



Universiteit
Leiden
The Netherlands

Simple Random Walk

Leenman, T.

Citation

Leenman, T. (2008). *Simple Random Walk*.

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository](#)

Downloaded from: <https://hdl.handle.net/1887/3596856>

Note: To cite this publication please use the final published version (if applicable).

Simple Random Walk

Timo Leenman, 0427004

June 3, 2008

Bachelor scription
Supervisor: Prof. dr. F den Hollander

Contents

1	Definition of the random walk	3
2	Recurrence of the random walk	3
3	Range of the random walk	10
4	Probability measures and stochastic convergence	15
5	Brownian motion	18

Preface

This treatise is on simple random walk, and on the way it gives rise to Brownian motion. It was written as my bachelor project, and it was written in such a way that it should serve as a good introduction into the subject for students that have as much knowledge as I when I began working on it. That is: a basic probability course, and a little bit of measure theory. To that end, the following track is followed:

In section 1, the simple random walk is defined.

In section 2, the first major limit property is studied: whether the walk be recurrent or not. Some calculus and the discrete Fourier transform are required to prove the result.

In section 3, a second limit property is studied: its range, or, the number of visited sites. In the full proof of the results, the notion of strong and weak convergence presents itself, and also the notion of tail events.

To understand these problems more precisely, and as a necessary preparation for Brownian motion, some measure theoretic foundations are treated in section 4. Emphasis is put, not on the formal derivation of the results, but on the right notion of them in our context.

In section 5, Brownian motion is studied. First, in what manner simple random walk gives rise to it, and secondly its formal definition. Special care is devoted to explain the exact steps that are needed for its construction, for that is something which I found rather difficult to understand from the texts I read on it.

Timo Leenman

1 Definition of the random walk

Random walk describes the motion on a lattice, say \mathbb{Z}^d , of a walker that jumps at discrete time steps $t = 1, 2, 3, \dots$ to a new, randomly chosen, site. Such a random walk can be defined in various ways, resulting in various properties. With each time a random variable is associated (the step), and the distribution of these random variables fixes the behaviour of the walk. One could for example define a walk that never visits a site twice, or one that never turns 180 degrees at once. But the random walk we are to look at here is *simple random walk*. Its successive steps are chosen independently, they can be of length 1 only and are chosen uniformly out of the $2d$ possible directions on \mathbb{Z}^d . Formally, define steps X_i as random variables on \mathbb{Z}^d as follows:

Definition. The discrete random variables X_1, X_2, \dots on \mathbb{Z}^d are called steps of the random walk and have the following probability distribution: $\forall i \in \mathbb{N} : P(X_i = e) = \frac{1}{2d}$ if $e \in \mathbb{Z}^d$ and $\|e\| = 1$, and $P(X_i = e) = 0$ otherwise.

Definition. $S_0 = 0 \in \mathbb{Z}^d$ and $S_n = X_1 + \dots + X_n$ for $n \in \mathbb{N}$ is called the position of the random walk at time n .

So the random walker begins his walk on the starting site $S_0 = 0$, and by taking i.i.d. steps X_i arrives at position S_n at time n . Because of the independence of the steps it is obvious that $\{S_n\}_{n \in \mathbb{N}_0}$ is a Markov process on the state space \mathbb{Z}^d , since the future positions only depend on the current position. Now that we have a Markov chain, we can talk about transition probabilities. Our notation shall be thus: call $p(l) = P(X_1 = l) = P(S_1 = l)$ the one-step transition probability, which is $\frac{1}{2d}$ for l a “neighbour”, and 0 otherwise. Call $P_n(l) = P(S_n = l)$ the n -step transition probability, which equals the probability that the walk is at site l at time n (starting in 0).

2 Recurrence of the random walk

In studying the long term behaviour of the random walk, one of the first questions one might be interested in is whether the random walker returns to its starting site. To this end, define F to be the probability that the random walker eventually returns to the starting site S_0 . If $F = 1$, then the site S_0 is called *recurrent*, if $F < 1$, then it is called *transient*. In the recurrent case, it is obvious that the random walker returns not only once but infinitely many times to S_0 , whereas in the transient case, the random walker may never return with positive probability $1 - F$. In which latter case the number of returns to S_0 is geometrically distributed with parameter $1 - F$, and therefore the expected number of returns to a transient starting site is $\frac{1}{1-F} - 1 = \frac{F}{1-F}$.

The state space of a general Markov chain can be partitioned into recurrent and transient classes of states. In the simple walk, however, it is clear that all states communicate (i.e. the walker has positive probability to reach any given site starting from any other given site), and hence that it consists of only one class. Therefore it makes sense to call the whole random walk recurrent or transient, whenever S_0 is so.

The certainty of returning to every visited site on the one hand, and the likelihood of not returning to them on the other hand, give recurrent and transient random walks completely different behaviour. A valuable result, due to Polya, tells which simple random walks are recurrent:

Polya's Theorem. Simple random walks of dimension $d = 1, 2$ are recurrent, and of $d \geq 3$ are transient.

The proof of this theorem is the object of this section. To that end, we first need a general criterion to see whether S_0 be recurrent or not.

Theorem. The state $S_0 = 0$ is transient if and only if $\sum_{n=1}^{\infty} P_n(0) < \infty$.

Proof. For $t \in \mathbb{N}$, define $I_t = 1$ if $S_t = 0$, and $I_t = 0$ otherwise. Note that $N = \sum_{t=1}^{\infty} I_t$ is the number of times that S_0 is revisited. For the expectation of that number the following holds:

$$E[N] = E\left[\sum_{t=1}^{\infty} I_t\right] = \sum_{t=1}^{\infty} E[I_t] = \sum_{t=1}^{\infty} P(S_t = 0) = \sum_{t=1}^{\infty} P_t(0).$$

Computing the expectation again, we have

$$\begin{aligned} E[N] &= \sum_{k=1}^{\infty} kP(N = k) = \sum_{k=1}^{\infty} [kP(N \geq k) - kP(N \geq k + 1)] \\ &= \sum_{k=1}^{\infty} P(N \geq k) = \sum_{k=1}^{\infty} F^k, \end{aligned}$$

where the last equation follows from the fact that every return occurs independently with probability F . Putting the results together we get

$$\sum_{t=1}^{\infty} P_t(0) = E[N] = \sum_{k=1}^{\infty} F^k,$$

which diverges if $F = 1$, and converges if $F < 1$. \diamond

In other words, the random walk is recurrent precisely when $\sum_{t=1}^{\infty} P_t(0)$ is infinity. It is this equivalence that we will use to prove Polya's theorem. First, we compute $P_n(0)$. After that we compute the sum over n to see

whether it diverges. There is a way to proceed that covers all dimensions at once, namely, by giving an integral expression for $P_n(0)$, and analyzing this expression for $n \rightarrow \infty$. We will need this procedure for dimensions $d \geq 3$, but to illustrate that the situation for $d = 1, 2$ is significantly easier, we will first carry out a simple computation for $d = 1, 2$.

