Cambridge University Press
978-1-108-47245-6 — Two-Dimensional Random Walk
Serguei Popov
Excerpt
More Information

1

Introduction

The main subject of this book is simple random walk (also abbreviated as SRW) on the integer lattice $\mathbb{Z}^d$ and we will pay special attention to the case $d = 2$. SRW is a discrete-time stochastic process which is defined in the following way: if at a given time the walker is at $x \in \mathbb{Z}^d$, then at the next time moment it will be at one of x's $2d$ neighbours chosen uniformly at random.[1] In other words, the probability that the walk follows a fixed length-n path of nearest-neighbour sites equals $(2d)^{-n}$. As a general fact, a random walk may be recurrent (i.e., almost surely it returns infinitely many times to its starting location) or transient (i.e., with positive probability it never returns to its starting location). A fundamental result about SRWs on integer lattices is Pólya's classical theorem [76]:

Theorem 1.1. *Simple random walk in dimension d is recurrent for* $d = 1, 2$ *and transient for* $d \geq 3$.

A well-known interpretation of this fact, attributed to Shizuo Kakutani, is: "a drunken man always returns home, but a drunken bird will eventually be lost". This observation may explain why birds do not drink vodka. Still, despite recurrence, the drunken man's life is not so easy either: as we will see, it may take him *quite* some time to return home.

Indeed, as we will see in (3.42), the probability that two-dimensional SRW gets more than distance n away from its starting position without revisiting it is approximately $(1.0293737 + \frac{2}{\pi} \ln n)^{-1}$ (and this formula becomes *very* precise as n grows). While this probability indeed converges to zero as $n \to \infty$, it is important to notice how slow this convergence is. To present a couple of concrete examples, assume that the size of the walker's step is equal to 1 metre. First of all, let us go to one of the most beautiful cities in the world, Paris, and start walking from its centre. The

[1] Here, the author had to resist the temptation of putting a picture of an SRW's trajectory in view of the huge number of animated versions easily available in the Internet, e.g., at `https://en.wikipedia.org/wiki/Random_walk`.

radius of Paris is around 5000m, and $(1.0293737 + \frac{2}{\pi}\ln 5000)^{-1}$ is approximately 0.155; that is, in roughly one occasion out of seven you would come to the Boulevard Périphérique before returning to your starting location. The next example is a bit more extreme: let us do the same walk on the *galactic plane* of our galaxy. (Yes, when one starts in the centre of our galaxy, there is a risk that the starting location could happen to be too close to a massive black hole;[2] we restrict ourselves to purely mathematical aspects of the preceding question, though.) The radius of the Milky Way galaxy is around 10^{21}m, and $(1.0293737 + \frac{2}{\pi}\ln 10^{21})^{-1} \approx 0.031$, which is *surprisingly* large. Indeed, this means that the walker[3] would revisit the origin only around 30 times on average, before leaving the galaxy; this is not something one would normally expect from a *recurrent* process.

Incidentally, these sorts of facts explain why it is difficult to verify conjectures about two-dimensional SRW using computer simulations. (For example, imagine that one needs to estimate how long we will wait until the walk returns to the origin, say, a hundred times.)

As we will see in Section 2.1, the recurrence of d-dimensional SRW is related to the divergence of the series $\sum_{n=1}^{\infty} n^{-d/2}$. Notice that this series diverges if and only if $d \le 2$, and for $d = 2$ it is the harmonic series that diverges quite slowly. This might explain why the two-dimensional case is, in some sense, *really* critical (and therefore gives rise to the previous "strange" examples). It is always interesting to study critical cases – they frequently exhibit behaviours not observable away from criticality. For this reason, in this book we dedicate more attention to dimension two than to other dimensions: two-dimensional SRW is a fascinating mathematical object indeed and this already justifies one's interest in exploring its properties (and also permits the author to keep this introduction short).

The next section is intentionally kept concise, since it is not really intended for *reading* but rather for occasional use as a reference.

