Proposition 4.1 (Duality)**.** *For each fixed $n \geqslant 0$, we have the following equality in distribution*

$$(0 = S_0, S_1, \ldots, S_n) \overset{(d)}{=} (S_n - S_n, S_n - S_{n-1}, S_n - S_{n-2}, \ldots, S_n - S_1, S_n - S_0).$$

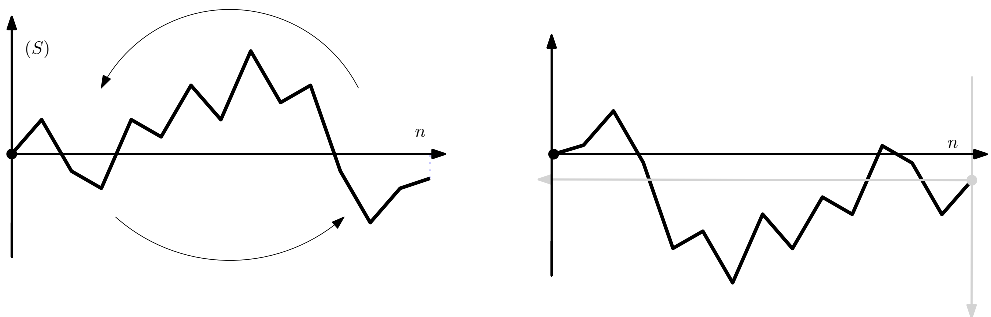


Figure 4.1: Geometric interpretation of the duality: the rotation by angle π of the first n steps of the walk (S) leaves the distribution invariant.

Proof. It suffices to notice that the increments of the walk $(S_n - S_{n-1}, S_n - S_{n-2}, \ldots, S_n - S_1, S_n - S_0)$ are just given by $(X_n, X_{n-1}, \ldots, X_1)$ which obviously has the same law as $(X_1, \ldots, X_n)$ since the $(X_i)_{i \geqslant 1}$ are i.i.d. hence exchangeable. $\square$

Beware the duality is an equality in distribution valid for a fixed time n and not as a process.

Exercise 4.1. Let (S) be a one-dimensional random walk drifting towards $-\infty$. Using duality show that $S_n - \inf_{0 \leqslant k \leqslant n} S_k$ converges in distribution as $n \to \infty$ towards $\sup_{k \geqslant 0} S_k$.

This innocent proposition enables us to connect the strict descending ladder variables to the weak ascending ones. Indeed, notice (on a drawing) that for any $n \geqslant 0$

$$\begin{aligned} \mathbb{P}(T_1^{<} > n) &= \mathbb{P}(S_0 = 0, S_1 \geqslant 0, \dots, S_n \geqslant 0) \\ &\underset{\text{duality}}{=} \mathbb{P}(S_n - S_n = 0, S_n - S_{n-1} \geqslant 0, \dots, S_n \geqslant 0) \\ &= \mathbb{P}(S_n \geqslant S_{n-1}, S_n \geqslant S_{n-2}, \dots, S_n \geqslant S_0) = \mathbb{P}(n \text{ is a weak ladder epoch}). \end{aligned}$$

Summing over $n \geqslant 0$ we deduce that

$$\begin{aligned} \sum_{n \geqslant 0} \mathbb{P}(T_1^{<} > n) &= \mathbb{E}[T_1^{<}] \\ &= \mathbb{E}[\text{number of weak ascending finite ladder epochs}] \\ &= \frac{1}{\mathbb{P}(T_1^{\geqslant} = \infty)}, \end{aligned} \tag{4.1}$$

because the total number of weak ascending finite ladder epochs follows a geometric distribution with success parameter $\mathbb{P}(T_1^{\geqslant} = \infty)$. We similarly establish that $\mathbb{E}[T_1^{\leqslant}] = 1/\mathbb{P}(T_1^{>} = \infty)$. From these observations we conclude:

Corollary 4.2. *We are in one of the three categories:*

- *Either* (S) *drifts towards* $+\infty$ *in which case we have*

$$\mathbb{P}(T_1^{\leqslant} = \infty) > 0, \quad \mathbb{P}(T_1^{<} = \infty) > 0, \quad \mathbb{E}[T_1^{>}] < \infty, \quad \mathbb{E}[T_1^{\geqslant}] < \infty,$$

- *Either* (S) *drifts towards* $-\infty$ *in which case we have*

$$\mathbb{P}(T_1^{\geqslant} = \infty) > 0, \quad \mathbb{P}(T_1^{>} = \infty) > 0, \quad \mathbb{E}[T_1^{<}] < \infty, \quad \mathbb{E}[T_1^{\leqslant}] < \infty,$$

- *Or* (S) *oscillates then the ladder epochs are finite but*

$$\mathbb{E}[T_1^{>}] = \mathbb{E}[T_1^{\geqslant}] = \mathbb{E}[T_1^{<}] = \mathbb{E}[T_1^{\leqslant}] = \infty.$$

Remark 4.1. The last corollary shows that for an oscillating random walk, although the walk will visit $\mathbb{R}_+$ and $\mathbb{R}_-$ infinitely many times, the time of the first visit to one of the half-spaces is of infinite expectation. This is a well-known fact for the simple symmetric random walk on $\mathbb{Z}$ (exercise!).

Proof. Let us suppose that (S) drifts towards $+\infty$. Then clearly (S) has a positive probability to stay positive for all positive times and so $T^{<}$ (as well as $T^{\leqslant}$) has a positive probability to be infinite. It follows from (4.1) (and its extension a few line below) that $\mathbb{E}[T_1^{>}]$ and $\mathbb{E}[T_1^{\geqslant}]$ are finite. The case when (S) drifts towards $-\infty$ is symmetric. When (S) oscillates then the ladder epoch are always finite and so their expectations are infinite by (4.1). $\square$

Remark 4.2. Combining the last corollary with Wald's equality (Theorem 3.6), we deduce that for a random walk with finite mean and positive drift then we have

$$\mathbb{E}[H_1^>] = \mathbb{E}[T_1^>] \cdot \mathbb{E}[X_1].$$

Exercise 4.2. Let (S) be a one-dimensional random walk with i.i.d. increments X_i. Suppose that $H_1^>, H_1^<$ are both finite and of finite expectation. Show that $\mathbb{E}[|X_1|] < \infty$ and $\mathbb{E}[X] = 0$.

Exercise 4.3. Show directly (without duality) that $\mathbb{E}[T_1^>]$ is finite if and only if $\mathbb{E}[T_1^\geqslant]$ is finite.

4.2 Cyclic lemma and Wiener–Hopf

In this section we prove the main formula of this chapter (Theorem 4.5) which is based on a particularly elegant combinatorial lemma due to Feller[1].

4.2.1 Feller's cyclic lemma

Let $x_1, x_2, \dots, x_n$ be real numbers which we consider as the increments of the walk (s) defined by

$$s_0 = 0, s_1 = x_1, s_2 = x_1 + x_2, \quad \dots \quad , s_n = x_1 + \cdots + x_n.$$

Recall that i is a strict ascending ladder epoch for (s) if $s_i > s_{i-1}, s_i > s_{i-2}, \dots, s_i > s_0$. If $\ell \in \{0, 1, 2, \dots, n-1\}$ we consider $(s^{(\ell)})$ the ℓ-th cyclic shift of the walk obtained by cyclically shifting its increments ℓ times, that is

$$s_0^{(\ell)} = 0, s_1^{(\ell)} = x_{\ell+1}, \quad \dots \quad , s_{n-\ell}^{(\ell)} = x_{\ell+1} + \cdots + x_n, \quad \dots \quad , s_n^{(\ell)} = x_{\ell+1} + \cdots + x_n + x_1 + \cdots + x_\ell.$$

Lemma 4.3 (Feller)**.** *Suppose that $s_n > 0$. We denote by $k \in \{0, 1, 2, \dots, n\}$ the number of cyclic shifts $(s^{(\ell)})$ with $\ell \in \{0, 1, 2, \dots, n-1\}$ for which n is a strict ascending ladder epoch. Then $k \geqslant 1$ and any of those cyclic shifts has exactly k strict ascending ladder epochs.*

Proof. Let us first prove that $k \geqslant 1$. For this consider the first time $\ell \in \{1, 2, \dots, n\}$ such that the walk (s) reaches its maximum. Then clearly (make a drawing) the time n is a strict ascending ladder epoch for $s^{(\ell)}$. We can thus suppose without loss of generality that n is a strict ascending ladder epoch for (s). It is now clear (see Fig. 4.2 below) that the only possible cyclic shifts of the walk such that the resulting walk admits a strict ascending ladder epoch at n correspond to the strict ascending ladder epochs of (s). Moreover these cyclic shifts do not change the number of strict ascending ladder epochs. □

[1]  William Feller 1906–1970, born Vilibald Srećko Feller

Remark 4.3. Beware, Feller's combinatorial lemma does not say that the cyclic shifts $(s^{(\ell)})$ are distinct. Indeed, in the action of $\mathbb{Z}/n\mathbb{Z}$ on $\{(s^{(\ell)} : \ell \in \{0, 1, \dots, n-1\}\}$ by cyclic shift, the size of the orbit is equal to n/j where $j|n$ is the cardinal of the subgroup stabilizing $(s^{(0)})$. In our case, it is easy to see that j must also divide k and in this case there are only j/k distinct cyclic shifts having n as the kth strict ascending ladder time.

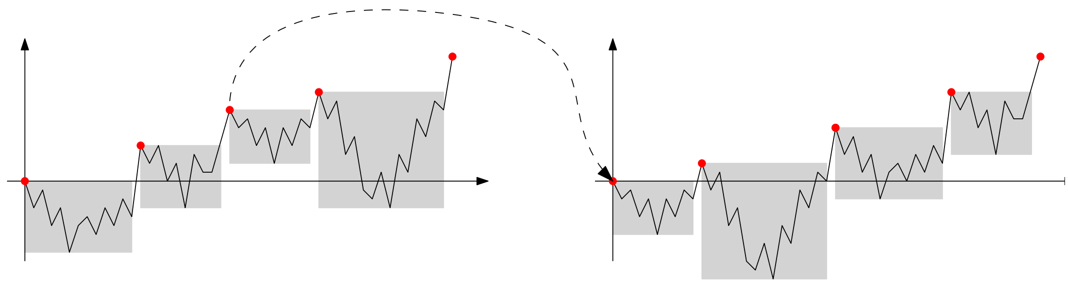


Figure 4.2: Illustration of Feller's combinatorial lemma. We show a walk such that n is a strict ascending ladder epoch and the cyclic shift corresponding to the second strict ascending ladder epoch.

The above lemma also holds if we replace strict ascending ladder epoch by weak/descending ladder epoch provided that $s_n \geqslant 0$ or $s_n \leqslant 0$ or $s_n < 0$ depending on the cases. Here is an exercise whose proof is similar to Feller's combinatorial lemma:

Exercise 4.4. Let (S) be a one-dimensional random walk with diffuse step distribution. Show that for every $n \geqslant 1$ the number of points of the walk lying strictly above the segment $(0,0) \to (n, S_n)$ is uniformly distributed on $\{0, 1, 2, \dots, n-1\}$.

Corollary 4.4. *For every $n \geqslant 1$ and any measurable subset $A \subset \mathbb{R}_+^*$ we have*

$$\frac{1}{n}\mathbb{P}(S_n \in A) = \sum_{k=1}^{\infty} \frac{1}{k}\mathbb{P}(T_k^> = n, H_k^> \in A).$$

Proof. Let us first re-write the last lemma in a single equation

$$\mathbf{1}_{s_n \in A} = \sum_{i=0}^{n-1} \sum_{k=1}^{\infty} \frac{1}{k} \mathbf{1}_{T_k^>(s^{(i)})=n} \mathbf{1}_{H_k^>(s^{(i)}) \in A}.$$

Indeed, if the walk (s) is such that $s_n \in A$ in particular $s_n > 0$ and there exists a unique k such that exactly k cyclic shifts do not annulate the indicator functions on the right-hand side. Since we divide by k the total sum is one. We take expectation when $(s) = (S)$ is a one-dimensional random walk with i.i.d. increments. Using the fact that for all $0 \leqslant i \leqslant n-1$ we have $(S_j^{(i)})_{0 \leqslant j \leqslant n} = (S_j)_{0 \leqslant j \leqslant n}$ in distribution, we deduce the statement of the corollary. □

We can rewrite the last corollary in terms of measures:

$$\mathbf{1}_{x>0} \frac{\mathbb{P}(S_n \in \mathrm{d}x)}{n} = \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{P}(H_k^> \in \mathrm{d}x, T_k^> = n) \mathbf{1}_{x>0}.$$

Exercise 4.5. For $n \geqslant 0$, let $G_n = \inf\{0 \leqslant k \leqslant n : S_k = \sup_{0 \leqslant i \leqslant n} S_i\}$ for the first time when the walk achieves its maximum over $[|0,n|]$. Show that conditionally on $S_n = 1$, the variable G_n is uniformly distributed over $\{1, 2, \ldots, n\}$.

4.2.2 Wiener–Hopf factorization

The following result is an analytic translation of our findings. In many application however, we shall come back to Feller's combinatorial lemma which is easier to remember!

Theorem 4.5 (***Spitzer–Baxter formula ; Wiener–Hopf factorization***)

For $r \in [0,1)$ and $\mu \in \mathbb{C}$ so that $\Re(\mu) \geqslant 0$ we have

$$\left(1 - \mathbb{E}\left[r^{T_1^>} \mathrm{e}^{-\mu H_1^>}\right]\right) = \exp\left(-\sum_{n=1}^{\infty} \frac{r^n}{n} \mathbb{E}\left[\mathrm{e}^{-\mu S_n} \mathbf{1}_{S_n > 0}\right]\right),$$

$$\left(1 - \mathbb{E}\left[r^{T_1^\leqslant} \mathrm{e}^{\mu H_1^\leqslant}\right]\right) = \exp\left(-\sum_{n=1}^{\infty} \frac{r^n}{n} \mathbb{E}\left[\mathrm{e}^{\mu S_n} \mathbf{1}_{S_n \leqslant 0}\right]\right).$$

Proof. First since $r \in [0,1)$ and $\Re(\mu) \geqslant 0$ all the quantities in the last two displays are well defined. We only prove the first display since the calculation is similar for the second one. Let us start from the right hand side and write

$$\begin{aligned}
\exp\left(-\sum_{n=1}^{\infty} \frac{r^n}{n} \mathbb{E}\left[\mathrm{e}^{-\mu S_n} \mathbf{1}_{S_n > 0}\right]\right) &\underset{\text{Cor.4.4}}{=} \exp\left(-\sum_{n=1}^{\infty} \frac{r^n}{n} \sum_{k=1}^{\infty} \frac{n}{k} \mathbb{E}\left[\mathrm{e}^{-\mu H_k^>} \mathbf{1}_{T_k^> = n}\right]\right) \\
&= \exp\left(-\sum_{k=1}^{\infty} \frac{1}{k} \underbrace{\mathbb{E}\left[\mathrm{e}^{-\mu H_k^>} r^{T_k^>}\right]}_{\left(\mathbb{E}\left[\mathrm{e}^{-\mu H_1^>} r^{T_1^>}\right]\right)^k}\right) \\
&= 1 - \mathbb{E}\left[\mathrm{e}^{-\mu H_1^>} r^{T_1^>}\right],
\end{aligned}$$

where in the last line we used the equality $\sum_{k=1}^{\infty} \frac{x^k}{k} = -\log(1-x)$ valid for $|x| < 1$. Note that we implicitly used the fact that $r < 1$ by putting $r^{T_k^>} = 0$ when $T_k^> = \infty$. This proves Spitzer's[2] formula □

Remark 4.4 (Explanation of the terminology of Wiener–Hopf factorization). If we write

$$\omega_r^>(\mu) = \exp\left(-\sum_{n=1}^{\infty} \frac{r^n}{n} \mathbb{E}\left[\mathrm{e}^{-\mu S_n} \mathbf{1}_{S_n > 0}\right]\right) \quad \text{and } \omega_r^\leqslant(\mu) = \exp\left(-\sum_{n=1}^{\infty} \frac{r^n}{n} \mathbb{E}\left[\mathrm{e}^{-\mu S_n} \mathbf{1}_{S_n \leqslant 0}\right]\right),$$

then $\omega_r^>$ is analytic on the half-space $\mathfrak{Re}(\mu) \geqslant 0$ whereas $\omega_r^\leqslant$ is analytic on $\mathfrak{Re}(\mu) \leqslant 0$. On the imaginary line where the two functions are well defined we have

$$\omega_r^>(it)\omega_r^\leqslant(it) = 1 - r\mathbb{E}[\mathrm{e}^{-itX_1}]. \tag{4.2}$$

[2]  Frank Ludvig Spitzer 1926–1992

Hence, the characteristic function of the increment of the walk (or a slight modification thereof) has been writing as a product of two analytic functions, each defined on a different half-space. The idea of writing a function on a line as a product of two functions defined on a half-space goes back to Wiener & Hopf and is often useful since we can use the tools of complex analysis for each of the factors.

Exercise 4.6. Show that for $r \in (0,1)$ we have

$$\sum_{n \geqslant 0} \mathbb{P}(T_1^{\leqslant} > n) r^n = \exp\left(\sum_{n \geqslant 1} \frac{r^n}{n} \mathbb{P}(S_n > 0) \right).$$

4.3 Applications

4.3.1 Direct applications

A first application of Theorem 4.5 (or more clearly of (4.2)) is that the law of $(T_1^>, H_1^>)$ and $(T_1^{\leqslant}, H_1^{\leqslant})$ are sufficient to recover the law of the increment (hence of the random walk), even better : the knowledge of the law of $H_1^>$ and that of $H_1^{\leqslant}$ is sufficient to recover μ since by taking $r \to 1$ in Theorem 4.5 we have

$$\left(1 - \mathbb{E}\left[\mathbf{1}_{H_1^> < \infty} \mathrm{e}^{\mathrm{i}tH_1^>}\right]\right)\left(1 - \mathbb{E}\left[\mathbf{1}_{H_1^{\leqslant} < \infty} \mathrm{e}^{\mathrm{i}tH_1^{\leqslant}}\right]\right) = 1 - \mathbb{E}[\mathrm{e}^{itX_1}].$$

This is not at all clear from the beginning! Actually a very recent theorem shows that the law of $(T_1^>, H_1^>)$ only, characterizes the step distribution of the walk:

Theorem 4.6 (*Kwaśnicki [23]*)

The law of $(T_1^>, H_1^>)$ characterizes the law of the underlying random walk.

See Exercise 4.15 for a proof of this theorem under relatively mild assumptions. Let us give another surprising corollary of the Wiener–Hopf formula:

Corollary 4.7. *Let (S) be a one-dimensional random walk with symmetric and diffuse step distribution. Hence the law of $T_1^>$ is given by*

$$\mathbb{E}[r^{T_1^>}] = 1 - \sqrt{1-r}, \; r \in [0,1), \qquad \textit{or equivalently} \qquad \mathbb{P}(T_1^> = n) = \frac{(2n-2)!}{2^{2n-1} n!(n-1)!}, \; n \geqslant 1.$$

Proof. It suffices to take the first display of Theorem 4.5 and to plug $\mu = 0$. Since by symmetry of the increments and the lack of atoms we have $\mathbb{P}(S_n > 0) = \mathbb{P}(S_n \geqslant 0) = \frac{1}{2}$ it follows that

$$\begin{aligned} 1 - \mathbb{E}[r^{T_1^>}] &= \exp\left(- \sum_{n \geqslant 1} \frac{r^n}{n} \mathbb{P}(S_n > 0) \right) \\ &= \exp\left(- \sum_{n \geqslant 1} \frac{r^n}{n} \frac{1}{2} \right) = \exp(-1/2 \log(1-r)) = \sqrt{1-r}. \end{aligned}$$

To get the exact values of $\mathbb{P}(T_1^> = n)$ it suffices to develop $1 - \sqrt{1-r}$ in power series and to identify the coefficients. $\square$

Remark 4.5. It is useful to notice the asymptotic $\mathbb{P}(T_1^> = n) \sim \frac{n^{-3/2}}{2\sqrt{\pi}}$ as $n \to \infty$.

Corollary 4.8. *The following conditions are equivalent*

(i) *the random walk* (S) *drifts towards* $-\infty$

(ii) *we have* $\sum_{n\geqslant 1} \frac{\mathbb{P}(S_n>0)}{n} < \infty$

(iii) *we have* $\sum_{n\geqslant 1} \frac{\mathbb{P}(S_n\geqslant 0)}{n} < \infty$.

In this case we have

$$\log \mathbb{E}[T_1^{\leqslant}] = \sum_{n\geqslant 1} \frac{\mathbb{P}(S_n > 0)}{n}.$$

Proof. From Theorem 4.5 with $\mu = 0$ we get for $r \in [0,1)$

$$1 - \mathbb{E}[r^{T_1^>}] = \exp\left(-\sum_{n\geqslant 1} \frac{r^n}{n}\mathbb{P}(S_n > 0)\right).$$

Letting $r \uparrow 1$ the left-hand side converges towards $1 - \mathbb{E}[\mathbf{1}_{T_1^> < \infty}] = \mathbb{P}(T_1^> = \infty)$ whereas the right-hand side converges towards $\exp(-\sum_{n\geqslant 1} \frac{\mathbb{P}(S_n>0)}{n})$. But clearly (S) drifts towards $-\infty$ if and only if $T_1^>$ may be infinite. In this case, recall that by (4.1) we have $\mathbb{E}[T_1^{\leqslant}] = 1/\mathbb{P}(T_1^> = \infty)$ which immediately implies the second claim. The equivalence with the large inequality is done similarly by considering $T_1^{\geqslant}$. □

Exercise 4.7. Show that we always have $\displaystyle\sum_{n\geqslant 1} \frac{1}{n}\mathbb{P}(S_n = 0) < \infty$.

Exercise 4.8. Suppose (S) is a one-dimensional random walk with integrable increments that drifts towards $-\infty$. Verify directly that $\sum_{n\geqslant 1} \frac{\mathbb{P}(S_n>0)}{n} < \infty$. (Hint: use the truncated increments $X_n^* = |X| \wedge n$).

Exercise 4.9 (Law of large numbers enhanced). Let $(S_n)_{n\geqslant 0}$ be a one-dimensional random walk with i.i.d. increments $X_1, X_2, \dots$. Show that the following propositions are equivalent:

(i) $\frac{S_n}{n} \to 0$ almost surely,

(ii) $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[X] = 0$,

(iii) for every $\varepsilon > 0$ we have $\displaystyle\sum_{n\geqslant 1} \frac{1}{n}\mathbb{P}(|S_n| > \varepsilon n) < \infty$.

4.3.2 Skip-free walks

Definition 4.1. *Let* (S) *be a one-dimensional random walk whose step distribution* μ *is supported by* $\mathbb{Z}$. *We say that* (S) *is skip-free ascending (resp. descending) when* $\mu(\{1,2,3,\dots\}) = \mu_1$ *(resp.* $\mu(\{\dots,-3,-2,-1\}) = \mu_{-1}$*); or in words when the only positive (resp. negative) jumps of* S *are jumps of* $+1$ *(resp.* -1*).*

The best examples of skip-free walks are simple random walks where the step distribution is supported by ± 1 (they are both skip-free ascending and descending). The nice thing with skip-free ascending walk is the fact that the k-th ladder height $H_k^>$ must be equal to k when it is finite. This simple observation turns out to have many implications. First, Lemma 4.4 reduces to the well-known Kemperman's (a.k.a. Otter-Dwass' formula) [3] formula

Proposition 4.9 (Kemperman's formula)**.** *Let (S) be a skip-free ascending walk. Then for every $n \geqslant 1$ and every $k \geqslant 1$ we have*

$$\frac{1}{n}\mathbb{P}(S_n = k) = \frac{1}{k}\mathbb{P}(T_k^> = n).$$

Let us give a first application of this formula in the case of the symmetric simple random walk whose step distribution is $\frac{1}{2}(\delta_1 + \delta_{-1})$. Since this walk is both skip-free ascending and descending we have for $n \geqslant 1$ (due to parity reason $T_1^<$ and $T_1^>$ have to be odd)

$$\mathbb{P}(T_1^< = 2n-1) = \frac{1}{2n-1}\mathbb{P}(S_{2n-1} = -1) = \frac{1}{2n-1} 2^{-(2n-1)} \binom{2n-1}{n} = 2^{-2n+1}\frac{(2n-2)!}{n!(n-1)!}. \quad (4.3)$$

We recover the probability that a symmetric continuous random walk first hits $\mathbb{R}_-$ at time n. Surprising isn't it ? Do you have a simple explanation of this phenomenon?

Exercise 4.10 (Borel distribution). Consider (S) the one-dimensional random walk with step distribution given by the law of $\mathcal{P}_1 - 1$ where $\mathcal{P}_1$ is a Poisson random variable of parameter 1. Compute the distribution of $T_1^<$ and deduce that

$$\sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} \mathrm{e}^{-n} = 1.$$

(Do you have a elementary way to deduce the last display?)

Ballot theorem

Lemma 4.10. *Let (S) be a skip-free ascending random walk. Then for every $n \geqslant 1$ and every $k \geqslant 1$ we have*

$$\mathbb{P}(S_i > 0, \forall 1 \leqslant i \leqslant n \mid S_n = k) = \frac{k}{n}.$$

Proof. By duality we have

$$\begin{aligned}
\mathbb{P}(S_i > 0, \forall 1 \leqslant i \leqslant n \text{ and } S_n = k) &\underset{\text{duality}}{=} \mathbb{P}(n \text{ is a strict ascending ladder epoch for } S \text{ and } S_n = k) \\
&\underset{\text{skip free}}{=} \mathbb{P}(T_k^> = n) \\
&\underset{\text{Prop.4.9}}{=} \frac{k}{n}\mathbb{P}(S_n = k).
\end{aligned}$$

$\square$

Let us give an immediate application which is useful during election days:

[3]  Johannes Henricus Bernardus Kemperman 1924–2011

Theorem 4.11 (***Ballot theorem***)

During an election, candidates A and B respectively have $a > b$ votes. Suppose that votes are spread uniformly in the urn. What is the chance that during the counting of votes, candidate A is always ahead?

$$\text{answer:} \qquad \frac{a-b}{a+b}.$$

Proof. Let us model the scenario by a uniform path making only $+1$ or -1 steps which starts at $(0,0)$ and ends at $(a+b, a-b)$. The $+1$ steps correspond to votes for candidate A and the -1 steps for votes for B. Hence we ask about the probability that such a path stay positive between time 1 and $a+b$. We just remark that such a walk has the same distribution as the first $a+b$ steps of $(S_k)_{0 \leqslant k \leqslant a+b}$ a simple random walk (with $\pm$ steps with equal probability) conditioned on $S_{a+b} = a-b$: indeed each $\pm$ path starting at $(0,0)$ and ending at $(a+b, a-b)$ has the same probability under that distribution. The conclusion follows from Lemma 4.10. $\square$

Staying positive

If (S) is a one-dimensional random walk with integrable increments with positive mean then by the law of large numbers, the probability that the walk stays positive after time 1 is strictly positive. We compute below this probability in the case of skip-free ascending and skip-free descending walks:

Corollary 4.12. *If (S) is skip-free ascending such that $\mathbb{E}[S_1] > 0$ then we have*

$$\mathbb{P}(S_i > 0 : \forall i \geqslant 1) = \mathbb{E}[S_1].$$

Proof. We have

$$\begin{aligned} \mathbb{P}(S_i > 0 : \forall i \geqslant 1) &= \lim_{n\to\infty} \mathbb{P}(S_i > 0 : \forall 1 \leqslant i \leqslant n) \\ &= \lim_{n\to\infty} \mathbb{E}\left[\mathbb{P}(S_i > 0 : \forall 1 \leqslant i \leqslant n \mid S_n)\right] \\ &\underset{\text{Lem.4.10}}{=} \lim_{n\to\infty} \mathbb{E}\left[\frac{S_n}{n}\mathbf{1}_{S_n>0}\right] \to \mathbb{E}[S_1], \end{aligned}$$

by the strong law of large numbers (since $S_n/n \to \mathbb{E}[S_1]$ almost surely and in $\mathbb{L}^1$). $\square$

Proposition 4.13. *If (S) is skip-free descending (with $\mu \neq \delta_0$) then $\mathbb{P}(S_n \geqslant 0, \forall n \geqslant 0) = 1 - \alpha$ where α is the smallest solution in $\alpha \in [0,1]$ to the equation:*

$$\alpha = \sum_{k=-1}^{\infty} \mu_k \alpha^{k+1}. \tag{4.4}$$

Proof. Since μ is supported by $\{-1, 0, 1, \dots\}$ its mean m is well-defined and belongs to $[-1, \infty]$. We already know from the previous chapter that $\mathbb{P}(S_n \geqslant 0, \forall n \geqslant 0) > 0$ if and only if $m > 0$ (we use here the fact that the walk is not constant since $\mu \neq \delta_0$). We denote by $\tau_{<0}$ the hitting time of $\{\dots,-3,-2,-1\}$ by the walk (S). Notice that by our assumptions if $\tau_{<0}$ is finite then necessarily $S_{\tau_{<0}} = -1$. To get the equation of the proposition we perform one step of the random walk S: if $S_1 = -1$ then $\tau_{<0} < \infty$. Otherwise if $S_1 \geqslant 0$ then consider the stopping times

$$\theta_0 = 0, \quad \theta_1 = \inf\{k \geqslant 1 : S_k = S_1 - 1\}, \quad \theta_2 = \inf\{k \geqslant \theta_1 : S_k = S_1 - 2\}, \dots.$$

By the strong Markov property we see that $(\theta_{i+1} - \theta_i)_{i \geqslant 0}$ are i.i.d. of law $\tau_{<0}$. Furthermore on the event $S_1 \geqslant 0$ we have

$$\{\tau_{<0} < \infty\} = \bigcap_{n=0}^{S_1} \{\theta_{n+1} - \theta_n < \infty\}.$$

Taking expectation, we deduce that $\mathbb{P}(\tau_{<0} < \infty)$ is indeed solution to (4.4). Now, notice that $F : \alpha \mapsto \sum_{k=-1}^{\infty} \mu_k \alpha^{k+1}$ is a convex function on $[0,1]$ which always admits 1 as a fixed point. Since $F'(1) = m + 1$ we deduce that F admits two fixed points in the case $m > 0$. But when $m > 0$ we already know that $\alpha < 1$ and so α must be equal to the smallest solution of (4.4). $\square$

Exercise 4.11. Let (S) be a skip-free descending random walk which drifts towards $+\infty$. Compute the law of $\inf\{S_k : k \geqslant 0\}$.

4.3.3 Arcsine law

They are many different arcsine laws in the theory of random walk. We restrict to the usual one in the simplest case only to illustrate another application of our preceding results.

Proposition 4.14 (1st Arcsine law). *Let (S) be a one-dimensional random walk with a symmetric step distribution without atoms. We put $K_n = \inf\{0 \leqslant k \leqslant n : S_k = \sup_{0 \leqslant i \leqslant n} S_i\}$ then*

$$\frac{K_n}{n} \xrightarrow[n\to\infty]{(d)} \frac{\mathrm{d}x}{\pi\sqrt{x(1-x)}} \mathbf{1}_{x \in (0,1)}.$$

The name arcsine comes from the cumulative distribution function of the right-hand side which is $\frac{2}{\pi}\arcsin(\sqrt{x})$.

Remark 4.6. Quoting Feller "Contrary to intuition, the maximum accumulated gain is much more likely to occur towards the very beginning or the very end of a coin-tossing game than somewhere in the middle."

Proof. Using duality we can compute exactly

$$\begin{aligned}\mathbb{P}(K_n = k) &= \mathbb{P}(T_1^{>} \geqslant n-k)\mathbb{P}(T_1^{\geqslant} \geqslant k) \\ &\underset{\text{Rek.4.5}}{\sim} \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}},\end{aligned}$$

where the last asymptotic holds uniformly in $k >> 1$ and $n-k >> 1$. If we add a little blur to K_n and consider $\tilde{K_n} = K_n + U_n$ where U_n is independent of K_n and uniformly distributed over $[0,1]$. Then clearly $\tilde{K_n}/n$ has a density with respect to Lebesgue measure which converges pointwise

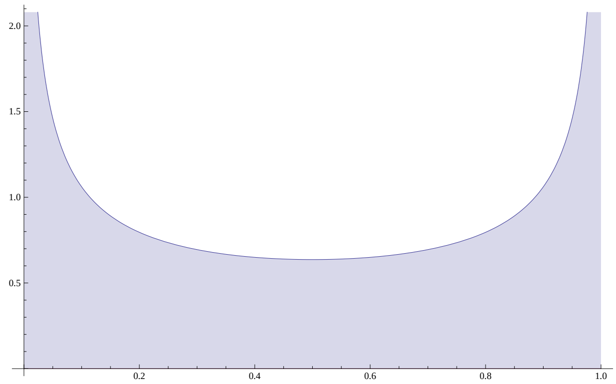


Figure 4.3: The arcsine distribution

towards the density of the arcsine law. It follows from Scheffé's lemma that $\tilde{K}_n/n$ converges in total variance towards the arcsine law and consequently K_n/n converges in distribution towards the arcsine law since $U_n/n \to 0$ in probability. □

Exercise 4.12 (Scheffé's lemma). Let X_n, X be random variables having densities f_n, f with respect to a background measure π. We suppose that $f_n \to f$ pointwise π-almost everywhere. Prove that

(i) $f_n \to f$ in $\mathbb{L}^1(\pi)$.

(ii) $\mathrm{d}_{TV}(f_n \mathrm{d}\pi, f\mathrm{d}\pi) \to 0$ where d_{TV} is the total variation distance,

(iii) deduce that $X_n \to X$ in distribution.

Exercise 4.13. Prove the arcsine law in the case of symmetric simple random walk.

4.3.4 Parking on the line

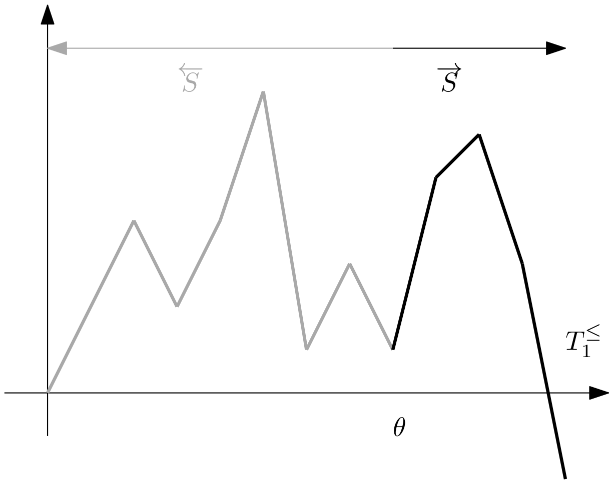
More exercises

Exercise 4.14 (Another approach to Wiener–Hopf factorization). Let $(S_n)_{n\geqslant 0}$ be a one-dimensional oscillating random walk. On the event $T_1^{\leqslant} \geqslant 2$, we put $Y = \inf\{S_i : 1 \leqslant i \leqslant T_1^{\leqslant} - 1\}$ and define the random instant $\theta = \sup\{i \leqslant T_1^{\leqslant} : S_\theta = Y\}$. We put $\overleftarrow{S}$ and $\overrightarrow{S}$ for the processes defined by

$$\overleftarrow{S}_i = S_\theta - S_{\theta-i}, \quad \text{for } 0 \leqslant i \leqslant \theta,$$

$$\overrightarrow{S}_i = S_{\theta+i} - S_\theta, \quad \text{for } 0 \leqslant i \leqslant T_1^{\leqslant} - \theta.$$

Otherwise if $T_1^{\leqslant} = 1$ we put $\overleftarrow{S} = \overrightarrow{S} = \dagger$ a cemetery point.