- $d = 1, 2$

$d = 1$: Imagine the one-dimensional lattice \mathbb{Z} lying horizontally. Any path that the random walker follows can be uniquely represented by an infinite sequence $(llrlrrrl\dots)$ (l standing for left, r for right). Conversely, any such sequence represents a path. Next, note that any path from 0 to itself has an even length, containing as many steps to the left as to the right. Therefore

$$P_{2n+1}(0) = 0; P_{2n}(0) = \binom{2n}{n} \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^n = \frac{(2n)!}{n!(2n-n)!} \frac{1}{2^{2n}}.$$

Now substitute Stirling's approximation of the factorial $n! \sim n^n e^{-n} \sqrt{2\pi n}$ as $n \rightarrow \infty$ to get

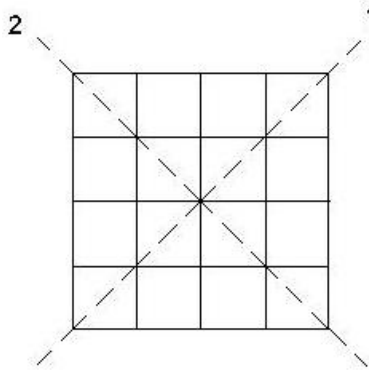
$$P_{2n}(0) \sim \frac{2^{2n} n^{2n} e^{-2n} \sqrt{4\pi n}}{n^{2n} e^{-2n} 2\pi n} \frac{1}{2^{2n}} = \frac{1}{\sqrt{\pi n}} \text{ as } n \rightarrow \infty.$$

So the infinite sum becomes

$$\begin{aligned} \sum_{n=1}^{\infty} P_n(0) &= \sum_{n=1}^{\infty} P_{2n}(0) = \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} [1 + o(1)] \\ &= \frac{1}{\sqrt{\pi}} [1 + o(1)] \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} > \frac{1}{\sqrt{\pi}} \sum_{n=1}^{\infty} \frac{1}{n} = \infty, \end{aligned}$$

and we see that the one-dimensional random walk is recurrent.

$d = 2$: For one two-dimensional walk, define two one-dimensional walks in the following way:



Let S_n be the two-dimensional position. Define for every n the positions of two one-dimensional walks S_n^1 and S_n^2 by the orthogonal projection of S_n on the respective axes 1 and 2. The steps of a random walk are the differences between two successive positions, and the two-dimensional step X_i can take the values north, east, south, west. The following table gives the relation between X_i , X_i^1 and X_i^2 :

	<i>N</i>	<i>E</i>	<i>S</i>	<i>W</i>
X^1	1	1	-1	-1
X^2	1	-1	-1	1

From this table it is obvious that the distribution of X^1 given X^2 is the same as the marginal distribution of X^1 , and $P(X^1 = 1) = P(X^1 = -1) = P(X^2 = 1) = P(X^2 = -1) = \frac{1}{2}$. So in this way any two one-dimensional random walks correspond precisely to one two-dimensional random walk, and the other way round. Therefore, in $d = 2$ we can write:

$$P_{2n}(0) = P(S_{2n} = 0) = P(S_{2n}^1 = 0)P(S_{2n}^2 = 0) \sim \left(\frac{1}{\sqrt{\pi n}}\right)^2 = \frac{1}{\pi n}$$

and, because still $P_{2n+1}(0) = 0$, the sum over n becomes

$$\sum_{n=1}^{\infty} P_n(0) = \sum_{n=1}^{\infty} P_{2n}(0) = \frac{1}{\pi} [1 + o(1)] \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

So the two-dimensional random walk is recurrent as well.

- $d \geq 3$

The general method to prove recurrence or transience needs some more computation. The whole method is based on the well-known theorem for Markov chains due to Chapman and Kolmogorov, which in our notation takes the form:

Theorem. In the above notation, the following holds for all $l \in \mathbb{Z}^d$:

$$P_{n+1}(l) = \sum_{l' \in \mathbb{Z}^d} p(l - l') P_n(l'). \quad (1)$$

In words this states that the probability of travelling to l in $n + 1$ steps can be found by summing over the positions the walker can occupy at time n . It is clear that the statement uses the translation invariance of the walk. The theorem can be seen as describing the evolution of the walk: a recurrence relation that expresses higher-step transition probabilities in terms of lower-step transition probabilities. The one-step transition probabilities are prescribed by the definition of the random walk. As many ordinary differential equations can be solved by applying the Fourier transform \mathcal{F} to them,

in a like manner we will solve our recurrence relation using the discrete Fourier transform \mathcal{F} , the properties of which are very much the same as of the continuous one. It takes the form

$$\mathcal{F}(P_n(l)) = \widehat{P}_n(k) = \sum_{l \in \mathbb{Z}^d} e^{il \cdot k} P_n(l), k \in [-\pi, \pi]^d.$$

For ease we define the structure function $\lambda(k)$ as the Fourier transform of the one-step transition probabilities: $\lambda(k) = \widehat{p}(k) = \sum_l e^{il \cdot k} p(l)$. We can now transform equation (1), and we get

$$\widehat{P}_{n+1}(k) = \sum_l e^{il \cdot k} \sum_{l'} p(l - l') P_n(l').$$

Note that $e^{il \cdot k} = e^{i(l-l') \cdot k} e^{il' \cdot k}$ is a constant (not depending on l') that can be placed behind the summation sign. Therefore

$$\widehat{P}_{n+1}(k) = \sum_l \sum_{l'} p(l - l') e^{i(l-l') \cdot k} e^{il' \cdot k} P_n(l').$$

Now call $m = l - l'$, and the above equals

$$\sum_{m+l} p(m) e^{im \cdot k} \sum_{l'} e^{il' \cdot k} P_n(l') = \lambda(k) \widehat{P}_n(k) = \widehat{P}_{n+1}(k).$$

The recurrence relation has now become easy to solve: we only need the initial condition. This is $P_0(l) = \delta_{0l}$ (at $t = 0$ the walker must be in the origin), and in Fourier transform this condition is $\widehat{P}_0(k) = \sum_l e^{il \cdot k} \delta_{0l} = 1$ (only for $l = 0$ does the delta-function not vanish). Substituting this initial condition, we see that the solution of (1) becomes $\widehat{P}_n(k) = \lambda(k)^n$. The inverse Fourier transform has the form of the d -dimensional integral

$$P_n(l) = \mathcal{F}^{-1}(\widehat{P}_n(k)) = \frac{1}{(2\pi)^d} \int \dots \int_{k \in [-\pi, \pi]^d} -e^{il \cdot k} \widehat{P}_n(k) dk.$$

So the formal solution (for any dimension) of the transition probabilities P_n is

$$P_n(l) = \frac{1}{(2\pi)^d} \int \dots \int_{k \in [-\pi, \pi]^d} -e^{il \cdot k} \lambda(k)^n dk.$$

Until now the calculation was purely formal. If we want to test the walk for transience, we must sum the transition probabilities $P_n(0)$ over n . For that, we need to know $\lambda(k)$ and then evaluate the multiple integral. Because we are only interested in whether the sum $\sum_{n=1}^{\infty} P_n(0)$ converge or not, it suffices for our purpose to approximate $\lambda(k)$, and then to determine the limiting behaviour of $P_n(0)$ for $n \rightarrow \infty$: if it approaches 0 fast enough, then the sum will converge, and hence the random walk will be transient.