1.1 Markov chains and martingales: basic definitions and facts

First, let us recall some basic definitions related to real-valued stochastic processes in discrete time. In the following, all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We write $\mathbb{E}$ for expectation corresponding to $\mathbb{P}$, which will be applied to real-valued random variables. Set $\mathbb{N} = \{1, 2, 3, \dots\}$, $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, $\overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{+\infty\}$.

[2] `https://en.wikipedia.org/wiki/Sagittarius_A*`.
[3] Given the circumstances, let me not say "you" here.

Definition 1.2 (Basic concepts for discrete-time stochastic processes)**.**

- A discrete-time real-valued *stochastic process* is a sequence of random variables $X_n : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ indexed by $n \in \mathbb{Z}_+$, where $\mathcal{B}$ is the Borel σ-field. We write such sequences as $(X_n, n \geq 0)$, with the understanding that the time index n is always an integer.
- A *filtration* is a sequence of σ-fields $(\mathcal{F}_n, n \geq 0)$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all $n \geq 0$. Let us also define $\mathcal{F}_\infty := \sigma\big(\bigcup_{n\geq 0} \mathcal{F}_n\big) \subset \mathcal{F}$.
- A stochastic process $(X_n, n \geq 0)$ is *adapted* to a filtration $(\mathcal{F}_n, n \geq 0)$ if X_n is $\mathcal{F}_n$-measurable for all $n \in \mathbb{Z}_+$.
- For a (possibly infinite) random variable $\tau \in \overline{\mathbb{Z}}_+$, the random variable X_τ is (as the notation suggests) equal to X_n on $\{\tau = n\}$ for finite $n \in \mathbb{Z}_+$ and equal to $X_\infty := \limsup_{n\to\infty} X_n$ on $\{\tau = \infty\}$.
- A (possibly infinite) random variable $\tau \in \overline{\mathbb{Z}}_+$ is a *stopping time* with respect to a filtration $(\mathcal{F}_n, n \geq 0)$ if $\{\tau = n\} \in \mathcal{F}_n$ for all $n \geq 0$.
- If τ is a stopping time, the corresponding σ-field $\mathcal{F}_\tau$ consists of all events $A \in \mathcal{F}_\infty$ such that $A \cap \{\tau \leq n\} \in \mathcal{F}_n$ for all $n \in \overline{\mathbb{Z}}_+$. Note that $\mathcal{F}_\tau \subset \mathcal{F}_\infty$; events in $\mathcal{F}_\tau$ include $\{\tau = \infty\}$, as well as $\{X_\tau \in B\}$ for all $B \in \mathcal{B}$.
- For $A \in \mathcal{B}$, let us define

$$\tau_A = \min\{n \geq 0 : X_n \in A\}, \tag{1.1}$$

and

$$\tau_A^+ = \min\{n \geq 1 : X_n \in A\}; \tag{1.2}$$

we may refer to either τ_A or τ_A^+ as the *hitting time* of A (also called the *passage time* into A). It is straightforward to check that both τ_A and τ_A^+ are stopping times.

Observe that, for any stochastic process $(X_n, n \geq 0)$, it is possible to define the minimal filtration to which this process is adapted via $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$. This is the so-called *natural filtration*.

To keep the notation concise, we will frequently write X_n and $\mathcal{F}_n$ instead of $(X_n, n \geq 0)$ and $(\mathcal{F}_n, n \geq 0)$ and so on, when no confusion will arise.

Next, we need to recall some martingale-related definitions and facts.

Definition 1.3 (Martingales, submartingales, supermartingales)**.** A real-valued stochastic process X_n adapted to a filtration $\mathcal{F}_n$ is a *martingale* (with respect to $\mathcal{F}_n$) if, for all $n \geq 0$,

(i) $\mathbb{E}|X_n| < \infty$, and
(ii) $\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = 0$.

If in (ii) "=" is replaced by "≥" (respectively, "≤"), then X_n is called a *submartingale* (respectively, *supermartingale*). If the filtration is not specified, that means that the natural filtration is used.

Clearly, if X_n is a submartingale, then $(-X_n)$ is a supermartingale, and vice versa; a martingale is both a submartingale and a supermartingale. Also, it is elementary to observe that if X_n is a (sub-, super-)martingale, then so is $X_{n\wedge\tau}$ for any stopping time τ.