1. Let $(S'_i)_{0 \leqslant i \leqslant \tau'}$ and $(S''_i)_{0 \leqslant i \leqslant \tau''}$ two independent random walks of law S stopped respectively at $T_1^<$ and $T_1^{\leqslant}$. Show that for any function ϕ so that $\phi(\dagger, \dagger) = 0$ we have
$$\mathbb{E}\left[\phi(\overleftarrow{S}, \overrightarrow{S})\right] = \mathbb{E}\left[\phi(S', S'')\mathbf{1}_{S'_{T_1^>} + S''_{T_1} \leqslant 0}\right],$$
and in particular that for any $t \in \mathbb{R}$ we have
$$\mathbb{E}[\mathrm{e}^{itH_1^{\leqslant}}] - \mathbb{E}[\mathrm{e}^{itH_1^{\leqslant}} \mathrm{e}^{itH_1^>} \mathbf{1}_{H_1^> \leqslant -H_1^{\leqslant}}] = \mathbb{E}[\mathrm{e}^{itX}\mathbf{1}_{X \leqslant 0}],$$
where X is a step of the random walk and $H_1^> = S_{T_1^>}$ and $H_1^{\leqslant}$ are two independent versions of its ascending and descending ladder heights.

2. By adapting the argument with $T_1^>$ recover the Wiener–Hopf factorization:
$$\left(1 - \mathbb{E}[\mathrm{e}^{itH_1^{\leqslant}}]\right)\left(1 - \mathbb{E}[\mathrm{e}^{itH_1^>}]\right) = 1 - \mathbb{E}[\mathrm{e}^{itX}].$$

Exercise 4.15 (From [12]). Let $(S_n)_{n \geqslant 0}$ be a one-dimensional random walk with iid increments of law μ not supported by $\mathbb{R}_-$ and with compact support. For $t \in \mathbb{R}$ we write
$$\phi(t) = \int_{\mathbb{R}} \mathrm{d}\mu(x)\ \mathrm{e}^{tx}$$
which is well-defined and analytic on $\mathbb{R}$ in our case. The goal is to prove that μ is characterized by the law of $(H^>, T^>) \equiv (H_1^>, T_1^>)$, the ascending space-time first ladder variables.

1. Show that the law of $(H^>, T^>)$ give access to $\mathbb{P}(S_n > 0)$ and to μ^{*n} restricted to $\mathbb{R}_+^*$.

2. Deduce that the law of $(H^>, T^>)$ enables us to decide oscillation, drift, recurrence or transience of (S).

3. In this question we suppose that μ has positive mean.

   (a) Using large deviations, show that there exists $\alpha > 0$ such that for $n \geqslant 1$ we have
$$\mathbb{P}(S_n \leqslant 0) \leqslant \mathrm{e}^{-\alpha n}.$$

(b) Show that for $t > 0$ small enough we have

$$\lim_{n \to \infty} \left(\mathbb{E}[\mathrm{e}^{tS_n} \mathbf{1}_{S_n > 0}] \right)^{1/n} = \phi(t).$$

(c) Deduce that the law of $(H^>, T^>)$ characterizes μ.

4. Back to the general case where μ has compact support.

   (a) Show that there exists $\lambda > 0$ with $\phi(\lambda) < \infty$ and $\phi'(\lambda) > 0$.

   (b) By considering $\mu_\lambda(\mathrm{d}x) = \frac{\mathrm{e}^{\lambda x}}{\phi(\lambda)} \mu(\mathrm{d}x)$ and the associated random walk $(S^{(\lambda)})$ show that

$$\mathbb{P}(S_n^{(\lambda)} > 0) = \frac{\mathbb{E}[\mathrm{e}^{\lambda S_n} \mathbf{1}_{S_n > 0}]}{(\phi(\lambda))^n}.$$

   (c) Prove that $(S^{(\lambda)})$ drift towards $+\infty$ and conclude.

Biliographical notes. The Wiener–Hopf factorisation can have several meanings and interpretations in the literature on random walks. We chose to focus on the most combinatorial one using Feller's cyclic lemma (see [24] for a more trajectorial approach). This chapter is adapted from [17, Chapter XII] and [13, Section 8.4] (in particular Exercise 4.4 is [17, Theorem 3, p423]). See [4] for much more about Ballot theorems. Exercise 4.14 is adapted from [28, 20] and Exercise 4.15 from [12]. Path transformation are very useful tools in fluctuation theory for random walk. In particular, we mention the Sparre-Andersen identity relating the position of the maximum and the time spent on the positive half-line for a random walk of length n, see [17, Chapter XII] for more details. More recent applications of fluctuation theory for random walks can be found e.g. in [**?**, **?**].

Chapter V: Conditioning random walks to stay positive

In this chapter we use the information gathered previously about the ladder processes to define and study the "random walk (S) conditioned to stay positive" (in fact non-negative). Unless in the case when (S) drifts towards $+\infty$, the previous conditioning is degenerate and one needs to work to make sense of it. We will see that even conditioned to stay non-negative for ever, the random walk remains a Markov process whose transitions probabilities are intimately connected to harmonic functions. This will also be a good opportunity to approach the general theory of Doob h-transformation.

5.1 h-transform of Markov chains

Let us first present the idea of Doob[1] to transform a Markov chain using harmonic functions. We restrict ourselves to the homogeneous case (no temporal parameter) and in the case of discrete time/space Markov chains.

5.1.1 h-transformation

Let $p(\cdot,\cdot)$ be a Markov kernel, i.e. probability transitions of a discrete Markov chain $(X_i : i \geqslant 0)$ on a countable state space Ω. Suppose that $h : \Omega \to \mathbb{R}_+$ is a non-negative function which is harmonic and positive on $A \subset \Omega$ i.e.

$$h(x) > 0, \quad \text{and} \quad h(x) = \sum_{y \in \Omega} p(x,y)h(y), \qquad \forall x \in A.$$

Under these circumstances, one can define a new transition kernel q on A by the formula:

$$q(x,y) = \frac{h(y)}{h(x)} p(x,y), \quad x \in A, y \in \Omega.$$

It is plain from the harmonicity of h on A that q indeed defines a transition kernel on A, that is $\sum_y q(x,y) = 1$ for any $x \in A$. For any $x_0 \in A$ we can thus consider the Markov process $(Y_i : i \geqslant 0)$ governed by the kernel q until (Y) possibly enters $\Omega \backslash A$ where it is stopped. This process is called the **the Doob h-transform of p**. The law of the q-Markov chain is equivalently

[1]  Joseph Leo Doob (1910-2004)

characterized as follows: for any given path $x_0, x_1, \ldots, x_{n-1}$ in A and $x_n \in \Omega$ we have

$$\prod_{i=0}^{n-1} q(x_i, x_{i+1}) = \frac{h(x_n)}{h(x_0)} \prod_{i=0}^{n-1} p(x_i, x_{i+1}). \tag{5.1}$$

In particular if $y \in \Omega$ is such that $h(y) = 0$ then the process (Y) never hits y. In the case when h is zero on $\Omega \backslash A$, the q-Markov chain never escapes A and so can be interpreted as a way to *condition the p-chain to stay in A* (see below for a justification of this name in the case of random walks).

Remark 5.1. The h-transformation is equivalent to the change of measure operated by the martingale biasing via the positive martingale $(h(X_{n\wedge\tau_{\Omega\backslash A}}) : n \geqslant 0)$. We shall not enter the details and refer to [6].

5.1.2 Examples

Exiting through a particular state. A convenient way to build harmonic functions is via Dirichlet problem (see Section 1.1.2). For example, let us fix $x' \notin A$, a neighbor of some point in A, and consider the harmonic function on A defined by

$$h(x) = \mathbb{P}_x(\tau_{\Omega\backslash A} < \infty \text{ and } X_{\tau_{\Omega\backslash A}} = x'), \quad x \in A$$

where as usual $\tau_{\Omega\backslash A}$ is the hitting time of $\Omega\backslash A$ by the p-chain $(X_i : i \geqslant 0)$. As in the case when A is finite, one easily check that h is indeed harmonic on A and let us suppose it is positive on A (this is true if for any starting point in A the p-chain can exit A through x' with positive probability). Then one can consider the h-transform (Y) of A.

Proposition 5.1. *Under the above hypotheses, the q-chain starting from $x_0 \in A$ has the same law as the p-chain started from x_0, stopped when touching x' and conditioned to exit A through x' in finite time (an event of positive probability).*

Proof. For any finite path $x_0, x_1, \ldots, x_n = x'$ in A exiting at x', by (5.1) the probability to see this entire path for the q-chain is equal to

$$\prod_{i=0}^{n-1} q(x_i, x_{i+1}) = \frac{h(x_n)}{h(x_0)} \prod_{i=0}^{n-1} p(x_i, x_{i+1}) = \frac{\prod_{i=0}^{n-1} p(x_i, x_{i+1})}{\mathbb{P}_{x_0}(\tau_{\Omega\backslash A} < \infty \text{ and } X_{\tau_{\Omega\backslash A}} = x')}.$$

The statement of the proposition is then clear on the last display. □

Not touching a set in the transient case. Suppose now that from any $x \in A$ there is positive probability that starting from x, the chain (X) never escapes from A. We can thus form the harmonic function on A

$$h(x) = \mathbb{P}_x(\tau_{\Omega\backslash A} = \infty),$$

and again consider the associated h-transform. In this case, since $h \equiv 0$ on $\Omega\backslash A$, the q-chain (Y) will never escapes from A and we have:

Proposition 5.2. *Under the above hypotheses, the q-chain starting from $x_0 \in A$ has the same law as the p-chain started from x_0 and conditioned to never escape from A (an event of positive probability).*

Proof. For any path $x_0, x_1, \ldots, x_n$ staying in A, by the Markov property applied at time n we have

$$\begin{aligned}\mathbb{P}_{x_0}(X_0 = x_0, \ldots, X_n = x_n \mid \tau_{\Omega\backslash A} = \infty) &\underset{\text{Markov}}{=} \frac{\prod_{i=0}^{n-1} p(x_i, x_{i+1})}{\mathbb{P}_{x_0}(\tau_{\Omega\backslash A} = \infty)} \mathbb{P}_{x_n}(\tau_{\Omega\backslash A} = \infty) \\ &= \frac{h(x_n)}{h(x_0)} \prod_{i=0}^{n-1} p(x_i, x_{i+1}) \underset{(5.1)}{=} \prod_{i=0}^{n-1} q(x_i, x_{i+1}).\end{aligned}$$

Not touching a set in the recurrent case. The last two examples reinterpret Markov chains conditioned on some event of positive probability as h-transform chains. Let us show an example where the q-chain is singular with respect to the initial one. For the simple symmetric random walk on $\mathbb{Z}$, consider $A = \mathbb{Z}_{>0} = \{1, 2, 3, \ldots\}$ and let h be the harmonic function

$$h(i) = i, \qquad \text{for } i \geqslant 1.$$

The h-transform (Y) is thus a Markov chain on $\mathbb{Z}_{>0}$, with ± 1 steps which never touches 0 (and is even transient, see Proposition 5.5). The law of $(Y_i : i \geqslant 0)$ is thus singular with respect to the law of simple symmetric random walk. As we will see below, this chain can be interpreted as the random walk (X) conditioned to stay non-negative (an event of probability 0). For the connoisseurs, it is also a discrete version of the Bessel process of dimension 3.

5.2 Renewal function

We now come back to our one-dimensional random walk setting. We suppose that μ is a step distribution on $\mathbb{Z}$ (whose supported is not included in $\mathbb{Z}_{\geqslant 0}$ nor $\mathbb{Z}_{\leqslant 0}$) and consider the random walk $(S_i : i \geqslant 0)$ which under $\mathbb{P}_x$ starts from $x \in \mathbb{Z}$ and has i.i.d. increments of law μ. We write $\mathbb{P}$ for $\mathbb{P}_0$ to lighten notation. Using the ladder processes, we shall construct two (super) harmonic functions on $\mathbb{Z}_{>0}$ and $\mathbb{Z}_{\geqslant 0}$ for S which we will then use to h-transform it. In this section, in order to lighten notation, for $i \geqslant 0$ we write $H_i = H_i^<$ and $T_i^< = T_i$ for the strict descending ladder height and epoch processes.

5.2.1 Pre-renewal and renewal functions

Recalling the definition of the ladder processes (Section 3.2.2) we define two functions for $\ell \geqslant 0$

$$h^{\downarrow}(\ell) = \sum_{i \geqslant 0} \mathbb{P}(H_i = -\ell) = \mathbb{P}(\tau_{\mathbb{Z}_{\leqslant -\ell}} = \tau_{-\ell} < \infty),$$

$$h^{\uparrow}(\ell) = h^{\downarrow}(0) + \cdots + h^{\downarrow}(\ell) = \mathbb{E}\left[\sum_{i \geqslant 0} \mathbf{1}_{H_i \geqslant -\ell}\right].$$

The function $h^{\downarrow}$ is called the *pre-renewal* function and $h^{\uparrow}$ is called the *renewal function* of the walk S. Using (4.1), it is easy to see that $h^{\uparrow}$ is bounded if and only if (S) drifts towards $+\infty$. Actually, the functions $h^{\downarrow}$ and $h^{\uparrow}$ have harmonic properties with respect to the walk (S):

Proposition 5.3. *The functions $h^{\downarrow}$ and $h^{\uparrow}$ are respectively harmonic on $\mathbb{Z}_{>0}$ and super-harmonic on $\mathbb{Z}_{\geqslant 0}$ for the walk (S). Moreover the function $h^{\uparrow}$ is harmonic on $\mathbb{Z}_{\geqslant 0}$ (not only super-harmonic) if and only if (S) does not drift towards $-\infty$.*

Proof. By writing the Markov property at time 1 under $\mathbb{P}_x$ for $x \in \{1, 2, \dots\}$ we have $h^{\downarrow}(x) = \sum_{k\in\mathbb{Z}} \mu(k)h^{\downarrow}(x+k)$ which is the required harmonicity of $h^{\downarrow}$ on $\mathbb{Z}_{>0}$. By summing-up these equations for $x = 1, 2, \dots, y$ we get that

$$h^{\uparrow}(y) - 1 = \sum_{k\geqslant 0} \mu(k)h^{\uparrow}(y+k) - \sum_{k\geqslant 0} h^{\uparrow}(k)\mu(k).$$

Hence the (super-)harmonicity of $h^{\uparrow}$ is tied to the value of $\sum_{k\geqslant 0} h^{\uparrow}(k)\mu(k)$. To evaluate the latter, we use duality:

$$\begin{aligned}
\sum_{k\geqslant 0} h^{\uparrow}(k)\mu(k) \quad &= \quad \sum_{k\geqslant 0} \mu(k)\mathbb{E}\left[\sum_{i\geqslant 0} \mathbf{1}\{H_i \geqslant -k\}\right] \\
&= \quad \sum_{k\geqslant 0} \mu(k) \sum_{i\geqslant 0}\sum_{n\geqslant 0} \mathbb{P}\left(H_i \geqslant -k \text{ and } T_i = n\right) \\
&\underset{\text{duality}}{=} \quad \sum_{k\geqslant 0} \mu(k) \sum_{n\geqslant 0} \mathbb{P}\left(S_1, \dots, S_n < 0 \text{ and } S_n \in [0, -k]\right) \\
&= \quad \sum_{n\geqslant 0}\sum_{k\geqslant 0} \mu(k)\mathbb{P}\left(S_1, \dots, S_n < 0 \text{ and } S_n \in [0, -k]\right) \\
&= \quad \sum_{n\geqslant 0} \mathbb{P}(T_1^{\geqslant} = n+1) = \mathbb{P}(T_1^{\geqslant} < \infty).
\end{aligned}$$

The last probability is equal to 1 if and only if (S) does not drift towards $-\infty$. $\square$

Remark 5.2. In the case of a skip-free descending random walk, the pre-renewal and the renewal functions take a particularly simple form. Indeed, since $H_i = -i$ on the event when $T_i < \infty$ we deduce that, as soon as the walk (S) oscillates or drifts to $-\infty$ (i.e. if the mean of the increment is less than or equal to zero), we have

$$\forall i \geqslant 0, \quad h^{\downarrow}(i) = 1, \quad \text{and} \quad h^{\uparrow}(i) = i + 1.$$

When (S) drifts to $+\infty$ (i.e. if the mean of the increment is negative), as in Proposition 4.13 we let $\alpha = \mathbb{P}_0(T_1 < \infty)$ and we check that $h^{\downarrow}(i) = \alpha^i$ and so $h^{\uparrow}(i) = 1 + \alpha + \cdots + \alpha^i$.

Exercise 5.1. Suppose that the walk S drifts towards $-\infty$. Prove that in this case we have for $x \geqslant 0$

$$h^{\uparrow}(x) = \frac{\mathbb{E}_x[\tau_{\mathbb{Z}_<}]}{\mathbb{E}_0[\tau_{\mathbb{Z}_<}]}.$$

5.2.2 $h^{\downarrow}$-transform

By Proposition 5.3 the function $h^{\downarrow}$ is harmonic on $\mathbb{Z}_{>0}$ and so one can consider the $h^{\downarrow}$-transform of the random walk (S) started from $x \geqslant 1$ which we denote by $(S_i^{\downarrow} : i \geqslant 0)$. This is actually a particular instance of our example "exiting through a particular state" of Section 5.1.2. Indeed, in the previous notation we have $A = \mathbb{Z}_{>0}$, $x' = 0$ and $(X) = (S)$. The harmonic function $h(x) = \mathbb{P}_x(\tau_{\Omega\backslash A} < \infty \text{ and } X_{\tau_{\Omega\backslash A}} = x')$ reduces to $\mathbb{P}_x(\tau_{\mathbb{Z}_{\leqslant 0}} = \tau_{\{0\}} < \infty) = h^{\downarrow}(x)$ for $x \geqslant 1$. We deduce from Proposition 5.1 that the $h^{\downarrow}$-transform of S has the law of the random walk S conditioned on the event of positive probability to hit the negative integers exactly at 0 (an event of positive probability). This is sometimes called the random walk conditioned to die "continuously" when touching $\mathbb{Z}_{\leqslant 0}$.

5.2.3 $h^{\uparrow}$-transform when drifting to $+\infty$

Suppose next that (S) drifts to $+\infty$ so that $h^{\uparrow}$ is harmonic on $\mathbb{Z}_{\geqslant 0}$ and denote by $S^{\uparrow}$ the $h^{\uparrow}$-transform of S. Here also, this is a special case of our general example "not touching a set in the transient case". Indeed, fixing $A = \mathbb{Z}_{\geqslant 0}$, the function $h(x) = \mathbb{P}_x(\tau_{\Omega\backslash A} = \infty)$ coincides with a scaled version of $h^{\uparrow}$:

Proposition 5.4. *When (S) drift to $+\infty$ we have*

$$h^{\uparrow}(x) = \frac{\mathbb{P}_x(\tau_{\mathbb{Z}_{<0}} = \infty)}{\mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty)}, \quad x \geqslant 0$$

where $\tau_{\mathbb{Z}_{<0}}$ is the first hitting time of $\mathbb{Z}_{<0}$ by the walk.

Proof. Start from $x \geqslant$ and denote by $0 = T_0, T_1, \dots, T_k, \dots$ and $x = H_0 > H_1 > \dots$ the strict minimal record times and heights of the walk (S). Since (S) drifts towards $+\infty$, notice that there is a finite number of those record times. It is easy to see (Figure 5.1) that under $\mathbb{P}_x$ we have

$$\mathbf{1}_{\tau_{\mathbb{Z}_{<0}}=\infty} = \sum_{k \geqslant 0} \mathbf{1}_{T_k<\infty} \mathbf{1}_{H_k \geqslant 0} \mathbf{1}_{\tau_{\mathbb{Z}_{<0}}(S^{(T_k)})=\infty},$$

where $S^{(T_k)} = (S_{T_k+i} - S_{T_k} : i \geqslant 0)$ is the shifted walk. Taking expectations and using the strong Markov property at time T_k we deduce that

$$\mathbb{P}_x(\tau_{\mathbb{Z}_{<0}} = \infty) = \mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty) \mathbb{E}_x \left[\sum_{k \geqslant 0} \mathbf{1}_{H_k \geqslant 0} \right] = \mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty) h^{\uparrow}(x).$$

□

We deduce from Proposition 5.2 that when S drifts to $+\infty$, the process $S^{\uparrow}$ has the law of S conditioned on $\{\tau_{\mathbb{Z}_{<0}} = \infty\}$, an event of positive probability.

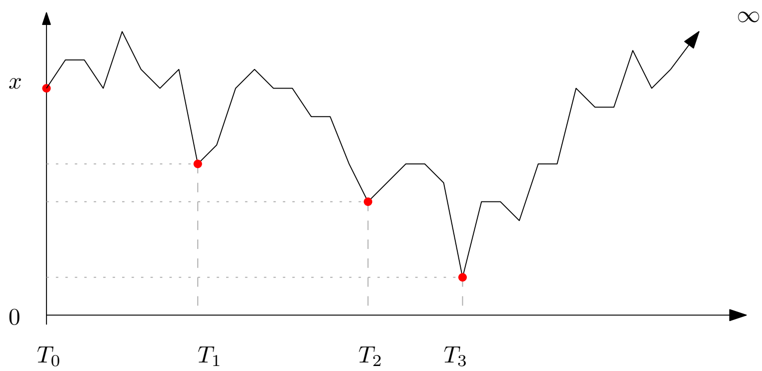


Figure 5.1: Illustration of the proof. If the walk does not touch the negative half-line, then at least one of the excursions above the minima should escape to ∞ before the current minimal record drops below 0.

5.3 The $h^{\uparrow}$-transform when (S) does not drift to $-\infty$

In this section, we suppose that S does not drift towards $-\infty$ so that $h^{\uparrow}$ is harmonic and we denote by $(S_i^{\uparrow} : i \geqslant 0)$ its $h^{\uparrow}$-transform which is a well-defined Markov chain. In particular, this covers the case when (S) oscillates. We first prove that this chain is transient:

Proposition 5.5 (Transience of the $h^{\uparrow}$-transform)**.** *Suppose that S does not drift towards $-\infty$. Then the Markov chain $(S_i^{\uparrow})_{i\geqslant 0}$ is transient.*

Proof. Let us consider the Markov chain $S^{\uparrow}$ started from $\ell \geqslant 1$ and consider the first time $\tau_{<\ell}$ it reaches a value strictly lower than ℓ. By (5.1) we can write

$$\mathbb{P}_\ell(\tau_{<\ell}(S^{\uparrow}) < \infty) = \frac{1}{h^{\uparrow}(\ell)} \mathbb{E}_\ell[h^{\uparrow}(S_{\tau_{<\ell}}) \mathbf{1}_{\tau_\ell < \infty}] \leqslant \frac{\sup_{\ell' < \ell} h^{\uparrow}(\ell')}{h^{\uparrow}(\ell)},$$

unless the walk has non-negative increments (in which case the statement of the proposition is plain), the function $h^{\uparrow}$ is strictly increasing so that the last fraction is < 1. This implies that $S^{\uparrow}$ is transient. □

5.3.1 Limit of large conditionings

We shall now see that the process $S^{\uparrow}$ arises as a limit in distribution of the random walk S conditioned on staying non-negative for large time, hence justifying the terminology "random walk conditioned to stay non-negative for ever". Specifically, we consider the event

$$\Lambda_n = \{S_k \geqslant 0 \text{ for all } 0 \leqslant k \leqslant n\}.$$

Theorem 5.6 (*Bertoin & Doney*)

Suppose that S does not drift towards $-\infty$. The process $(S^{\uparrow})$ appears as the limit in distribution of the random walk S conditioned on Λ_n as $n \to \infty$, specifically for $0 = s_0, s_1, \dots, s_k \geqslant 0$ we have

$$\mathbb{P}_0 \left(S_0 = s_0, \dots, S_k = s_k \mid \Lambda_n\right) \xrightarrow[n\to\infty]{} \mathbb{P}_0(S_0^{\uparrow} = s_0, \dots, S_k^{\uparrow} = s_k).$$

Proof. The technical key is:

Lemma 5.7. *We have*

$$\liminf_{n\to\infty} \frac{\mathbb{P}_x(\Lambda_n)}{\mathbb{P}_0(\Lambda_n)} \geqslant \frac{h^{\uparrow}(x)}{h^{\uparrow}(0)},$$

Given the lemma, the proof of the theorem is rather easy: by the Markov property applied at time k we have

$$\begin{aligned}
\mathbb{P}_0 \left(S_0 = s_0, \dots, S_k = s_k \mid \Lambda_n\right) &= \mathbb{P}_0(S_0 = s_0, \dots, S_k = s_k \mathbf{1}_{s_i \geqslant 0, \forall 0 \leqslant i \leqslant k}) \frac{\mathbb{P}_{s_k}(\Lambda_{n-k})}{\mathbb{P}_0(\Lambda_n)} \\
&\geqslant \mathbf{1}_{s_i \geqslant 0, \forall 0 \leqslant i \leqslant k} \prod_{i=0}^{k-1} \mu(s_{i+1} - s_i) \frac{\mathbb{P}_{s_k}(\Lambda_n)}{\mathbb{P}_0(\Lambda_n)} \\
\liminf \quad &\underset{\text{Lem.5.7}}{\geqslant} \mathbf{1}_{s_i \geqslant 0, \forall 0 \leqslant i \leqslant k} \prod_{i=0}^{k-1} \mu(s_{i+1} - s_i) \frac{h^{\uparrow}(s_k)}{h^{\uparrow}(s_0)} \\
&= \mathbb{P}_0(S_0^{\uparrow} = s_0, \dots, S_k^{\uparrow} = s_k).
\end{aligned}$$

To get the converse inequality we use the fact that the last expression is a probability measure on positive paths, and Fatou's lemma:

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}_0 \left(S_0 = s_0, \dots, S_k = s_k \mid \Lambda_n\right) \\
= \ &\limsup_{n\to\infty} \left(1 - \sum_{(\tilde{s}_i : 0 \leqslant i \leqslant k) \neq (s_i : 0 \leqslant i \leqslant k)} \mathbb{P}_0 \left(S_0 = \tilde{s}_0, \dots, S_k = \tilde{s}_k \mid \Lambda_n\right)\right) \\
\underset{\text{Fatou}}{\leqslant} \ & 1 - \sum_{(\tilde{s}_i : 0 \leqslant i \leqslant k) \neq (s_i : 0 \leqslant i \leqslant k)} \liminf_{n\to\infty} \mathbb{P}_0 \left(S_0 = \tilde{s}_0, \dots, S_k = \tilde{s}_k \mid \Lambda_n\right) \\
= \ & 1 - \sum_{(\tilde{s}_i : 0 \leqslant i \leqslant k) \neq (s_i : 0 \leqslant i \leqslant k)} \mathbb{P}_0 \left(S_0^{\uparrow} = \tilde{s}_0, \dots, S_k^{\uparrow} = \tilde{s}_k\right) = \mathbb{P}_0(S_0^{\uparrow} = s_0, \dots, S_k^{\uparrow} = s_k).
\end{aligned}$$

This completes the proof, given Lemma 5.7. □

Let us now prove Lemma 5.7. The idea is similar to the proof of Proposition 5.4. Start from $x \geqslant 0$, denote by $0 = T_0, T_1, \dots$ and $x = H_0 > H_1 \dots$ the strict minimal record times and heights of the walk (S) and by $\mathrm{Exc}_k = (S_{T_k}, S_{T_k+1} \dots, S_{T_{k+1}-1})$ the associated excursions above the running minimum. Denote by $K \geqslant 0$ the index of the first excursion such that $\mathrm{Exc}_K \in \Lambda_n$.

Then it is easy to see that if $H_K \geqslant 0$ then the event Λ_n happens for the walk (S). Hence by the Markov property applied at time T_K we have

$$\mathbb{P}_x(\Lambda_n) \geqslant \mathbb{P}_0(\Lambda_n) \sum_{k \geqslant 0} \mathbb{P}_x(\mathrm{Exc}_j \notin \Lambda_n, \text{ for } j < k \text{ and } H_k \geqslant 0).$$

By monotone convergence the last sum converges to $\sum_{k \geqslant 0} \mathbb{P}_0(H_k \geqslant -x) = h^{\uparrow}(x)$. □

Exercise 5.2. Show that Theorem 5.6 holds true if we replace the event $\Lambda_n = \{S_k \geqslant 0 : 0 \leqslant k \leqslant n\}$ by the event $\widetilde{\Lambda}_n = \{\tau_{[n,\infty)} < \tau_{\mathbb{Z}_{<0}}\}$.

5.3.2 Tanaka's construction

In this section, we give a direct and very neat construction of the random walk $(S^{\uparrow})$ conditioned to stay non-negative due to Tanaka. To start with, let Exc be the time and space reversal of a negative excursion of S

$$\mathrm{Exc} = (0, S_{T^{\geqslant}} - S_{T^{\geqslant}-1}, S_{T^{\geqslant}} - S_{T^{\geqslant}-2}, \dots, S_{T^{\geqslant}} - S_1, S_{T^{\geqslant}})$$

where we recall that $T^{\geqslant} = \inf\{k \geqslant 0 : S_k \geqslant 0\}$. One then considers independent copies $\mathrm{Exc}_1, \mathrm{Exc}_2, \dots$ of Exc obtained by running the walk S and looking at its strict ascending ladder process, which we glue together in the most natural way to get an infinite walk. Tanaka proved that the process obtained has the law of $S^{\uparrow}$.

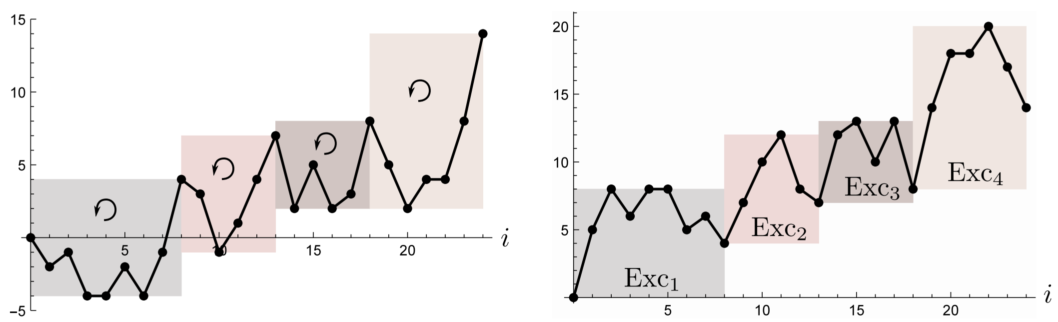


Figure 5.2: Illustration of Tanaka's construction of the process $S^{\uparrow}$.

Proposition 5.8 (Tanaka). *Suppose that (S) does not drift towards $-\infty$. Then the process obtained by concatenating i.i.d. time and space reversals of negative excursions of S has the same law as $S^{\uparrow}$.*

Proof. Denote by $(\mathcal{S}_n : n \geqslant 0)$ the process obtained by Tanaka's construction. Under our hypothesis, $\mathcal{S} \to +\infty$ and stays non-negative for all times. Fix s_0 and $s_1, s_2, \dots, s_k \in \mathbb{Z}_{\geqslant 0}$ and let us try to compute directly

$$\mathbb{P}(\mathcal{S}_1 = s_1, \dots, \mathcal{S}_k = s_k).$$

The problem is that if we are only given the path of the process up to time k, we do not know to which excursions it corresponds within the walk S. However, if we assume that the walk after time k never drops below $\mathcal{S}_k$ (large inequality), then we can reverse Tanaka's construction, in the sense that time k is a large ascending ladder height for S and $S_1 = \tilde{s}_1, \dots, S_k = \tilde{s}_k$ are obtained by reversing the excursion of the $s_1, \dots, s_k$ when read backwards in time and space. Hence decomposing according to the future infimum $I_k = \min\{\mathcal{S}_i : i \geqslant k\}$ and denoting $\theta_k = \inf\{i \geqslant k : \mathcal{S}_i = I_k\}$ we have

$$\begin{aligned}\mathbb{P}(\mathcal{S}_1 = s_1, \dots, \mathcal{S}_k = s_k) &= \sum_{i=0}^{s_k} \sum_{t=k}^{\infty} \sum_{s_{k+1},\dots,s_t = i \in \mathbb{Z}_{\geqslant 0}} \underbrace{\mathbb{P}(\mathcal{S}_1 = s_1, \dots, \mathcal{S}_t = i \text{ and } \theta_k = t \text{ and } I_k = i)}_{\text{product of the increments over the path}} \\ &= \prod_{i=0}^{k-1} \mu(s_{i+1} - s_i) \sum_{i=0}^{s_k} \sum_{\gamma} \mathbb{P}(\gamma),\end{aligned}$$

where the sum runs over all paths $\gamma : s_k \to i$ reaching i for the first time at its endpoint. After summing over $0 \leqslant i \leqslant s_k$ this is nothing else but $h^{\uparrow}(s_k)$. This proves the desired result. $\square$

In the case of the simple symmetric random walk, Tanaka's construction yields the construction that Pitman used to prove his famous theorem on Brownian motion and Bessel(3) process [?]:

Exercise 5.3 (Pitman's theorem in the discrete). Let (S) be the simple symmetric random on $\mathbb{Z}$ and denote by $I_n = \inf\{S_i : 0 \leqslant i \leqslant n\}$, the running infimum process. Show that the process

$$S_n - 2I_n, \quad n \geqslant 0,$$

has the same law as $(S_n^{\uparrow} : n \geqslant 0)$.

Exercise 5.4. Suppose that (S) drifts to $+\infty$ and denote by $J = \min\{n \geqslant 0 : S_n = \min_{i \geqslant 0} S_i\}$. Show that the process $(S_{J+n} - S_J : n \geqslant 0)$ has the same law as $S^{\uparrow}$ started from 0.

Exercise 5.5. Let $(S_n : n \geqslant 0)$ be a random walk with finite mean. We suppose that S is centered, in particular it is recurrent and thus oscillates. Show that $(S_n^{\uparrow} : n \geqslant 0)$ satisfies the law of large numbers:

$$\frac{S_n^{\uparrow}}{n} \xrightarrow[n\to\infty]{a.s.} 0.$$

5.4 Drift to $-\infty$ and Cramér's condition

Let us now suppose that we are in the annoying case when S drifts towards $-\infty$. In full generality it might be impossible to define a good notion of walk conditioned to stay non-negative since it might not exist any harmonic functions on $\mathbb{Z}_{\geqslant 0}$ and null on $\mathbb{Z}_{<0}$, see Doney [14]. However, in the so-called Cramér[2] case, things are under control:

[2]Harald Cramér (Swedish 1893 – 1985) not to confuse with Gabriel Cramer (Swiss 1704 – 1752)

Definition 5.1 (Cramér's condition)**.** *Suppose that S drifts towards $-\infty$. We say that μ satisfies Cramér's condition if there exists $\omega > 1$ so that*

$$\sum_{k\in\mathbb{Z}} \omega^k \mu(k) = 1.$$

It is easy to check that if such an ω exists, it must be unique by convexity. When Cramér's condition holds, one can define a step distribution $\tilde{\mu}(k) = \omega^k \mu(k)$ and the associated random walk $\tilde{S}$. It is easy to establish the following Radon-Nikodym derivative

$$\mathbb{E}[f(S_0,\dots,S_n)\omega^{S_n}] = \mathbb{E}[f(\tilde{S}_0,\dots,\tilde{S}_n)].$$

This walk $\tilde{S}$ can be seen as a first h-transformation of the walk (S) by the harmonic function $h(x) = \omega^x$. Since $\tilde{\mu}$ has an integrable left tail, its expectation is well defined and it is easy to check that $\mathbb{E}[\tilde{S}_1] > 0$ so that $\tilde{S}$ drifts towards $+\infty$. It thus admits a renewal function $\tilde{h}^{\uparrow}$ and a version $\tilde{S}^{\uparrow}$ conditioned to stay non-negative. By the previous display and the results of Section 5.2 we can write

$$\tilde{h}^{\uparrow}(\ell) = \sum_{i=0}^{\ell} \tilde{h}^{\downarrow}(i) = \sum_{i=0}^{\ell} \omega^{-i} h^{\downarrow}(i).$$

In total, the process $(\tilde{S}^{\uparrow})$ can be seen as the h-transformation of the walk (S) by the function $h(x) = \omega^x \tilde{h}^{\uparrow}(x)$ which is indeed harmonic on $\mathbb{Z}_{\geqslant 0}$. This process can be seen as the walk (S) conditioned to stay non-negative, see the next proposition (whose proof is omitted in these lecture notes, see [9]):

Proposition 5.9. *Assume that Cramér's condition holds and that $\sum_{k\in\mathbb{Z}} \tilde{\mu}(k)|k| < \infty$. Then we have*

$$\mathbb{P}\left(S_0 = s_0, \dots, S_k = s_k \mid \tau_{[n,\infty)} < \tau_{\mathbb{Z}_{<0}}\right) \xrightarrow[n\to\infty]{} \mathbb{P}(\tilde{S}^{\uparrow}_0 = s_0, \dots, \tilde{S}^{\uparrow}_k = s_k).$$

Bibliographical notes. As in the preceding chapter, the role of path transformation is ubiquitous when studying random walk conditioned to stay positive. Most of this chapter is adapted from the beautiful paper [9] of Bertoin & Doney. See [9] for a solution to Exercise 5.2. Tanaka's construction was first explained in [40]. The reference [6] is very nice survey of the applications of size-biasing. Doob's h-transformation via positive harmonic functions is also connected to the topic of Martin boundary, we refer the interested reader to [41, Chapter IV]. For applications of fluctuation theory to Lévy processes, we refer to the Saint-Flour course of Ron Doney [15].