So, what is $\lambda(k)$ for our simple random walk? The answer follows from a short computation. Write the vector $k = (k_1, k_2, \dots, k_d)^T$, and fill in the formula for the structure function:

$$\lambda(k) = \widehat{p}(k) = \sum_l e^{il \cdot k} p(l) = \sum_{l: \|l\|=1} e^{il \cdot k} \frac{1}{2d}.$$

Here we use that the one-step transition probability $p(l)$ equals $\frac{1}{2d}$ if l is a unit vector, and equals 0 otherwise. Denote these $2d$ unit vectors by e_j and $-e_j$, for $j = 1, \dots, d$. Then

$$\begin{aligned} \lambda(k) &= \frac{1}{2d} \sum_{j=1}^d \left[e^{ie_j \cdot k} + e^{i(-e_j) \cdot k} \right] = \frac{1}{2d} \sum_{j=1}^d \left[e^{ik_j} + e^{-ik_j} \right] \\ &= \frac{1}{2d} \sum_{j=1}^d [\cos(k_j) + \cos(-k_j)] + i \frac{1}{2d} \sum_{j=1}^d [\sin(k_j) + \sin(-k_j)] \\ &= \frac{1}{d} \sum_{j=1}^d \cos(k_j). \end{aligned}$$

For the sake of approximating this with a function that is more easily integrated over, recall the following two Taylor expansions for $x \rightarrow 0$:

$$\cos(x) = 1 - \frac{x^2}{2!} + h.o. \text{ and } e^x = 1 + x - \frac{x^2}{2!} + h.o.,$$

where *h.o.* means higher order terms. Substitute this into the formula for $\lambda(k)$ to obtain

$$\begin{aligned} \lambda(k) &= \frac{1}{d} \sum_{j=1}^d \left(1 - \frac{k_j^2}{2!} + h.o. \right) \\ &= 1 - \frac{1}{2d} \|k\|^2 + h.o. \sim e^{-\frac{1}{2d} \|k\|^2} \text{ for } k \rightarrow 0. \end{aligned}$$

The integral we are to calculate has now become

$$P_n(0) \sim \frac{1}{(2\pi)^d} \int \dots \int_{k \in [-\pi, \pi]^d} -e^{il \cdot k} e^{-\frac{n}{2d} \|k\|^2} dk \text{ for } k \rightarrow 0$$

but this approximation only holds for small k . Yet this is all we need, because we are interested in the limiting behaviour as $n \rightarrow \infty$, in which case the second factor in the integrand clearly becomes very small, unless k is taken to be so small as to compensate for the large n . Thus it can be seen that the dominating contribution to the integral is contained in the region where $\|k\| = O\left(\frac{1}{\sqrt{n}}\right)$. In the limit $n \rightarrow \infty$ this means that $\|k\|$ approaches

0, and the value of the integral is not affected when we integrate over the whole of \mathbb{R}^d in stead of only $[-\pi, \pi]^d$. Hence

$$P_n(0) \sim \frac{1}{(2\pi)^d} \int \dots \int_{\mathbb{R}^d} e^{-\frac{n}{2d}\|k\|^2} dk \text{ for } n \rightarrow \infty.$$

Now observe that the integrand only depends on the length of k . This suggests a transformation to spherical coordinates, because then the integrand only depends on the radius $r = \|k\|$. The above integral equals

$$\frac{1}{(2\pi)^d} \int_0^\infty \frac{d}{dr} B_r^d e^{-\frac{n}{2d}r^2} dr,$$

where B_r^d is the volume of the ball of radius r in d dimensions. For computing B_r^d , define the stepfunction $\theta = 1_{[0, \infty)}$. Then, for all $d \geq 1$,

$$B_r^d = \int_{\mathbb{R}^d} \theta(r^2 - \|x\|^2) dx.$$

Now substitute $y = \frac{x}{r}$, and hence $\|x\|^2 = r^2\|y\|^2$, and $dx = r^d dy$. This yields

$$B_r^d = \int_{\mathbb{R}^d} \theta(r^2 - r^2\|y\|^2) r^d dy = r^d \int_{\mathbb{R}^d} \theta(1 - \|y\|^2) dy = \omega_d r^d,$$

where ω_d represents the volume of the unit sphere in d dimensions. Because $\frac{d}{dr} B_r^d = d r^{d-1} \omega_d$, our asymptotic integral expression for $P_n(0)$ becomes

$$\frac{1}{(2\pi)^d} \int_0^\infty d\omega_d r^{d-1} e^{-\frac{n}{2d}r^2} dr.$$

To evaluate further, substitute $x = \frac{n}{2d}r^2$. Then $\frac{dx}{dr} = \frac{n}{d}r$, so $r dr = \frac{d}{n} dx$, and

$$r^{d-2} = \left(\sqrt{\frac{2d}{n}x} \right)^{d-2} = \left(\frac{2d}{n} \right)^{\frac{1}{2}d-1} x^{\frac{1}{2}d-1},$$

and the boundaries 0 and ∞ remain the same. Now the integral has become, for $n \rightarrow \infty$,

$$\begin{aligned} P_n(0) &\sim \frac{1}{(2\pi)^d} \int_0^\infty d\omega_d \left(\frac{2d}{n} \right)^{\frac{1}{2}d-1} x^{\frac{1}{2}d-1} e^{-x} \frac{d}{n} dx \\ &= \frac{1}{(2\pi)^d} \omega_d (2d)^{\frac{1}{2}d-1} \Gamma\left(\frac{1}{2}d\right) \frac{1}{n^{\frac{1}{2}d}} = \frac{C(d)}{n^{\frac{1}{2}d}}, \end{aligned}$$

where $\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx$ for $a > 0$, and $C(d)$ is a constant that depends on the dimension. At last we can sum over the transition probabilities. Note that

$$\sum_{n=1}^\infty \frac{C(d)}{n^{\frac{1}{2}d}} = C(d) \sum_{n=1}^\infty \frac{1}{n^{\frac{1}{2}d}} < \infty \text{ if } d \geq 3.$$

Because for n large enough $P_n(0) \leq 2 \frac{C(d)}{n^{\frac{1}{2}d}}$, we conclude that for $d \geq 3$ it must hold that $\sum_{n=1}^\infty P_n(0) < \infty$, and therefore simple random walk in dimension $d \geq 3$ is transient.

3 Range of the random walk

Intuitively, it is clear that a transient random walker is much more likely to visit new sites of the lattice than a recurrent one. To make this precise, define the range R_n of a random walk at time n as the number of distinct points visited within time n :

Definition. $\forall n \geq 0 : R_n = \text{card}(\{0 = S_0, S_1, \dots, S_n\})$.

Defining F to be the probability that the walker will eventually return to its starting site S_0 , the behaviour of the range is stated in the following theorem:

Theorem. $\forall \epsilon > 0 : \lim_{n \rightarrow \infty} P(|\frac{R_n}{n} - (1 - F)| > \epsilon) = 0$.