Martingales have a number of remarkable properties, which we will not even try to elaborate on here. Let us only cite the paper [75], whose title speaks for itself. In the following, we mention only the results needed in this book.

We start with

Theorem 1.4 (Martingale convergence theorem)**.** *Assume that X_n is a submartingale such that $\sup_n \mathbb{E}X_n^+ < \infty$. Then there is an integrable random variable X such that $X_n \to X$ a.s. as $n \to \infty$.*

Observe that, under the hypotheses of Theorem 1.4, the sequence $\mathbb{E}X_n$ is non-decreasing (by the submartingale property) and bounded above by $\sup_n \mathbb{E}[X_n^+]$; so $\lim_{n\to\infty} \mathbb{E}X_n$ exists and is finite. However, it is not necessarily equal to $\mathbb{E}X$.

Using Theorem 1.4 and Fatou's lemma, it is straightforward to obtain the following result.

Theorem 1.5 (Convergence of non-negative supermartingales)**.** *Assume that $X_n \geq 0$ is a supermartingale. Then there is an integrable random variable X such that $X_n \to X$ a.s. as $n \to \infty$, and $\mathbb{E}X \leq \mathbb{E}X_0$.*

Another fundamental result that we will use frequently is the following:

Theorem 1.6 (Optional stopping theorem)**.** *Suppose that $\sigma \leq \tau$ are stopping times, and $X_{\tau\wedge n}$ is a uniformly integrable submartingale. Then $\mathbb{E}X_\sigma \leq \mathbb{E}X_\tau < \infty$ and $X_\sigma \leq \mathbb{E}[X_\tau \mid \mathcal{F}_\sigma]$ a.s.*

Note that, if X_n is a uniformly integrable submartingale and τ is any stopping time, then it can be shown that $X_{\tau\wedge n}$ is also uniformly integrable: see, e.g., section 5.7 of [44]. Also, observe that two applications of Theorem 1.6, one with $\sigma = 0$ and one with $\tau = \infty$, show that for any uniformly integrable submartingale X_n and any stopping time τ, it holds that $\mathbb{E}X_0 \leq \mathbb{E}X_\tau \leq \mathbb{E}X_\infty < \infty$, where $X_\infty := \limsup_{n\to\infty} X_n = \lim_{n\to\infty} X_n$ exists and is integrable, by Theorem 1.4.

Theorem 1.6 has the following corollary, obtained by considering $\sigma = 0$

and using well-known sufficient conditions for uniform integrability (e.g., sections 4.5 and 4.7 of [44]).

Corollary 1.7. *Let X_n be a submartingale and τ a finite stopping time. For a constant $c > 0$, suppose that at least one of the following conditions holds:*

(i) $\tau \le c$ *a.s.;*
(ii) $|X_{n\wedge\tau}| \le c$ *a.s. for all* $n \ge 0$*;*
(iii) $\mathbb{E}\tau < \infty$ *and* $\mathbb{E}[|X_{n+1} - X_n| \mid \mathcal{F}_n] \le c$ *a.s. for all* $n \ge 0$.

Then $\mathbb{E}X_\tau \ge \mathbb{E}X_0$. *If* X_n *is a martingale and at least one of the conditions (i) through (iii) holds, then* $\mathbb{E}X_\tau = \mathbb{E}X_0$.

Next, we recall some fundamental definitions and facts for Markov processes in discrete time and with countable state space, also known as *countable Markov chains*. In the following, $(X_n, n \ge 0)$ is a sequence of random variables taking values on a countable set Σ.