Chapter VI: Renewal theory (Exercices de style)

Disclaimer: As usual in these notes and even more in this section, the main goal is rather to present a couple of ideas and techniques which are common in probability theory (generating function method, coupling, recursive distributional equations, harmonic functions...) rather than the shortest proof to well-known and useful results (see e.g. [25]). We will present 4 different proofs of Theorem 6.1 (in its simplest version).

In this chapter we study the behavior of a one-dimensional random walk (S) with step distribution μ supported by $\mathbb{R}_+$ and whose mean we denote by $m \in [0, \infty]$. Unless in the trivial case $\mu = \delta_0$ the walk drifts to ∞ and our goal is this chapter is to understand the asymptotic density of the random set

$$\mathcal{R} = \{S_0, S_1, \dots, S_n, \dots\}.$$

Such a process is often used to model the breakdown of different machines, then the random times X_i represent the time between one machine breaking down before another one does. The random set $\mathcal{R}$ then correspond to the times when a machine needs to be replaced. We first suppose that that μ is supported by $\{1, 2, \cdots\}$ and that $\gcd(\mathrm{supp}(\mu)) = 1$. Hence $\mathcal{R}$ is a random set of points of $\mathbb{Z}_+$ so that $\mathbb{P}(n \in \mathcal{R})$ is positive for n large enough. The main result of this chapter is the following:

Theorem 6.1 (*Erdös–Feller–Pollard*)

Under the above hypotheses we have

$$\lim_{n\to\infty} \mathbb{P}(n \in \mathcal{R}) = \frac{1}{m}.$$

Notice that even if $m = \int \mu(\mathrm{d}x)x = \infty$ the above statement has a well-defined meaning. Remark also that by the strong law of large numbers we have $n^{-1}S_n \to m$ almost surely as $n \to \infty$. It follows that $N_n = \sup\{k \geqslant 0 : S_k \leqslant n\}$ satisfies $n^{-1}N_n \to \frac{1}{m}$ almost surely. This easily implies a weaker "integrated" version of the last result:

$$\frac{1}{n}\sum_{i=0}^{n} \mathbb{P}(i \in \mathcal{R}) = \frac{1}{n}\mathbb{E}[\#(\mathcal{R} \cap [[0, n]])] = \frac{1}{n}\mathbb{E}[N_n] \xrightarrow[n\to\infty]{} \frac{1}{m},$$

where in the last convergence we used dominated convergence.

6.1 Proof via analytic combinatorics

In this section we give a proof of Theorem 6.1 when $m < \infty$ using analytical method based on generating series. We start with the case when μ has bounded support as a training.

6.1.1 Bounded support

We suppose here that μ is supported by $\{1, 2, \dots, n_0\}$ for some $n_0 \geqslant 1$. Hence μ can be encoded by the polynomial

$$P(z) = z\mu_1 + z^2\mu_2 + \cdots + z^{n_0}\mu_{n_0} \in \mathbb{R}[X].$$

Clearly the generating function of the law of $S_k = X_1 + \cdots + X_k$ is given by $(P(z))^k$. Summing-up for all $k \geqslant 0$ we deduce the following equality of formal series

$$\begin{aligned} \frac{1}{1 - P(z)} &= \sum_{k \geqslant 0} (P(z))^k \\ &= \sum_{k \geqslant 0} \sum_{n \geqslant 0} z^n \mathbb{P}(S_k = n) \\ &= \sum_{n \geqslant 0} \mathbb{E}[\#\{k \geqslant 0 : S_k = n\}] z^n \\ &= \sum_{n \geqslant 0} \mathbb{P}(n \in \mathcal{R}) z^n, \end{aligned} \tag{6.1}$$

where we have used the fact that $X \geqslant 1$ to argue that if n is visited by the walk S, then it is visited once. The last equality actuality makes perfect sense for all $z \in \mathbb{C}$ such that $|z| < 1$. We know use partial fraction decomposition on $(1 - P(z))^{-1}$:

- Notice first that $z = 1$ is a trivial pole of this function. It is of order 1 since $P'(1) = m > 0$.
- All the zeros of $1 - P(\cdot)$ must have modulus larger than 1 since otherwise

  $$|z\mu_1 + z^2\mu_2 + \cdots + z^{n_0}\mu_{n_0}| \leqslant |z|\mu_1 + |z|^2\mu_2 + \cdots + |z|^{n_0}\mu_{n_0} < 1$$

  (this can also directly be see from (6.1) since the equality makes sense for all $|z| < 1$).
- We also claim that the aperiodicity condition on μ ensures that $z = 1$ is the only root of $1 - P$ on $\mathbb{S}^1$. Indeed, if $z = \mathrm{e}^{i\theta}$ with $\theta \in (0, 2\pi)$ was another root of $1 - P$, by the equality case of the triangle inequality it would follow that $\mu_k z^k = \mu_k \mathrm{e}^{ik\theta}$ must be real and positive and in fact the same is true for each $k \in \cup_{i \geqslant 1} \mathrm{Supp}(\mu^{*i})$. Since $\gcd(\mathrm{Supp}(\mu)) = 1$ we can find two consecutive such k in $\cup_{i \geqslant 1} \mathrm{Supp}(\mu^{*i})$ and so $\theta = 0 \bmod 2\pi$ which is a contradiction.

Hence the partial fraction decomposition of $1/(1 - P)$ reads as follows

$$\frac{1}{1 - P(z)} = \frac{(P'(1))^{-1}}{z - 1} + \sum_i \frac{\alpha_i}{(z - \beta_i)^{\ell_i}},$$

where $\ell_i \geqslant 1, \alpha_i \in \mathbb{C}$ and β_i are the other roots of $1-P$ in particular $|\beta_i| > 1$. If we now expand each of these terms in series, we see that the expansion of the first term gives $\frac{1}{m}\sum_{n\geqslant 0} z^n$ whereas the expansion of the other terms contribute with a series of the form

$$\sum_{n\geqslant 0} z^n \zeta_{n,i},$$

where $\zeta_{n,i} = o(1)$ as $n \to \infty$ (more precisely $|\zeta_{n,i}| = O(|\beta_i|^{-n} n^{\ell_i})$). Taking the coefficient in front of z^n in both sides of (6.1) we indeed deduce the desired result i.e.

$$[z^n]\frac{1}{1-P(z)} = \mathbb{P}(n \in \mathcal{R}) = \frac{1}{m} + o(1).$$

6.1.2 Unbounded support with $m < \infty$

We now suppose that μ has a possibly unbounded support on $\{1, 2, \dots\}$ but that $m = \sum_{k\geqslant 0} k\mu_k < \infty$. Our surrogate to the explicit partial fraction decomposition will be a theorem of Wiener on power series whose proof can be found in [?, Theorem 18.21]:

Lemma 6.2 (Wiener). *If $g(z) = \sum_{n\geqslant 0} a_n z^n$ is a power series with $\sum_{n\geqslant 0} |a_n| < \infty$ which has no zeros inside $\overline{\mathbb{D}}$ then $1/g(z)$ has a power series expansion*

$$\frac{1}{g(z)} = \sum_{n\geqslant 0} b_n z^n \qquad \textit{with} \qquad \sum_{n\geqslant 0} |b_n| < \infty.$$

Following the last section, if $P(z) = \sum_{k\geqslant 1} z^k \mu_k$ is the generating function of μ (which is not necessarily a polynomial function) then

$$\frac{1-z}{1-P(z)} = (1-z)\sum_{n\geqslant 0} z^n \mathbb{P}(n \in \mathcal{R}) = \sum_{n\geqslant 0} z^n \left(\mathbb{P}(n \in \mathcal{R}) - \mathbb{P}(n-1 \in \mathcal{R})\right).$$

It is easy then to check that $\frac{1-P(z)}{1-z} = \sum_{n\geqslant 0} z^n \sum_{k=n+1}^{\infty} \mu_k$ and so this function satisfies the requirements of the lemma because

$$\sum_{n\geqslant 0} \sum_{k\geqslant n+1} \mu_k = m < \infty.$$

Applying the above lemma we deduce that $\sum |\mathbb{P}(n \in \mathcal{R}) - \mathbb{P}(n-1 \in \mathcal{R})| < \infty$ and consequently by Abel's theorem that $\mathbb{P}(n \in \mathcal{R})$ converges as $n \to \infty$ towards

$$\lim_{n\to\infty} \mathbb{P}(n \in \mathcal{R}) = \lim_{z\to 1} \frac{1-z}{1-P(z)} = \frac{1}{P'(1)} = \frac{1}{m}.$$

Remark 6.1. In passing we have proved a slightly stronger version of Theorem 6.1 since when $m < \infty$ we have

$$\sum_{n\geqslant 0} |\mathbb{P}(n \in \mathcal{R}) - \mathbb{P}(n+1 \in \mathcal{R})| < \infty.$$

6.2 Finite mean case via stationarity and coupling

The probabilistic reader may be disappointed by the last proofs based on analytical methods. In this section, we provide an alternative proof of Theorem 6.1 in the case when $m < \infty$ based on the probabilistic concepts of coupling and stationarity.

6.2.1 Point processes and stationarity

Definition 6.1. *A* renewal set $\mathcal{S}$ *is a set of points* $\{s_0 < s_1 < \dots\}$ *in* $\mathbb{Z}_+$. *Such a set is naturally associated with its inter-times* $x_0 = s_0 - 0, x_1 = s_1 - s_0, x_2 = s_2 - s_1$ *etc. We can define a translation operation* θ *on the set of all renewal sets by setting* $\theta\mathcal{S} = \mathcal{S}'$ *where* $\mathcal{S}'$ *is described by its inter-times* $(x'_i)_{i \geqslant 0}$ *where*

$$\text{if } x_0 = 0 \text{ then } \begin{cases} x'_0 = x_1 - 1 \\ x'_i = x_{i+1} \text{ for } i \geqslant 1 \end{cases} \qquad \text{otherwise if } x_0 > 0 \text{ then } \begin{cases} x'_0 = x_0 - 1 \\ x'_i = x_i \text{ for } i \geqslant 1 \end{cases}$$

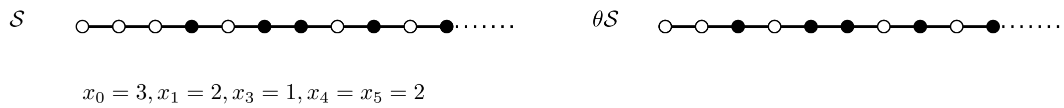


Figure 6.1: Illustration of the definition of a renewal set and its translate.

In words, $\theta\mathcal{S}$ is just obtained by erasing the first point 0 of $\mathbb{Z}_+$ (and possibly the point of $\mathcal{S}$ at this position) and translating all other values and points by 1. Our strategy to prove Theorem 6.1 is to show the following stronger convergence

$$\theta^n \mathcal{R} \xrightarrow[n\to\infty]{(d)} \tilde{\mathcal{R}}, \tag{6.2}$$

where $\tilde{\mathcal{R}}$ is a renewal set whose law will be described later on and satisfies $\mathbb{P}(0 \in \tilde{\mathcal{R}}) = \frac{1}{m}$ (recall that we focused first on the case $m < \infty$). The convergence in distribution of renewal set in the last display simply means finite-dimensional convergence of its associated inter-times. If (6.2) is granted, we have in particular

$$\mathbb{P}(n \in \mathcal{R}) = \mathbb{P}(0 \in \theta^n \mathcal{R}) \xrightarrow[n\to\infty]{} \mathbb{P}(0 \in \tilde{\mathcal{R}}) = \frac{1}{m},$$

as desired.

A stationary renewal set

We first describe explicitly the law of $\theta^n\mathcal{R}$.

Proposition 6.3. *The law of the renewal set* $\theta^n\mathcal{R}$ *is described by its inter-times* $(X_i^{(n)})_{i \geqslant 0}$ *whose law is characterized as follows:*

- $(X_i^{(n)})_{i\geqslant 1}$ *are i.i.d. random variables of law μ independent of $X_0^{(n)}$,*
- *the law of $X_0^{(n)}$ is prescribed by the following recursive distributional equation: $X_0^{(0)} = 0$ almost surely and for $n \geqslant 1$ we have*

$$X_0^{(n+1)} \overset{(d)}{=} (X_0^{(n)} - 1)\mathbf{1}_{X_0^{(n)}>0} + (X-1)\mathbf{1}_{X_0^{(n)}=0}, \tag{6.3}$$

where X is of law μ and independent of $X_0^{(n)}$.

Proof. This is merely a writing exercise. The case $n = 1$ is granted by definition of the renewal set $\mathcal{R}$. Consider now $f_0, f_1, f_2, \dots, f_k$ bounded measurable functions. Then by definition of θ we have

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=0}^{k} f_i(X_i^{(n+1)})\right] &= \mathbb{E}\left[\mathbf{1}_{X_0^{(n)}>0} f_0(X_0^{(n)}-1)\prod_{i=1}^{k} f_i(X_i^{(n)})\right] + \mathbb{E}\left[\mathbf{1}_{X_0^{(n)}=0} f_0(X_1^{(n)}-1)\prod_{i=1}^{k} f_i(X_{i+1}^{(n)})\right] \\
&\underset{\text{ind. hyp}}{=} \mathbb{E}[\mathbf{1}_{X_0^{(n)}>0} f_0(X_0^{(n)}-1)]\prod_{i=1}^{k}\mathbb{E}[f_i(X)] + \mathbb{E}[\mathbf{1}_{X_0^{(n)}=0} f_0(X_1^{(n)}-1)]\prod_{i=1}^{k}\mathbb{E}[f_i(X)] \\
&= \mathbb{E}[\mathbf{1}_{X_0^{(n)}>0} f_0(X_0^{(n)}-1) + \mathbf{1}_{X_0^{(n)}=0} f_0(X_1^{(n)}-1)]\prod_{i=1}^{k}\mathbb{E}[f_i(X)].
\end{aligned}$$

This exactly tells us that $(X_i^{(n+1)})_{i\geqslant 0}$ has the desired law. □

Equation (6.3) is an example of a recursive distributional equation. Indeed, if μ_n is the law of $X_0^{(n)}$ then this equation states that

$$\mu_{n+1} = \phi(\mu_n), \tag{6.4}$$

where the function ϕ maps a law ν supported by $\mathbb{Z}_+$ to the law $\phi(\nu)$ of $(Y-1)\mathbf{1}_{Y>0}+(X-1)\mathbf{1}_{Y=0}$ where Y has law ν and is independent of X which has law μ. If we interpret the last display as a classical iteration scheme (but in the space of measures) we are naturally yield to consider fixed point (and contraction property) for the mapping ϕ. This is simple in our case:

Lemma 6.4. *Suppose $m < \infty$ then Equation (6.4) has a unique fixed point ν whose law is given by*

$$\nu_k = \frac{1}{m}\mu((k,\infty)), \qquad \textit{for } k \geqslant 0.$$

Proof. We can rewrite the equation $\nu = \phi(\nu)$ equivalently as

$$\nu_k = \nu_{k+1} + \nu_0\mu_{k+1}, \qquad \text{for } k \geqslant 0.$$

Summing these equation for $k = n, n+1, \dots$ we find that $\nu_k = \nu_0\mu((k,\infty))$ for all $k \geqslant 0$. Since ν has to be a probability this forces ν_0 to be the inverse of $\sum_{k\geqslant 0}\mu((k,\infty)) = m$. In this case the last calculation is reversible and this proves the lemma. □

Remark 6.2. One can interpret the law ν of the last lemma from a probabilistic point of view: Let $\overline{\mu}$ denote the size biased distribution of μ given by

$$\overline{\mu}_k = \frac{k\mu_k}{m} \qquad \text{for } k \geqslant 1.$$

If we first sample Z according to $\overline{\mu}$ and next conditionally on Z sample $Y \in \{0, 1, 2, \dots, Z\}$ uniformly at random then the law of Y follows ν. This has the following interpretation: when n is very large the point n lies in an interval between two points of the renewal set whose law is that of X_1 *biaised* by its length. Furthermore, conditionally on the length of this interval, the point n is asymptotically uniformly distributed in it.

Definition 6.2. *Suppose that the mean m of μ is finite. Then we let $\tilde{\mathcal{R}}$ be the renewal set whose inter-times are independent and given by $X_0 \sim \nu$ and $X_i \sim \mu$ for $i \geqslant 1$.*

Proposition 6.5. *The renewal set $\tilde{\mathcal{R}}$ is stationary in the sense that $\theta\tilde{\mathcal{R}} = \tilde{\mathcal{R}}$ in distribution.*

Proof. This is a consequence of the proof of Proposition 6.3 as well as the fact that ν is a fixed point for Equation (6.4). □

One way to finish the proof of Theorem 6.1 (more precisely (6.2)) would be to use Proposition 6.3 and doing the following exercise:

Exercise 6.1 (Contraction for the recursive distributional equation (6.4)). If ν_0 is an arbitrary law on $\mathbb{Z}_+$, define the sequence of probability measures $(\nu_n)_{n\geqslant 0}$ recursively by $\nu_{n+1} = \phi(\nu_n)$ as in (6.4) for $n \geqslant 0$.

1. Show that if $\int \mu(\mathrm{d}x)x < \infty$ then $\nu_n \to \nu$ weakly (ν defined in Proposition 6.4).

2. Otherwise show that if $m = \infty$ then for all $k \geqslant 0$ we have $\nu_n(k) \to 0$ as $n \to \infty$.

Exercise 6.2. Adapt the last section in the case when we define a renewal set to be a two-sided set of points $\{\dots < s_{-2} < s_{-1} < s_0 < s_1 < s_2 < \dots\}$ of $\mathbb{Z}$. The translation θ consists then in moving the underlying points of $\mathbb{Z}$ by $+1$.

6.2.2 Coupling argument

If we were working with the stationary renewal set $\tilde{\mathcal{R}}$ instead of $\mathcal{R}$ then the proof of Theorem 6.1 would be a piece of cake since for $n \geqslant 0$ we have

$$\mathbb{P}(n \in \tilde{\mathcal{R}}) = \mathbb{P}(0 \in \theta^n \tilde{\mathcal{R}}) \underset{\text{stat.}}{=} \mathbb{P}(0 \in \tilde{\mathcal{R}}) = \nu_0 = \frac{1}{m}. \tag{6.5}$$

The idea to prove Eq. (6.2) is to couple $\mathcal{R}$ and $\tilde{\mathcal{R}}$ (i.e. to construct both of them on the same probability space) in such a way that for n large enough (this large enough being random) we have $\theta^n \tilde{\mathcal{R}} = \theta^n \mathcal{R}$. Let us proceed. We start with $(Y_i)_{i\geqslant 0}$ and $(Y_i')_{i\geqslant 0}$ be independent variables so that $Y_0 = 0$ almost surely, $Y_0' \sim \nu$ and for $i \geqslant 1$ we have $Y_i \sim Y_i' \sim \mu$. We associate with these variables the random walk

$$\Delta_n = (Y_0 + \cdots + Y_n) - (Y_0' + \cdots + Y_n').$$

Hence (Δ) is a centered random walk and so is recurrent by Theorem 3.5. We denote $\tau = \inf\{k \geqslant 0 : \Delta_k = 0\}$. We then construct two renewal sets $\mathcal{S}$ and $\tilde{\mathcal{S}}$ whose inter-times (X_i) and $(\tilde{X}_i)$ are described as follows:

$$\text{for } 0 \leqslant i \leqslant \tau \text{ we put } X_i = Y_i \text{ and } \tilde{X}_i = Y_i',$$

$$\text{whereas for } i \geqslant \tau + 1 \text{ we put } X_i = \tilde{X}_i = Y_i.$$

Proposition 6.6. *The above construction $(\mathcal{S}, \tilde{\mathcal{S}})$ is indeed a coupling of $\mathcal{R}$ and $\tilde{\mathcal{R}}$, in other words we do have $\mathcal{R} = \mathcal{S}$ and $\tilde{\mathcal{R}} = \tilde{\mathcal{S}}$ in law.*

Proof. It is clear that $\mathcal{R} = \mathcal{S}$ in distribution since the inter-times of $\mathcal{R}$ are given by the X_i no matter τ. In the case of $\tilde{\mathcal{R}}$ we have to show that $(\tilde{X}_i)_{i \geqslant 0}$ are i.i.d. random variables of law ν for $i = 0$ and μ for $i \geqslant 1$. Let $f_0, \ldots, f_k$ be bounded measurable functions and let us compute

$$\mathbb{E}\left[\prod_{i=0}^{k} f_i(\tilde{X}_i)\right] = \mathbb{E}\left[\prod_{i=0}^{k-1} f_i(\tilde{X}_i)\mathbf{1}_{\tau<k} f_k(Y_k)\right] + \mathbb{E}\left[\prod_{i=0}^{k-1} f_i(\tilde{X}_i)\mathbf{1}_{\tau\geqslant k} f_k(Y_k')\right].$$

Noticing that $\{\tau \geqslant k\}$ is measurable with respect to $Y_0, \ldots, Y_{k-1}, Y_0', \ldots, Y_{k-1}'$ we get by independence that

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=0}^{k} f_i(\tilde{X}_i)\right] &= \mathbb{E}\left[\prod_{i=0}^{k-1} f_i(\tilde{X}_i)\mathbf{1}_{\tau<k}\right]\mathbb{E}[f_k(Y_k)] + \mathbb{E}\left[\prod_{i=0}^{k-1} f_i(\tilde{X}_i)\mathbf{1}_{\tau\geqslant k}\right]\mathbb{E}[f_k(Y_k')] \\
&= \left(\mathbb{E}\left[\prod_{i=0}^{k-1} f_i(\tilde{X}_i)\mathbf{1}_{\tau<k}\right] + \mathbb{E}\left[\prod_{i=0}^{k-1} f_i(\tilde{X}_i)\mathbf{1}_{\tau\geqslant k}\right]\right) \cdot \mathbb{E}[f_k(Y_1)] \\
&= \mathbb{E}\left[\prod_{i=0}^{k-1} f_i(\tilde{X}_i)\right]\mathbb{E}[f_k(Y_1)].
\end{aligned}$$

Iterating the argument until we reach $k = 0$ we have proved the proposition. □

Proof of Eq. (6.2). We consider the last coupling $(\mathcal{S}, \tilde{\mathcal{S}})$ of $\mathcal{R}$ and $\tilde{\mathcal{R}}$. If f is a bounded measurable function we can write

$$\begin{aligned}
\mathbb{E}[f(\theta^n \mathcal{R})] &= \mathbb{E}[f(\theta^n \mathcal{S})] \\
&= \mathbb{E}[f(\theta^n \mathcal{S})\mathbf{1}_{\tau\geqslant n}] + \mathbb{E}[f(\theta^n \tilde{\mathcal{S}})\mathbf{1}_{\tau<n}] \\
&= \mathbb{E}[f(\theta^n \mathcal{S})\mathbf{1}_{\tau\geqslant n}] - \mathbb{E}[f(\theta^n \tilde{\mathcal{S}})\mathbf{1}_{\tau\geqslant n}] + \mathbb{E}[f(\theta^n \tilde{\mathcal{S}})].
\end{aligned}$$

Since f is bounded and because $\mathbb{P}(\tau \geqslant n) \to 0$ the first two terms in the last display tend to 0 as $n \to \infty$. As for the third term, arguing as in (6.5) we see that it is equal to $\mathbb{E}[f(\tilde{\mathcal{S}})]$ no matter the value of n. Hence the whole thing tends to $\mathbb{E}[f(\tilde{\mathcal{S}})]$ as wanted. □

6.3 A deceptively simple, analytic and tricky proof in the general case

In this section we finally give an elementary proof of Theorem 6.1 which is also valid in the case $m = \infty$. To simplify notation we write

$$u_n = \mathbb{P}(n \in \mathcal{R}) \quad \text{for } n \geqslant 0, \tag{6.6}$$

with $u_0 = 1$ and $u_n = 0$ for $n < 0$. By the Markov property applied after the first step of the random walk we naturally get the following equation for $n \geqslant 1$

$$u_n = \sum_{k=1}^{\infty} \mu_k u_{n-k} \quad \text{for } n \geqslant 1. \tag{6.7}$$

Let $\lambda \leqslant \Lambda$ respectively be the $\liminf$ and $\limsup$ of the sequence u_n. Suppose that $\psi(n)$ is a subsequence such that $u_{\psi(n)} \to \Lambda$, then we claim that for all $i \geqslant 1$ we have $u_{\psi(n)-i} \to \Lambda$ as well when $n \to \infty$. This is clearly implied by (6.7) for all values of i such that $\mu_i > 0$ and is later extended to all $i \geqslant 1$ by aperiodicity (Exercise). Summing (6.7) for $n = \psi(N), \psi(N-1), \dots$ and re-arranging the terms we get that

$$u_{\psi(N)}\mu((0,\infty)) + u_{\psi(N)-1}\mu((1,\infty)) + \cdots + u_0\mu((\psi(N),\infty)) = 1.$$

Sending $N \to \infty$ we get by the above remark that $\Lambda \cdot \sum_{k=0}^{\infty} \mu((k,\infty)) \leqslant 1$. This proves that $\Lambda \leqslant \frac{1}{m}$ and this makes sense even if $m = \infty$. Repeating the argument with the $\liminf$ we get similarly that $\lambda \geqslant \frac{1}{m}$ and this completes the proof of the theorem in the general case.

6.4 Extensions

6.4.1 Non increasing case

In this section we use a few results of the last chapter in order to investigate the case when μ is not necessary supported by positive integers. We suppose now that μ is supported by $\mathbb{Z}$ (and not $\mathbb{Z}_{>0}$ anymore), that its support generates $\mathbb{Z}$ and furthermore that μ admits a (finite and) strictly positive expectation $0 < m < \infty$. As above we consider the random walk (S) whose increments are i.i.d. of law μ (which may not be strictly increasing anymore).

Proposition 6.7. *Under the above hypotheses we have*

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{S_i = n}\right] \xrightarrow[n\to\infty]{} \frac{1}{m}.$$

Proof. The idea is to decompose the walk along the strict ascending ladder variables. Remember that $(H_i^+)_{i\geqslant 0}$ and (T_i^+) are the strict ascending ladder heights and epochs. Then by the result of the preceding chapter (H^+) is a random walk with i.i.d. strictly positive increments of law H_1^+ and since (S) drifts towards ∞ we deduce that $\mathbb{E}[H_1^+] < \infty$. If we denote by $\mathcal{H}$ the renewal set $\{0 = H_0^+ < H_1^+ < H_2^+ < \cdots\}$ then we can apply the result of the last section and deduce that

$$\lim_{n\to\infty} \mathbb{P}(n \in \mathcal{H}) = \frac{1}{\mathbb{E}[H_1^+]}.$$

On the other hand, we can write

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{S_i=n}\right] = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_{H_k^+=j} \sum_{i=T_k^+}^{T_{k+1}^+-1} \mathbf{1}_{S_i=n}\right].$$

By the strong Markov property and translation invariance the last expectation which we now denote by $\varphi(j-n)$ only depends on $j-n$ and equal 0 as long as $n > j$. By resuming over k, the last display is also equal to

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{S_i=n}\right] = \sum_{j=n}^{\infty} \mathbb{P}(j \in \mathcal{H})\varphi(n-j).$$

Now notice that $\sum_{k\geqslant 0} \varphi(k) = \mathbb{E}[T_1^+]$ which is equal to $\mathbb{E}[H_1^+]/m$ by Wald's identity. Letting $n \to \infty$, we have from Theorem 6.1 that $\mathbb{P}(j \in \mathcal{H}) \to \frac{1}{\mathbb{E}[H_1^+]}$. Hence we can use dominate convergence to finally get

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{S_i=n}\right] \xrightarrow[n\to\infty]{} \frac{1}{\mathbb{E}[H_1^+]} \times \frac{\mathbb{E}[H_1^+]}{m} = \frac{1}{m}. \quad \square$$

6.4.2 Continuous case

In this section we state the analog of Theorem 6.1 in the continuous case and only sketch the main difference in the proof. Let μ be a distribution over $\mathbb{R}_+^*$ which is non-lattice. We denote by m the mean of μ. As usual let $(S_i)_{i\geqslant 0}$ be a random walk with i.i.d. increments of law μ started from 0 and set $\mathcal{R} = \{S_0, S_1, S_2, \cdots\}$.

Theorem 6.8 (*Blackwell*)

For any $h > 0$ we have

$$\lim_{t\to\infty} \mathbb{E}\Big[\#\left(\mathcal{R} \cap [t, t+h)\right)\Big] = \frac{h}{m}.$$

The strategy of the proof of the above theorem in the case $m < \infty$ can be modeled on Section 6.2. We first build a renewal set $\tilde{\mathcal{R}}$ whose inter-times as i.i.d. of law μ except for the first time X_0 whose distribution is characterized as follows:

$$\mathbb{E}[f(X_0)] = \frac{1}{m}\int \mu(\mathrm{d}x)x \int_0^x \mathrm{d}s f(s).$$

One can then check that the law of $\tilde{\mathcal{R}}$ is invariant under translation by any fixed time $t > 0$. Using this stationarity we deduce that

$$\mathbb{E}\left[\#\left(\tilde{\mathcal{R}} \cap [t, t+h)\right)\right] = C \cdot h,$$

for some constant $C \geqslant 0$. A argument similar to the one presented just after Theorem 6.1 using the law of large numbers shows that $C = \frac{1}{m}$. One then use a similar coupling argument to transfer the result from $\tilde{\mathcal{R}}$ to $\mathcal{R}$. In this case by Theorem 3.5 the random walk (Δ) of the last section is again recurrent. However it does not come back *exactly* to 0 (because the walk is non-lattice) but it approaches 0 arbitrarily close. Hence, for any $\varepsilon > 0$ one can couple $\mathcal{R}$ and $\tilde{\mathcal{R}}$ so that for t large enough their t-translates are the same up to a small shift of length smaller than $\varepsilon > 0$. Provided that $\varepsilon > 0$ is small enough in front of h this is sufficient to deduce Theorem 6.8.

Bibliographical notes. The first proof based on analytic combinatorics is close to the original proof of Erdös–Feller and Pollard [**?**] named "A property of power series with positive coefficients". This theorem was later revisited in [**?**] using the coupling method (a must-have in nowadays probability toolbox) and highlights the importance of the stationary distribution obtained by size-biaising the first inter-time.

Part III:
d-dimensional random walks

In this chapter we fix a distribution μ on $\mathbb{Z}^d$ for $d \geqslant 1$ and as usual put

$$S_n = X_1 + \cdots + X_n,$$

where $X_1, X_2, \ldots$ are i.i.d. copies of distribution μ. This random walk (S) can thus be seen as a Markov chain (with homogeneous and translation invariant transitions) on $\mathbb{Z}^d$. We will usually assume that this is a "true" random walk on $\mathbb{Z}^d$ in the following sense:

Definition 6.3. *We say that the walk is aperiodic if the Markov chain (S) is irreducible and aperiodic on $\mathbb{Z}^d$.*

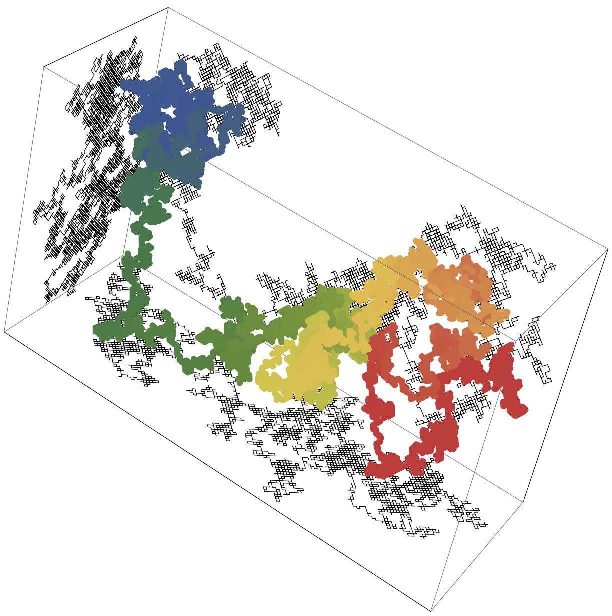


Figure 6.2: A random walk in three dimensions (the color varies from blue to red as time passes) and its projections onto the three planes.

Chapter VII: Applications of the Fourier transform

Recall that μ is the step distribution of an aperiodic random walk (S) on $\mathbb{Z}^d$. This chapter is devoted to the use of the Fourier[1] transform of the measure μ in order to study the walk (S). Recall that this is defined by

$$\hat{\mu}(\xi) = \mathbb{E}[\mathrm{e}^{i\xi\cdot X_1}], \qquad \text{for } \xi \in \mathbb{R}^d.$$

The main idea being of course Cauchy's formula relating probability estimation on the random walk to estimating integrals of powers of the Fourier transform. Namely we have

$$\begin{aligned}\forall x \in \mathbb{Z}^d, \quad \mathbb{P}(S_n = x) &= \sum_{k\in\mathbb{Z}^d} \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \mathrm{d}\xi\, \mathrm{e}^{-\mathrm{i}\xi\cdot x} \mathrm{e}^{\mathrm{i}\xi\cdot k} \mathbb{P}(S_n = k) \\ &= \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \mathrm{d}\xi\, \mathrm{e}^{-\mathrm{i}\xi\cdot x} \mathbb{E}[\mathrm{e}^{\mathrm{i}\xi\cdot S_n}] = \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \mathrm{d}\xi\, \mathrm{e}^{-\mathrm{i}\xi\cdot x} \left(\hat{\mu}(\xi)\right)^n,\end{aligned} \tag{7.1}$$

where we used the fact that $\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi S_n}] = (\hat{\mu}(\xi))^n$ by independence of the increments and where the interversion of series and integral is easily justified by dominated convergence. Before drawing important consequences of this formula, let us first recall a few basic properties of $\hat{\mu}$.