Proof. Define $\phi_0 = 1$ and, for $k \in \mathbb{N}$ $\phi_k = 1$ if $S_i \neq S_k$ for all $i = 1, \dots, k - 1$ and $\phi_k = 0$ otherwise. In other words, $\phi_k = 1$ if and only if the random walker visits a new site on its k 'th step. It is obvious that $R_n = \sum_{k=0}^n \phi_k$, and because of linearity of expectations also $E[R_n] = \sum_{k=0}^n E[\phi_k]$ holds. For any $k \in \mathbb{N}$ we can write:

$$\begin{aligned} E[\phi_k] &= P(\phi_k = 1) = P(S_k \neq S_{k-1}, S_k \neq S_{k-2}, \dots, S_k \neq S_0 = 0) \\ &= P(S_k - S_{k-1} \neq 0, S_k - S_{k-2} \neq 0, \dots, S_k \neq 0) \\ &= P(X_k \neq 0, X_k + X_{k-1} \neq 0, \dots, X_k + \dots + X_1 \neq 0) \\ &= P(X_1 \neq 0, X_1 + X_2 \neq 0, \dots, X_1 + \dots + X_k \neq 0) \\ &= P(S_j \neq 0 \text{ for } j = 1, \dots, k) \\ &= 1 - \sum_{j=1}^k F_j(0, 0), \end{aligned}$$

where in the fourth line we reverse the indices, and in the sixth line $F_j(0, 0)$ is the probability that the random walker, starting in 0, returns to 0 for the first time on its j 'th step. Taking the limit $k \rightarrow \infty$, we get

$$\lim_{k \rightarrow \infty} E[\phi_k] = 1 - F$$

and this equals 0 if and only if the random walk is recurrent. Consequently,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n E[\phi_k] = \lim_{n \rightarrow \infty} \frac{1}{n} E[R_n] = 1 - F.$$

To proceed, we consider the recurrent and the transient case separately. For $\epsilon \geq 0$, write

$$\begin{aligned} P\left(\frac{R_n}{n} > \epsilon\right) &= \sum_{k:k > n\epsilon} P(R_n = k) \leq \sum_{k:k > n\epsilon} \frac{k}{n\epsilon} P(R_n = k) \\ &\leq \frac{1}{n\epsilon} \sum_{k=0}^{\infty} k P(R_n = k) = \frac{1}{n\epsilon} E[R_n]. \end{aligned}$$

For the recurrent case it therefore follows that $\lim_{n \rightarrow \infty} P\left(\frac{R_n}{n} > \epsilon\right) = 0$ for all $\epsilon > 0$, and we are done.

The transient case is more complicated. First, notice that for any continuous random variable X on $\mathbb{R}_{\geq 0}$ the following holds: $E[X] = \int_0^\infty xP(X = x)dx = \int_0^\epsilon xP(X = x)dx + \int_\epsilon^\infty xP(X = x)dx \geq \epsilon \int_\epsilon^\infty f(x)dx = \epsilon P(X > \epsilon)$ where $f(x)$ is the probability density function, and this inequality also holds for discrete random variables on $\mathbb{R}_{\geq 0}$. Secondly, use this inequality for the random variable $|R_n - n(1 - F)|^2$. Starting from the probability we are interested in, we get

$$\begin{aligned}
& P\left(\left|\frac{R_n}{n} - (1 - F)\right| > \epsilon\right) = P(|R_n - n(1 - F)|^2 > n^2\epsilon^2) \\
& \leq \frac{1}{n^2\epsilon^2} E[|R_n - n(1 - F)|^2] \\
& = \frac{1}{n^2\epsilon^2} E[R_n^2 - 2n(1 - F)R_n + n^2(1 - F)^2] \\
& = \frac{1}{n^2\epsilon^2} E[R_n^2 - 2R_n E[R_n] + 2(E[R_n])^2 - 2n(1 - F)E[R_n] + n^2(1 - F)^2] \\
& = \frac{1}{n^2\epsilon^2} E[R_n^2 - 2R_n E[R_n] + (E[R_n])^2] \\
& \quad + \frac{1}{n^2\epsilon^2} (n^2(1 - F)^2 - 2n(1 - F)E[R_n] + (E[R_n])^2) \\
& = \frac{1}{n^2\epsilon^2} E[(R_n - E[R_n])^2] + \frac{1}{\epsilon^2} \left(1 - F - E\left[\frac{R_n}{n}\right]\right)^2.
\end{aligned}$$

Write $E[(R_n - E[R_n])^2] = \text{var}(R_n)$. To prove that the above probability converges to zero as $n \rightarrow \infty$, it suffices to prove that $\lim_{n \rightarrow \infty} \frac{1}{n^2} \text{var}(R_n) = 0$, because it was already shown that $\lim_{n \rightarrow \infty} E\left[\frac{R_n}{n}\right] = 1 - F$, so that the second term of the expression vanishes in the limit.

Continue by computing $\text{var}(R_n)$ as follows (using the linearity of expectations):

$$\begin{aligned}
\text{var}(R_n) &= E[(R_n)^2 - (E[R_n])^2] = E\left[\sum_{j=0}^n \phi_j \sum_{k=0}^n \phi_k\right] - E\left[\sum_{j=0}^n \phi_j\right]^2 \\
&= E\left[\sum_{j=0}^n \sum_{k=0}^n \phi_j \phi_k\right] - E\left[\sum_{j=0}^n \phi_j\right] E\left[\sum_{k=0}^n \phi_k\right] \\
&= \sum_{j=0}^n \sum_{k=0}^n (E[\phi_j \phi_k] - E[\phi_j] E[\phi_k]) \\
&= 2 \sum_{0 \leq j < k \leq n} (E[\phi_j \phi_k] - E[\phi_j] E[\phi_k]) + \sum_{j=0}^n E[\phi_j - E[\phi_j]]^2.
\end{aligned}$$

This last equality follows from the fact that summing over the elements of a symmetric (square) matrix, one may as well take twice the sum over the elements under the diagonal, and add the diagonal elements (notice that $\phi_j^2 = \phi_j$). Because $E[\phi_j - E[\phi_j]] \leq E[\phi_j]$, $\text{var}(R_n)$ can be estimated by

$$\text{var}(R_n) \leq 2 \sum_{0 \leq j \leq k \leq n} (E[\phi_j \phi_k] - E[\phi_j]E[\phi_k]) + \sum_{j=0}^n E[\phi_j].$$