Definition 1.8 (Markov chains)**.**

- A process X_n is a *Markov chain* if, for any $y \in \Sigma$, any $n \ge 0$, and any $m \ge 1$,

$$\mathbb{P}[X_{n+m} = y \mid X_0, \dots, X_n] = \mathbb{P}[X_{n+m} = y \mid X_n], \text{ a.s..} \tag{1.3}$$

  This is the *Markov property*.
- If there is no dependence on n in (1.3), the Markov chain is *homogeneous in time* (or *time homogeneous*). Unless explicitly stated otherwise, all Markov chains considered in this book are assumed to be time homogeneous. In this case, the Markov property (1.3) becomes

$$\mathbb{P}[X_{n+m} = y \mid \mathcal{F}_n] = p_m(X_n, y), \text{ a.s.,} \tag{1.4}$$

  where $p_m : \Sigma \times \Sigma \to [0, 1]$ are the m-step Markov *transition probabilities*, for which the Chapman–Kolmogorov equation holds: $p_{n+m}(x, y) = \sum_{z\in\Sigma} p_n(x, z)p_m(z, y)$. Also, we write $p(x, y) := \mathbb{P}[X_1 = y \mid X_0 = x] = p_1(x, y)$ for the one-step transition probabilities of the Markov chain.
- We use the shorthand notation $\mathbb{P}_x[\,\cdot\,] = \mathbb{P}[\,\cdot \mid X_0 = x]$ and $\mathbb{E}_x[\,\cdot\,] = \mathbb{E}[\,\cdot \mid X_0 = x]$ for probability and expectation for the time homogeneous Markov chain starting from initial state $x \in \Sigma$.
- A time homogeneous, countable Markov chain is *irreducible* if for all $x, y \in \Sigma$ there exists $n_0 = n_0(x, y) \ge 1$ such that $p_{n_0}(x, y) > 0$.

- For an irreducible Markov chain, we define its *period* as the greatest common divisor of $\{n \geq 1 : p_n(x,x) > 0\}$ (it is not difficult to show that it does not depend on the choice of $x \in \Sigma$). An irreducible Markov chain with period 1 is called *aperiodic*.
- Let X_n be a Markov chain and τ be a stopping time with respect to the natural filtration of X_n. Then for all $x, y_1, \ldots, y_k \in \Sigma$, $n_1, \ldots, n_k \geq 1$, it holds that

$$\mathbb{P}[X_{\tau+n_j} = y_j, j = 1, \ldots, k \mid \mathcal{F}_\tau, X_\tau = x] = \mathbb{P}_x[X_{\tau+n_j} = y_j, j = 1, \ldots, k]$$

  (this is the *strong Markov property*).
- For a Markov chain, a probability measure $(\pi(x), x \in \Sigma)$ is called an *invariant measure* if $\sum_{x\in\Sigma} \pi(x)p(x,y) = \pi(y)$ for all $y \in \Sigma$. It then holds that $\mathbb{P}_\pi[X_n = y] = \pi(y)$ for all n and y (where $\mathbb{P}_\pi$ means that the initial state of the process is chosen according to π).

Suppose now that X_n is a countable Markov chain. Recall the definitions of hitting times τ_A and τ_A^+ from (1.1)–(1.2). For $x \in \Sigma$, we use the notation $\tau_x^+ := \tau_{\{x\}}^+$ and $\tau_x := \tau_{\{x\}}$ for hitting times of one-point sets. Note that for any $x \in A$ it holds that $\mathbb{P}_x[\tau_A = 0] = 1$, while $\tau_A^+ \geq 1$ is then the *return time* to A. Also note that $\mathbb{P}_x[\tau_A = \tau_A^+] = 1$ for all $x \in \Sigma \setminus A$.

Definition 1.9. For a countable Markov chain X_n, a state $x \in \Sigma$ is called

- *recurrent* if $\mathbb{P}_x[\tau_x^+ < \infty] = 1$;
- *transient* if $\mathbb{P}_x[\tau_x^+ < \infty] < 1$.

A recurrent state x is classified further as

- *positive recurrent* if $\mathbb{E}_x\tau_x^+ < \infty$;
- *null recurrent* if $\mathbb{E}_x\tau_x^+ = \infty$.

It is straightforward to see that the four properties in Definition 1.9 are *class properties*, which entails the following statement.

Proposition 1.10. *For an irreducible Markov chain, if a state $x \in \Sigma$ is recurrent (respectively, positive recurrent, null recurrent, transient), then all states in Σ are recurrent (respectively, positive recurrent, null recurrent, transient).*

By the preceding fact, it is legitimate to call an irreducible Markov chain itself recurrent (positive recurrent, null recurrent, transient).