7.1 Estimates on Fourier transform

First $\hat{\mu}$ is continuous and 2π periodic in each coordinate and it characterizes the distribution μ (Lévy's theorem). Clearly its modulus is less than or equal to 1. Actually, under our aperiodicity assumption we even have

Lemma 7.1. *When μ is aperiodic we have*

$$|\hat{\mu}(\xi)| < 1, \qquad \text{for } \xi \in [0, 2\pi)^d \backslash 0_{\mathbb{Z}^d}.$$

Proof. Indeed, if we have $|\hat{\mu}(\xi)| = |\mathbb{E}[\mathrm{e}^{i\xi\cdot X_1}]| = 1$ we have also $|\big(\hat{\mu}(\xi)\big)^n| = |\mathbb{E}[\mathrm{e}^{i\xi\cdot S_n}]| = 1$. This implies by the equality case in the triangle inequality that all $\mathrm{e}^{i\xi\cdot x}$ for $x \in \mathrm{Supp}(\mathcal{L}(S_n))$ are (positively) aligned. Using the aperiodicity assumption, one can choose n large enough so that the support of the law of S_n contains 0 and the basis vectors $(1, 0, 0, \dots)$ $(0, 1, \dots)$ up to $(0, 0, \dots, 1)$. This shows that ξ must have all its coordinates equal to 0 modulo 2π. $\square$

[1]  Jean Baptiste Joseph Fourier 1768–1830

Also recall that when μ has a finite first moment $m \in \mathbb{R}^d$ then we have $\hat{\mu}(\xi) = 1 + \mathrm{i}m\cdot\xi + o(\xi)$ as $\xi \to 0$. If in addition μ admits a second moment (i.e. $\int \mu(\mathrm{d}x)|x|^2 < \infty$) we can define the covariance matrix $Q = (\mathbb{E}[(X)_i(X)_j])_{1\leqslant i,j\leqslant d}$ where $(X)_i$ represents the i-th coordinate of the vector X which follows the law μ. Then we have

$$\hat{\mu}(\xi) = 1 + \mathrm{i}m \cdot \xi - \frac{{}^t\xi \cdot Q \cdot \xi}{2} + o(|\xi|^2). \tag{7.2}$$

Exercise 7.1. We work in the one-dimensional case to simplify:

1. Prove that if the kth moment m_k of μ exists then $\hat{\mu}^{(k)}(0)$ exists and is equal to $i^k m_k$.
2. Prove that if $\hat{\mu}$ admits a second derivative then m_2 is finite.
3. Prove that $\hat{\mu}$ may admit a derivative at 0 without m_1 being finite (take $\mu_k = c/(k^2\ln(k))$ for $k \geqslant 2$ and an appropriate $c > 0$).

Lemma 7.2. *Under the aperiodicity assumption there exists* $\lambda > 0$ *such that*

$$|\hat{\mu}(\xi)| \leqslant 1 - \lambda|\xi|^2, \quad \forall \xi \in [-\pi/2, \pi/2]^d.$$

Proof. Let us suppose first that μ has finite support. Then from (7.2) we deduce that

$$|\hat{\mu}(\xi)|^2 = \left(1 - \frac{{}^t\xi \cdot Q \cdot \xi}{2}\right)^2 + |m\cdot\xi|^2 + o(|\xi|^2) = 1 - {}^t\xi\cdot Q\cdot\xi + |m\cdot\xi|^2 + o(|\xi|^2).$$

By the aperiodicity condition, there is no $\xi \in [-\pi/2, \pi/2]^d$ so that $\xi \cdot X$ is constant, hence by the Cauchy–Schwarz we have the strict inequality between the two quadratic forms ${}^t\xi \cdot Q \cdot \xi > |m\cdot\xi|^2$. By compactness we deduce that ${}^t\xi\cdot Q\cdot\xi - |m\cdot\xi|^2 > \lambda|\xi|^2$ for some $\lambda > 0$ for all $\xi \in \mathbb{R}^d$. The statement of the lemma follows from this local estimate around 0 combined with Lemma 7.1. To deduce the general case, pick a finite subset $A \subset \mathbb{Z}^d$ so that $\{x \in A : \mu_x > 0\}$ is already aperiodic and generates $\mathbb{R}^d$ as a vector space. We then bound

$$|\hat{\mu}(\xi)| = |\mu(A)\cdot\hat{\mu}_A(\xi) + (1-\mu(A))\hat{\mu}_{A^c}(\xi)| \leqslant \mu(A)\cdot|\hat{\mu}_A(\xi)| + (1-\mu(A)),$$

where μ_A and μ_{A^c} are the conditional probabilities on A and A^c respectively. Using the statement of the lemma on μ_A we deduce the general form for arbitrary aperiodic measures μ. □

7.2 Anti-concentration inequalities

The concentration function of a real random variable $X \in \mathbb{R}$ is a convenient way to encode how much the distribution of X is spread out it has been introduced by Paul Lévy as

$$Q(X;\lambda) = \sup_{x\in\mathbb{R}} \mathbb{P}(X \in [x, x+\lambda]), \qquad \lambda \geqslant 0.$$

In our case of discrete random variables, we shall be interested in $\sup_{k\in\mathbb{Z}} \mathbb{P}(X = k)$. The following theorem roughly shows that the distribution of a n step random walk is always more spread out than in the simple random walk case:

Theorem 7.3 (*Anti-concentration*)

If $(S_n)_{n\geqslant 0}$ is an aperiodic d-dimensional random walk then there exists some constant $C>0$ so that for every $n \geqslant 1$ we have

$$\sup_{x\in\mathbb{Z}} \mathbb{P}(S_n = x) \leqslant C n^{-d/2}.$$

Proof. Using Cauchy Formula (7.1) and a trivial bound we have

$$\mathbb{P}(S_n = x) = \frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d} \mathrm{d}\xi\, \mathrm{e}^{-\mathrm{i}\xi\cdot x}\,(\hat{\mu}(\xi))^n \leqslant \frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d} \mathrm{d}\xi\ |\hat{\mu}(\xi)|^n\,.$$

Since $|\hat{\mu}(\xi)| < 1$ outside of $\xi = 0$, the main contribution of the previous integral is located around 0 and so we may use Lemma 7.2 to upper-bound it (up to neglecting an exponentially small factor) by

$$\int_{(-\pi/2;\pi/2)^d} \mathrm{d}\xi\ |\hat{\mu}(\xi)|^n \leqslant \int_{(-\pi/2;\pi/2)^d} \mathrm{d}\xi\ \left(1-\lambda|\xi|^2\right)^n.$$

Performing the change of variable $\sqrt{\lambda}\xi = z/\sqrt{n}$ and passing to polar coordinates, the previous integral is upper bounded by

$$\frac{C}{n^{d/2}}\int_0^\infty \mathrm{d}r\, r^{d-1}(1-\frac{r^2}{n})^n \leqslant \frac{C}{n^{d/2}}\int_0^\infty \mathrm{d}r\, r^{d-1}\mathrm{e}^{-r^2} = \frac{C'}{n^{d/2}},$$

for some constants $C, C' > 0$ depending on λ (hence on μ) but not on n. This proves the theorem □

Remark 7.1. The previous result gives an anti-concentration result in the sense that the distribution of S_n cannot put too much mass on a single point. Anti-concentrations results such as Littlewood–Offord theorem, or Kolmogorov-Rogozin inequality are very useful e.g. in additive combinatorics or in the theory of random matrices and random polynomials.

The upper-bounds on the previous theorem are attained for random walk with first and second moments, see the forthcoming local central limit theorem. Here is a (non-trivial) corollary whose proof is immediate from the previous theorem:

Corollary 7.4. *Any aperiodic random walk in $\mathbb{Z}^d$ with $d \geqslant 3$ is transient.*

Proof. Using the previous theorem, the expected number of return to $0_{\mathbb{Z}^d}$ by such a walk is upper-bounded by $C\sum_{n\geqslant 1} n^{-d/2} < \infty$ if $d \geqslant 3$. It follows that the walk must be recurrent. □

Exercise 7.2. Let $X_i, i \geqslant 1$ be i.i.d. random variables with $\mathbb{P}(X = \pm 1) = \frac{1}{2}$. Prove that we have

$$\mathbb{P}(X_1 + 2X_2 + \cdots + nX_n = 0) \underset{n\to\infty}{\sim} \sqrt{\frac{6}{\pi}}\cdot n^{-3/2},$$

when we restrict to n divisible by 4.

7.3 Recurrence criterion

In this section, we give criterion for recurrence of a d-dimensional random walk based on its Fourier transform. We also present the proof in the case of random walks with values in $\mathbb{R}^d$ to demonstrate the power of the Fourier point of view.

7.3.1 Chung-Fuchs

Theorem 7.5 (*Easy version of Chung–Fuchs*)

The d-dimensional walk (S) is recurrent if and only if we have

$$\lim_{r\uparrow 1} \int_{[-\pi,\pi]^d} \mathrm{d}\xi \, \mathfrak{Re}\left(\frac{1}{1-r\hat{\mu}(\xi)}\right) = \infty.$$

For clarity, we will prove the theorem in the case of one-dimensional random walk, the multi-dimensional case being *mutadis mutandis* the same.

Proof when $d = 1$. We suppose that μ is supported by $\mathbb{Z}$. In this setup, (S) is recurrent if and only if the series $\sum_{n\geqslant 0} \mathbb{P}(S_n = 0)$ diverges. Recall from (7.1) that

$$\mathbb{P}(S_n = 0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{d}t \, \mathbb{E}[\mathrm{e}^{itS_n}] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{d}t \, (\hat{\mu}(t))^n .$$

We are lead to sum the last equality for $n \geqslant 0$, but before that we first multiply by r^n for some $r \in [0, 1)$ in order to be sure that we can exchange series, expectation and integral. One gets

$$\sum_{n\geqslant 0} r^n \mathbb{P}(S_n = 0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{d}t \sum_{n\geqslant 0} r^n \, (\hat{\mu}(t))^n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\mathrm{d}t}{1 - r\hat{\mu}(t)}.$$

Since the left-hand side is real, one can take the real part in the integral. Letting $r \uparrow 1$, the first series diverges if and only if $\sum_{n\geqslant 0} \mathbb{P}(S_n = 0) = \infty$. This completes the proof of the theorem in the lattice case. □

Actually, the Chung-Fuchs criterion is also valid for general step distribution over $\mathbb{R}^d$ (thus outside of the general setup of this chapter). Let us provide the proof so that the reader may compare them:

Proof for non-lattice step distribution ($d = 1$). When the walk in not lattice we may have $\mathbb{P}(S_n = 0) = 0$ for $n \geqslant 1$ but thanks to Definition-Proposition 3.1 one rather needs to express $\mathbb{P}(|S_n| < 1)$ in terms of the Fourier transform. This is done thanks to this lemma:

Lemma 7.6. *Consider the function $f(x) = (1 - |x|)_+$ for $x \in \mathbb{R}$ then we have*

$$\hat{f}(t) = \int_{\mathbb{R}} \mathrm{d}x \, f(x)\mathrm{e}^{itx} = \frac{2}{t^2}(1 - \cos(t)) = \frac{1}{2}\left(\frac{\sin(t/2)}{t/2}\right)^2,$$

$$\hat{\hat{f}}(t) = \int_{\mathbb{R}} \mathrm{d}x \, \frac{2}{x^2}(1 - \cos(x))\mathrm{e}^{itx} = 2\pi f(t).$$

Proof of the lemma. The first display is an easy calculation. The second one can be seen as a particular case of the inversion formula for Fourier transform[2]. $\square$

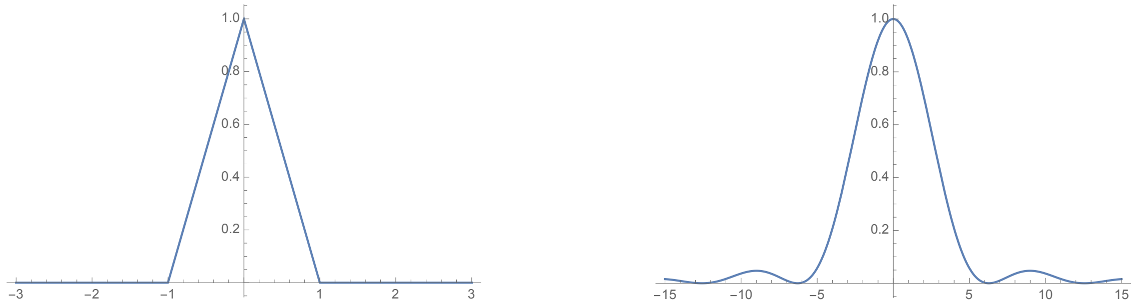


Figure 7.1: The functions f and $\hat{f}$.

Back to the proof of the theorem, we will use these *non-negative* functions as surrogates for the indicator function $\mathbf{1}_{|x|<1}$. More precisely, for some constant $c > 0$ we can write:

$$\begin{aligned}\mathbb{P}(|S_n| < 1) &\leqslant c\mathbb{E}\left[\hat{f}(S_n)\right] \\ &= c\int_{\mathbb{R}} \mathrm{d}t\, \mathbb{E}[\mathrm{e}^{itS_n}] f(t) \\ &= c\int_{\mathbb{R}} \mathrm{d}t\, f(t)\, (\hat{\mu}(t))^n .\end{aligned}$$

We proceed as in the last proof and multiply by r^n before summing and taking the real part to get that

$$\sum_{n\geqslant 0} r^n \mathbb{P}(|S_n| < 1) \leqslant c \int_{\mathbb{R}} \mathrm{d}t\ \mathfrak{Re}\left(\frac{f(t)}{1 - r\hat{\mu}(t)}\right) \leqslant c \int_{-1}^{1} \mathrm{d}t\ \mathfrak{Re}\left(\frac{1}{1 - r\hat{\mu}(t)}\right).$$

If the limit as $r \uparrow 1$ of the integral in the theorem is finite, then so is the limit of the integral of the last display and consequently the limit of the series in the left-hand side is finite. This proves transience of the walk thanks to Definition-Proposition 3.1. For the other direction we use $\hat{f}$ instead of f and write

$$\begin{aligned}\mathbb{P}(|S_n| < 1) &\geqslant \mathbb{E}\left[f(S_n)\right] = \frac{1}{2\pi}\mathbb{E}\left[\hat{\hat{f}}(S_n)\right] \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{d}t\, \mathbb{E}[\mathrm{e}^{itS_n}] \hat{f}(t) \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{d}t\, \hat{f}(t)\, (\hat{\mu}(t))^n ,\end{aligned}$$

[2]Here it is as an exercise: Let $\gamma_\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$ be the standard Gaussian density with variance σ^2. We recall that $\hat{\gamma}_1(t) = \sqrt{2\pi}\gamma_1(t)$. Let $f \in \mathbb{L}^1(\mathbb{R})$ and suppose that $\hat{f} \in \mathbb{L}^1(\mathbb{R})$. By convoluting f with γ_σ as $\sigma \to 0$ show that for almost every x we have

$$2\pi f(x) = \int_{\mathbb{R}} \mathrm{d}t\ \mathrm{e}^{-itx} \hat{f}(t).$$

so that in the end of the day we get for $r \in [0,1)$

$$\sum_{n \geqslant 0} r^n \mathbb{P}(|S_n| < 1) \geqslant \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}t \, \mathfrak{Re} \left(\frac{\hat{f}(t)}{1 - r\hat{\mu}(t)} \right) \geqslant c' \int_{-\pi}^{\pi} \mathrm{d}t \, \mathfrak{Re} \left(\frac{1}{1 - r\hat{\mu}(t)} \right),$$

for some $c' > 0$. It is now easy to see that if the limit as $r \to 1$ of the integral in the statement of the theorem is infinite then so is the series $\sum_{n \geqslant 0} \mathbb{P}(|S_n| < 1)$ implying recurrence of the walk. This completes the proof. □

In fact, there is a stronger version of Theorem 7.5 which is obtained by formally interchanging the limit and the integral in the last theorem: the random walk (S) is transient or recurrent according as to whether the real part of $(1 - \hat{\mu}(t))^{-1}$ is integrable or not near 0 (we do not give the proof). Notice that in the case when the law μ is symmetric (i.e. $X \sim -X$ when $X \sim \mu$) then $\hat{\mu}$ is real valued and the monotone convergence theorem shows that the recurrence criterion of Theorem 7.5 indeed reduces to

$$\int_{0 \in \mathbb{R}^d} \frac{\mathrm{d}\xi}{1 - \hat{\mu}(\xi)} = \infty.$$

Exercise 7.3. Let $(S_n)_{n \geqslant 0}$ be the simple random walk on $\mathbb{Z}^3$, i.e. the increments are uniform among the 6 directions $(\pm 1, 0, 0), (0, \pm 1, 0)$ and $(0, 0, \pm 1)$. Show that the probability p that S never returns to the origin is equal to $\frac{1}{m}$ where

$$m = \frac{3}{(2\pi)^3} \int_{[0,2\pi]^3} \mathrm{d}\xi_1 \mathrm{d}\xi_2 \mathrm{d}\xi_3 \, \frac{1}{3 - (\cos(\xi_1) + \cos(\xi_2) + \cos(\xi_3))}.$$

In particular $p \approx 0.659463\ldots$.

Let us finish this section with two exercises using Fourier techniques to compute probabilities:

Exercise 7.4. If $(S_n)_{n \geqslant 0}$ is a one-dimensional random walk, its symmetrized version is the random walk $(\tilde{S}_n)_{n \geqslant 0} = (S_n - S'_n)_{n \geqslant 0}$ where (S) and (S') are independent copies of the walk (S). Using Exercise 3.1 show that if (S) is recurrent then so is its symmetrized version $(\tilde{S})$ (is that an equivalence?).

7.3.2 Applications

Let us now give a few applications of the Chung-Fuchs theorem 7.5.

Recurrence and transience in various dimensions

Corollary 7.7. *An aperiodic random walk on $\mathbb{Z}^d$ is*

(i) recurrent if $d = 1$ and μ has finite first moment and is centered,

(ii) recurrent if $d = 2$ and μ is centered with finite variance,

(iii) always transient if $d \geqslant 3$.

Proof. We have already given a proof of point (i) in Chapter 3, and a proof of point (iii) in the last section. But let us give other proofs based on the Fourier criterion which generalize easily to random walks on $\mathbb{R}^d$. In this case since μ is centered, it is classical that we have

$$\hat{\mu}(t) = 1 + o(t) \qquad \text{as } t \to 0. \tag{7.3}$$

Writing $\hat{\mu}(t) = a(t) + ib(t)$ we have

$$\mathfrak{Re}\left(\frac{1}{1 - r\hat{\mu}(t)}\right) = \frac{1 - r\,a(t)}{(1 - r\,a(t))^2 + (r\,b(t))^2} \geqslant \frac{1 - r}{(1 - r\,a(t))^2 + (r\,b(t))^2}.$$

Fix $\varepsilon > 0$, using (7.3) we can find $\delta > 0$ such that for $t \in [0, \delta]$ we have $|b(t)| \leqslant \varepsilon t$ and $|1 - a(t)| \leqslant \varepsilon t$. Hence for these values of t we have $(r\,b(t))^2 \leqslant \varepsilon^2 r^2 t^2$ and $(1 - r\,a(t))^2 = ((1 - r) + r(1 - a(t)))^2 \leqslant 2(1 - r)^2 + 2\varepsilon^2 r^2 t^2$ (we used $(x + y)^2 \leqslant 2x^2 + 2y^2$). Finally, using the positivity of the integrand we deduce that

$$\int_{-\pi}^{\pi} \frac{\mathrm{d}t}{1 - r\hat{\mu}(t)} \geqslant \int_0^{\delta} \mathrm{d}t\, \frac{1 - r}{2(1 - r)^2 + 3t^2\varepsilon^2 r^2} = \int_0^{\delta} \frac{\mathrm{d}t}{1 - r}\, \frac{1}{2 + 3\left(\frac{t\varepsilon r}{1 - r}\right)^2}.$$

Performing the change of variable $y = \frac{t\varepsilon r}{1-r}$ the previous integral is easily computed and seen to be of order ε^{-1} as $r \uparrow 1$. Since $\varepsilon > 0$ was arbitrary, the initial integral diverges as desired.

The proof of point (ii) is similar. First notice that $\mathfrak{Re}\left(\frac{1}{1 - r\hat{\mu}(\xi)}\right)$ is always positive so that Fatou's lemma implies that

$$\int_{[-\pi,\pi]^d} \mathrm{d}\xi\, \mathfrak{Re}\left(\frac{1}{1 - \hat{\mu}(\xi)}\right) = \int_{[-\pi,\pi]^d} \mathrm{d}\xi \liminf_{r \uparrow 1} \mathfrak{Re}\left(\frac{1}{1 - r\hat{\mu}(\xi)}\right) \leqslant \liminf_{r \uparrow 1} \int_{[-\pi,\pi]^d} \mathrm{d}\xi\, \mathfrak{Re}\left(\frac{1}{1 - r\hat{\mu}(\xi)}\right).$$

But because of (7.2) the integrand in the left-hand size is equivalent to $({}^t\xi Q \xi/2)^{-1}$ which after an orthogonal change of variable is just equal to $\|\xi\|_2^2$. Using the polar coordinates we have

$$\int_{\text{around } 0 \in \mathbb{R}^2} \frac{\mathrm{d}\xi}{\|\xi\|_2^2} = \int_{0+} \frac{2\pi r \mathrm{d}r}{r^2} = \infty,$$

and so the limit of the integral in the Chung–Fuchs criterion is indeed infinite as desired.
For point (iii) we can write for $0 < r < 1$ and ξ close to 0:

$$\begin{aligned}
\mathfrak{Re}\left(\frac{1}{1 - r\hat{\mu}(\xi)}\right) &= \frac{\mathfrak{Re}(1 - r\hat{\mu}(\xi))}{(\mathfrak{Re}(1 - r\hat{\mu}(\xi)))^2 + (\mathfrak{Im}(1 - r\hat{\mu}(\xi)))^2} \\
&\leqslant \frac{1}{\mathfrak{Re}(1 - r\hat{\mu}(\xi))} \\
&\leqslant \frac{1}{\mathfrak{Re}(1 - \hat{\mu}(\xi))} \\
&\leqslant \frac{1}{\lambda|\xi|^2},
\end{aligned}$$

where we used Lemma 7.2 for the last inequality and the fact that $\mathfrak{Re}(\hat{\mu}(\xi)) > 0$ when ξ is close to 0 in the third line. It remains to notice using polar coordinates that $|x|^{-2}$ is always integrable in the neighborhood of 0 in $\mathbb{R}^d$ for $d \geqslant 3$. $\square$

Heavy-tailed random walks. The Chung–Fuchs criterion can also be used to give an alternative proof of the transience of the heavy-tailed random walks considered in Theorem 3.7 and is central in the proof of Theorem 3.8.

Exercise 7.5. Reprove Theorem 3.7 using the Chung–Fuchs criterion.

7.4 The local central limit theorem

The central limit theorem is one of the most important theorems in probability theory and says in our context that the rescaled random walk $S_n/\sqrt{n}$ converges in distribution towards a normal law provided that μ is centered and has finite variance. There are many proofs of this result, the most standard being through the use of Fourier transform and Lévy's criterion for convergence in law[3]. We will see below that the central limit theorem can be "desintegrated" to get a more powerful "local" version of it. The proof is again based on (7.1).

7.4.1 Local CLT

When a one-dimensional random walk with mean m and variance σ^2 satisfies a central limit theorem we mean that for any $a < b$ we have

$$\mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}} \in [a,b]\right) \xrightarrow[n\to\infty]{} \int_a^b \frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-x^2/(2\sigma^2)}.$$

We say that we have a local central limit theorem if we can diminish the interval $[a,b]$ as a function of n until it contains just one point of the lattice, that is if for $x \in \mathbb{Z}$ we have

$$\mathbb{P}(S_n = x) = \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}} \in \left[\frac{x-nm}{\sqrt{n}}, \frac{x-nm+1}{\sqrt{n}}\right)\right) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-nm)^2}{2n\sigma^2}} \frac{1}{\sqrt{n}}.$$

It turns out that the necessary conditions for the central limit theorem are already sufficient to get the local central limit theorem:

Theorem 7.8 (*Local central limit theorem, Gnedenko*)

Let μ be a distribution supported on $\mathbb{Z}$, aperiodic, with mean $m \in \mathbb{R}$ and with a finite variance $\sigma^2 > 0$. Then if we denote by $\gamma_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-x^2/(2\sigma^2)}$ the density distribution of the centered normal law of variance σ^2 then we have

$$\lim_{n\to\infty} \sup_{x\in\mathbb{Z}} n^{1/2} \left| \mathbb{P}(S_n = x) - n^{-1/2}\gamma_\sigma\left(\frac{x-nm}{\sqrt{n}}\right)\right| = 0.$$

The usual central limit theorem follows from its local version. Indeed, if we consider the random variable $\tilde{S}_n = S_n + n^{-1/2}U_n$ where U_n is uniform over $[0,1]$ and independent of S_n. Then the

[3] Here are a couple of other proofs: Lindeberg swapping trick, method of moments, Stein method, contraction method and Zolotarev metric. See the beautiful page by Terence Tao on this subject: https://terrytao.wordpress.com/2010/01/05/254a-notes-2-the-central-limit-theorem/

local central limit theorem shows that the law of $(\tilde{S}_n - nm)/\sqrt{n}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ whose density f_n converges pointwise towards the density of γ_σ. Scheffe's lemma (see Exercise 4.12) then implies that $(\tilde{S}_n - nm)/\sqrt{n}$ converges in law towards $\gamma_\sigma(\mathrm{d}x)$ and similarly after removing the tilde.

Proof. The starting point is again Cauchy formula's relating probabilities to Fourier transform:

$$\mathbb{P}(S_n = x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}t\, \mathrm{e}^{-ixt}\mathbb{E}[\mathrm{e}^{iS_n t}] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}t\, \mathrm{e}^{-ixt}(\hat{\mu}(t))^n.$$

Since $|\hat{\mu}(t)| < 1$ when $t \neq 0$ the main contribution of the integral comes from the integration near 0, it is then an application of Laplace's method. Since we want to use the series expansion of the Fourier transform near 0 it is natural to introduce ν the image measure of μ after translation of $-m$ so that ν is centered and has finite variance: we can write $\hat{\nu}(t) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2)$ for t small. The last display then becomes

$$\begin{aligned}
\mathbb{P}(S_n = x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}t\, \mathrm{e}^{-ixt}\mathrm{e}^{inm}(\hat{\nu}(t))^n \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}t\, \mathrm{e}^{-ixt}\mathrm{e}^{inmt}(1 - \frac{\sigma^2}{2}t^2 + o(t^2))^n \\
&= \frac{1}{\sqrt{n}}\frac{1}{2\pi}\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \mathrm{d}u\, \mathrm{e}^{-iux/\sqrt{n}}\mathrm{e}^{i\sqrt{n}mu}\underbrace{(1 - \frac{\sigma^2}{2n}u^2 + o(u^2/n))^n}_{\approx\gamma_{1/\sigma}(u)}.
\end{aligned}$$

We can then approximate the last integral by

$$\frac{1}{\sqrt{n}}\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}u\, \mathrm{e}^{-iux/\sqrt{n}}\mathrm{e}^{i\sqrt{n}mu}\gamma_{1/\sigma}(u) = \frac{1}{\sqrt{2\pi n}}\mathbb{E}\left[\exp\left(i\left(\sqrt{n}m - \frac{x}{\sqrt{n}}\right)\frac{\mathcal{N}}{\sigma}\right)\right],$$

where $\mathcal{N}$ denote a standard normal variable. Using the identity $\mathbb{E}[\mathrm{e}^{it\mathcal{N}}] = \mathrm{e}^{-t^2/2}$ the last display is indeed equal to $\gamma_\sigma\left(\frac{x-nm}{\sqrt{n}}\right)/\sqrt{n}$ as desired. It remains to quantify the last approximation. The error made in the approximation is clearly bounded above by the sum of the two terms:

$$A = \frac{1}{\sqrt{n}}\frac{1}{2\pi}\int_{|u|>\pi\sqrt{n}} \mathrm{d}u\, \gamma_{1/\sigma}(u),$$

$$B = \frac{1}{\sqrt{n}}\frac{1}{2\pi}\int_{|u|<\pi\sqrt{n}} \mathrm{d}u\, \left|\gamma_{1/\sigma}(u) - \left(\hat{\nu}(\frac{u}{\sqrt{n}})\right)^n\right|.$$

The first term A causes no problem since it is exponentially small (of the order of e^{-n}) hence negligible in front of $1/\sqrt{n}$. The second term may be further bounded above by the sum of three terms

$$B \leqslant \frac{1}{\sqrt{n}}\int_{|u|<n^{1/4}} \mathrm{d}u\, \left|\gamma_{1/\sigma}(u) - \left(\hat{\nu}(\frac{u}{\sqrt{n}})\right)^n\right| + \int_{n^{1/4}<|u|<\pi\sqrt{n}} \mathrm{d}u\, \gamma_{1/\sigma}(u) + \int_{n^{1/4}<|u|<\pi\sqrt{n}} \mathrm{d}u\, \left|\hat{\nu}(\frac{u}{\sqrt{n}})\right|^n.$$

The first of this term is shown to be $o(n^{-1/2})$ using dominated convergence: in the region considered for u, the integrand converges pointwise to 0; for the domination we may use the fact for $|u| < \varepsilon\sqrt{n}$ we have by the expansion of $\hat{\nu}$ that $\left|\hat{\nu}(\frac{u}{\sqrt{n}})\right|^n \leqslant (1 - \frac{\sigma^2 u^2}{4n})^n \leqslant \mathrm{e}^{-\sigma^2 u^2/4}$. The second term of the sum is handle as above and seen to be of order $\mathrm{e}^{-\sqrt{n}}$. For the third term, we bound the integrand by $\left|\hat{\nu}(\frac{u}{\sqrt{n}})\right|^n \leqslant \mathrm{e}^{-\sigma^2 u^2/4}$ for $|u| < \varepsilon\sqrt{n}$, as for $\varepsilon\sqrt{n} < |u| < \pi\sqrt{n}$ we use Lemma 7.1 to bound the integrand by some $c < 1$. The sum of the three terms is then of negligible order compared to $n^{-1/2}$ as desired. □

The result extends to higher dimension (and to the case of random walks converging towards stable Lévy process) with *mutatis mutandis* the same proof. We only give the multidimensional statement in the finite variance case:

Theorem 7.9 (*Local central limit theorem, Gnedenko*)

Let μ be distribution supported on $\mathbb{Z}^d$, aperiodic, of mean $m \in \mathbb{R}^d$ and with covariance matrix Q. If we write

$$\gamma_Q(x) = \frac{1}{\sqrt{2\pi}^d \sqrt{\det(Q)}} \exp\left(-\frac{{}^t x Q^{-1} x}{2}\right)$$

for the density of the centered Gaussian distribution with covariance Q on $\mathbb{R}^d$ then we have

$$\lim_{n\to\infty} \sup_{x \in \mathbb{Z}^d} n^{d/2} \left| \mathbb{P}(S_n = x) - n^{-d/2} \gamma_Q\left(\frac{x - nm}{\sqrt{n}}\right) \right| = 0.$$

A closer look at the proof of the local central limit theorem shows that the main term which causes the the largest discrepancy between $\mathbb{P}(S_n = x)$ and its ideal version $n^{-d/2}\gamma_Q\left(\frac{x-nm}{\sqrt{n}}\right)$ is caused by the term "B" in the last proof (and we can even replace $n^{1/4}$ by n^{ε} with no harm). If we assume further regularity on μ we can have a better control on this term.

7.4.2 Applications

As an application of the local central limit theorem we can prove straight away that an aperiodic random walk on $\mathbb{Z}^2$ with zero mean and finite variance is recurrent since the return probability $\mathbb{P}(S_n = (0,0))$ decays as $1/n$ and so its series diverges. We can also use the local central limit theorem to get a precise asymptotic for the tail of the size of Galton–Watson trees in the critical case:

Proposition 7.10. *Let μ be an offspring distribution law which is aperiodic, of mean 1 and with a finite variance $\sigma^2 \in (0, \infty)$. If $\mathcal{T}$ is a μ-Galton–Watson tree then we have*

$$\mathbb{P}(|\mathcal{T}| = n) \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \frac{1}{n^{3/2}}.$$

Proof. Taking the same notation as in Proposition 9.1 we have

$$\begin{aligned}\mathbb{P}(|\mathcal{T}| = n) &\underset{\text{Prop.9.1}}{=} \mathbb{P}(\theta_{-1} = n)\\ &\underset{\text{Prop.4.9}}{=} \frac{1}{n}\mathbb{P}(S_n = -1)\\ &\underset{\text{Thm.7.8}}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \frac{1}{n^{3/2}} + o(n^{-3/2}).\end{aligned}$$

□

Exercise 7.6 (One-dimensional of the Brownian bridge). Suppose S is a one-dimensional centered random walk with finite non-zero variance σ^2. Prove that for $x \in (0,1)$ that $S_{[nx]}/\sqrt{n}$, conditioned on $S_n = 0$ converges in law towards $\mathcal{N}(0, \sigma^2 x(1-x))$.

Bibliographical notes. The Fourier transform is a remarkable tool (whose efficiency is sometimes a bit mysterious) to study random walk with independent increments. The results of this chapter are mostly based on [39, Chapter II]. Exercise 7.2 is taken from [**?**]. The local central limit theorem is valid is the much broader context of random walks converging towards stable Lévy processes, see Gnedenko's local limit theorem in [**?**, Theorem 4.2.1], and can be sharpened when we have further moment assumptions, see [27]. There are other proofs of the local central limit theorem for example using the "Bernoulli part" of a random variable (see [**?**]) which are easier to adapt in a context with less independance.

Chapter VIII: Ranges and Intersections of random walks

As before we suppose that we are given an aperiodic random walk (S) on $\mathbb{Z}^d$.

8.1 Range and recurrence

Definition 8.1. *The* range *of the random walk (S) up to time $n \geqslant 0$ is*

$$R_n = \#\{S_0, S_1, \ldots, S_n\}.$$

8.1.1 The Kesten–Spitzer–Whitman theorem

Theorem 8.1 (*Kesten–Spitzer–Whitman*)

We have the following convergence in probability

$$\frac{R_n}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} \mathbb{P}(S \text{ does not come back to its starting point}).$$

In particular the range grows linearly (in probability) if and only if (S) is transient.

Proof. We write $c = \mathbb{P}(S_i \neq 0_{\mathbb{Z}^d} : \forall i \geqslant 1)$ to simplify notation. Let us first compute the expectation of the range. To do this we count each *last* visit to a given vertex between time 0 and time n:

$$\begin{aligned}
\mathbb{E}[R_n] &= \mathbb{E}\left[\sum_{i=0}^{n} \mathbf{1}_{S_j \neq S_i : \forall i < j \leqslant n}\right] \\
&= \sum_{i=0}^{n} \mathbb{P}(S_j \neq S_0 : \forall 0 < j \leqslant n - i) \\
&\sim n \cdot \mathbb{P}(S_i \neq S_0 : \forall i \geqslant 1) = c \cdot n,
\end{aligned}$$

by Cesàro[1] summation. If $c = 0$, i.e. when the walk is recurrent, we directly deduce that $R_n/n \to 0$ in probability by Markov's inequality. Let us suppose now that $c > 0$ (i.e. that the walk is transient) and let us estimate the variance of the range. To simplify notation we write $\gamma_{i,n} = \mathbf{1}_{S_j \neq S_i : \forall i < j \leqslant n}$ so that we can write

$$\mathbb{E}\left[(R_n)^2\right] = \sum_{0 \leqslant i,j \leqslant n} \mathbb{E}[\gamma_{i,n}\gamma_{j,n}] = 2 \sum_{0 \leqslant i < j \leqslant n} \mathbb{E}[\gamma_{i,n}\gamma_{j,n}] + O(n).$$

[1]  Ernesto Cesàro (1859-1906)

Furthermore,

$$\begin{aligned}
\mathbb{E}[\gamma_{i,n}\gamma_{j,n}] &= \mathbb{P}\left(S_k \neq S_0, \forall 0 < k \leqslant n-i \text{ and } S_k \neq S_{j-i}, \forall j-i < k \leqslant n-i\right) \\
&\leqslant \mathbb{P}\left(S_k \neq S_0, \forall 0 < k \leqslant j-i \text{ and } S_k \neq S_{j-i}, \forall j-i < k \leqslant n-i\right) \\
&\underset{\text{Markov}}{=} \mathbb{P}\left(S_k \neq S_0, \forall 0 < k \leqslant j-i\right)\mathbb{P}\left(S_k \neq S_0, \forall 0 < k \leqslant n-j\right) \qquad (8.1) \\
&\to c^2,
\end{aligned}$$

when both $j-i$ and $n-j$ tend to ∞. By Cesàro summation again we deduce that $\mathbb{E}[(R_n)^2]$ is asymptotically smaller than $(cn)^2$ and since it is anyway larger than $\mathbb{E}[R_n]^2 \sim (cn)^2$ (by Cauchy–Schwarz) we deduce that $\mathrm{Var}(R_n) = \mathbb{E}[(R_n)^2] - \mathbb{E}[R_n]^2 = o(n^2)$. The desired convergence in probability then follows from a classical application of Markov's inequality: for all $\varepsilon > 0$ we have

$$\mathbb{P}\left(\left|\frac{R_n}{n} - c\right| \geqslant \varepsilon\right) \leqslant \frac{\mathrm{Var}(R_n)}{\varepsilon^2 n^2} \xrightarrow[n\to\infty]{} 0.$$

□

Remark 8.1. There exist a sharper version of the last theorem where the convergence is improved to an almost sure convergence. One way to prove it is to use ergodic theory and more precisely Kingman's subadditive ergodic theorem (which is an useful extension of the well-known theorem of Birkhoff). The key observation is to notice that the range is subadditive in the sense that

$$\#\{S_0, \dots, S_{n+m}\} \leqslant \#\{S_0, \dots, S_n\} + \#\{S_n, \dots, S_{n+m}\}$$

and that $\{S_n, \dots, S_{n+m}\}$ is exactly the function R_m applied on the walk after shifting by the first n steps.