But we can estimate it by a yet simpler expression: Notice that for $0 \leq j < k$,

$$\begin{aligned} E[\phi_j \phi_k] &= P(\phi_j \phi_k = 1) = P(S_j \neq S_\alpha \text{ for } 0 \leq \alpha < j, S_k \neq S_\beta \text{ for } 0 \leq \beta < k) \\ &\leq P(S_j \neq S_\alpha \text{ for } 0 \leq \alpha < j, S_k \neq S_\beta \text{ for } j < \beta < k) \\ &= P(X_j \neq 0, X_j + X_{j-1} \neq 0, \dots, X_j + \dots + X_1 \neq 0; \\ &\quad X_k \neq 0, X_k + X_{k-1} \neq 0, \dots, X_k + \dots + X_{j+1} \neq 0) \\ &= P(X_1 \neq 0, \dots, X_1 + \dots + X_j \neq 0) \\ &\quad P(X_1 \neq 0, \dots, X_1 + \dots + X_{k-j} \neq 0). \end{aligned}$$

The factorization and mixing of indices is allowed because the X_i are i.i.d. Now recall that $E[\phi_k] = P(X_1 \neq 0, \dots, X_1 + \dots + X_k \neq 0)$, so the inequality says that $E[\phi_j \phi_k] \leq E[\phi_j]E[\phi_{k-j}]$ for $0 \leq j < k$. Substitution into the former estimate of $\text{var}(R_n)$ yields

$$\text{var}(R_n) \leq 2 \sum_{j=0}^n E[\phi_j] \sum_{k=j+1}^n (E[\phi_{k-j} - E[\phi_k]] + E[R_n])$$

Since $E_k[\phi_k] = 1 - \sum_{j=1}^k F_j(0, 0)$, we have $\{E[\phi_k]\}_{k=0}^n$ is a monotone non-increasing sequence. But for any such sequence $a_1 \geq a_2 \geq \dots \geq a_n$ the sum

$$\sum_{k=j+1}^n (a_{k-j} - a_k) = (a_1 + a_2 + \dots + a_{n-j}) - (a_{j+1} + a_{j+2} + \dots + a_n)$$

is maximized by taking $j = \lfloor \frac{n}{2} \rfloor$ (that is, round off $\frac{n}{2}$ downward). Indeed, by taking j smaller, the left term increases less than the right one and, by taking j larger, the left term decreases more than the right one. Its maximum value is $(a_1 + \dots + a_{n-\lfloor \frac{n}{2} \rfloor}) - [(a_1 + \dots + a_n) - a_1 + \dots + a_{\lfloor \frac{n}{2} \rfloor}]$. Taking $a_k = E[\phi_k]$, and recalling that $\sum_{k=0}^n E[\phi_k] = E[R_n]$, it therefore holds that

$$\text{var}(R_n) \leq 2 \sum_{j=0}^n E[\phi_j] E[R_{n-\lfloor \frac{n}{2} \rfloor} + R_{\lfloor \frac{n}{2} \rfloor} + R_n] + E[R_n].$$

Because we already showed that $\lim_{n \rightarrow \infty} \frac{1}{n} E[\sum_{j=0}^n \phi_j] = \lim_{n \rightarrow \infty} E[\frac{R_n}{n}] = 1 - F$, we get

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \text{var}(R_n) \leq 2(1 - F) \lim_{n \rightarrow \infty} E \left[\frac{E[R_{n-\lfloor \frac{n}{2} \rfloor}] + R_{\lfloor \frac{n}{2} \rfloor} + R_n}{n} \right] + 0$$

$$= 2(1 - F) \left(\frac{1 - F}{2} + \frac{1 - F}{2} - (1 - F) \right) = 0,$$

which was still left to be proved. \diamond

The above theorem states that the random variable $\frac{R_n}{n}$ converges to $1 - F$ in probability, but in fact a stronger statement also holds: $\frac{R_n}{n}$ converges to $1 - F$ almost surely. The proof thereof requires the ergodic theorem, which we cannot prove here. The difference between these two types of convergence of random variables is defined in section 4, but intuitively it means the following.

Consider the collection of all possible paths (of infinite length) a random walker might take. For each of those paths with every time n a value $\frac{R_n}{n}$ is associated. Strong convergence means that, if one of those paths is selected, then it is with probability one a path for which the sequence $(\frac{R_n}{n})_{n \in \mathbb{N}}$ converges to $1 - F$.

If some arbitrary deviation $\epsilon \geq 0$ is given, then the probability (when selecting a path) $p_n = P(|\frac{R_n}{n} - (1 - F)| > \epsilon)$ depends on n . Convergence in probability means that the sequence $(p_n)_{n \in \mathbb{N}}$ converges to 0.

Theorem. $P(\lim_{n \rightarrow \infty} \frac{R_n}{n} = 1 - F) = 1$.

Proof. We will define two sequences $(D_n)_{n \in \mathbb{N}}$ and $(R_{n,M})_{n \in \mathbb{N}}$ such that $D_n \leq R_n \leq R_{n,M}$, and use these to determine the value of the limit.

First, define $R_{n,M}$ like R_n , but at every M 'th step forget which sites were already visited (but do not reset the counter to 0). Formally, define the ranges of subsequent M -step walks

$$Z_k(M) = \text{card}\{S_{kM}, S_{kM+1}, \dots, S_{k(M+1)-1}\},$$

and now add these up to get

$$R_{n,M} = \sum_{k=0}^{\lfloor \frac{n}{M} \rfloor} Z_k(M),$$

which is the sequence described above. It is obvious that $R_{n,M} \leq R_n$. Note that it is not clear yet that $\lim_{n \rightarrow \infty} \frac{R_n}{n}$ exists at all, so that we will estimate downwards by taking the limit inferior and upwards by taking the limit superior (because these exist for every sequence). Thus it must hold that

$$\limsup_{n \rightarrow \infty} \frac{R_n}{n} \leq \limsup_{n \rightarrow \infty} \frac{\sum_{k=0}^{\lfloor \frac{n}{M} \rfloor} Z_k(M)}{n}.$$

Now, the $Z_k(M)$ are i.i.d., and so the strong law of large numbers can be applied to obtain

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \left\lfloor \frac{n}{M} \right\rfloor \frac{1}{n} \frac{1}{\left\lfloor \frac{n}{M} \right\rfloor} \sum_{k=0}^{\left\lfloor \frac{n}{M} \right\rfloor} Z_k(M) \\ & \leq \frac{1}{M} E[Z_0(M)] = \frac{1}{M} E[R_M] \text{ a.s.} \end{aligned}$$

Now take the limit $M \rightarrow \infty$. We already saw that $\frac{1}{M} E[R_M]$ converges to $1 - F$, and therefore we get

$$\limsup_{n \rightarrow \infty} \frac{R_n}{n} \leq 1 - F \text{ a.s.}$$

Secondly, define D_n as the number of distinct sites visited in time n that the walker never visits again. Formally, let

$$D_n = \sum_{k=0}^n \psi_k \quad \text{with } \psi_k = 1 \text{ if } X_{k+1} + \dots + X_{k+i} \neq 0 \text{ for all } i \in \mathbb{N}$$

and $\psi_k = 0$ otherwise.

So $\psi_k = 1$ only if after the k 'th step the walker never returns to the site where it is at time k , i.e. if the walker visits the current site for the last time. The number of times that the walker visits a site for the last time is evidently smaller than the number of sites visited, so $D_n \leq R_n$.