Next, the following proposition is an easy consequence of the strong Markov property.

Proposition 1.11. *For an irreducible Markov chain, if a state $x \in \Sigma$ is recurrent (respectively, transient), then, regardless of the initial position of the process, it will be visited infinitely (respectively, finitely) many times almost surely.*

Finally, let us state the following simple result which sometimes helps in proving recurrence or transience of Markov chains.

Lemma 1.12. *Let X_n be an irreducible Markov chain on a countable state space Σ.*

(i) If for some $x \in \Sigma$ and some nonempty $A \subset \Sigma$ it holds that $\mathbb{P}_x[\tau_A < \infty] < 1$, then X_n is transient.
(ii) If for some finite nonempty $A \subset \Sigma$ and all $x \in \Sigma \setminus A$ it holds that $\mathbb{P}_x[\tau_A < \infty] = 1$, then X_n is recurrent.

(For the proof, cf. e.g. lemma 2.5.1 of [71].)

2

Recurrence of two-dimensional simple random walk

This chapter is mainly devoted to the proof of the recurrence part of Theorem 1.1 (although we still discuss the transience in higher dimensions later in the exercises). We first present a direct "path-counting" proof, and then discuss the well-known correspondence between reversible Markov chains and electrical networks, which also yields a beautiful proof of recurrence of SRW in dimensions one and two. Then, we go for a side-quest: we do a basic exploration of the Lyapunov function method, a powerful tool for proving recurrence or transience of general Markov chains. With this method, we add yet another proof of recurrence of two-dimensional SRW to our collection.

2.1 Classical proof

In this section, we present the classical combinatorial proof of recurrence of two-dimensional simple random walk.

Let us start with some general observations on recurrence and transience of random walks, which, in fact, are valid in a much broader setting. Namely, we will prove that the number of visits to the origin is a.s. finite if and only if the *expected* number of visits to the origin is finite (note that this is something which is *not* true for general random variables). This is a useful fact, because, as it frequently happens, it is easier to control the expectation than the random variable itself.

Let $p_m(x, y) = \mathbb{P}_x[S_m = y]$ be the transition probability from x to y in m steps for the simple random walk in d dimensions. Let $q_d = \mathbb{P}_0[\tau_0^+ < \infty]$ be the probability that, starting at the origin, the walk eventually returns to the origin. If $q_d < 1$, then the total number of visits (counting the initial instance $S_0 = 0$ as a visit) is a geometric random variable with success probability $1 - q_d$, which has expectation $(1 - q_d)^{-1} < \infty$. If $q_d = 1$, then, clearly, the walk visits the origin infinitely many times a.s.. So, random walk is transient (i.e., $q_d < 1$) if and only if the *expected* number of visits to the origin is finite. This expected number equals[1]

$$\mathbb{E}_0 \sum_{k=0}^{\infty} \mathbf{1}\{S_k = 0\} = \sum_{k=0}^{\infty} \mathbb{E}_0 \mathbf{1}\{S_k = 0\} = \sum_{n=0}^{\infty} \mathbb{P}_0[S_{2n} = 0]$$

(observe that the walk can be at the starting point only after an even number of steps). We thus obtain that the recurrence of the walk is *equivalent* to

$$\sum_{n=0}^{\infty} p_{2n}(0,0) = \infty. \tag{2.1}$$

Before actually proving anything, let us try to *understand* why Theorem 1.1 should hold. One can represent the d-dimensional simple random walk S as

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

where $(X_k, k \geq 1)$ are independent and identically distributed (i.i.d.) random vectors, uniformly distributed on the set $\{\pm e_j, j = 1, \ldots, d\}$, where $e_1, \ldots, e_d$ is the canonical basis of $\mathbb{R}^d$. Since these random vectors are centred (expectation is equal to 0, component-wise), one can apply the (multivariate) Central Limit Theorem (CLT) to obtain that $S_n/\sqrt{n}$ converges in distribution to a (multivariate) centred Normal random vector with a diagonal covariance matrix. That is, it is reasonable to expect that S_n should be at distance of order $\sqrt{n}$ from the origin.