8.1.2 Back on recurrence in $d = 1$ and $d = 2$

We can use Theorem 8.1 in order to give new proofs (or to re-interpret those already given) of the recurrence of centered random walk in dimension 1 and centered random walk on $\mathbb{Z}^2$ with finite variance. Indeed, in the first case we have by the strong law of large numbers (or more precisely its functional version) that

$$\left(\frac{S_{[nt]}}{n}\right)_{t\geqslant 0} \xrightarrow[n\to\infty]{p.s.} 0,$$

for the topology of uniform convergence on every compact of $\mathbb{R}_+$. It follows that the range of the random walk is sublinear and hence the walk must be recurrent by Theorem 8.1. The second case is a bit more subtle. Suppose $(S_n)_{n\geqslant 0}$ is a centered random walk on $\mathbb{Z}^2$ with finite covariance matrix Σ^2. By Donsker's invariance principle we have

$$\left(\frac{S_{[nt]}}{\sqrt{n}}\right)_{t\in[0,1]} \xrightarrow[n\to\infty]{(d)} (B_t^{(2)})_{t\in[0,1]}$$

where $B^{(2)}$ is a two-dimensional Brownian motion with covariance matrix Σ^2. Since the range R_n during the first n steps is at most the number of points in the smallest disk centered at the origin containing all $S_0, \dots, S_n$ we deduce that

$$R_n \leqslant \pi \cdot (\max\{\|S_k\|_2 : 0 \leqslant k \leqslant n\} + 2)^2.$$

But the convergence towards $B^{(2)}$ implies that $\max\{\|S_k\|_2 : 0 \leqslant k \leqslant n\}/\sqrt{n}$ converges in distribution towards χ, the maximal L^2 norm of $B^{(2)}$ over the time interval $[0,1]$. Whatever this random variable is, it is easy to see that for every $\varepsilon > 0$ we have $\mathbb{P}(\chi < \varepsilon) > 0$. By Portmanteau theorem we deduce that

$$\liminf_{n\to\infty} \mathbb{P}\left(\frac{R_n}{n} < \pi\varepsilon^2\right) \geqslant \liminf_{n\to\infty} \mathbb{P}\left(\frac{\max\{\|S_k\|_2 : 0 \leqslant k \leqslant n\}}{\sqrt{n}} < \varepsilon\right) > \mathbb{P}(\chi < \varepsilon) > 0.$$

The only way the last inequality is compatible with Theorem 8.1 is if $\mathbb{P}(S_0 \neq S_i : \forall i \geqslant 1) = 0$ i.e. if the walk is recurrent as desired.

8.1.3 The critical case $d = 2$

Let us study in more details the range in the case of a simple symmetric random walk $S^{(2)}$ on $\mathbb{Z}^2$, i.e. when the step distribution μ is the uniform measure over the 4 basis vectors $\{(\pm 1, 0), (0, \pm 1)\}$.

Remark 8.2 (The rotation trick when $d = 2$). Let $(\mathbb{S}^{(2)}_n)_{n\geqslant 0}$ be the random walk on $\mathbb{Z}^2$ such that its coordinates are independent and distributed as simple symmetric random walks on $\mathbb{Z}$. Then this walk is not aperiodic since it lives on the lattice $\mathbb{L}^2 = \{(x_1, \dots, x_d) \in \mathbb{Z}^d : x_1 + \dots + x_d \equiv 0[2]\}$. But this lattice is easily seen to be equivalent to the initial $\mathbb{Z}^2$ after a dilation by $1/\sqrt{2}$ and a $\pi/4$-rotation. We deduce that

$$\begin{array}{ccc} \underbrace{\mathbb{P}(S^{(2)}_{2n} = (0,0))} & = & \underbrace{\mathbb{P}(\mathbb{S}^{(2)}_{2n} = (0,0))} \\ \| & & \| \\ \overbrace{4^{-2n}\sum_{k=0}^{n}\binom{2n}{2k}\binom{2k}{k}\binom{2(n-k)}{n-k}} & = & \overbrace{4^{-2n}\binom{2n}{n}\binom{2n}{n}} \end{array},$$

because for the random walk (S) to come back to the origin after $2n$ steps, it must operate exactly $2k$ steps along the x-direction among which k are in the positive direction and similarly $2(n-k)$ steps along the y-direction among which $n-k$ in the positive direction, for some $k \in \{0, 1, 2, \dots, n\}$.

Exercise 8.1. Give a direct proof of the equality $\sum_{k=0}^{n}\binom{2n}{2k}\binom{2k}{k}\binom{2(n-k)}{n-k} = \binom{2n}{n}\binom{2n}{n}$.

Theorem 8.2 (*Dvoretzky & Erdös*)

If $(S_n)_{n\geqslant 0}$ is the simple symmetric random walk on $\mathbb{Z}^2$ we have

$$\frac{\log n}{n} \cdot R_n \xrightarrow[n\to\infty]{(\mathbb{P})} \pi.$$

Proof. The proof is similar to that of Theorem 8.1 and goes through first and second moments calculation done with more care. First when estimating $\mathbb{E}[R_n]$, the expectation of the range, we need to compute

$$\mathbb{P}(S_j \neq S_0 : \forall 0 < j \leqslant n)$$

hence the probability that the two-dimensional simple random walk does not come back to its starting point in the first n steps. This probability tends to 0 but how slow? To simplify the exposition we write τ for the first return time to the origin by the simple symmetric two-dimensional random walk.

Lemma 8.3. *We have* $\mathbb{P}(\tau > n) \sim \frac{\pi}{\log n}$ *as* $n \to \infty$

Proof of the lemma. By the calculation done in Remark 8.2 we have

$$P(S_{2n} = (0,0)) = 4^{-2n}\binom{2n}{n}\binom{2n}{n} \underset{n\to\infty}{\sim} \frac{1}{\pi n}.$$

The asymptotic can also directly be derived from the local central limit theorem (after a reduction to the aperiodic case, exercise!). Hence its generating series satisfies

$$G(x) = \sum_{n\geqslant 0} x^n P(S_{2n} = (0,0)) \sim -\frac{1}{\pi}\log(1-x) \text{ as } x \to 1^-.$$

Easy calculus exercises with formal series show that if we introduce the two other generating series

$$F(x) = \sum_{n\geqslant 1} x^n \mathbb{P}(\tau = 2n) \quad \text{and} \quad H(x) = \sum_{n\geqslant 0} \mathbb{P}(\tau > 2n)x^n$$

then we have the relations

$$H(x)(1-x) = (1-F(x)) \quad \text{and} \quad G(x) = 1 + F(x) + F(x)^2 + \cdots = \frac{1}{1-F(x)}.$$

Combining the last displays we deduce that

$$H(x) \sim \frac{-\pi}{(1-x)\log(1-x)} \quad \text{as } x \to 1^-.$$

Since the coefficient of H are all positive and non-increasing, we are in the setup to apply Tauberian theorems to transfer estimates on H into estimates on its coefficients, specifically, applying Theorem VI.13 in [18] (more precisely its easy extension in the case of non-increasing coefficients) we deduce the desired estimate

$$\mathbb{P}(\tau > n) \sim \frac{\pi}{\log(n)}, \quad \text{as} \quad n \to \infty.$$

$\square$

Coming back to the proof of the theorem we deduce by adapting (8.1) that

$$\begin{aligned}\mathbb{E}[R_n] &= \sum_{i=0}^{n} \mathbb{P}(\tau > i) \\ &\sim \sum_{i=2}^{n} \frac{\pi}{\log(i)} = \pi\frac{n}{\log n}.\end{aligned}$$

Hence it remains to estimate the variance of the range and show that $\mathbb{E}[(R_n)^2] \sim \mathbb{E}[R_n]^2$ to conclude similarly as in the case $d \geqslant 3$. Since we always have $\mathbb{E}[(R_n)^2] \geqslant \mathbb{E}[R_n]^2$ we focus on the other inequality. We again proceed as above and write

$$\begin{aligned}\mathbb{E}\left[(R_n)^2\right] = \sum_{0\leqslant i,j\leqslant n} \mathbb{E}[\gamma_{i,n}\gamma_{j,n}] &= 2\sum_{0\leqslant i<j\leqslant n} \mathbb{E}[\gamma_{i,n}\gamma_{j,n}] + O(n)\\ &\leqslant 2\sum_{0\leqslant i<j\leqslant n} \mathbb{P}(\tau > j-i)\mathbb{P}(\tau > n-j) + O(n).\end{aligned}$$

We now exclude the terms i, j such that either $|j-i| < n^{1-\varepsilon}$ or $|n-j| \leqslant n^{1-\varepsilon}$. They are at most $n^{2-\varepsilon/2}$ such couples (for large n's) and so using Lemma 8.3 we get that

$$\mathbb{E}\left[(R_n)^2\right] \leqslant O(n^{2-\varepsilon/2}) + n^2\left(\frac{\pi}{(1-\varepsilon)\log(n)}\right)^2.$$

Hence we deduce that asymptotically we have $\mathbb{E}[(R_n)^2]/\mathbb{E}[R_n]^2 \leqslant 1$ and this suffices to complete the proof. □

Remark 8.3. The convergence in probability in the last theorem can also be improved into an almost sure convergence.

We see that dimension 2 is *critical* for simple random walks in the sense that the behavior of the range is in-between being linear (as in the transient case when $d \geqslant 3$) and really sublinear as in the case $d = 1$ (with finite variance say). This is usually referred to by saying that "$d = 2$ is the critical dimension for recurrence of the simple random walk". In other words, although the walk in recurrent in dimension 2 it is only barely recurrent.

8.2 Intersection of random walks

In this section we consider that the step distribution μ is aperiodic on $\mathbb{Z}^d$, is centered, and has bounded support. We then consider independent random walks trace $S^{(d)}(i) = (S_n^{(d)}(i))_{n\geqslant 0}$ on $\mathbb{Z}^d$ with step distribution μ which we see as random subsets of $\mathbb{Z}^d$ and ask whether these subsets intersect (i.e. whether the random walk paths intersect). For $k \geqslant 1$ and $d \geqslant 1$ we write

$$\mathcal{I}^{(d)}(k) = \sum_{x\in\mathbb{Z}^d}\prod_{i=1}^{k} \mathbf{1}_{x\in S^{(d)}(i)},$$

for the number of points in the intersection of the ranges. Notice that the event $\{\mathcal{I}^{(d)}(k) = \infty\}$ is invariant by any finite permutation of the increments of each of these walks, hence by Theorem 3.3 this event is independent of each σ-field generated by a fixed walk. Consequently, this event is independent of the total σ-field generated by the k walks and is thus of probability 0 or 1. The main result is then:

Theorem 8.4 (*Erdös & Taylor*)

We have the following scenario according to the dimension of the ambient space:

(i) *If $d \geqslant 5$ we have $\mathcal{I}^{(d)}(2) < \infty$ almost surely,*

(ii) *If $d = 4$ we have $\mathcal{I}^{(4)}(2) = \infty$ but $\mathcal{I}^{(4)}(3) < \infty$ almost surely,*

(iii) *If $d = 3$ we have $\mathcal{I}^{(3)}(3) = \infty$ but $\mathcal{I}^{(3)}(4) < \infty$ almost surely,*

(iv) *If $d \leqslant 2$ for any $i \geqslant 2$ we have $\mathcal{I}^{(d)}(i) = \infty$ almost surely.*

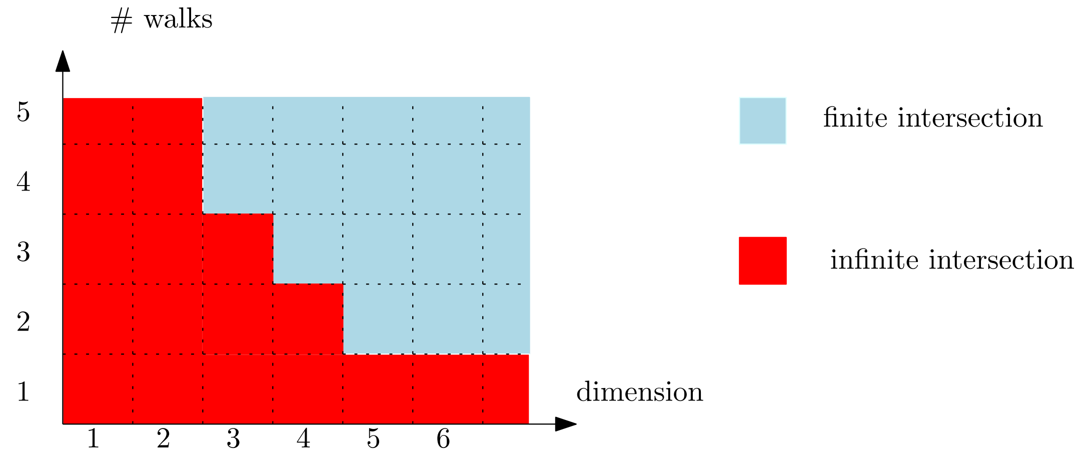


Figure 8.1: Illustration of the theorem.

8.2.1 Estimate on Green's function

For $x \in \mathbb{Z}^d$ we denote by $q(x)$ the probability that the random walk S ever visits x and $G(x)$ the expected number of such visits. Clearly we have $q(x) \leqslant G(x)$ but in dimension $d \geqslant 3$ by transience of the walk and the Markov property we have

$$G(x) \leqslant q(x)G(0),$$

where $G(0) < \infty$. Hence $q(x)$ and $G(x)$ are comparable up to multiplicative constants.

Lemma 8.5 (Green's function)**.** *Let $x \in \mathbb{Z}^d$ for $d \geqslant 3$ and denote by $G(x)$ the expected number of visits to x by $S^{(d)}$. Then there exist two constants $0 < c_1 < c_2 < \infty$ such that for all $x \in \mathbb{Z}^d$ we have*

$$c_1|x|^{2-d} \leqslant G(x) \leqslant c_2|x|^{2-d}.$$

Sketch of proof. Our main tool in this proof will be the local limit theorem (notice that since

μ has finite support it automatically has finite variance). First write

$$\begin{aligned} G(x) &= \sum_{n=0}^{\infty} \mathbb{P}(S_n^{(d)} = x) \\ &= \sum_{n=0}^{|x|^2} \mathbb{P}(S_n^{(d)} = x) + \sum_{n>|x|^2}^{\infty} f\left(\frac{|x^2|}{n}\right) n^{-d/2} + o(n^{-d/2}), \end{aligned}$$

where $f : \mathbb{R}^+ \to \mathbb{R}_+$ is bounded and roughly of the form $f(x) \approx \mathrm{e}^{-x}$ by the local limit theorem. This already suffices for the lower bound: Considering the sum $\sum_{|x|^2<n<2|x|^2} f(\frac{|x^2|}{n})n^{-d/2} + o(n^{-d/2})$ yields to roughly $|x|^2$ values of order $n^{-d/2} \approx |x|^{-d}$ hence a total larger than $c_1|x|^{2-d}$. For the upper bound, we need to combine an improved version of the local limit theorem (notice that since μ has bounded support, it has all its moments) together with large deviation estimates. We do not give the details and refer to [27, Theorem 4.3.4] for details. □

8.2.2 Proof of the Erdös–Taylor theorem

Proof of the theorem. There is no problem for dimension $d \leqslant 2$ since the walks are recurrent and so each $S^{(d)}(i)$ is the whole $\mathbb{Z}^d$. The next easy case is dimension $d \geqslant 5$ for two walks, $d \geqslant 4$ for three walks and dimension 3 for four walks, since a first moment calculation does the job. Indeed we have

$$\begin{aligned} \mathbb{E}[\mathcal{I}^{(d)}(k)] &= \sum_{x\in\mathbb{Z}^d} (q(x))^k \approx \sum_{x\in\mathbb{Z}^d} (G(x))^k \\ &\approx \sum_{x\in\mathbb{Z}^d} (|x|^{2-d})^k \\ &\approx \sum_{n\geqslant 1} n^{k(2-d)} n^{d-1}, \end{aligned}$$

and the series is finite if $k = 2$ and $d \geqslant 5$, if $k = 3$ and $d \geqslant 4$ or if $k = 4$ and $d \geqslant 3$. It is just barely diverging for $(k,d) \in \{(2,4),(3,3)\}$. But to prove points (ii) and (iii) the fact that $\mathbb{E}[\mathcal{I}^{(d)}(k)] = \infty$ does not imply that $\mathcal{I}^{(d)}(k) = \infty$, if we want to provide a *lower bound* on a non-negative random variable (saying that it must stay close to its expectation) one needs to use the second moment method (Definition 11.3). Here, we use the second moment method, more precisely the fact that if $X > 0$ be a non-negative random variable and if $C > 0$ is such that $\mathbb{E}[X]^2 \leqslant \mathbb{E}[X^2] \leqslant C\mathbb{E}[X]^2$ then[2]

$$\mathbb{P}\left(X > \frac{\mathbb{E}[X]}{2}\right) > \frac{1}{4C}$$

We use this method with

$$\mathcal{I}_n^{(d)}(k) = \sum_{x\in\mathbb{Z}^d} \prod_{j=1}^{k} \mathbf{1}_{S_i^{(d)}(j)=x, \text{ for some } i\leqslant n}.$$

[2]this is just a simple Cauchy–Schwarz: $\mathbb{E}[X] = \mathbb{E}[X\mathbf{1}_{X>\mathbb{E}[X]/2}] + \mathbb{E}[X\mathbf{1}_{X<\mathbb{E}[X]/2}] \leqslant \mathbb{E}[X^2]^{1/2}\mathbb{P}(X > \mathbb{E}[X]/2)^{1/2} + \mathbb{E}[X]/2$.

Recall that $q_n(x)$ is the probability that the point x has been visited by the walk by time n. We have $\mathbb{E}[\mathcal{I}_n^{(d)}(k)] = \sum_{x\in\mathbb{Z}^d} q_n(x)^k$. Using the fact that $(a+b)^k \leqslant 2^k(a^k+b^k)$ we can also write

$$\begin{aligned}
\mathbb{E}\left[\left(\mathcal{I}_n^{(d)}(k)\right)^2\right] &= \sum_{x,y\in\mathbb{Z}^d} \mathbb{P}(x \text{ and } y \text{ visited by all } S_i^{(d)} \text{ by time } n) \\
&= \sum_{x,y\in\mathbb{Z}^d} \mathbb{P}(x \text{ and } y \text{ visited by } S_1^{(d)} \text{ by time } n)^k \\
&\underset{\text{Markov}}{\leqslant} \sum_{x,y\in\mathbb{Z}^d} (q_n(x)q_n(y-x) + q_n(y)q_n(x-y))^k \\
&\leqslant 2^{k+1} \sum_{x,y\in\mathbb{Z}^d} q_n(x)^k q_n(y-x)^k \\
&= 2^{k+1}\left(\sum_{x\in\mathbb{Z}^d} q_n(x)^k\right)^2 = 2^{k+1}\mathbb{E}[\mathcal{I}_n^{(d)}(k)]^2.
\end{aligned}$$

We now focus on the case $(k,d) \in \{(2,4),(3,3)\}$. Since $\mathbb{E}[\mathcal{I}^{(d)}(k)] = \infty$ we get that $\mathbb{E}[\mathcal{I}_n^{(d)}(k)] \to \infty$ as $n\to\infty$ and thanks the the above calculation and the second moment method we deduce that $\mathbb{P}(\mathcal{I}_n^{(d)}(k) > \mathbb{E}[\mathcal{I}_n^{(d)}(k)]/2) > \frac{1}{4\cdot 2^k}$ for all n and so $\mathbb{P}(\mathcal{I}^{(d)}(k) = \infty) \geqslant \frac{1}{2^{k+2}}$. Since the latter probability is equal to 0 or 1 by the Hewitt-Savage law, it must be 1! □

Remark 8.4. As in the case of dimension 2 for recurrence, the dimension 4 is critical of ∞ intersection of two independent random walk paths as dimension 3 is for the intersection of 3 independent random walk paths. With little more work one can for example show that the number of common intersection points within distance n is of order $\log(n)$. Similarly, in dimension 2 the number of returns of the walk by time n is also logarithmic.

Remark 8.5. Although two random walks in dimension 2, 3 and 4 almost surely intersect, one can ask what is the probability they actually do not intersect within the first n steps. In dimension 2 and 3 this probability decays as a polynomial $n^{-\alpha_d}$ for some $\alpha_d > 0$. In dimension 3 the value α_3 is not explicitly known and is not expected to be an especially nice number whereas in dimension 2 we have $\alpha_2 = \frac{5}{8}$. Although simple looking, the derivation of this so-called intersection exponent has required Lawler, Schramm and Werner to use a deep mixture of complex analysis and probability theory. This was celebrated by Werner's Fields medal in 2006 !

8.2.3 Point of increase

We now turn to a problem for which dimension one is already "critical": existence of point of increase for the simple random walk. In the rest of this section $(S_i : i \geqslant 0)$ is a simple symmetric random walk on $\mathbb{Z}$.

Definition 8.2. *A time $k \in \{0,1,\dots,n\}$ is a point of increase for $S_0, S_1, \dots, S_n$*

$$\forall 0 \leqslant i \leqslant k, \quad S_i \leqslant S_k \quad \text{and} \quad \forall k \leqslant i \leqslant n, \quad S_k \leqslant S_i.$$

Theorem 8.6 (*Peres*)

There exists $C > 0$ such that for all $n \geqslant 1$ we have

$$\mathbb{P}\left(S_0, \dots, S_n \text{ has a point of increase}\right) \leqslant \frac{C}{\log n}.$$

Proof. Let us denote $\mathcal{I}_n$ the total number of point of increase of $S_0, \dots, S_n$. If we denote by P_n the event $\{S_0 \geqslant 0, \dots, S_n \geqslant 0\}$ and p_n its probability, then the probability that time k is a time of increase for $S_0, \dots, S_n$ is equal to $p_k p_{n-k}$. Using the asymptotic $p_k \sim ck^{-1/2}$ deduced from (4.3) we see that for some constants $c', c'' > 0$ for all $n \geqslant 1$ we have

$$\mathbb{E}[\mathcal{I}_n] \leqslant c' \sum_{k=1}^{n-1} \frac{1}{\sqrt{k(n-k)}} \leqslant c''.$$

Hence, the number of point of increase is bounded is expectation. To show that it is actually unlikely to have a single point of increase, we need to show that point of increase "come in pack". More precisely, we use the simple observation:

$$\mathbb{P}(\mathcal{I}_n > 0) = \frac{\mathbb{E}[\mathcal{I}_n]}{\mathbb{E}[\mathcal{I}_n \mid \mathcal{I}_n > 0]}, \tag{8.2}$$

since the event $\mathcal{I}_n = 0$ does not contribute to the expectation of $\mathcal{I}_n$. Let $\mathcal{A}_k$ be the event that $k \in \{0, 1, \dots, n\}$ is the first point of increase of $S_0, \dots, S_n$. On the event $\mathcal{A}_k$ notice that $S_k - S_k, S_{k+1} - S_k, \dots, S_n - S_k$ is path of length $n-k$ that stays non-negative. Notice also that we could change this path for any other path that stays non-negative and we would still belong to $\mathcal{A}_k$. This shows that conditionally on $\mathcal{A}_k$, the law of $(S_{k+i} - S_k : 0 \leqslant i \leqslant n-k)$ is that of the simple random walk S conditioned on $\mathcal{P}_{n-k}$. Given the last display, we write

$$\mathbb{E}[\mathcal{I}_n \mid \mathcal{I}_n > 0] \quad = \quad \frac{1}{\mathbb{P}(\mathcal{I}_n > 0)} \sum_{k=0}^{n} \mathbb{E}[\mathcal{I}_n \mid \mathcal{A}_k] \cdot \mathbb{P}(\mathcal{A}_k).$$

Given (8.2) and since $\sum_k \mathbb{P}(\mathcal{A}_k) = \mathbb{P}(\mathcal{I}_n > 0)$ it suffices to prove that there exists $c > 0$ such that for all $0 \leqslant k \leqslant n$ we have $\mathbb{E}(\mathcal{I}_n \mid \mathcal{A}_k) \geqslant c \cdot \log n$ to complete the proof of the theorem. By time-reversal we may even restrict to $0 \leqslant k \leqslant n/2$. Writing $m = n - k \geqslant n/2$ by the above remark we can compute

$$\begin{aligned}
\mathbb{E}[\mathcal{I}_n \mid \mathcal{A}_k] \quad &= \quad \mathbb{E}[\mathcal{I}_{n-k} \mid \mathcal{P}_{n-k}] \quad = \quad \mathbb{E}[\mathcal{I}_m \mid \mathcal{P}_m] \\
&= \quad \frac{1}{p_m} \sum_{i=0}^{m} \mathbb{P}(i \text{ is a p.o.i. for } S_0, \dots, S_m \text{ and } S_j \geqslant 0, \forall 0 \leqslant j \leqslant m) \\
&= \quad \frac{1}{p_m} \sum_{i=0}^{m} \mathbb{P}(0 \leqslant S_j \leqslant S_i, \forall 0 \leqslant j \leqslant i) \cdot p_{m-i}. \qquad (8.3)
\end{aligned}$$

Lemma 8.7. *There exist constant $c_1, c_2 > 0$ such that for all $n \geqslant 1$ we have*

$$\frac{c_1}{n} \leqslant \mathbb{P}(S_0 \leqslant S_i \leqslant S_n : \forall 0 \leqslant i \leqslant n) \leqslant \frac{c_2}{n}.$$

Given the last lemma we can finish the proof of the theorem. Indeed, using again $p_k \sim ck^{-1/2}$ and the last lemma, we easily see that (8.3) is bounded from below by $c' \cdot \log m \geqslant c \cdot \log \frac{n}{2}$ for some constant $c > 0$ which imply the desired estimate. $\square$

Proof of Lemma 8.7. The upper bound (which has not been used in our application) follows easily from the fact that

$$\begin{aligned}\mathbb{P}(S_0 \leqslant S_i \leqslant S_n : \forall 0 \leqslant i \leqslant n) &\leqslant \mathbb{P}(S_i \geqslant 0 : \forall 0 \leqslant i < n/2) \cdot \mathbb{P}(S_i \leqslant S_n : \forall n/2 < i \leqslant n) \\ &\underset{\text{duality}}{=} \mathbb{P}(S_i \geqslant 0 : \forall 0 \leqslant i < n/2)^2 = \left(p_{[n/2]}\right)^2 \underset{(4.3)}{\sim} \frac{2c}{n},\end{aligned}$$

for some $c > 0$. For the lower bound, notice that the events $A = \{S_i \geqslant 0 : \forall 0 \leqslant i \leqslant n\}$ and $B = \{S_i \leqslant S_n : \forall 0 \leqslant i \leqslant n\}$ are positively correlated. Indeed, they are *increasing events* in the sense that if A or B is realized for a path s then it realized for all the paths we can get from s by changing a minus step into a plus step. An application of the FKG inequality then shows that $\mathbb{P}(A \cap B) \geqslant \mathbb{P}(A) \cdot \mathbb{P}(B)$ which in our case yields

$$\mathbb{P}(S_0 \leqslant S_i \leqslant S_n : \forall 0 \leqslant i \leqslant n) \geqslant \mathbb{P}(S_i \geqslant 0 : \forall 0 \leqslant i \leqslant n) \cdot \mathbb{P}(S_i \leqslant S_n : \forall 0 \leqslant i \leqslant n) \underset{\text{duality}}{=} p_n^2 \underset{(4.3)}{\sim} \frac{c}{n}.$$

Bibliographical notes. Ranges and intersection of (simple) random walks is a classic subject in the study of random walks, see [27]. After the initial works of Erdös and collaborators, the two-dimensional case became very active in the 80 in connection with the Brownian intersection local time (Dynkin, Geman–Horowitz–Rosen, Le Gall...). Brownian intersection exponentswere famously computed by Lawler–Schramm–Werner using the SLE processes [**?**]. The proof of Theorem 8.6 is due to Y. Peres and the presentation here is adapted from [35, Chapter 5.2]. Increasing points of stable Lévy processes have been studied by Bertoin in the 90's, see [**?**].

Part IV:

Random trees and graphs

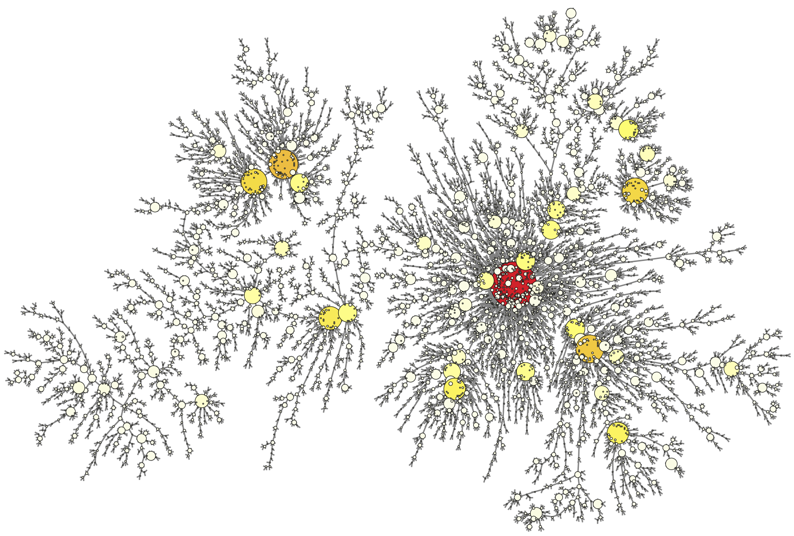


Figure 8.2: A large random plane tree with vertices of high degrees.

Chapter IX: Galton-Watson trees

In this chapter we use our knowledge on one-dimensional random walk to study random trees coding for the genealogy of a population where individuals reproduce independently of each other according to the same offspring distribution. These are the famous Galton–Watson trees.

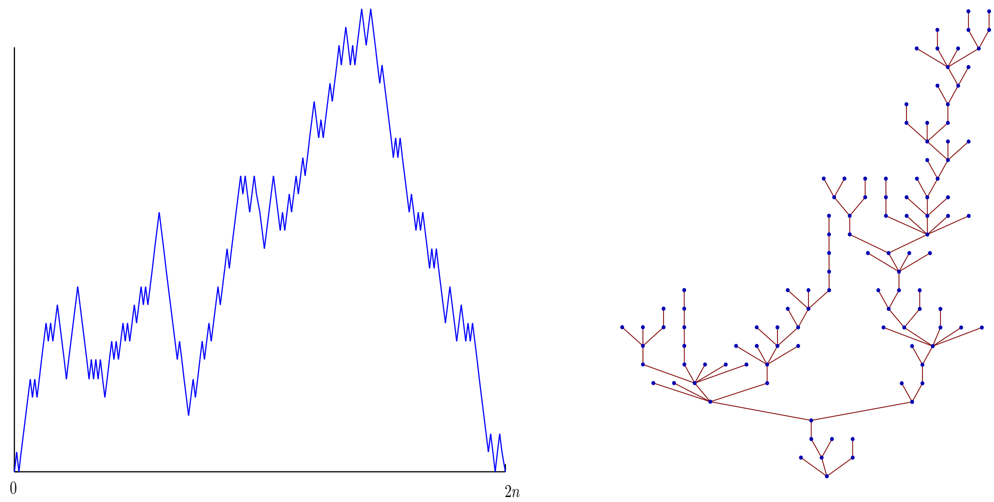


Figure 9.1: A large Galton–Watson tree and its contour function

9.1 Plane trees and Galton–Watson processes

9.1.1 Plane trees

Throughout this work we will use the standard formalism for plane trees as found in [37]. Let

$$\mathcal{U} \;=\; \bigcup_{n=0}^{\infty} (\mathbb{N}^*)^n$$

where $\mathbb{N}^* = \{1, 2, \ldots\}$ and $(\mathbb{N}^*)^0 = \{\varnothing\}$ by convention. An element u of $\mathcal{U}$ is thus a finite sequence of positive integers. We let $|u|$ be the length of the word u. If $u, v \in \mathcal{U}$, uv denotes the concatenation of u and v. If v is of the form uj with $j \in \mathbb{N}$, we say that u is the *parent* of v or that v is a *child* of u. More generally, if v is of the form uw, for $u, w \in \mathcal{U}$, we say that u is an *ancestor* of v or that v is a *descendant* of u.

Definition 9.1. *A* plane tree τ *is a (finite or infinite) subset of* $\mathcal{U}$ *such that*

1. $\varnothing \in \tau$ *($\varnothing$ is called the* root *of* τ*),*
2. *if* $v \in \tau$ *and* $v \neq \varnothing$*, the parent of* v *belongs to* τ
3. *for every* $u \in \mathcal{U}$ *there exists* $k_u(\tau) \in \{0, 1, 2, \dots\} \cup \{\infty\}$ *such that* $uj \in \tau$ *if and only if* $j \leqslant k_u(\tau)$*.*

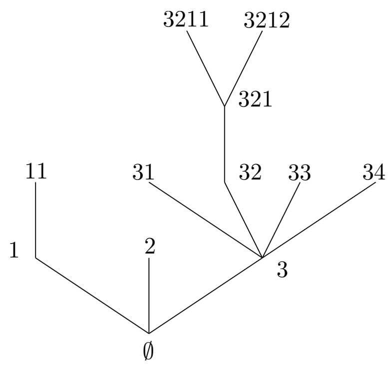


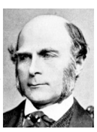
Figure 9.2: A finite plane tree.

A plane tree can be seen as a graph, in which an edge links two vertices u, v such that u is the parent of v or vice-versa. Notice that with our definition, vertices of infinite degree are allowed since k_u may be infinite. When all degrees are finite, the tree is said to be *locally finite*. In this case, this graph is of course a tree in the graph-theoretic sense, and has a natural embedding in the plane, in which the edges from a vertex u to its children $u1, \dots, uk_u(\tau)$ are drawn from left to right. All the trees considered in these pages are plane trees. The integer $|\tau|$ denotes the number of vertices of τ and is called the size of τ. For any vertex $u \in \tau$, we denote the shifted tree at u by $\sigma_u(\tau) := \{v \in \tau : uv \in \tau\}$. If a and b are two vertices of τ, we denote the set of vertices along the unique geodesic path going from a to b in τ by $[\![a, b]\!]$.

Definition 9.2. *The set* $\mathcal{U}$ *is a plane tree where* $k_u = \infty, \forall u \in \mathcal{U}$*. It is called Ulam's tree.*

9.1.2 Galton–Watson trees

Let μ be a distribution on $\{0, 1, 2, \dots\}$ which we usually suppose to be different from δ_1. Informally speaking, a Galton–Watson[1] tree with offspring distribution μ is a random (plane) tree coding the genealogy of a population starting with one individual and where all individuals reproduce independently of each other according to the distribution μ. Here is the proper definition:

[1]  Francis Galton (1822-1911) and  Henry William Watson (1827-1903)

Definition 9.3. *Let K_u for $u \in \mathcal{U}$ be independent and identically distributed random variables of law μ. We let $\mathcal{T}$ be the random plane tree made of all $u = u_1 u_2 \dots u_n \in \mathcal{U}$ such that $u_i \leqslant K_{u_1 \dots u_{i-1}}$ for all $1 \leqslant i \leqslant n$. Then the law of $\mathcal{T}$ is the μ-Galton–Watson distribution.*

Notice that the random tree defined above may very well be infinite.