Y_0, Y_1, \dots is said to be a *stationary sequence* of random variables if, for every k , the sequence Y_k, Y_{k+1}, \dots has the same distribution, that is, for every n , the $(n+1)$ -tuples (Y_0, Y_1, \dots, Y_n) and $(Y_k, Y_{k+1}, \dots, Y_{k+n})$ have the same joint probability distribution. In particular, the Y_i 's are identically distributed, but they may be dependent. Because of symmetry and the Markov property of the simple random walk, it is clear that $(\psi_k)_{k \in \mathbb{N}}$ is a stationary sequence. Therefore, by the ergodic theorem the limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \psi_k = \lim_{n \rightarrow \infty} \frac{D_n}{n}$$

exists with probability 1 (but it may still be a random variable). But $\lim_{n \rightarrow \infty} \frac{D_n}{n}$ assuming a certain value is a so-called *tail event*. Intuitively, tail events are those events for a sequence of random variables that would still have occurred if some finite number of those random variables would have had a different realisation. Tail events are treated more thoroughly in section 4. Indeed, all events of the form $\lim_{n \rightarrow \infty} \frac{D_n}{n} < C$, $\lim_{n \rightarrow \infty} \frac{D_n}{n} \geq C$ etc., are tail events, and must therefore, according to Kolmogorov's 0-1 law,

occur with probability 0 or 1. Consequently, the limit must needs be equal to a constant. This constant can only be the natural candidate, namely,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \psi_k = E[\psi_0] = P(X_1 + \dots + X_i \neq 0 \text{ for all } i \geq 1) = 1 - F \text{ a.s.}$$

Consequently,

$$\liminf_{n \rightarrow \infty} \frac{R_n}{n} \geq \liminf_{n \rightarrow \infty} \frac{D_n}{n} = \lim_{n \rightarrow \infty} \frac{D_n}{n} = 1 - F \text{ a.s.}$$

Finally, note that the first statement ($\limsup_{n \rightarrow \infty} \frac{R_n}{n} \leq 1 - F$ a.s.) and the second statement ($\liminf_{n \rightarrow \infty} \frac{R_n}{n} \geq 1 - F$ a.s.) together imply the statement of the theorem ($\lim_{n \rightarrow \infty} \frac{R_n}{n} = 1 - F$ a.s.). \diamond

4 Probability measures and stochastic convergence

The purpose of this section is to define more precisely what is meant by random variables and their convergence. This is done in measure-theoretic terms, because that is the only way to make precise our construction of so-called Brownian motion in section 5. While this treatise is on simple random walk, and not on measure-theoretic probability, we will put more emphasis on the intuitive interpretation of the definitions than on the proof of their properties, which can be found in [1].

We use a *probability space* to model experiments involving randomness. A probability space $(\Omega, \Sigma, \mathbf{P})$ is defined as follows:

Sample space. Ω is a set, called the sample space, whereof the points $\omega \in \Omega$ are called *sample points*.

Events. Σ is a σ -algebra of subsets of Ω , that is, a collection of subsets with the property that, firstly, $\Omega \in \Sigma$; secondly, whenever $F \in \Sigma$ then $F^C = \Omega \setminus F \in \Sigma$; thirdly, whenever $F_n \in \Sigma (n \in \mathbb{N})$, then $\bigcup_n F_n \in \Sigma$. Notice that these imply that Σ contains the empty set, and is closed under countable intersections. All $F \in \Sigma$ are called *measurable subsets*, or (when talking about probability spaces) *events*.

Probability. \mathbf{P} is called a *probability measure* on (Ω, Σ) , that is, a function $\mathbf{P} : \Sigma \rightarrow [0, 1]$ that assigns to every event a number between 0 and 1. \mathbf{P} must satisfy: firstly, $\mathbf{P}(\emptyset) = 0$ and $\mathbf{P}(\Omega) = 1$; secondly, whenever $(F_n)_{n \in \mathbb{N}}$ is a sequence of disjoint events with union $F = \bigcup_n F_n$, then $\mathbf{P}(F) = \sum_{n \in \mathbb{N}} \mathbf{P}(F_n)$ (σ -additivity).

Randomness is contained in this model in the following way: When performing an experiment, some $\omega \in \Omega$ is chosen in such a way that for every $F \in \Sigma$, $\mathbf{P}(F)$ represents the probability that the chosen sample point ω belongs to F (in which case the event F is said to occur).

Some statement S about the outcomes is said to be true *almost surely* (*a.s.*), or *with probability 1*, if

$$F = \{\omega : S(\omega) \text{ is true}\} \in \Sigma \text{ and } \mathbf{P}(F) = 1.$$

If R is a collection of subsets of S , then $\sigma(R)$, the σ -algebra generated by R , is defined to be the smallest σ -algebra contained in Σ .

The *Borel σ -algebra* \mathcal{B} is the smallest σ -algebra that contains all open sets in \mathbb{R} . A function $h : \Omega \rightarrow \mathbb{R}$ is called Σ -measurable if $h^{-1} : \mathcal{B} \rightarrow \Sigma$, that is, if $h(A) \in \Sigma$ for all $A \in \mathcal{B}$. (Compare continuous maps in topology: a map is called continuous if the inverse image $F^{-1}(G)$ is open for all open G . A map is called measurable if the inverse image of every measurable subset is measurable.)

Given (Ω, Σ) , a *random variable* is a Σ -measurable function. So, for a random variable X :

$$X : \Omega \rightarrow \mathbb{R} \text{ and } X^{-1} : \mathcal{B} \rightarrow \Sigma$$

Given a collection $(Y_\gamma)_{\gamma \in C}$ of maps $Y_\gamma : \Omega \rightarrow \mathbb{R}$, its *generated σ -algebra*

$$\mathcal{Y} = \sigma(Y_\gamma : \gamma \in C)$$

is defined to be the smallest σ -algebra \mathcal{Y} on Ω such that each map Y_γ is \mathcal{Y} -measurable (that is, such that each map is a random variable). The following holds:

$$\mathcal{Y} = \sigma(Y_\gamma : \gamma \in C) = \sigma(\{\omega \in \Omega : Y_\gamma(\omega) \in B\} : \gamma \in C, B \in \mathcal{B}).$$

If X is a random variable for some (Ω, Σ) , then obviously $\sigma(X) \subset \Sigma$.

Suppose $(\Omega, \Sigma, \mathbf{P})$ is a model for some experiment, and that the experiment has been performed, that is, some $\omega \in \Omega$ has been selected. Suppose further that $(Y_\gamma)_{\gamma \in C}$ is a collection of random variables associated with the experiment. Now consider the values $Y_\gamma(\omega)$, that is, the observed values (realisations) of the random variables. Then the intuitive significance of the σ -algebra $\sigma(Y_\gamma : \gamma \in C)$ is that it consists precisely of those events F for which it is possible to decide whether or not F has occurred (i.e. whether or not $\omega \in F$) on the basis of the values $Y_\gamma(\omega)$ only. Moreover, this must be possible for every $\omega \in \Omega$.