So, what about $p_{2n}(0,0)$? Well, if $x, y \in \mathbb{Z}^d$ are two *even* sites[2] at distance of order at most $\sqrt{n}$ from the origin, then our CLT intuition tell us that $p_{2n}(0,x)$ and $p_{2n}(0,y)$ should be *comparable*, i.e., their ratio should be bounded away from 0 and ∞. In fact, this statement can be made rigorous by using the *local* Central Limit Theorem (e.g., theorem 2.1.1 from [63]). Now, if there are $O(n^{d/2})$ sites where $p_{2n}(0,\cdot)$ are comparable, then the value of these probabilities (including $p_{2n}(0,0)$) should be of order $n^{-d/2}$. It remains only to observe that $\sum_{n=1}^{\infty} n^{-d/2}$ diverges only for $d = 1$ and 2 to convince oneself that Pólya's theorem indeed holds. Notice, by the way, that for $d = 2$ we have the harmonic series which diverges just barely; its partial sums have only logarithmic growth.[3]

Now, let us *prove* that (2.1) holds for one- and two-dimensional simple random walks. In the one-dimensional case, it is quite simple to calculate $p_{2n}(0,0)$: it is the probability that a Binomial$(2n, \frac{1}{2})$-random variable

[1] Note that we can put the expectation inside the sum because of the Monotone Convergence Theorem.
[2] A site is called even if the sum of its coordinates is even; observe that the origin is even.
[3] As some physicists say, "in practice, logarithm is a constant!"

equals 0, so it is $2^{-2n}\binom{2n}{n}$. Certainly, this expression is concise and beautiful; it is, however, not a priori clear which asymptotic behaviour it has (as it frequently happens with concise and beautiful formulas). To clarify this, we use Stirling's approximation[4], $n! = \sqrt{2\pi n}(n/e)^n(1+o(1))$, to obtain that

$$\begin{aligned}
2^{-2n}\binom{2n}{n} &= 2^{-2n}\frac{(2n)!}{(n!)^2}\\
&= 2^{-2n}\frac{\sqrt{4\pi n}(2n/e)^{2n}}{2\pi n(n/e)^{2n}}(1+o(1))\\
&\qquad \text{(fortunately, almost everything cancels)}\\
&= \frac{1}{\sqrt{\pi n}}(1+o(1)). \qquad (2.2)
\end{aligned}$$

The series $\sum_{k=1}^{\infty} k^{-1/2}$ diverges, so (2.1) holds, and this implies recurrence in dimension 1.

Let us now deal with the two-dimensional case. For this, we first count the number of paths N_{2n} of length $2n$ that start and end at the origin. For such a path, the number of steps up must be equal to the number of steps down, and the number of steps to the right must be equal to the number of steps to the left. The total number of steps up (and, also, down) can be any integer k between 0 and n; in this case, the trajectory must have $n-k$ steps to the left and $n-k$ steps to the right. So, if the number of steps up is k, the total number of trajectories starting and ending at the origin is the polynomial coefficient $\binom{2n}{k,k,n-k,n-k}$. This means that

$$\mathsf{N}_{2n} = \sum_{k=0}^{n}\binom{2n}{k,k,n-k,n-k} = \sum_{k=0}^{n}\frac{(2n)!}{(k!)^2((n-k)!)^2}.$$

Note that

$$\frac{(2n)!}{(k!)^2((n-k)!)^2} = \binom{2n}{n}\binom{n}{k}\binom{n}{n-k};$$

the last two factors are clearly equal, but in a few lines it will become clear why we have chosen to write it this way. Since the probability of any particular trajectory of length m is 4^{-m}, we have

$$\begin{aligned}
p_{2n}(0,0) &= 4^{-2n}\mathsf{N}_{2n}\\
&= 4^{-2n}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k}. \qquad (2.3)
\end{aligned}$$

[4] See, e.g., http://mathworld.wolfram.com/StirlingsApproximation.html.