Exercise 9.1. Show that the the random tree $\mathcal{T}$ constructed above has the following branching property: Conditionally on $k_\varnothing(\mathcal{T}) = \ell \geqslant 0$ then the ℓ random trees $\sigma_i(\mathcal{T})$ for $1 \leqslant i \leqslant \ell$ are independent and distributed as $\mathcal{T}$.

Exercise 9.2. If τ_0 is a *finite* plane tree and if $\mathcal{T}$ is a μ-Galton–Watson tree then

$$\mathbb{P}(\mathcal{T} = \tau_0) = \prod_{u \in \tau_0} \mu_{k_u(\tau_0)}.$$

We now link the Galton–Watson tree to the well-known Galton–Watson process. We first recall its construction. Let $(\xi_{i,j} : i \geqslant 0, j \geqslant 1)$ be i.i.d. random variables of law μ. The μ-Galton–Walton process is defined by setting $Z_0 = 1$ and for $i \geqslant 0$

$$Z_{i+1} = \sum_{j=1}^{Z_i} \xi_{i,j}.$$

It is then clear from our constructions that if $\mathcal{T}$ is a μ-Galton–Watson tree, then the process $X_n = \#\{u \in \mathcal{T} : |u| = n\}$ has the law of a μ-Galton–Watson process.

9.2 Łukasiewicz walk and direct applications

In this section we will encode (finite) trees via one-dimensional walks. This will enable us to get information on random Galton–Watson trees from our previous study of one-dimensional random walks.

9.2.1 Łukasiewicz walk

The *lexicographical* order $<$ on $\mathcal{U}$ is defined as the reader may imagine: if $u = i_1 i_2 \dots i_k$ and $v = j_1 j_2 \dots j_k$ are two words of the same length then $u < v$ if $i_\ell < j_\ell$ where ℓ is the first index where $i_\ell \neq j_\ell$. The *breadth first* order on $\mathcal{U}$ is defined by $u \prec v$ if $|u| < |v|$ and if the two words are of the same length then we require $u < v$ (for the lexicographical order).

Definition 9.4. *Let τ be a locally finite tree (i.e. $k_u(\tau) < \infty$ for every $u \in \tau$). Write $u_0, u_1, \dots$ for its vertices listed in in the breadth first order. The* Łukasiewicz walk *$\mathcal{W}(\tau) = (\mathcal{W}_n(\tau), 0 \leqslant n \leqslant |\tau|)$ associated to τ is given by $\mathcal{W}_0(\tau) = 0$ and for $0 \leqslant n \leqslant |\tau| - 1$:*

$$\mathcal{W}_{n+1}(\tau) = \mathcal{W}_n(\tau) + k_{u_n}(\tau) - 1.$$

In words, the Łukasiewicz walk consists in listing the vertices in breadth first order and making a stack by adding the number of children of each vertex and subtracting one (for the

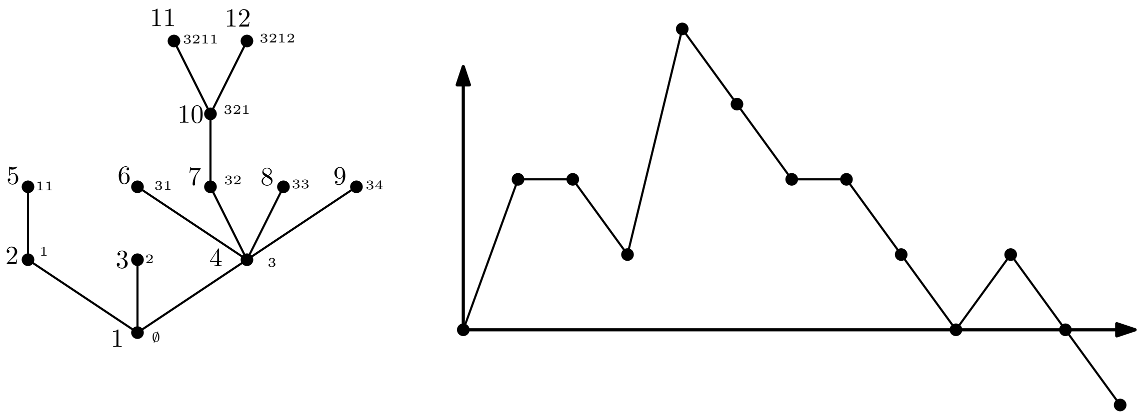


Figure 9.3: Left: a finite plane tree and its vertices listed in breadth-first order. Right: its associated Łukasiewicz walk.

current vertex). In the case of a finite plane tree, the total number of children is equal to the number of vertices minus one, it should be clear that the Łukasiewicz walk which starts at 0 stays non-negative until it finishes at the first time it touches the value -1. Note also that this walk (or more precise its opposite) is skip-free in the sense of Chapter 4 since $\mathcal{W}_{i+1}(\tau) - \mathcal{W}_i(\tau) \geqslant -1$ for any $0 \leqslant i \leqslant |\tau| - 1$. When the tree is infinite but locally finite, every vertex of the tree will appear in the breadth first ordering[2] and the Łukasiewicz path stays non-negative for ever. We leave the following exercise to the reader:

Exercise 9.3. Let $\mathbf{T}_{lf}$ the set of all finite and infinite locally finite plane trees. Let $\mathbf{W}_{lf}$ the set of all finite and infinite paths $(w_0, w_1, \dots, w_n)$ with $n \in \{1, 2, \dots\} \cup \{\infty\}$ which starts at $w_0 = 0$ and ends at $w_n = -1$ and such that $w_{i+1} - w_i \geqslant -1$ as well as $w_i \geqslant 0$ for any $0 \leqslant i \leqslant n-1$. Then taking the Łukasiewicz walk creates a bijection between $\mathbf{T}_{lf}$ and $\mathbf{W}_{lf}$.

9.2.2 Łukasiewicz walk of a Galton–Watson tree

As it turns out the Łukasiewicz walk associated to a μ-Galton–Watson tree is roughly speaking a random walk. Recall that the offspring distribution μ is supported by $\{0, 1, 2, \dots\}$ so a μ-Galton–Watson tree is locally finite a.s.

Proposition 9.1. *Let $\mathcal{T}$ be a μ-Galton–Watson tree, and let $(S_n)_{n\geqslant 0}$ be a random walk with i.i.d. increments of law $\mathbb{P}(S_1 = k) = \mu(k+1)$ for $k \geqslant -1$. If $T^<$ is the first hitting time of -1 by the walk S then we have*

$$\big(\mathcal{W}_0(\mathcal{T}), \mathcal{W}_1(\mathcal{T}), \dots, \mathcal{W}_{|\mathcal{T}|}(\mathcal{T})\big) \quad \overset{(d)}{=} \quad (S_0, S_1, \dots, S_{T^<})$$

Proof. Let $(\omega_i^0 : 0 \leqslant i \leqslant n)$ be the first n steps of a skip-free random walk so that n is less than or equal than the hitting time of -1 by this walk. By reversing the Łukasiewicz construction

[2] this is not true if we had chosen to explore the tree in the lexicographical order.

we see that in order that the first n steps of the Łukasiewicz walk of the tree $\mathcal{T}$ matches with $(\omega_i^0 : 0 \leqslant i \leqslant n)$ the first n vertices of the trees in breadth first order as well as their number of children are fixed by $(\omega_i^0 : 0 \leqslant i \leqslant n)$, see Figure 9.4.

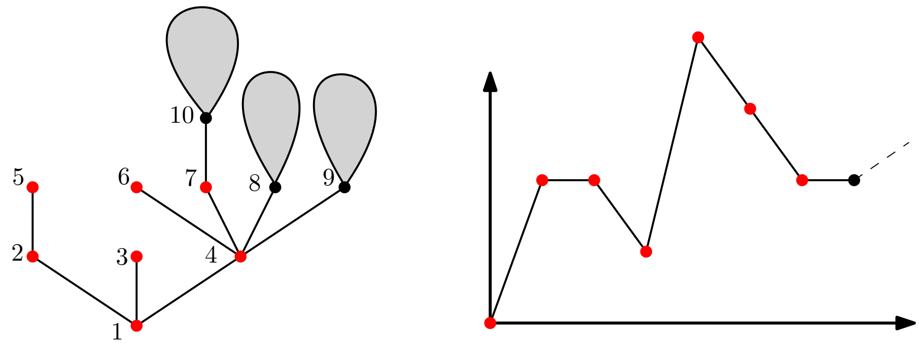


Figure 9.4: Fixing the first n vertices explored (in red) during the breadth first exploration of a Galton–Watson tree. The black vertices and their subtrees (in gray) have not been explored yet.

The probability under the μ-Galton–Watson to see this event is seen to be

$$\prod_{u \in \tau_0} \mu_{k_u(\mathcal{T})} = \prod_{i=0}^{n-1} \mu_{\omega_{i+1}^0 - \omega_i^0 + 1} = \mathbb{P}\left((S_i)_{0 \leqslant i \leqslant n} = (\omega_i^0)_{0 \leqslant i \leqslant n}\right),$$

where τ_0 is the tree made by the first n vertices in breadth first order in $\mathcal{T}$. The proposition easily follows. □

Extinction probability. Here is a direct application of the last result: a random walk proof of the following well-known criterion for survival of a Galton–Watson process.

Theorem 9.2 (*Extinction probability*)

Let μ be an offspring distribution of mean $m \geqslant 0$ such that $\mu \neq \delta_1$. Recall that $\mathcal{T}$ denotes a μ-Galton–Watson tree and $(Z_n)_{n \geqslant 0}$ a μ-Galton–Watson process. Then the probability that the tree is finite is equal to the smallest solution $\alpha \in [0, 1]$ to the equation

$$\alpha = \sum_{k \geqslant 0} \mu_k \alpha^k, \tag{9.1}$$

in particular it is equal to 1 if $m \leqslant 1$.

Proof. The equivalence between the first two assertions is obvious by the discussion at the end of Section 9.1.2. We then use Proposition 9.1 to deduce that $\mathbb{P}(|\mathcal{T}| = \infty) = \mathbb{P}(T^{<} = \infty)$. Since the walk S is non trivial (i.e. not constant), we know from Chapter 3 that the last probability is positive if and only if its drift is strictly positive i.e. if $m > 1$. The computation of the probability that the walks stays positive has been carried out in Proposition 4.13 and its translation in our context yields the statement. □

Exercise 9.4. The standard proof of the last theorem is usually done as follows: Let $g(z) = \sum_{k\geqslant 0} \mu_k z^k$ be the generating function of the offspring distribution μ.

1. Show that the generating function of Z_n is given by $g \circ g \circ \cdots \circ g$ (n-fold composition).
2. Deduce that $u_n = \mathbb{P}(Z_n = 0)$ follows the recurrence relation $u_0 = 0$ and $u_{n+1} = g(u_n)$.
3. Re-deduce Theorem 9.2.

Remark 9.1 (A historical remark). We usually attribute to Galton and Watson the introduction and study of the so-called Galton–Watson trees in 1873 in order to study the survival of family names among British lords. However, in their initial paper devoted to the calculation of the extinction probability they conclude hastily that the latter is a fixed point of Equation 9.1 and since 1 is always a fixed point, then the extinction is almost sure whatever the offspring distribution. Too bad. This is even more surprising since almost thirty years before, in 1845 Irénée-Jules Bienaymé[3] considered the very same model and derived correctly the extinction probability. This is yet just another illustration of Stigler's law of eponymy!

9.2.3 Lagrange inversion formula

The Lagrange inversion is a close formula for the coefficients of the inverse of a power series. More precisely, imagine that $f(z) = \sum_{i\geqslant 0} f_i z^i \in \mathbb{C}[[z]]$ is a formal power series in the indeterminate z (no convergence conditions are assumed) so that $f_0 = 0$ and $f_1 \neq 0$. One then would like to invert f i.e. finding a power series $\phi \in \mathbb{C}[[z]]$ such that $z = \phi(f(z)) = f(\phi(z))$. In combinatorics the above equation is usually written in the "Lagrange formulation" by supposing that $f(z) = \frac{z}{R(z)}$ with $R(z) \in \mathbb{C}[[z]]$ with $R(0) \neq 0$ so that the equation becomes

$$\phi(z) = z \cdot R(\phi(z)). \tag{9.2}$$

Theorem 9.3 (*Lagrange inversion formula*)

Let $R \in \mathbb{C}[[z]]$ be a formal power series in z such that $[z^0]R \neq 0$. Then there exists a unique formal power series ϕ satisfying (9.2) and we have for all $k \geqslant 0$ and all $n \geqslant 1$

$$[z^n]\big(\phi(z)\big)^k = \frac{k}{n}[z^{n-1}]\left(z^{k-1}R(z)^n\right),$$

where $[z^n]f(z)$ in the coefficient in front of z^n in the formal power series $f \in \mathbb{C}[[z]]$.

Proof. The idea is to interpret combinatorially the weights in the formal expansion $z \cdot R(\phi(z))$, where $R(z) = \sum_{i\geqslant 0} r_i z^i$. Indeed, by expanding recursively the equation (9.2), it can be seen that the coefficient in front of z^n in ϕ can be interpreted as a sum over all plane trees with n

[3]  Irénée-Jules Bienaymé (1796-1878)

vertices where the weight of a tree τ is given by

$$\mathrm{w}(\tau) \quad = \quad \prod_{u \in \tau} r_{k_u(\tau)}.$$

Similarly for $k \geqslant 1$, the coefficient of z^n in ϕ^k is the weight of forest of k trees having n vertices in total. Now, using the Łukasiewicz encoding, a forest can be encoded by a skip-free descending path with n steps and reaching $-k$ for the first time at time n and the weight of such paths become $\mathrm{w}(S) = \prod_{i \geqslant 0}^{n-1} r_{S_{i+1}-S_i+1}$. By Feller's combinatorial lemma, for a skip-free descending walk (S) of length n such that $S_n = -k$ there are exactly k cyclic shifts so that n is the k-th strict descending ladder time. So if we partition the set of all walks of length n so that $S_n = -k$ using the cyclic shift as an equivalence relation, we know that in each equivalence classes, the proportion of walks so that $T_k^< = n$ is $\frac{k}{n}$ (most of the classes actually have n elements in it, but it could be the case that the subgroup of cyclic shifts fixing the walk is non-trivial and has order $\ell | k$, in which case there are n/ℓ elements in the orbit and k/ℓ are such that $T_k^< = n$). Since the weight $\mathrm{w}(\cdot)$ is constant over all equivalence classes we deduce that:

$$\sum_{\substack{(S) \text{ walks of length } n \\ S_n = -k}} \mathrm{w}(S) = \frac{n}{k} \sum_{\substack{(S) \text{ walks of length } n \\ S_n = -k \text{ and } T_k^< = n}} \mathrm{w}(S).$$

It remains to notice that

$$[z^{n-1}] \left(z^{k-1} R(z)^n \right),$$

is exactly the weight of all paths (S) of length n such that $S_n = -k$. □

Here are two recreative (but surprising) applications of Lagrange inversion formula:

Exercise 9.5. Let $F(x)$ be the be the unique power series with rational and positive coefficients such that for all $n \geqslant 0$ the coefficient of x^n in $F^{n+1}(x)$ is equal to 1. Show that $F(x) = \frac{x}{1-\mathrm{e}^{-x}}$.

Exercise 9.6. For $a \in (0,1)$ shows that the only positive solution $x = x(a)$ of $x^5 - x - a = 0$ can be written as

$$x = -\sum_{k \geqslant 0} \binom{5k}{k} \frac{a^{4k+1}}{4k+1}.$$

9.3 Probabilistic counting of trees

In this section we illustrate how to enumerate certain classes of trees using our knowledge on random walks. One underlying idea is to design a random variable which is uniformly distributed on the set we wish to count.

9.3.1 Prescribed degrees

Theorem 9.4 (*Harary & Prins & Tutte (1964)*)

The number of plane trees with d_i vertices with $i \geqslant 0$ children, and with $n = 1 + \sum i d_i = \sum d_i$

vertices is equal to

$$\frac{(n-1)!}{d_0!d_1!\cdots d_i!\cdots}.$$

Proof. Fix d_i, k and n as in the theorem. Notice that we must have $n = 1 + \sum id_i = \sum d_i$. By the encoding of plane trees into their Łukaciewicz path it suffices to enumerate the number of paths starting from 0, ending at -1 at n and with d_i steps of $i-1$ and which stay non-negative until time $n-1$. Clearly, if one removes the last assumption there are

$$\binom{n}{d_0, \dots, d_i, \dots} = \frac{n!}{d_0!d_1!\cdots}$$

such paths. If we partition those paths according to the cyclic shift equivalence relation, then by Lemma 4.3 (see Remark 4.3) we know that each equivalence class has cardinal n and has a unique element which stays non-negative until time $n-1$. Hence the quantity we wanted to enumerate is equal to

$$\frac{1}{n}\binom{n}{d_0, \dots, d_i, \dots} = (n-1)! \prod_i \frac{1}{d_i!}.$$

□

Corollary 9.5 (Catalan's counting). *We have*

$$\#\{\text{plane trees with } n+1 \textit{ vertices}\} = \frac{1}{n+1}\binom{2n}{n}.$$

Proof. With the same notation as in the preceding theorem, the number of trees with $n \geqslant 1$ vertices is equal to

$$\sum_{\substack{d_0, d_1, d_2, \dots \\ 1+\sum id_i = n = \sum d_i}} \frac{(n-1)!}{d_0!d_1!\cdots} = \frac{1}{n} \sum_{\substack{d_0, d_1, d_2, \dots \\ 1+\sum id_i = n = \sum d_i}} \binom{n}{d_0, d_1, \dots} = \frac{1}{n}[z^{n-1}]\left(1 + z + z^2 + z^3 + \cdots\right)^n.$$

Using Lagrange inversion formula (Theorem 9.3) the last quantity can be expressed as $[z^n]\phi(z)$ where $\phi(z)$ is the formal series solution to $\phi(z) = \frac{z}{1-\phi(z)}$. Solving we get $\phi(z) = \frac{1}{2}(1 - \sqrt{1-4z})$ and a coefficient extraction yields the desired formula. □

9.3.2 Uniform trees

We denote by $\mathbf{T}_n$ the set of all plane trees with n edges and by $\mathcal{T}_n$ a uniform plane tree taken in $\mathbf{T}_n$. As we shall see $\mathcal{T}_n$ can be seen as a conditioned version of a Galton–Watson tree:

Proposition 9.6. *Let $\mathcal{T}$ be the Galton–Watson tree with geometric offspring distribution of parameter $1/2$, i.e. $\mu_k = \left(\frac{1}{2}\right)^{k+1}$ for $k \geqslant 0$. Then $\mathcal{T}_n$ has the law of $\mathcal{T}$ conditioned on having n edges.*

Proof. Let τ_0 be a tree with n edges. Then by Exercise 9.2 we have

$$\mathbb{P}(\mathcal{T} = \tau_0) = \prod_{u \in \tau_0} 2^{-k_u(\tau)-1}.$$

However, it is easy to see that in a plane tree with n edges we have $\sum_{u \in \tau_0} k_u(\tau_0) = |\tau_0| - 1 = n$ so that the last display is equal to $\frac{1}{2}4^{-n}$. The point is that this probability does not depend on τ_0 as long as it has n edges. Hence, the conditional law of $\mathcal{T}$ on $\mathbf{T}_n$ is the uniform law. $\square$

Notice that the above proposition and its proof hold for any non trivial parameter of the geometric offspring distribution. However, we chose 1/2 because in this case the offspring distribution is critical, i.e. it has mean 1. We can give another proof of Corollary 9.5:

Proof. Combining by the last proposition with Proposition 9.1 and Kemperman formula yields

$$\mathbb{P}(|\mathcal{T}| = n+1) = \mathbb{P}(T^{<} = n+1) \underset{\text{Prop.4.9}}{=} \frac{1}{n+1}\mathbb{P}(S_{n+1} = -1),$$

where (S) is the random walk whose increments are distributed as $\mathbb{P}(S_1 = k) = 2^{-k-2}$ for $k \geqslant -1$ or equivalently as $G-1$ where G is the geometric offspring distribution of parameter $1/2$. Recall that G is also the number of failures before the first success in a series of independent coin flips: this is the negative Binomial distribution with parameter $(1, 1/2)$. Hence $\mathbb{P}(S_{n+1} = -1) = \mathbb{P}(\mathrm{Binneg}(n+1, 1/2) = n)$ where $\mathrm{Binneg}(n,p)$ is the negative Binomial distribution with parameter (n,p). This distribution is explicit and we have $\mathbb{P}(\mathrm{Binneg}(n,p) = k) = \binom{n+k-1}{n-1}p^n(1-p)^k$ which is our case reduces to

$$\frac{1}{n+1}\mathbb{P}(S_{n+1} = -1) = \frac{1}{n+1}\mathbb{P}(\mathrm{Binneg}(n+1, 1/2) = n) = \frac{1}{2}4^{-n}\frac{1}{n+1}\binom{2n}{n}.$$

By the previous proposition (and its proof) we have on the other hand

$$\mathbb{P}(|\mathcal{T}| = n+1) = \#\{\text{plane trees with } n+1 \text{ vertices}\} \cdot \frac{1}{2}4^{-n}.$$

The result follows by comparing the last two displays. $\square$

Exercise 9.7. Extend the above proof to show that the number of forest of $f \geqslant 1$ trees whose total number of edges is n is equal to

$$\frac{f}{2n+f}\binom{2n+f}{n}.$$

Exercise 9.8. Give another proof of the last display using Theorem 9.3.

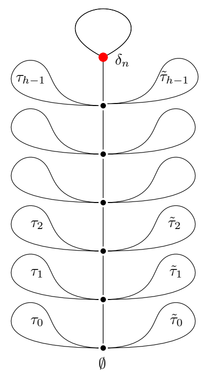
Theorem 9.7 (***Height of a uniform point***)

Let $\mathcal{T}_n$ be a uniform plane tree with n edges. Conditionally on $\mathcal{T}_n$ let δ_n be a uniformly chosen vertex of $\mathcal{T}_n$ and denote H_n its height (the distance to the ancestor of $\mathcal{T}_n$) then we have

$$\mathbb{P}(H_n = h) = \frac{2h+1}{2n+1}\frac{\binom{2n+1}{n-h}}{\binom{2n}{n}}.$$

In particular $H_n/\sqrt{n}$ converges towards the Rayleigh distribution with density $2x\exp(-x^2)\mathbf{1}_{x>0}$.

Proof. We compute exactly the probability that the point δ_n is located at height $h \geqslant 0$.



If so, the tree $\mathcal{T}_n$ is obtained from the line joining $\varnothing$ to δ_n by grafting $2h+1$ plane trees on its left and on its right. Obviously, the total number of edges of these trees must be equal to $n-h$. Using Exercise 9.7 we deduce that

$$\mathbb{P}(H_n = h) = \frac{\frac{2h+1}{2n+1}\binom{2n+1}{n-h}}{\binom{2n}{n}}.$$

The second item of the theorem follows after applying Stirling formula and using Exercise 4.12. □

9.3.3 Poisson Galton–Watson trees and Cayley trees

In this section we focus on a different type of trees first studied by Cayley [4]

Definition 9.5. *A* Cayley tree *is a labeled tree with n vertices without any orientation nor distinguished point. In other words it is a spanning tree on $\mathbb{K}_n$, the complete graph over n vertices. See Fig. 9.5.*

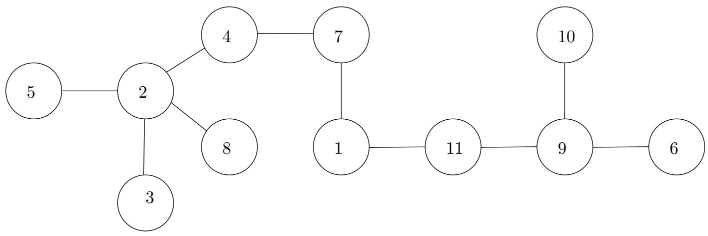
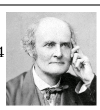


Figure 9.5: A Cayley tree over $\{1, 2, 3, 4, \dots, 11\}$.

Let $\mathcal{T}$ be a Galton-Watson (plane) tree with Poisson offspring distribution of parameter 1 (in particular, the mean number of children is 1 and we are in the critical case). As above we denote by $\mathcal{T}_n$ the random tree $\mathcal{T}$ conditioned on having n vertices.

Proposition 9.8. *Consider $\mathcal{T}_n$ and assign the labels $\{1, \dots, n\}$ uniformly at random to the vertices of $\mathcal{T}_n$. After forgetting the plane ordering $\mathcal{T}_n$ this produces a Cayley tree which we denote by $\mathfrak{T}_n$. Then $\mathfrak{T}_n$ is uniformly distributed over all Cayley trees.*

Proof. Let us first compute the probability that $\mathcal{T}$ has n vertices. Using the Łukasiewicz walk and the cyclic lemma we get that $\mathbb{P}(|\mathcal{T}| = n) = \frac{1}{n}\mathbb{P}(S_n = -1)$, where S is the random walk

[4]  Arthur Cayley (1821-1895) receiving a phone call

whose increments are centered distributed according to Poisson(1) − 1 variable. It follows that

$$\mathbb{P}(|\mathcal{T}| = n) = \frac{1}{n}\mathbb{P}(\text{Poisson}(n) = n-1) = \mathrm{e}^{-n}\frac{n^{n-2}}{(n-1)!}.$$

Fix a Cayley tree $\mathfrak{t}$ and let us study the possible ways to obtain $\mathfrak{t}$ by the above process. We first choose the root of the tree among the n possibles vertices. Once the origin is distinguished, there are $\prod_{u\in\mathfrak{t}} k_u!$ possible ways to give a planar orientation to the tree, where k_u is the number of children of the vertex u (for this we needed to distinguish the ancestor vertex). After these operations, each of the labeled, rooted, plane trees obtained appears with a probability (under the Galton–Watson measure) equal to

$$\frac{1}{n!}\mathrm{e}^{-n}\prod_{u\in\mathfrak{t}}\frac{1}{k_u!}.$$

Performing the summation, the symmetry factors involving the $k_u!$ conveniently disappear and we get

$$\mathbb{P}(\mathcal{T}_n \to \mathfrak{t}) = n \times \frac{\mathrm{e}^{-n}}{n!}\left(\mathrm{e}^{-n}\frac{n^{n-2}}{(n-1)!}\right)^{-1} = \frac{1}{n^{n-2}}.$$

Since the result of the last display does not depend on the shape of $\mathfrak{t}$, the induced law is indeed uniform over all Cayley trees and we have even proved:

Corollary 9.9 (Cayley's formula)**.** *The total number of Cayley trees on n vertices is n^{n-2}.*

Exercise 9.9. For any $p \geqslant 2$ a p-tree is a plane tree such that the number of children of each vertex is either 0 or p. When $p = 2$ we speak of binary trees. In particular the number of edges of a p-tree must be a multiple of p. Show that for any $k \geqslant 1$ we have

$$\#\{p-\text{trees with } kp \text{ edges }\} = \frac{1}{kp+1}\binom{kp+1}{k},$$

in three ways: using a direct application of Theorem 9.4, using a probabilistic approach via a certain class of random Galton–Watson trees, or via Lagrange inversion's formula Theorem 9.3.

9.4 Contour function

We finish this chapter by mentioning another more geometrical encoding of plane trees which is less convenient in the general case but very useful in the case of geometric Galton–Watson trees.

Let τ be a finite plane tree. The contour function $\mathcal{C}_\tau$ associated with τ is heuristically obtained by recording the height of a particle that climbs the tree and makes its contour at unit speed. More formally, to define it properly one needs the definition of a corner: We view τ as embedded in the plane then a *corner* of a vertex in τ is an angular sector formed by two consecutive edges in clockwise order around this vertex. Note that a vertex of degree k in τ has exactly k corners. If c is a corner of τ, $\mathrm{Ver}(c)$ denotes the vertex incident to c.

The corners are ordered clockwise cyclically around the tree in the so-called *contour order*. If τ has $n \geqslant 2$ vertices we index the corners by letting $(c_0, c_1, c_2, \ldots, c_{2n-3})$ be the sequence of corners visited during the contour process of τ, starting from the corner c_0 incident to $\varnothing$ that is located to the left of the oriented edge going from $\varnothing$ to 1 in τ.

Definition 9.6. *Let τ be a finite plane tree with $n \geqslant 2$ vertices and let $(c_0, c_1, c_2, \ldots, c_{2n-3})$ be the sequence of corners visited during the contour process of τ. We put $c_{2n-2} = c_0$ for notational convenience. The contour function of τ is the walk defined by*

$$\mathcal{C}_\tau(i) = |\mathrm{Ver}(c_i)|, \qquad \textit{for } 0 \leqslant i \leqslant 2n-2.$$

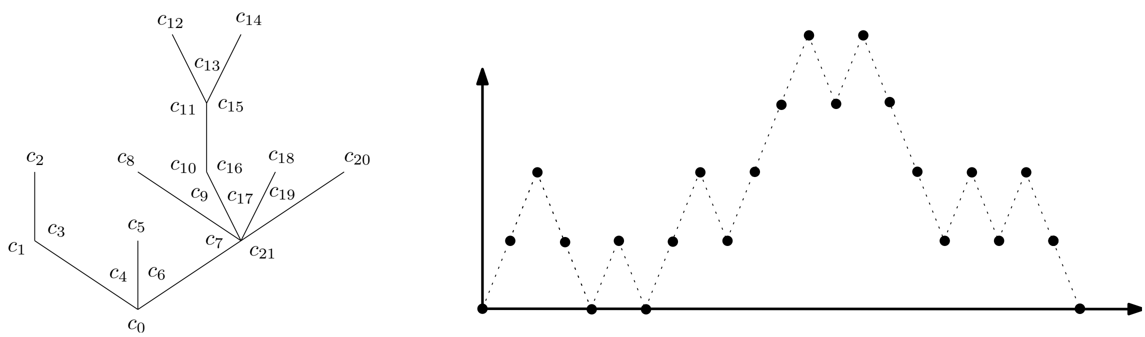


Figure 9.6: The contour function associated with a plane tree.

Clearly, the contour function of a finite plane tree is a finite non-negative walk of length $2|\tau| - 1$ which only makes ± 1 jumps. Here as well, the encoding of a tree into its contour function is invertible:

Exercise 9.10. Show that taking the contour function creates a bijection between the set of all finite plane trees and the set of all non-negative finite walks with ± 1 steps which start and end at 0.

Now, we give a probabilistic description of the law of the contour function of $\mathcal{T}$ when $\mathcal{T}$ is distributed as a geometric(1/2)-Galton–Watson tree (i.e. has the same law as in Section 9.3.2).

Proposition 9.10. *Let $\mathcal{T}$ as above. Then its contour function $\mathcal{C}_\mathcal{T}$ has the same law as*

$$(S_0, S_1, \ldots, S_{T^< - 1}),$$

where (S) is a simple symmetric random walk and $T^<$ is the first hitting time of -1.

Proof. Notice first that $\mathcal{T}$ is almost surely finite by Theorem 9.2 and so all the objects considered above are well defined. Let τ_0 be a plane tree with n edges. We have seen in the previous proposition that $\mathbb{P}(\mathcal{T} = \tau_0) = \frac{1}{2} 4^{-n}$. On the other hand, the contour function of τ_0 has length $2n$ and the probability that the first $2n$ steps of (S) coincide with this function and that $T^< = 2n+1$ is equal to $2^{-2n} \cdot \frac{1}{2} = \frac{1}{2} 4^{-n}$. This concludes the proof. $\square$

Exercise 9.11. Give a new proof of Corollary 9.5 using the contour function.

Exercise 9.12. Let $\mathcal{T}$ be a Galton–Watson tree with geometric(1/2) offspring distribution. The height of $\mathcal{T}$ is the maximal length of one of its vertices. Prove that

$$\mathbb{P}(\mathrm{Height}(\mathcal{T}) \geqslant n) = \frac{1}{n+1}.$$

Additional exercises.

Exercise 9.13. Let $\mathcal{T}$ be a Galton–Watson tree with offspring distribution μ such that $\mu(0) = 0$ and $\mu(1) \neq 1$. In particular the mean of μ is strictly larger than 1 and $\mathcal{T}$ is almost surely infinite. The tree $\mathcal{T}$ induces an infinite network (that we still denote by $\mathcal{T}$) by putting conductance 1 on every edge. The goal of the exercise is to show that this tree is almost surely transient. For this we build a unit flux on this graph entering at the origin vertex (the image of the ancestor vertex) and escaping at infinity, the flux being split equally likely at each branching point. Let θ_n be the expected energy of that flux over the first n generations of the tree.

1. Show that if $\sup_{n\geqslant 0} \mathbb{E}[\theta_n] < \infty$ then $\mathcal{T}$ is transient.

2. Using the branching property of $\mathcal{T}$ show that

$$\mathbb{E}[\theta_n] = \sum_{k=1}^{\infty} \mu_k k \left(\frac{1}{k^2} (\mathbb{E}[\theta_{n-1}] + 1) \right).$$

3. Conclude.

4. How would you extend the result to the case when $\mu(0) > 0$ but $\mathbb{E}[\mu] > 1$ and $\mathcal{T}$ is conditioned to be infinite (this is an event of positive probability) ?

Bibliographical notes. The material about Galton–Watson tree is rather classical. The coding of trees and the formalism for plane trees (the so-called Neveu's notation [37]) can be found in [29]. The lecture notes of Igor Kortchemski [22] is a very good introduction accessible to the first years of undergraduate studies in math. Beware some authors prefer to take the lexicographical order rather than the breadth first order to define the Łukasiewicz walk (in the finite case this causes no problem but this is not a bijection if the trees can be infinite). The two exercices illustrating Lagrange inversion formula are taken from the MathOverFlow post "What is Lagrange inversion good for?".

Chapter X: Galton-Watson trees conditioned to survive

In this chapter we study the infinite random trees arising by conditioning a μ-Galton-Watson tree to survive forever. Thanks to the coding of Galton–Watson trees via their Łukasiewicz path, this is equivalently obtained by conditioning the path to stay non-negative. We shall however give an equivalent description of the Galton–Watson trees conditioned to survive via the so-called spine decomposition which is now ubiquitous in probability theory. In this chapter the offspring distribution μ supported by $\{0, 1, 2, \dots\}$ is fixed and its mean is denoted by $m = \sum_{k=0}^{\infty} k\mu_k$.

10.1 Construction of the Galton–Watson trees conditioned to survive

In this section we define the infinite μ-Galton–Watson $\mathcal{T}^{\uparrow}$ conditioned to survive. After giving an "abstract" construction based on the the results of Chapter 5, we give an alternative description of the random tree due to Kesten in the critical case $m = 1$.

10.1.1 The abstract construction by $h^{\uparrow}$-transform

Recall from Proposition 9.1 that any locally finite plane tree can be bijectively associated with a non-negative excursion of a skip-free walk. In the case of a μ-Galton–Watson plane tree, this path is nothing but a random walk with i.i.d. increments of law $(\mu(k+1) : k \geqslant -1)$ stopped at $T^{<}$, the first instant when it touches -1. When the mean offspring m is larger than or equal to 1, the above random walk drifts to $+\infty$ or oscillates and by the results of Chapter 5 (see Remark 5.2) the renewal function $h^{\uparrow}$ of this walk is equal to

$$h^{\uparrow}(i) = 1 + \alpha + \cdots + \alpha^i, \quad \text{for } i \geqslant 0,$$

where α is the probability that our Galton–Watson process becomes extinct, which is the smallest solution in $[0, 1]$ to $\alpha = \sum_{k \geqslant 0}^{\infty} \mu_k \alpha^k$ by Proposition 9.2. An abstract way to build the μ-Galton–Watson tree conditioned to survive is to consider the locally finite plane tree $\mathcal{T}^{\uparrow}$ such that $\mathcal{W}(\mathcal{T}^{\uparrow}) = (S_i^{\uparrow} : i \geqslant 0)$, the version of (S) conditioned to survive by $h^{\uparrow}$-transformation. In particular, the law of such a random infinite plane tree is characterized as follows. For any plane tree τ we write by $[\tau]_n = \{u \in \tau : |u| \leqslant n\}$ the tree made by the first n generations.