Given a sequence of random variables $(X_n)_{n \in \mathbb{N}}$ and the generated σ -algebras $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$, define the *tail σ -algebra* \mathcal{T} of the sequence $(X_n)_{n \in \mathbb{N}}$ as follows:

$$\mathcal{T} = \bigcap_{n \in \mathbb{N}} \mathcal{T}_n.$$

So \mathcal{T} consists of those events which can be said to occur (or not to occur) on the basis of the realisations of the random variables, beyond any finite index. For such events the following theorem states, that they will either occur with certainty, or not at all.

Kolmogorov's 0-1 law. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent random variables, and \mathcal{T} the tail σ -algebra thereof. Then $F \in \mathcal{T} \Rightarrow \mathbf{P}(F) = 0$ or $\mathbf{P}(F) = 1$.

Suppose X is a random variable carried by the probability space $(\Omega, \Sigma, \mathbf{P})$. We have

$$\begin{aligned} \Omega &\xrightarrow{X} \mathbb{R} \\ [0, 1] &\xleftarrow{\mathbf{P}} \Sigma \xleftarrow{X^{-1}} \mathcal{B} \\ [0, 1] &\xleftarrow{\mathbf{P}} \sigma(X) \xleftarrow{X^{-1}} \mathcal{B}. \end{aligned}$$

Define the *law* \mathcal{L}_X of X by $\mathcal{L}_X = \mathbf{P} \circ X^{-1}$, so $\mathcal{L}_X : \mathcal{B} \rightarrow [0, 1]$. The law can be shown to be a probability measure on $(\mathbb{R}, \mathcal{B})$.

The *distribution function* of a random variable X is a function $F_X : \mathbb{R} \rightarrow [0, 1]$ defined by:

$$F_X(c) = \mathcal{L}_x(-\infty, c) = \mathbf{P}(X \leq c) = \mathbf{P}(\{\omega : X(\omega) \leq c\}).$$

Because we have defined a random variable as a function from Ω to \mathbb{R} , we now have several notions of a converging sequence of random variables. The usual modes of convergence for functions are all well-defined for any sequence $(X_n)_{n \in \mathbb{N}}$ of random variables, and we may for example consider uniform convergence or pointwise convergence to some random variable X . The latter one is weaker, and we mean by it: $\forall \omega \in \Omega : \lim_{n \rightarrow \infty} (X_n(\omega)) = X(\omega)$ or, shortly, $X_n(\omega) \rightarrow X(\omega)$. But in practice, for random variables we are only interested in yet weaker modes of convergence, which we define below.

A sequence $(X_n)_{n \in \mathbb{N}}$ of random variables is said to converge to X :

almost surely (or, *with probability 1*), if $\mathbf{P}(\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}) = 1$.

Note that not for all $\omega \in \Omega$ needs $(X_n(\omega))_{n \in \mathbb{N}}$ converge to $X(\omega)$, but the set of ω 's for which it does converge, has probability one. Which is the same as saying that if for some random $\omega \in \Omega$, the sequence of real numbers $(X_n(\omega))_{n \in \mathbb{N}}$ is considered, it is certain to converge to $X(\omega)$.

in probability, if $\forall \epsilon > 0 : \mathbf{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}) \rightarrow 0$. Note that $(X_n(\omega))_{n \in \mathbb{N}}$ may not converge to $X(\omega)$ for all $\omega \in \Omega$, but for any $\epsilon > 0$ the probability that X_n deviates from X more than ϵ for a fixed n tends to 0 as $n \rightarrow \infty$.

in distribution, if $F_{X_n}(c) \rightarrow F_X(c)$ for all continuity points c of F_X . Which expresses nothing but the pointwise convergence of the distribution functions, and tells nothing about the random variables themselves.

Convergence almost surely is also called *strong convergence* and is denoted \rightarrow^{as} , convergence in probability is also called *weak convergence* and is denoted \rightarrow^P , convergence in distribution is also called *convergence in law* and is denoted \rightsquigarrow . Note that for convergence in law, the sequence of random variables needs not be defined on the same probability space as its limit. In particular, the X_n 's may be defined on a discrete space, while X may be defined on a continuous space. For example, if $\mathbf{P}(X_n = \frac{i}{n}) = \frac{1}{n}$ for $i = 1, 2, \dots, n$, then $X_n \rightsquigarrow X$, with X uniformly distributed on $[0, 1]$.

It can be shown that strong convergence implies weak convergence, and weak convergence implies convergence in law, but none of the three are equivalent in general.

5 Brownian motion

Imagine the following: instead of executing a random walk on the standard lattice with site distance 1 at time steps of size 1, we make the lattice ever narrower and reduce our time steps appropriately. Eventually we will get some random process in continuous space and time. Loosely speaking, making the lattice narrower and reducing the time steps should happen in some harmonized manner: when the distance travelled per step tends to 0, the number of steps per time unit should tend to infinity in the right way to have the proportion of the visited area be constant. We proceed to show how the above idea can be made precise.

For $t \geq 0$, consider the sequence $\left(\frac{1}{\sqrt{n}}S_{[tn]}\right)_{n \in \mathbb{N}}$. This is a sequence of positions, rescaled such that it converges to a random variable as $n \rightarrow \infty$, by the central limit theorem. To apply this theorem, we must know expectation and variance of the steps X_i . Obviously $E[X_i] = 0$, and so $var(X_i) = E[X_i^2] - E[X_i]^2 = E[X_i^2] = \frac{1}{2d} \sum_{i=1}^{2d} 1^2 = 1$.

With $E[X_i] = \mu$ and $var(X_i) = \sigma^2$ the central limit theorem states

$$\frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0, 1),$$

and therefore, for $\mu = 0$ and $\sigma = 1$,

$$\sqrt{n} \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \frac{1}{\sqrt{n}} S_n \rightsquigarrow N(0, 1).$$

Now fix $t \geq 0$ and observe that

$$\begin{aligned} \frac{1}{\sqrt{\lceil tn \rceil}} S_{\lceil tn \rceil} &= \frac{1}{\sqrt{\lceil tn \rceil}} \frac{1}{\sqrt{n}} S_{\lceil tn \rceil} \\ &\sim \frac{1}{\sqrt{t}} \frac{1}{\sqrt{n}} S_{\lceil tn \rceil} \rightsquigarrow N(0, 1). \end{aligned}$$

Multiplying by \sqrt{t} we can now see that $\frac{1}{\sqrt{n}} S_{\lceil tn \rceil}$ must for $t \geq 0$ converge in distribution to the normal distribution $N(0, t)$ of expectation 0 and variance t . Because of the independence of the steps X_i , it is also clear that for any $t \geq 0$ and $s \geq 0$ with $t - s \geq 0$, it must hold that $\left(\frac{1}{\sqrt{n}} S_{\lceil tn \rceil} - \frac{1}{\sqrt{n}} S_{\lceil sn \rceil} \right) \rightsquigarrow N(0, t - s)$.

Recall that convergence in distribution does not mean convergence of the sequence of actual values when an experiment is performed (in fact, the limit $\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} S_{\lceil tn \rceil}$ does not exist with probability 1). Therefore it is impossible to treat the desired continuous process as an actual limit of some rescaled random walk, but the convergence in distribution strongly suggests a definition of the limiting process by using the acquired normal distributions, that is, a process with independent increments that are normally distributed.