Proposition 10.1. *For any $n \geqslant 0$ and for any plane tree τ_0 of height exactly n with ℓ vertices* *at maximal height, we have*

$$\mathbb{P}([\mathcal{T}^{\uparrow}]_n = \tau_0) = h^{\uparrow}(\ell) \cdot \prod_{\substack{u \in \tau_0 \\ |u| \neq n}} \mu_{k_u(\tau_0)}.$$

Proof. Fix a tree τ_0 of height exactly n and with ℓ vertices at maximal height. Denote by $(\omega_i^0 : 0 \leqslant i \leqslant m)$ the Łukasiewicz path obtained by exploring all the vertices of the tree in breadth first order except for the vertices at maximal height. By the Łukasiewicz endcoding, the event $\{[\mathcal{T}^{\uparrow}]_n = \tau_0\}$ occurs if and only if the first m steps of $\mathcal{W}(\mathcal{T}^{\uparrow})$ are equal to $(\omega_i^0 : 0 \leqslant i \leqslant m)$. However, by construction of the path, the height ω_m^0 is the number of vertices in the stack to be further explored, that is ℓ. We conclude using the properties of the $h^{\uparrow}$-transform that

$$\begin{aligned} \mathbb{P}([\mathcal{T}^{\uparrow}]_n = \tau_0) &= \mathbb{P}(\mathcal{W}_i(\mathcal{T}^{\uparrow}) = \omega_i^0, \forall 0 \leqslant i \leqslant m) \\ &\underset{(5.1)}{=} \frac{h^{\uparrow}(\ell)}{h^{\uparrow}(0)} \mathbb{P}(S_i = \omega_i^0, \forall 0 \leqslant i \leqslant m) \\ &= h^{\uparrow}(\ell) \cdot \prod_{\substack{u \in \tau_0 \\ |u| \neq n}} \mu_{k_u(\tau_0)}. \end{aligned}$$

□

It is easy to see that the previous proposition indeed characterizes the law of $\mathcal{T}^{\uparrow}$.

10.1.2 Kesten's tree and spine decomposition

We now construct Kesten's[1] tree, an infinite random plane tree denoted by $\mathcal{K}$ using two types of reproduction laws. We denote by $\overline{\mu}$ the size biased distribution of μ obtained by putting for $k \geqslant 0$

$$\overline{\mu}_k = k\mu_k \frac{1}{m},$$

which is indeed a probability distribution thanks to our normalization. Notice that $\overline{\mu}$ is supported by (strictly) positive integers. We construct a random infinite plane tree $\mathcal{K}$ which is the genealogical tree made of two sorts of particles: standard and mutant particles. Initially there is only one mutant particle. All particles reproduce independently of each other, and standard particles produce a random number of standard particles distributed as μ. Mutant particles however, reproduce according to $\overline{\mu}$. Among the descendant of a mutant particle, a uniform child is picked (independently of the past) and is declared "mutant" whereas the other children are standard particles. In particular, there is exactly one mutant at each generation. We do not give the formal definition of this random plane tree and refer to Figure 10.1 for intuition.

Clearly, in the above construction, the tree $\mathcal{K}$ comes with a distinguished infinite ray corresponding to the genealogical line of the mutant particles. This spine can be recovered from $\mathcal{K}$ only in the critical case $m = 1$ as the only infinite line of descent. Let us now describe the law of $\mathcal{K}$:

---

[1]  Harry Kesten 1931-2019

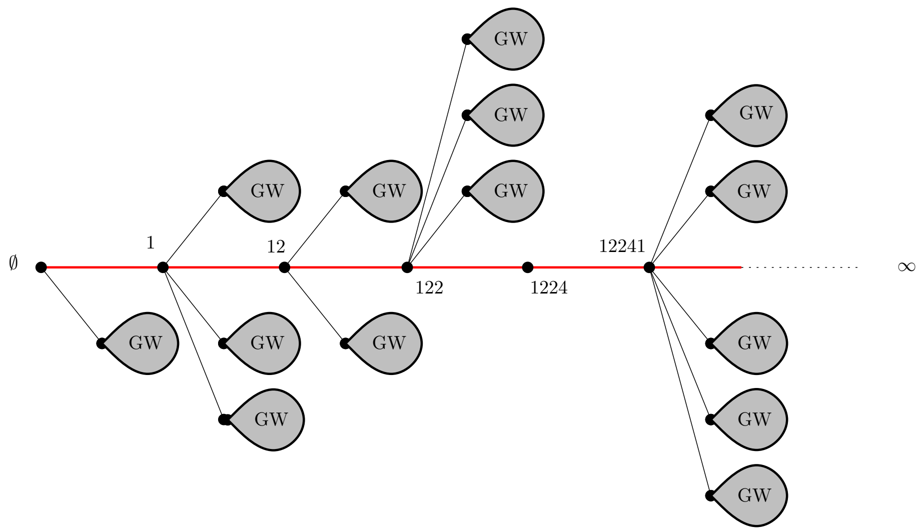


Figure 10.1: The construction of the tree $\mathcal{K}$ from a spine of mutant particles (in red) and μ-Galton–Watson trees produced by the standard children of the mutants.

Proposition 10.2. *For any $n \geqslant 0$ and for any plane tree τ_0 of height exactly n with ℓ vertices at maximal height, we have*

$$\mathbb{P}([\mathcal{K}]_n = \tau_0) = \frac{\ell}{m^n} \cdot \prod_{\substack{u \in \tau_0 \\ |u| \neq n}} \mu_{k_u(\tau_0)}.$$

Proof. Let us split the cases according to which $x \in \tau_0$ is the head of the spine of $\mathcal{K}$ at height n. Once it is fixed, we can compute

$$\mathbb{P}([\mathcal{K}]_n = \tau_0 \text{ and Spine ends at } x) = \prod_{\substack{u \in \tau_0 \\ |u| \neq n \\ u \notin [[\varnothing, x]]}} \mu_{k_u(\tau_0)} \prod_{u \in [[\varnothing, x]] \setminus \{x\}} \overline{\mu}_{k_u(\tau_0)} \frac{1}{k_u(\tau_0)} = \frac{1}{m^n} \prod_{\substack{u \in \tau_0 \\ |u| \neq n}} \mu_{k_u(\tau_0)},$$

where $[[\varnothing, x]]$ is the geodesic line in the tree τ_0 between vertices $\varnothing$ and x. Summing over all possible x at height n yields the result. $\square$

Combining the previous proposition with Proposition 10.1, we see that $\mathcal{T}^{\uparrow} = \mathcal{K}$ in the critical case $m = 1$. In the supercritical case $m > 1$ however, those laws are not identical, but under mild assumptions the random trees $\mathcal{K}$ and $\mathcal{T}^{\uparrow}$ are actually mutually absolutely continuous (see next section).

10.1.3 More in the supercritical case

X Log X criterion Tree of descendants. Case of the geometric distribution

10.2 Local limit of critical Galton–Watson trees

In this section we restrict ourselves to the critical case $m = 1$ and study various conditionings of the μ-Galton–Watson tree which give rise to the infinite random tree $\mathcal{T}^{\uparrow} = \mathcal{K}$ in the limit. Before stating the results, we shall need a little background on local convergence for plane trees.

10.2.1 Trees with one spine

On the set of all locally finite plane trees $\mathbb{T}_{lc}$ we shall consider the following distance (called the local distance):

$$\mathrm{d}_{\mathrm{loc}}(\tau, \tau') = \left(1 + \sup\{n \geqslant 0 : [\tau]_n = [\tau']_n\}\right)^{-1},$$

where we recall that $[\tau]_n$ is the tree made of the first n generations of τ. It is easy to see that $\mathrm{d}_{\mathrm{loc}}$ is a distance and moreover the set of locally finite plane trees becomes Polish:

Proposition 10.3. *The metric space* $(\mathbb{T}_{lc}, \mathrm{d}_{\mathrm{loc}})$ *is Polish, that is, separable and complete.*

Proof. As we said, it is easy to see that $\mathrm{d}_{\mathrm{loc}}$ is a distance. Consider $(\tau_n)_{n\geqslant 0}$ a Cauchy sequence for $\mathrm{d}_{\mathrm{loc}}$. By definition of the local distance for any $m \geqslant 1$, the subtree $[\tau_n]_m$ is eventually constant, denote by θ_m its limit. By coherence we have $[\theta_{n+1}]_n = \theta_n$ and θ_n can be seen of the first n generations of a unique locally finite (possibly infinite) plane tree θ. It is plain that $\tau_n \to \theta$ as $n \to \infty$ for the local distance. Separability in $\mathbb{T}_{lc}$ is easy since finite trees are dense in $\mathbb{T}_{lc}$. □

Exercise 10.1. Find the compact subsets in $(\mathbb{T}_{lc}, \mathrm{d}_{\mathrm{loc}})$.

The above setup enables us to speak about convergence in distribution of random plane trees as random variables taking values in the Polish space $(\mathbb{T}_{lc}, \mathrm{d}_{\mathrm{loc}})$. For our purpose we shall restrict to $\mathbb{T}_f$, the set of all finite plane trees, and to $\mathbb{T}_1$ the set of all infinite plane trees with only one end (i.e. a unique infinite path starting from the origin of the tree). Note that when $m = 1$ we have $\mathcal{K} \in \mathbb{T}_1$ almost surely (hence $\mathcal{T}^{\uparrow} \in \mathbb{T}_1$ a.s.) since all the trees grafted to the spine are a.s. finite. This remark is important since it restricts our state space a lot and thus to check convergence in distribution it is sufficient to check convergence on a much smaller set of events.

If t, s are plane trees and x is a leaf of t (that is a vertex with no child), we denote by $\mathsf{t} \circledast (\mathsf{s}, x)$ the tree obtained by grafting s on the vertex x of t, or formally the set $\{u \in \mathsf{t}\} \cup \{xv : v \in \mathsf{s}\}$. We also introduce the set

$$\mathbb{T}(\mathsf{t}, x) = \{\mathsf{t} \circledast (\mathsf{s}, x) : \mathsf{s} \text{ plane tree}\}.$$

These sets are nice since they generate the Borel σ-field on $\mathbb{T}_f \cup \mathbb{T}_1$.

Proposition 10.4. *Let θ_n for $n \geqslant 1$ and θ be random variables taking values in $\mathbb{T}_f \cup \mathbb{T}_1$ almost surely. For θ_n to converge in distribution for the local distance towards θ, it is sufficient to prove that for any $\mathsf{t} \in \mathbb{T}_f$ and any leaf x of t we have*

$$\lim_{n\to\infty} \mathbb{P}(\theta_n \in \mathbb{T}(\mathsf{t}, x)) = \mathbb{P}(\theta \in \mathbb{T}(\mathsf{t}, x)) \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}(\theta_n = \mathsf{t}) = \mathbb{P}(\theta = \mathsf{t}).$$

Proof. The family of all events of the form $\{\theta \in \mathbb{T}(\mathsf{t}, x)\}$ or $\{\theta = \mathsf{t}\}$ is a π-system and generate the Borel σ-field for the local topology on $\mathbb{T}_f \cup \mathbb{T}_1$. Hence by the monotone class theorem, the knowledge of a probability distribution on this class determines it completely. To see that these events form a convergence determining class, it is sufficient to check that every open set for the local topology on $\mathbb{T}_f \cup \mathbb{T}_1$ is obtained as a finite or countable union of those elements (see [10, Theorem 2.2]). This is easily seen. $\square$

10.2.2 Conditioning a tree to be large

There are different ways to say that a tree is large : either by considering its number of vertices, its height, its number of leaves, or more exotic thoughts. Abraham & Delmas [1] unified these notions as follows. Let A be an integer-valued function which is finite on the set of all finite plane trees and which satisfies an "asymptotic additive property": for any finite tree t with a leaf x, as soon as $A(\mathsf{t} \circledast (\mathsf{s}, x))$ is large enough we have

$$A(\mathsf{t} \circledast (\mathsf{s}, x)) = A(\mathsf{s}) + D(\mathsf{t}, x), \tag{10.1}$$

where $D(\mathsf{t}, x)$ is some function of t and x. Let us right away give examples of such functions:

- $A(\mathsf{t})$ is the size (number of vertices) of t, in this case $D(\mathsf{t}, x) = \#\mathsf{t} - 1$,
- $A(\mathsf{t})$ is the height of t, in this case $D(\mathsf{t}, x) = \mathrm{Height}(x)$,
- $A(\mathsf{t})$ is the number of leaves of t, in this case $D(\mathsf{t}, x) = \#\mathrm{Leaves}(\mathsf{t}) - 1$.

We also denote $\mathcal{A}_n$ the set of all trees in $\mathbb{T}_f \cup \mathbb{T}_1$ such that $A(\mathsf{t}) \in [n, n + n_0)$ where $n_0 \in \{1, 2, 3, \dots\} \cup \{\infty\}$ is fixed. Usually we think of $n_0 = 1$ or $n_0 = \infty$.

Theorem 10.5 (*Limit of large conditionings, critical case $m = 1$*)

Let $\mathcal{T}_n$ be a μ-Galton–Watson tree $\mathcal{T}$ conditioned on the event $\mathcal{T} \in \mathcal{A}_n$ (we restrict our attention to the values of n such that the latter event has positive probability) then as soon as

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathcal{T} \in \mathcal{A}_{n+1})}{\mathbb{P}(\mathcal{T} \in \mathcal{A}_n)} = 1,$$

we have $\mathcal{T}_n \to \mathcal{K}$ in distribution for the local distance as $n \to \infty$.

Proof. Since $\mathcal{K}$ is almost surely infinite and $\mathbb{P}(\mathcal{T}_n = \mathsf{t}) \to 0$ for any finite plane tree t, using Proposition 10.4 it is sufficient to check that for any finite tree t with a leaf x we have

$$\mathbb{P}(\mathcal{T}_n \in \mathbb{T}(\mathsf{t}, x)) \xrightarrow[n\to\infty]{} \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathsf{t}, x)) = \prod_{u \in \mathsf{t} \setminus \{x\}} \mu_{k_u(\mathsf{t})},$$

where the last equality has been shown in the proof of Proposition 10.2. On the event $\mathcal{T} \in \mathbb{T}(\mathsf{t}, x)$ we denote by s the tree grafted on x. Using the assumption (10.1) we made on the "size" function

A we can write for all n large enough

$$\begin{aligned}
\mathbb{P}(\mathcal{T}_n \in \mathbb{T}(\mathsf{t},x)) &= \frac{1}{\mathbb{P}(\mathcal{T} \in \mathcal{A}_n)} \mathbb{P}(\mathcal{T} \in \mathbb{T}(\mathsf{t},x) \text{ and } A(\mathcal{T}) \in [n, n+n_0)) \\
&\underset{\text{for large } n}{=} \frac{1}{\mathbb{P}(\mathcal{T} \in \mathcal{A}_n)} \mathbb{P}(\mathcal{T} \in \mathbb{T}(\mathsf{t},x) \text{ and } \mathsf{s} \in \mathcal{A}_{n-D(\mathsf{t},x)}) \\
&\underset{\text{branching}}{=} \frac{\mathbb{P}(\mathcal{T} \in \mathcal{A}_{n-D(\mathsf{t},x)})}{\mathbb{P}(\mathcal{T} \in \mathcal{A}_n)} \mathbb{P}(\mathcal{T} \in \mathbb{T}(\mathsf{t},x)).
\end{aligned}$$

Since $D(\mathsf{t},x)$ is a fixed number, by our assumption on $\mathbb{P}(\mathcal{T} \in \mathcal{A}_n)$ the fraction in the last display tends to 1 as $n \to \infty$. The second term is easily seen to be equal to $\prod_{u \in \mathsf{t} \setminus \{x\}} \mu_{k_u(\mathsf{t})}$ as desired. $\square$

10.2.3 Applications

Conditioning at large heights We consider the size function $A(\mathsf{t})$ to be the height, i.e. the maximal generation attained by the tree t. Clearly this function satisfies (10.1). So in order to apply the last result, one needs to verify that $\mathbb{P}(T \in \mathcal{A}_{n+1})/\mathbb{P}(T \in \mathcal{A}_n) \to 1$. We first treat the case when $n_0 = \infty$. In this case $\mathcal{T} \in \mathcal{A}_n$ if and only if $\mathcal{T}$ is not extinct at generation n. However, the extinction probability for a Galton–Watson tree is known to obey (see Exercise 9.4)

$$\mathbb{P}(\mathcal{T} \text{ is extinct after } n \text{ generations}) = g^{(n)}(0),$$

where $g(z) = \sum_{k \geqslant 0} z^k \mu_k$ is the generating function of the offspring distribution and $g^{(n)}$ is its n-fold composition. Recall that when $m = 1$ (and $\mu_1 \neq 1$) we have $g^{(n)}(0) \to 1$ (the Galton–Watson tree almost surely dies out). Since $g^{(n)}(0)$ is a sequence defined by iterations of g we have

$$\frac{\mathbb{P}(\mathcal{A}_{n+1})}{\mathbb{P}(\mathcal{A}_n)} = \frac{(1 - g^{(n+1)}(0))}{(1 - g^{(n)}(0))} \to g'(1) = 1.$$

Now let us treat the case $n_0 = 1$ meaning that $T \in \mathcal{A}_n$ is the height of T is exactly n. Using the last arguments in this case we have

$$\begin{aligned}
\frac{\mathbb{P}(\mathcal{T} \in \mathcal{A}_{n+1})}{\mathbb{P}(\mathcal{T} \in \mathcal{A}_n)} &= \frac{\mathbb{P}(\mathrm{Height}(\mathcal{T}) \geqslant n+1) - \mathbb{P}(\mathrm{Height}(\mathcal{T}) \geqslant n+2)}{\mathbb{P}(\mathrm{Height}(\mathcal{T}) \geqslant n) - \mathbb{P}(\mathrm{Height}(\mathcal{T}) \geqslant n+1)} \\
&= \frac{(1 - g^{(n+1)}(0)) - (1 - g^{(n+2)}(0))}{(1 - g^{(n)}(0)) - (1 - g^{(n+1)}(0))} \to 1.
\end{aligned}$$

We can thus apply Theorem 10.5 in both cases $n_0 = 1$ or $n_0 = \infty$ and get that Galton–Watson tree conditioned on having an extinction after a large height or at an exact large height converge towards the infinite Galton–Watson tree conditioned to survive.

Conditioning at large size and the strong ratio limit theorem If the function $A(t)$ is the number of vertices it is a little more difficult to prove the required condition on $\mathbb{P}(T \in \mathcal{A}_n)$ demanded by Theorem 10.5 especially when $n_0 = 1$. This is essentially equivalent to the so-called strong ratio limit theorem (see below) since by Kemperman's formula we have

$$\frac{\mathbb{P}(\mathcal{T} \in \mathcal{A}_{n+1})}{\mathbb{P}(\mathcal{T} \in \mathcal{A}_n)} \underset{\text{Lukasiewicz}}{=} \frac{\mathbb{P}(T^< = n+1)}{\mathbb{P}(T^< = n)} \underset{\text{Kemperman}}{=} \frac{n}{n+1} \frac{\mathbb{P}(S_{n+1} = -1)}{\mathbb{P}(S_n = -1)} \xrightarrow[n \to \infty]{\text{strong ratio}} 1,$$

when μ is aperiodic. To enjoy the strong ratio limit theorem, recall the classical ratio limit theorem for an aperiodic recurrent Markov chain (X) on a countable state space

$$\frac{\sum_{k=1}^{n} \mathbb{P}(X_n = x)}{\sum_{k=1}^{n} \mathbb{P}(X_n = y)} \xrightarrow[n\to\infty]{} \frac{\pi(x)}{\pi(y)},$$

where π is the invariant measure of the chain (unique up to multiplicative constant). The following theorem is a considerable reinforcement of the previous convergence which holds for random walks on $\mathbb{Z}^d$ (not necessarily recurrent). As the reader will see, the proof is short and elegant:

Theorem 10.6 (*Strong ratio limit theorem*)

Let μ be an aperiodic distribution on $\mathbb{Z}^d$. Let $(X_i)_{i\geqslant 0}$ be i.id. random variables with law μ and form the random walk $S_n = \sum_{k=1}^{n} X_i \in \mathbb{Z}^d$. Then for all $m \geqslant 0$ and $b \in \mathbb{Z}^d$ we have

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_{n-m} = s_n - b)}{\mathbb{P}(S_n = s_n)} = 1, \quad \text{as soon as} \quad \mathbb{P}(S_n = s_n)^{1/n} \to 1.$$

Remark 10.1. Notice that $\mathbb{P}(S_n = 0)$ is a super-multiplicative function i.e. $\mathbb{P}(S_{n+m} = 0) \geqslant \mathbb{P}(S_n = 0)\mathbb{P}(S_m = 0)$. By the subadditive theorem (Fekete's lemma) we deduce that $\mathbb{P}(S_n = 0)^{1/n}$ always has a limit which must of course be 1 in the case of a recurrent random walk.

Proof. Using the aperiodicity of μ it is easy to see that it is sufficient to restrict our attention to $m = 1$ and $b \in \mathbb{Z}^d$ such that $\mathbb{P}(X_1 = b) > 0$. For such m, b let us introduce $N_n = \#\{1 \leqslant i \leqslant n : X_i = b\}$ and notice that by symmetry of the increments of the walk we have

$$\mathbb{E}\left[\frac{N_n}{n} \,\middle|\, S_n = s_n\right] \underset{\text{symmetry}}{=} \mathbb{P}(X_1 = b \mid S_n = s_n) \underset{\text{Markov}}{=} \mathbb{P}(X_1 = b) \cdot \frac{\mathbb{P}(S_{n-1} = s_n - b)}{\mathbb{P}(S_n = s_n)}.$$

On the other hand, N_n has binomial $\mathrm{Bin}(\mathbb{P}(X_1 = b), n)$ distribution and in particular an easy large deviation estimate shows that for any $\varepsilon > 0$ we have $\mathbb{P}(|\frac{N_n}{n} - \mathbb{P}(X_1 = b)| \geqslant \varepsilon) \leqslant \mathrm{e}^{-c_\varepsilon n}$ for some $c_\varepsilon > 0$. Writing

$$\begin{aligned}
\left|\mathbb{E}\left[\frac{N_n}{n} \mid S_n = s_n\right] - \mu(b)\right| &\leqslant \varepsilon + \mathbb{P}\Big(\big|\frac{N_n}{n} - \mu(b)\big| \geqslant \varepsilon \;\Big|\; S_n = s_n\Big) \\
&\leqslant \varepsilon + \frac{\mathbb{P}(|N_n/n - \mu(b)| \geqslant \varepsilon)}{\mathbb{P}(S_n = s_n)} \\
&\leqslant \varepsilon + \frac{\mathrm{e}^{-c_\varepsilon n}}{\mathrm{e}^{-o(n)}} \leqslant 2\varepsilon,
\end{aligned}$$

as $n \to \infty$. In other words, $\mathbb{E}\left[\frac{N_n}{n} \mid S_n = s_n\right] \to \mathbb{P}(X_1 = b)$ as $n \to \infty$ and thanks to the first display of the proof the theorem is proved. □

One can give a sufficient condition on the sequence s_n so that $\mathbb{P}(S_n = s_n)^{1/n} \to 1$. Let us recall the classical definition of the log-Laplace function and its Legendre transform, for $x, \theta \in \mathbb{R}^d$ put

$$\phi(\theta) = \log \int \mathrm{d}\mu(x)\ \exp(\theta \cdot x) \in \mathbb{R} \cup \{\infty\} \quad \text{and} \quad \psi(x) = \left(\sup_{\theta \in \mathbb{R}^d} \theta \cdot x - \phi(\theta)\right) \in \mathbb{R}_+.$$

In [3, Section 4] it is shown that as soon as

$$\sup |s_n/n| < \infty \text{ and } \lim_{n\to\infty} \psi(s_n/n) = 0 \quad \text{then} \quad \mathbb{P}(S_n = s_n)^{1/n} \to 1. \tag{10.2}$$

We shall not reproduce the argument here but invite the reader to connect this with the classical large deviations Cramér theory for sums of independent random variables in $\mathbb{R}^d$. The proof of the strong ratio limit theorem can be adapted to many situation when a "large deviation effect" competes with a subexponential conditioning, e.g:

Exercise 10.2. Let μ be a non-trivial aperiodic offspring distribution with mean 1. Denote by $\mathcal{T}_n$ a μ-Galton–Watson tree conditioned to have n vertices (for all n large enough so that the conditioning is non-degenerate). As usual we denote by (S) the associated random walk with i.i.d. increments of law $\mu(k+1)$ for $k \geqslant -1$.

1. Recall why $\frac{1}{n} \log \mathbb{P}(T_1^< = n) \to 0$ as $n \to \infty$.
2. For all $n \geqslant 0$ we denote by $D_n = \#\{0 \leqslant i \leqslant n-1 : S_{i+1} = S_i - 1\}$ the number of down-steps of the walk S up to time n. Show that for all $\varepsilon > 0$, there exists some constant $c_\varepsilon > 0$ such that for all n large enough we have
$$\mathbb{P}(|D_n - \mu(0)n| \geqslant \varepsilon n) \leqslant \mathrm{e}^{-c_\varepsilon n}.$$
3. Conclude that we have
$$\frac{\#\mathrm{Leaves}(\mathcal{T}_n)}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} \mu(0),$$
where a vertex u of $\mathcal{T}_n$ is a leaf if it has no descendant.

Bibliographical notes. Most of the material of this chapter is adapted from the nice lecture notes [2] of Abraham & Delmas, see also [1]. See also [21] for the original work of Kesten on conditioning Galton–Watson trees to survive. The spine decomposition techniques have been popularized by Lyons, Pemantle & Peres in [32] where they give a conceptual proof of the $L \log L$ criterion of Kesten and Stigum, see also [33, Chapter 17]. The proof of the strong ratio limit theorem follows the argument of Neveu [36, 1], whereas the result is initially due to Kesten.

Chapter XI: Erdös-Rényi random graph

The goal of this chapter is to use the tools on random walks and Galton–Watson trees in order to study the phase transition for the component sizes in the famous model of random graph due to Erdös[1] and Rényi[2]:

Definition 11.1 ($\mathcal{G}(n,p)$ model)**.** *The Erdös–Rényi random graph $\mathcal{G}(n,p)$ with parameters $n \geqslant 1$ and $p \in [0,1]$ is the (distribution of a) random graph whose vertex set is $V = \{1, 2, \ldots, n\}$ and where for each pair $i \neq j \in V$ the edge $i \leftrightarrow j$ is present with probability p independently of all the other pairs.*

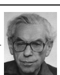
Notice that contrary to the setup of Part III, the random graph $\mathcal{G}(n,p)$ may be disconnected. We will see that the geometry of $\mathcal{G}(n,p)$ undergoes various phase transitions depending on the parameters n, p.

11.1 Sharp threshold transitions

11.1.1 History

The "random graph" model was introduced by Erdös and Rényi in 1959 who wanted to probe randomly a graph with n (labeled) vertices. Actually Definition 11.1 is not really due to Erdös and Rényi who consider a fixed number $m \leqslant \binom{n}{2}$ of edges but rather to Edgar Gilbert. This random graph model has since then become ubiquitous in probability and commonly referred to as the "*mean field model*". This means that the initial geometry of the model is trivial: one could permute all the vertices and get the same model. This is due to the fact that the isometry group of the underlying graph, the complete graph on n vertices $\mathbb{K}_n$ is the full permutation group Σ_n. For the connoisseurs, the $\mathcal{G}(n,p)$ model can also be seen as a Bernoulli bond percolation model on $\mathbb{K}_n$.

Exercise 11.1 (Infinite Erdös–Rényi random graph, Rado graph). $(*)$ Consider the random graph G whose vertex set is $\mathbb{N}$ and where independently for each $i, j \in \mathbb{N}$ the edge $i \leftrightarrow j$ is present in the graph with probability $1/2$. Show that almost surely two samples of this random graph are homomorphic.

[1]  Paul Erdös 1913-1996

[2]  Alfréd Rényi 1921-1970

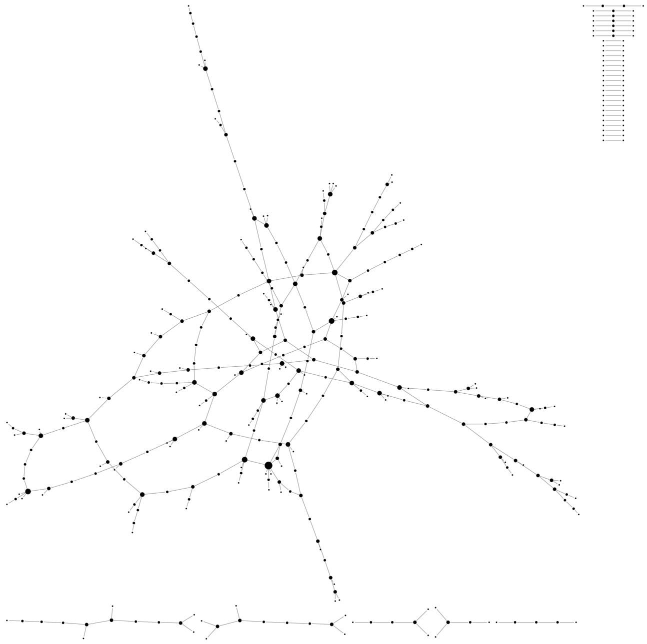


Figure 11.1: A sample of $\mathcal{G}(n,p)$ when $n = 500$ and $p = \frac{2}{n}$, in the so-called supercritical regime. The dot representing the vertices have radii proportional to their degrees. We clearly see that there is a unique "giant" component in the graph and all the other components are very tiny.

Beyond its apparent simplicity the $\mathcal{G}(n,p)$ model is connected to many interesting and present-day topics in probability theory: the Brownian tree of Aldous, mixing time for transpositions dynamics, expander graphs... The rough structure of the graph $\mathcal{G}(n,p)$ is quite well understood. When $p = 0$ obviously the graph is totally disconnected. As far as vertex degrees are concerned, the situation is quite trivial by the Binomial-Poisson approximation theorem:

Proposition 11.1. *If $n \cdot p(n) \to \lambda \in [0,\infty]$ then the vertex degree of the node 1 converges in distribution towards a Poisson random variable with mean λ.*

In the following, if (A_n) is an event referring to $\mathcal{G}(n,p_n)$ we write A_n w.h.p. "with high probability" if $\mathbb{P}(\mathcal{G}(n,p_n) \in A_n) \to 1$ as $n \to \infty$. When we are in presence of events A_n for which their apparition in $\mathcal{G}(n,p)$ depends on p in a drastic way we speak of *sharp threshold phenomena:*

Definition 11.2 (Sharp thresholds for graph properties)**.** *Let A_n be an increasing graph property (i.e. adding edges to a graph only helps verifying A_n) for graphs on n vertices. We say that A_n has a* sharp threshold *at $p \equiv p_n$ if for every $\varepsilon > 0$ we have*

$$\mathbb{P}(\mathcal{G}(n,(1-\varepsilon)p_n) \in A_n) \to 0 \qquad \textit{whereas} \qquad \mathbb{P}(\mathcal{G}(n,(1+\varepsilon)p_n) \in A_n) \to 1 \qquad \textit{as } n \to \infty.$$

11.1.2 Sharp threshold for connectivity at $\frac{\log n}{n}$

The first natural question is about the overall connectedness of the graph, and this property turns out has a sharp threshold ruled by the presence or not of isolated vertices.

Theorem 11.2 (*Erdös–Rényi*)

Connectedness and existence of isolated vertices have both a sharp threshold at $p_n = \frac{\log n}{n}$.

Proof. We start by studying isolated vertices using the method of first and second moment. Since the degree of any single vertex in $\mathcal{G}(n, p_n)$ follows a $\mathrm{Bin}(n-1, p_n)$ distribution, the first moment method shows that

$$\mathbb{P}(\mathcal{G}(n,p_n) \text{ has an isolated vertex}) \leqslant \mathbb{E}[\# \text{ isolated vertices}] = n\mathbb{P}(\mathrm{Bin}(n-1,p_n)=0) = n(1-p_n)^{n-1}.$$

If $p_n \geqslant (1+\varepsilon)\frac{\log n}{n}$ then the right-hand size clearly tends to 0 as $n \to \infty$ and this shows that $\mathcal{G}(n,p_n)$ has no isolated vertices in this regime. If now $p_n \leqslant (1-\varepsilon)\frac{\log n}{n}$ we deduce from the last display that $\mathbb{E}[\# \text{ isolated vertices}] \to \infty$ but this is not sufficient to guarantee that $\mathbb{P}(\mathcal{G}(n,p_n) \text{ has an isolated vertex}) \to 1$. To do this, we use the second moment method:

Lemma 11.3 (Second moment method)**.** *Let $X \in \{0,1,2,\dots\}$ be an non-negative integer valued random variable. Then we have*

$$\mathbb{P}(X \geqslant 1) \geqslant \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

One-line proof: Use Cauchy-Schwarz $\mathbb{E}[X]^2 = \mathbb{E}[X\mathbf{1}_{X>0}]^2 \leqslant \mathbb{E}[X^2]\mathbb{P}(X>0)$. □

Using this method with $X = \#$ isolated vertices, we have $\mathbb{E}[X] = n(1-p_n)^{n-1}$ and

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{1 \leqslant i,j \leqslant n} \mathbb{E}[i \text{ and } j \text{ are isolated}] \\
&= n\mathbb{P}(1 \text{ is isolated}) + n(n-1)\mathbb{P}(1 \text{ and } 2 \text{ are isolated}) \\
&= n(1-p_n)^{n-1} + n(n-1)(1-p_n)^{2n-3}.
\end{aligned}$$

In the regime $p_n \leqslant (1-\varepsilon)\frac{\log n}{n}$ it is easy to see that $\mathbb{E}[X]^2 \sim \mathbb{E}[X^2]$ and so there are isolated vertices with high probability. To complete the proof of the theorem, one has to show that as long as there are no isolated vertices the graph is actually connected with high probability. Indeed, if there are no isolated vertices, but if the graph is not connected, that means that there is a connected component in $\mathcal{G}(n,p_n)$ whose size is between 2 and $n/2$. The number of ways to cut the graph into two disconnected parts is then

$$\sum_{k=1}^{[n/2]} \binom{n}{k} \left((1-p_n)^{n-k}\right)^k.$$

Taking $p_n = (1+\varepsilon)\frac{\log n}{n}$ with $\varepsilon \in (0,1)$, it is then an easy but tedious exercise to prove that the above series is asymptotically dominated by its first term and so converges to 0 as $n \to \infty$.This finishes the proof of the theorem. □

A much more refined analysis show that the number of isolated vertices in $\mathcal{G}(n, \frac{\log n+c}{n})$ for $c \in \mathbb{R}$ converges in distribution towards a Poisson variable of parameter e^{-c} and similarly as above the presence of isolated vertices is the main obstacle to connectedness so that we have the nice formula

$$\mathbb{P}\left(\mathcal{G}\left(n, \frac{\log n + c}{n}\right) \text{ is connected}\right) \xrightarrow[n\to\infty]{} \mathrm{e}^{-\mathrm{e}^{-c}}.$$

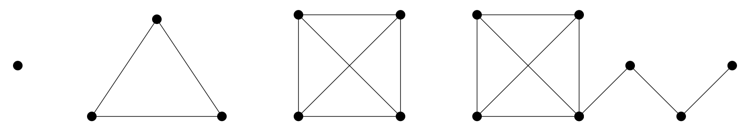
Exercise 11.2. How would you attack such a result?

Exercise 11.3. Show a sharp threshold phenomenon for the event {having diameter $\leqslant 2$} at

$$p_n = \sqrt{\frac{2\log n}{n}}.$$

As seen in the previous theorem, the global property of connectedness is actually ruled by a local property, namely the existence of isolated vertices. One can also focus on other "local observables" such existence of triangles (i.e. triplets $\{i, j, k\}$ connected to each other and not to other vertices of the graph) or more generally of a given graph K as a subgraph. These properties does not necessarily exhibit sharp thresholds:

Exercise 11.4 (threshold function of existence of subgraph)*.* Find the (sharp?) threshold functions for existence of a fixed subgraph K inside $\mathcal{G}(n,p)$ for K ranging in :



11.1.3 Sharp threshold for the giant

Actually, if one does not demand that the graph is fully connected but only requires that a large connected component exists then one obtains another sharp threshold transition at a much smaller parameter. More precisely we will denote by $|\mathcal{C}_1(n)|, |\mathcal{C}_2(n)|, \dots$ the sizes of the connected components of $\mathcal{G}(n, p_n)$ in decreasing order. The main theorem is the following:

Theorem 11.4 (*Double phase transition, Erdös & Rényi, Aldous*)

There is a sharp threshold transition for the existence of a giant connected component at $p_n = \frac{1}{n}$. More precisely, if $p = p_n = \frac{c}{n}$ then we have

- ***Subcritical:** If $\mathbf{c} < \mathbf{1}$ then there exists $A > 0$ depending on $c > 0$ such that w.h.p. we have $|\mathcal{C}_1(n)| \leqslant A \log n$.*

- ***Supercritical:** If $\mathbf{c} > \mathbf{1}$ then there exists $A > 0$ such that w.h.p. we have $|\mathcal{C}_2(n)| \leqslant A \log n$ whereas $|\mathcal{C}_1(n)| \sim (1 - \alpha(c)) \cdot n$ where $\alpha(c)$ is the solution in $(0, 1)$ of the equation*

$$\alpha(c) = \mathrm{e}^{-c(1-\alpha(c))}.$$

- ***Critical:*** ***If*** $\mathbf{c = 1}$ *then the vector* $(n^{-2/3}|\mathcal{C}_i(n)|)_{i \geqslant 1}$ *converges in law in the finite dimensional sense towards a positive infinite vector.*

The goal of the next sections is to prove the above result. We will actually only prove points (i) and (ii) and just indicate how to recover the scaling in point (iii).

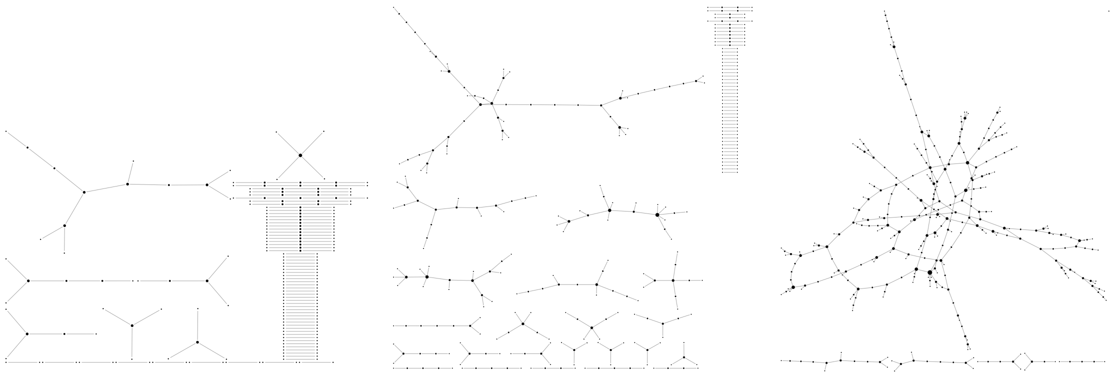


Figure 11.2: Illustration of the structure of components in $\mathcal{G}(n,p)$ for $n = 500$ in the subcritical regime, critical and supercritical regimes from left to right. Notice in particular that in the supercritical regime there is a unique giant component whereas in the other regime the first and second largest components are "comparable".

11.2 Exploration process and large deviation inequality

In this section we explain the main idea below Theorem 11.4 which consists in remarking that "locally" the $\mathcal{G}(n, c/n)$ model resembles a Poisson(c)-Galton–Watson trees and can thus be analyzed via random walk with Poisson(c)-1 increments.

11.2.1 Exploration process

Fix a given vertex, say 1, of $\mathcal{G}(n,p)$. We will explore the connected component of 1 using the following algorithm. At any given time $t \geqslant 0$ there are three type of vertices: the untouched vertices $\mathcal{U}_t$, the vertices totally *explored* $\mathcal{E}_t$ and the vertices in the current stack $\mathcal{S}_t$ whose neighborhoods remain to be explored. The algorithm evolves as follows:

- at time $t = 0$ we have $\mathcal{E} = \varnothing$, the untouched vertices are $\mathcal{U}_0 = \{2, 3, \dots\}$ and the only vertex in the stack is 1.

- suppose $t \geqslant 0$ is given and such that $\mathcal{S}_t \neq \varnothing$. We then select the vertex $x \in \mathcal{S}_t$ with minimal label and reveal all the neighbors $y_1, \dots, y_k$ of x among $\mathcal{U}_t$ (this could be an empty set!). We then put

$$\mathcal{U}_{t+1} = \mathcal{U}_t \backslash \{y_1, \dots, y_k\}, \quad \mathcal{S}_{t+1} = \big(\mathcal{S}_t \backslash \{x\}\big) \cup \{y_1, \dots, y_k\}, \quad \mathcal{E}_{t+1} = \mathcal{E}_t \cup \{x\}.$$

- if at time $t \geqslant 0$ we have $\mathcal{S}_t = \varnothing$ then the algorithm stops and outputs $\mathcal{E}_\infty := \mathcal{E}_t$ and $\mathcal{U}_\infty = \mathcal{U}_t$.

It should be clear from the above exploration that deterministically the set $\mathcal{E}_\infty$ is precisely the connected component of 1 in $\mathcal{G}(n,p)$. In the following we write $\mathbb{S}_t = \#\mathcal{S}_t - 1$, we also denote by $\tau = \tau(\mathbb{S}) = \inf\{t \geqslant 0 : \mathcal{S}_t = \varnothing\} = \inf\{t \geqslant 0 : \mathbb{S}_t = -1\}$ and notice that $\#\mathcal{E}_\infty = \tau$. When this exploration is performed on $\mathcal{G}(n,p)$ the process

$$(\#\mathcal{U}_t, \#\mathcal{S}_t, \#\mathcal{E}_t)_{t\geqslant 0}$$

is actually a Markov chain whose transition probabilities are described by the following proposition.

Proposition 11.5. *Let $\mathcal{F}_t$ be the σ-field generated by the first $t \geqslant 0$ steps of the exploration process of 1 in $\mathcal{G}(n,p)$. Conditionally on $\mathcal{F}_t$ and on $\mathcal{S}_t \neq \varnothing$ the edges between vertices of $\mathcal{S}_t$ and $\mathcal{U}_t$ or between two vertices of $\mathcal{U}_t$ are all independent and present with probability p. In particular, conditionally on $\mathcal{F}_t$ the increment $\Delta\mathbb{S}_t = \mathbb{S}_{t+1} - \mathbb{S}_t$ is distributed as*

$$\Delta\mathbb{S}_t \overset{(d)}{=} \mathrm{Bin}(\#\mathcal{U}_t, p) - 1.$$

Proof. Notice that given the status of the edges and vertices revealed by time t, one could deterministically change the status of all the edges between $\mathcal{S}_t$ and $\mathcal{U}_t$ or in-between vertices of $\mathcal{U}_t$ and this would not have affected the exploration up to time t (because these edges have not been explored by the algorithm). It is easy to see from this that those edges are indeed i.i.d. present with probability p. The last point of the theorem follows easily. An alternative, and more "algorithmic" way to see this is to imagine that all the edges of the graph $\mathcal{G}(n,p)$ carry a question mark "?" which means that its status is currently unknown, present with probability p and absent with probability $1-p$. When performing the exploration of the cluster of the vertex 1 we reveal the status of certain edges. The key point is to notice that since we are not allowed to use the randomness of unrevealed edges, at any (stopping) time t, conditionally on the past exploration, the edges which still carry "?" are independent and open/closed with probability $p/1-p$. □

The key idea is then the following: When n is large we can approximate the increments of $\mathbb{S}_t$ by independent Poisson$(c)-1$ random variables, at least at the beginning of the exploration when $\#\mathcal{U}_t \approx n$. When $c < 1$ the mean increment of the exploration process is negative and so it finishes quickly. However, when $c > 1$, there is a positive probability $\alpha(c)$, computed via Proposition 4.13, such that the "ideal" exploration process continues forever. In the finite world of $\mathcal{G}(n, c/n)$ this translates into the fact that the vertex from which we started the exploration belongs to the infinite giant component. To turn these heuristics into real proofs the main technical tool is a large deviations inequality for the walk associate to the ideal model: Let $(S^{(n)})$ be a random walk $(S^{(n)}_t : t \geqslant 0)$ starting from 0 with i.i.d. increments of law $\mathrm{Bin}(n, c/n) - 1$. Notice that $(S^{(n)})$ is skip-free descending.

Lemma 11.6 (Large deviations)**.** *Fix $c > 0$ and consider the random walk $S^{(n)}$ having* $\mathrm{Bin}(n, c/n) - 1$ *increments. For any $\varepsilon > 0$ there exists $\eta > 0$ such that for all $n \geqslant 1$ and all $k \geqslant 0$*

$$\mathbb{P}(|S_k^{(n)} - (c-1)k| \geqslant \varepsilon k) \leqslant \exp(-\eta \cdot k).$$

Proof. Notice that $c - 1$ is the mean of the increment of $S^{(n)}$: This is an example of a large deviation inequality where we show that the probability that a sum of i.i.d. random variables deviates linearly from its mean decays exponentially fast. A subtlety here, however, is that we seek for an estimate which is uniform in n (involved in the law of the increments). We proceed as usual with an exponential Markov inequality. We start with the upper bound

$$\mathbb{P}(S_k^{(n)} - (c-1)k \geqslant \varepsilon k) = \mathbb{P}(\mathrm{e}^{\theta(S_k^{(n)} - k(c-1))} \geqslant \mathrm{e}^{\theta \varepsilon k}) \leqslant \left(\mathbb{E}\big[\mathrm{e}^{\theta(S_1^{(n)} - (c-1))}\big]\mathrm{e}^{-\theta\varepsilon}\right)^k,$$

for any $\theta > 0$. The Laplace transform of $S_1^{(n)} - (c-1)$ can be evaluated explicitly and we have

$$\begin{aligned}\mathbb{E}\big[\mathrm{e}^{\theta(S_1^{(n)} - (c-1))}\big] &= \mathrm{e}^{-\theta c} \sum_{k=0}^{n} \mathrm{e}^{k\theta} \binom{n}{k} \left(\frac{c}{n}\right)^k \left(1 - \frac{c}{n}\right)^{n-k} = \mathrm{e}^{-\theta c}\left(1 + (\mathrm{e}^{\theta} - 1)\frac{c}{n}\right)^n \\ &\leqslant \exp(c(\mathrm{e}^{\theta} - 1 - \theta)) \leqslant \exp(c\theta^2),\end{aligned}$$

for all $\theta \in (0, 1)$. Multiplying by $\mathrm{e}^{-\theta\varepsilon}$, we can thus choose $\theta_0 \in (0, 1)$ small enough so that

$$\mathbb{E}\big[\mathrm{e}^{\theta_0(S_1^{(n)} - (c-1))}\big]\mathrm{e}^{-\theta_0\varepsilon} < 1$$

for all $n \geqslant 1$ and then the above exponential Markov inequality for $\theta = \theta_0$ yields that $\mathbb{P}(S_k^{(n)} - (c-1)k \geqslant \varepsilon k)$ decays exponentially (uniformly in $n \geqslant 1$) as $k \to \infty$ as desired. The lower bound is similar except that we should use negative θ's. □

11.2.2 Upper bounds: subcritical and critical cases

Subcritical case. Suppose first that $c < 1$. We start the exploration of a given vertex in $\mathcal{G}(n, c/n)$ and use the above notation. Since we always have $\#\mathcal{U}_t \leqslant n$ we deduce that the increments of $\mathbb{S}_t$ are stochastically dominated by independent $\mathrm{Bin}(n, c/n)$ variables and so that the size of the connected component of any given vertex is stochastically dominated by $\tau^{(n)} - 1$ where $\tau^{(n)}$ is the hitting time of -1 by the random walk $(S_t^{(n)} : t \geqslant 0)$ starting from 0 with i.i.d. increments of law $\mathrm{Bin}(n, c/n) - 1$. Since $c < 1$ using the above lemma we have for all $A \geqslant 1$

$$\mathbb{P}(\tau^{(n)} > A) \leqslant \mathbb{P}(S_A^{(n)} \geqslant 0) \underset{\text{Lem. 11.6}}{\leqslant} \exp(-\eta A),$$

for some $\eta > 0$. Taking $A = \frac{2}{\eta}\log n$ we deduce using the union bound that

$$\mathbb{P}(|\mathcal{C}_1(n)| \geqslant \frac{2}{\eta}\log n) \leqslant n\mathbb{P}(\tau^{(n)} \geqslant \frac{2}{\eta}\log n) \leqslant n\exp(-2\log n) \to 0.$$

□

Critical case. The same strategy can be used in the critical case $c = 1$ (together with a little size-biasing trick). In this case the hitting time of 0 by $S^{(n)}$ can be evaluate via Kemperman's formula Proposition 4.9 and an explicit computation:

$$\begin{aligned}\mathbb{P}(\tau^{(n)} = k) &= \frac{1}{k}\mathbb{P}(\mathrm{Bin}(k \cdot n, c/n) = k-1)\\ &= \frac{1}{k}\binom{nk}{k-1}\left(\frac{1}{n}\right)^{k-1}\left(1-\frac{1}{n}\right)^{nk-(k-1)}\\ &\leqslant \frac{1}{k}\frac{k^{k-1}}{(k-1)!}\mathrm{e}^{-(k-1)} \leqslant Ck^{-3/2},\end{aligned}$$

for some universal constant $C > 0$ as long as n is large enough. In particular $\mathbb{P}(\tau^{(n)} > k)$ decays at least as $k^{-1/2}$ for k large. Now pick (independently of $\mathcal{G}(n, 1/n)$) a vertex U_n uniformly in $\{1, 2, \ldots, n\}$. Since the size of the cluster of U_n is stochastically dominated by $\tau^{(n)}$ we have

$$\mathbb{P}(\tau^{(n)} \geqslant A) \geqslant \mathbb{P}(\#\mathrm{Cluster}(U_n) \geqslant A) \geqslant \mathbb{P}(|\mathcal{C}_1(n)| \geqslant A \text{ and } U_n \in \mathcal{C}_1(n)) \geqslant \frac{A}{n}\mathbb{P}(|\mathcal{C}_1(n)| \geqslant A).$$

Now taking $A = \lambda n^{2/3}$ and using the above asymptotic for $\mathbb{P}(\tau \geqslant A)$ yields

$$\mathbb{P}(|\mathcal{C}_1(n)| \geqslant \lambda n^{2/3}) = O(\lambda^{-1/3}),$$

which already gives the good order of magnitude of the largest cluster in $\mathcal{G}(n, 1/n)$. Getting the full distributional convergence of $(n^{-2/3}|\mathcal{C}_i(n)|)_{i\geqslant 1}$ requires to understand in much more details the exploration process and this exits the scope of these lecture notes.

11.3 Supercritical case

We suppose now that $c > 1$. The random walk $(S^{(n)})$ starting from 0 and whose increments are given by independent $\mathrm{Bin}(n, c/n) - 1$ variables then drifts towards $+\infty$. According to Proposition 4.13, the probability $1 - \alpha_n(c)$ that $S^{(n)}$ stays non-negative for all time is solution to

$$\alpha_n(c) = \sum_{k=-1}^{\infty} (\alpha_n(c))^{k+1}\mathbb{P}(\mathrm{Bin}(n, c/n) - 1 = k).$$

Since $\mathrm{Bin}(n, c/n) \to \mathrm{Poi}(c)$ it is easy to check that $\alpha_n(c)$ converges towards $\alpha(c)$ which is solution to

$$\alpha(c) = \sum_{k=-1}^{\infty} (\alpha(c))^{k+1}\mathbb{P}(\mathrm{Poi}(c) = k+1) = \mathrm{e}^{-c(1-\alpha(c))},$$

so that $1 - \alpha(c)$ is the probability that a random walk starting from 0 with i.i.d. Poisson(c)-1 increments stays non-negative for ever. We recover here the value $\alpha(c)$ appearing in Theorem 11.4. Notice in passing that $\alpha(c)$ is continuous in c. For the supercritical case, the fact that we have explored vertices decreases the "super criticality" along the exploration and so we need some care. However, since $c > 1$, as long as $\#\mathcal{U}_t \geqslant n-m$ the increments of $\#\mathcal{E}_t$ are stochastically dominated by $\mathrm{Bin}(n-m, \frac{c}{n})$ whose mean is larger than 1, at least when $m = o(n)$. We will thus explore the cluster of a given vertex in $\mathcal{G}(n, c/n)$ but we freeze the exploration either at its death time $\theta^{(n)}$ or after $n^{2/3}$ exploration steps (which we assume to be integer to simplify our notation).

Lemma 11.7. *For any $\varepsilon > 0$, for all n large enough we can couple (i.e. realizing on the same probability space)*

- *the exploration process $(\mathbb{S}_t)_{0 \leqslant t \leqslant \tau}$,*
- *a random walk $(S_t^{(n)})_{t \geqslant 0}$ starting from 0 with i.i.d. increments of law $\mathrm{Bin}(n, c/n) - 1$,*
- *a random walk $(\tilde{S}_t^{(n)})_{t \geqslant 0}$ starting from 0 with i.i.d. increments of law $\mathrm{Bin}(n(1-\varepsilon), c/n) - 1$,*

such that with probability at least $1 - o(1/n)$ we have

$$\tilde{S}_t^{(n)} \leqslant \mathbb{S}_t \leqslant S_t^{(n)}, \qquad \textit{for every } 0 \leqslant t \leqslant \tau \wedge n^{2/3}.$$

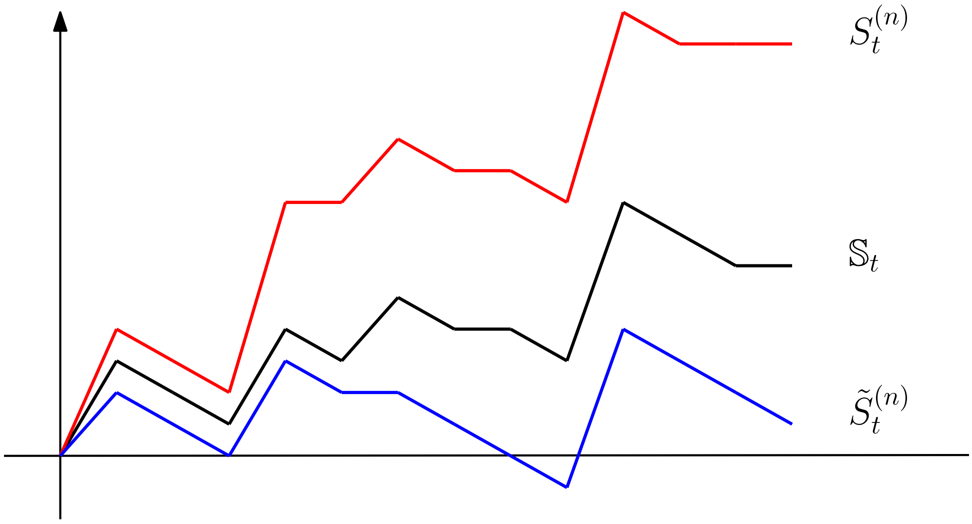


Figure 11.3: Illustration of the sandwich coupling of the exploration in-between two random walks.

Proof. The upper coupling of $\mathbb{S}$ and $S^{(n)}$ is already used in the last section and follows from the fact that the increments of $\mathbb{S}$ are always stochastically dominated by those of $S^{(n)}$. In passing, we can use the large deviations inequalities of Lemma 11.6 to deduce that the probability that $S^{(n)}$, and a fortiori $\mathbb{S}$, exceeds the level εn before time $n^{2/3}$ is of probability $\exp(-\delta n^{2/3}) = o(1/n)$. On this event, we can use the remark made before the lemma and deduce that the increments of $\mathbb{S}$ dominates stochastically those of $\tilde{S}^{(n)}$. This yields the desired coupling. □

Corollary 11.8. *Fix $c > 1$. There exist $A, \eta > 0$, such that during the exploration of the cluster of a given vertex in $\mathcal{G}(n, c/n)$, with a probability $1 - o(1/n)$ we are in one of the following two alternatives:*

- *Either the exploration process stops before time $A \log n$, and thus the size of the cluster is less than $A \log n$,*

- *Or the exploration process runs for at least $n^{2/3}$ steps and the current stack contains at least $\eta n^{2/3}$ vertices to be explored.*

Asymptotically, the probability of the second alternative is $1 - \alpha(c)$.

Proof. We use the coupling of the preceding lemma with $\varepsilon > 0$ small enough so that $(1-\varepsilon)c > 1 + \frac{c-1}{2} > 1$. On the event on which the size of the cluster τ, which is equal to the death time of the exploration, is equal to $k \leqslant n^{2/3}$ we must have $\tilde{S}^{(n)}_k \leqslant -1 \leqslant 0$ and so

$$\mathbb{P}(A \log n \leqslant \tau \leqslant n^{2/3}) \leqslant \sum_{k=A\log n}^{n^{2/3}} \mathbb{P}(\tilde{S}^{(n)}_k \leqslant 0).$$

Since we assume that the mean drift $(1-\varepsilon)c - 1$ of $\tilde{S}^{(n)}$ is positive, we can then apply the large deviation estimate of Lemma 11.6 to see that the right-hand side of the last display is bounded above by $n^{2/3} \mathrm{e}^{-\delta A \log n}$ for some $\delta > 0$. Picking $A > 0$ large enough we can make the last probability $o(1/n)$ establishing the desired dichotomy. The fact that on the second case $\mathbb{S}_{n^{2/3}} \geqslant \tilde{S}^{(n)}_{n^{2/3}} \geqslant \frac{c-1}{2} n^{2/3}$ with probability $1 - o(1/n)$ again follows from large deviations inequalities on $\tilde{S}^{(n)}_{n^{2/3}}$ and by our choice of ε.
Using the coupling of $\mathbb{S}, S^{(n)}$ and $\tilde{S}^{(n)}$ we see that the probability to be in the second case of the exploration implies that $S^{(n)}$ is non-negative until time $n^{2/3}$ and is implied by the fact that $\tilde{S}^{(n)}$ is non-negative until time $n^{2/3}$. It is easy to see that both these probabilities are very close when $n \to \infty$ to the probabilities that the respective walks with increments $\mathrm{Bin}(n, c/n) - 1$ and $\mathrm{Bin}((1-\varepsilon)n, c/n)$ are non-negative for all time. Those probabilities converge towards $1 - \alpha(c)$ and $1 - \alpha((1-\varepsilon)c)$. By continuity of $c \mapsto \alpha(c)$ for $c > 1$ we conclude that the probability of the last event is indeed asymptotic to $1 - \alpha(c)$. □

We can now complete the proof of Theorem 11.4 (iii):

Using the last corollary and the union bound we deduce that, w.h.p., for any $x \in \{1, 2, \ldots, n\}$ the cluster of x in $\mathcal{G}(n, c/n)$ either contains less than $A \log n$ vertices, or its exploration using the algorithm can be run for $n^{2/3}$ yielding to at least $\frac{c-1}{2} n^{2/3}$ vertices in the stack to be explored. We will show that all the vertices in the second case actually belong to the *same* cluster in the graph, that is

$$|\mathcal{C}_1(n)| > \frac{c-1}{2} n^{2/3} \text{ and } |\mathcal{C}_2(n)| < \frac{c-1}{2} n^{2/3} \quad \text{w.h.p.} \tag{11.1}$$

Indeed, take two vertices x_1, x_2 and assume that both exploration processes reach a stack of at least $\frac{c-1}{2} n^{2/3}$ vertices. Either during the first steps of these explorations we discovered a vertex in common to the cluster of x_1 and x_2 and so they belong to the same cluster, otherwise the stacks S_1 and S_2 of $\approx n^{2/3}$ vertices related to x_1 and x_2 are disjoint. In this case, the Markovian property of the exploration shows that any edge between any pair of vertices in S_1 and S_2 is present with probability c/n. The probability that there are no edge between S_1 and S_2 is then at most

$$(1-c/n)^{\left(\frac{c-1}{2}\right)^2 n^{4/3}} = o(n^{-2}).$$

Performing a union bound over the n^2 choices for x_1 and x_2, we get the desired claim (11.1).

It remains to show that the size of $|\mathcal{C}_1(n)|$ or alternatively of the essentially unique cluster of size larger than $n^{2/3}$ is of order $(1-\alpha(c)) \cdot n$ with high probability. We already know from the second point Lemma 11.7 that a typical point $x \in \{1, 2, \dots, n\}$ has a cluster of size larger than $n^{2/3}$ with probability tending to $1-\alpha(c)$ which implies that

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathbf{1}_{\#\mathrm{Cluster}(i) \geqslant n^{2/3}}\right] \sim (1-\alpha(c)) \cdot n, \qquad \text{as } n \to \infty.$$

To get concentration around the mean value, by Markov's inequality (it is again a second moment method) it suffices to show that

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \mathbf{1}_{\#\mathrm{Cluster}(i) \geqslant n^{2/3}}\right)^2\right] \sim \left(\mathbb{E}\left[\sum_{i=1}^{n} \mathbf{1}_{\#\mathrm{Cluster}(i) \geqslant n^{2/3}}\right]\right)^2,$$

or equivalently that the probability that two points $i, j \in \{1, 2, \dots, n\}$ both have a cluster of size $\geqslant n^{2/3}$ tends to $(1-\alpha(c))^2$ as $n \to \infty$. To see this, first perform the exploration of the cluster of i. As above, there is a chance of order $\alpha(c)$ that the exploration stops at $\tau \leqslant A \log n$ steps. Then either the vertex j is in the explored component, which has a probability less than $\frac{A \log n}{n}$ or it belongs to the remaining graph with conditionally on the first exploration has law $\mathcal{G}(n-\tau, \frac{c}{n})$. In the last case, conditionally on $\tau \leqslant A \log n$ it is easy to see that j belongs to a cluster of size $\geqslant n^{2/3}$ with probability asymptotic to $1-\alpha(c)$. This proves the claim and finishes the proof of the theorem. Of course, some details are left to the reader at that point. □

Bibliographical notes. The Erdös–Rényi model is probably the simplest and the most studied random graph. It is a wonderful playground for combinatorics and probability. It has many variations and descendants such as the stochastic block model, the rank 1 model, the configuration model which are more realistic models for real-life networks. The literature on this topic is vast and we simply refer to the beautiful monograph [?] for a detailed account and to [?] for very recent results on the scaling limit of components at criticality.

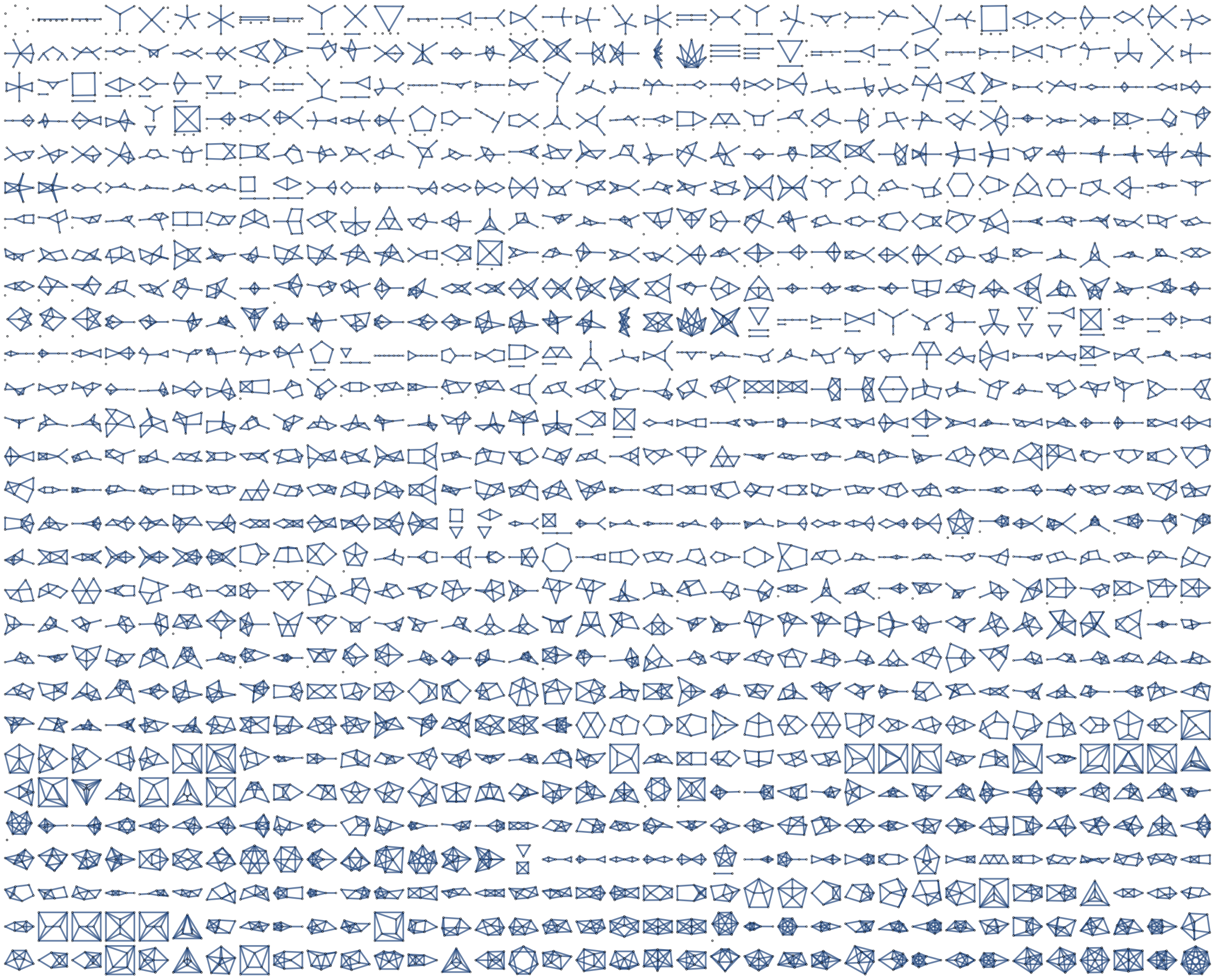


Figure 11.4: A list of the 1044 simple graph on 7 vertices.

Bibliography

[1] R. Abraham and J.-F. Delmas, *Local limits of conditioned Galton-Watson trees: the infinite spine case*, Electron. J. Probab., 19 (2014), pp. 1–19.

[2] R. Abraham and J.-F. Delmas, *An introduction to Galton-Watson trees and their local limits*, arxiv:1506.05571, (2015).

[3] R. Abraham, J.-F. Delmas, and H. Guo, *Critical multi-type Galton–Watson trees conditioned to be large*, Journal of Theoretical Probability, 31 (2018), pp. 757–788.

[4] L. Addario-Berry and B. A. Reed, *Ballot theorems, old and new*, in Horizons of combinatorics, vol. 17 of Bolyai Soc. Math. Stud., Springer, Berlin, 2008, pp. 9–35.

[5] O. Angel, G. Chapuy, N. Curien, and G. Ray, *The local limit of unicellular maps in high genus*, Electron. Commun. Probab., 18 (2013), pp. no. 86, 8.

[6] R. Arratia, L. Goldstein, and F. Kochman, *Size bias for one and all*, arXiv preprint arXiv:1308.2729, (2013).

[7] M. T. Barlow, *Random walks and heat kernels on graphs*, vol. 438 of London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 2017.

[8] I. Benjamini and A. Shamov, *Bi-lipschitz bijections of z*, Analysis and Geometry in Metric Spaces, 3 (2015).

[9] J. Bertoin and R. A. Doney, *On conditioning a random walk to stay nonnegative*, Ann. Probab., 22 (1994), pp. 2152–2167.

[10] P. Billingsley, *Convergence of probability measures, first edition*, Wiley, New-York, (1968).

[11] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, *The electrical resistance of a graph captures its commute and cover times*, Comput. Complexity, 6 (1996/97), pp. 312–340.

[12] L. Chaumont and R. Doney, *On distributions determined by their upward, space–time wiener–hopf factor*, Journal of Theoretical Probability, (2017), pp. 1–23.

[13] K. L. Chung, *A course in probability theory*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, second ed., 1974. Probability and Mathematical Statistics, Vol. 21.

[14] R. A. Doney, *The Martin boundary and ratio limit theorems for killed random walks*, J. London Math. Soc. (2), 58 (1998), pp. 761–768.

[15] R. A. Doney, *Fluctuation Theory for Levy Processes: Ecole D'Eté de Probabilités de Saint-Flour XXXV-2005*, Springer, 2007.

[16] P. G. Doyle and J. L. Snell, *Random walks and electric networks*, vol. 22 of Carus Mathematical Monographs, Mathematical Association of America, Washington, DC, 1984.

[17] W. Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[18] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[19] O. Kallenberg, *Foundations of Modern Probability*, Springer, New York, second ed., 2002.

[20] J. Kennedy, *Understanding the wiener–hopf factorization for the simple random walk*, Journal of Applied Probability, 31 (1994), pp. 561–563.

[21] H. Kesten, *Subdiffusive behavior of random walk on a random cluster*, Ann. Inst. H. Poincaré Probab. Statist., 22 (1986), pp. 425–487.

[22] I. Kortchemski, *Arbres et marches aléatoires*, Journées X-UPS, (2016).

[23] M. Kwaśnicki, *Random walks are determined by their trace on the positive half-line*, arXiv preprint arXiv:1902.08481, (2019).

[24] A. E. Kyprianou, *Wiener–hopf decomposition*, Encyclopedia of Quantitative Finance, (2010).

[25] S. Lalley, *Lecture notes on 1d random walk and renewal theory*, http://www.stat.uchicago.edu/ lalley/Courses/index.html.

[26] ——, *One-dimensional random walks (lecture notes)*, http://galton.uchicago.edu/ lalley/Courses/312/RW.pdf.

[27] G. F. Lawler and V. Limic, *Random walk: a modern introduction*, vol. 123 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2010.

[28] J.-F. Le Gall, *Marches aléatoires, mouvement brownien et processus de branchement*, in Séminaire de Probabilités, XXIII, vol. 1372 of Lecture Notes in Math., Springer, Berlin, 1989, pp. 258–274.

[29] ——, *Random trees and applications*, Probability Surveys, (2005).

[30] ——, *Random real trees*, Ann. Fac. Sci. Toulouse Math. (6), 15 (2006), pp. 35–62.

[31] Y. Le Jan, *Markov Paths, Loops and Fields: École D'Été de Probabilités de Saint-Flour XXXVIII–2008*, vol. 2026, Springer Science & Business Media, 2011.

[32] R. Lyons, R. Pemantle, and Y. Peres, *Conceptual proofs of l log l criteria for mean behavior of branching processes*, Ann. Probab., 23 (1995), pp. 1125–1138.

[33] R. Lyons and Y. Peres, *Probability on Trees and Networks*, Current version available at http://mypage.iu.edu/ rdlyons/, In preparation.

[34] G. Markowsky, *An electric-resistance approach to return time*, (2016).

[35] P. Mörters and Y. Peres, *Brownian motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[36] J. Neveu, *Sur le théorème ergodique de Chung–Erdös*, C. R. Acad. Sci. Paris Sér. I Math., 257 (1963), pp. 2953–2955.

[37] ——, *Arbres et processus de Galton-Watson*, Ann. Inst. H. Poincaré Probab. Statist., 22 (1986), pp. 199–207.

[38] L. Shepp, *Recurrent random walks with arbitrarily large steps*, Bulletin of the American Mathematical Society, 70 (1964), pp. 540–542.

[39] F. Spitzer, *Principles of random walk*, Springer-Verlag, New York-Heidelberg, second ed., 1976. Graduate Texts in Mathematics, Vol. 34.

[40] H. Tanaka, *Time reversal of random walks in one-dimension*, Tokyo J. Math., 12 (1989), pp. 159–174.

[41] W. Woess, *Random walks on infinite graphs and groups*, vol. 138 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2000.

[42] V. M. Zolotarev, *One-dimensional Stable Distributions*, vol. 65, American Mathematical Society, translations of mathematical monographs ed., 1986.