But first, let us define precisely what we mean by a continuous process, and state some of its properties. A *stochastic process* X is a parametrized collection of random variables

$$(X_t)_{t \in T}$$

defined on a probability space (Ω, Σ, P) and assuming values in \mathbb{R}^d . For our process, $T = [0, \infty)$, and hence it is called continuous.

The *finite-dimensional distributions* of a continuous process X are the measures μ_{t_1, \dots, t_k} defined on $(\mathbb{R}^d)^k$ by

$$\mu_{t_1, \dots, t_k}(B_1, \dots, B_k) = P(X_{t_1} \in B_1, \dots, X_{t_k} \in B_k),$$

where B_i are Borel sets, and $t_i \in T$, for $i = 1, \dots, k$. In other words, the finite-dimensional distributions are the joint laws of the finite collections of random variables out of the continuous process. Properties like continuity of the paths of the process are therefore not determined by the finite-dimensional distributions, and hence it is clear that a process X is not equivalent to its finite dimensional distributions.

Conversely, given a set $\{\mu_{t_1, \dots, t_k} : k \in \mathbb{N}, t_i \in T \text{ for } i = 1, \dots, k\}$ of probability measures on $(\mathbb{R}^d)^k$, under what conditions can a stochastic process be constructed, that has μ_{t_1, \dots, t_k} as its finite-dimensional distributions? Sufficient conditions are given in the following theorem.

Kolmogorov's extension theorem. For $t_1, \dots, t_k \in T$, let μ_{t_1, \dots, t_k} be probability measures on $(\mathbb{R}^d)^k$ such that

1) $\mu_{t_{\sigma(1)}, \dots, t_{\sigma(k)}}(B_1 \times \dots \times B_k) = \mu_{t_1, \dots, t_k}(B_{\sigma^{-1}(1)} \times \dots \times B_{\sigma^{-1}(k)})$ for all permutations σ on $\{1, 2, \dots, k\}$ and all $t_i \in T, i = 1, \dots, k$;

2) $\mu_{t_1, \dots, t_k}(B_1, \dots, B_k) = \mu_{t_1, \dots, t_k, t_{k+1}, \dots, t_{k+m}}(B_1 \times \dots \times B_k \times (\mathbb{R}^d)^m)$ for all $t_i \in T, i = 1, \dots, k$;

Then there exists a probability space (Ω, Σ, P) , and a continuous process $(X_t)_{t \in T}$ on Ω , such that for all $E_i \subset \mathbb{R}^d, i = 1, \dots, k$: $\mu_{t_1, \dots, t_k}(E_1 \times \dots \times E_k) = P(X_{t_1} \in E_1, \dots, X_{t_k} \in E_k)$.

This gives us the existence of some process, whereof only the finite dimensional distributions are known. It tells us nothing about the shape of Ω (which, of course, is not unique), yet the theorem is of crucial importance, since it allows us to consider joint laws of infinite collections of random variables drawn out of the process, and therefore questions on continuity of paths, etc.

Our consideration of the rescaling of the random walk yielded very natural candidates for a collection of measures on $(\mathbb{R}^d)^k$, namely, those to which $\left(\frac{1}{\sqrt{n}}S_{\lceil tn \rceil} - \frac{1}{\sqrt{n}}S_{\lceil sn \rceil}\right)_{n \in \mathbb{N}}$ converge in probability. The procedure for the formal definition of the desired process comprises therefore, firstly, the definition of such a collection of probability measures on $(\mathbb{R}^d)^k$ and secondly, the application of the extension theorem. This is carried out below.

Define $p(t, y) : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \rightarrow \mathbb{R}$ as the joint probability density function of d independent normal random variables with variance $t \geq 0$:

$$p(t, y) = \frac{1}{(2\pi t)^{-\frac{d}{2}}} \cdot e^{-\frac{\|y\|^2}{2t}} \text{ for } y \in \mathbb{R}^d, t \geq 0.$$

In order to define the required probability measures on $(\mathbb{R}^d)^k$, first define for $k \in \mathbb{N}$ and $0 \leq t_1 \leq \dots \leq t_k$ the measure P_{t_1, \dots, t_k} by

$$P_{t_1, \dots, t_k}(E_1 \times \dots \times E_k) = \int_{E_1 \times \dots \times E_k} p(t_1, x_1)p(t_2 - t_1, x_2 - x_1) \times \dots \times p(t_k - t_{k-1}, x_k - x_{k-1}) dx_1 \dots dx_k,$$

where as a convention $p(0, y)dy = \delta_0$ to avoid inconsistencies. Secondly, extend this definition to all finite sequences t_1, \dots, t_k by using the first condition in Kolmogorov's extension theorem. Then also the second condition is satisfied, because p is a probability density (that integrates to 1). So there exists a probability space (Ω, Σ, P) and a continuous process $(B_t)_{t \geq 0}$ on Ω such that the finite-dimensional distributions of B_t are given by the prescribed ones.

This process $(B_t)_{t \geq 0}$ is called *Brownian motion*. The fact that it has independent and normally distributed increments can be easily seen from our definition. The third essential property of Brownian motion is its continuity. To show this, we use another theorem of Kolmogorov.

Kolmogorov's continuity theorem. Let $X = (X_t)_{t \geq 0}$ be a continuous-time process. If for all $T > 0$ there exist $\alpha, \beta, D > 0$ such that

$$E [|X_t - X_s|^\alpha] \leq D|t - s|^{1+\beta} \text{ for } 0 \leq s, t \leq T,$$

then the paths of X are continuous with probability 1, that is, $P(t \rightarrow X_t \text{ is continuous}) = 1$.

We will use this result to show continuity of Brownian motion in dimension $d = 1$. Because $(B_t - B_s) \sim N(0, t - s)$, partial integration gives

$$\begin{aligned} E [|B_t - B_s|^\alpha] &= \int_{-\infty}^{\infty} x^\alpha \frac{1}{\sqrt{t-s}\sqrt{2\pi}} e^{-\frac{x^2}{2(t-s)}} dx \\ &= \frac{1}{\sqrt{t-s}\sqrt{2\pi}} \cdot \frac{\alpha-1}{2\frac{1}{2(t-s)}} \int_{-\infty}^{\infty} x^{\alpha-2} e^{-\frac{x^2}{2(t-s)}} dx \\ &= \frac{1}{\sqrt{t-s}\sqrt{2\pi}} \cdot \frac{\alpha-1}{2\frac{1}{2(t-s)}} \frac{\alpha-3}{2\frac{1}{2(t-s)}} \int_{-\infty}^{\infty} x^{\alpha-4} e^{-\frac{x^2}{2(t-s)}} dx. \end{aligned}$$

Take $\alpha = 4$ to get

$$\begin{aligned} E [|B_t - B_s|^4] &= \frac{1}{\sqrt{t-s}\sqrt{2\pi}} \cdot 3(t-s)(t-s) \sqrt{\frac{\pi}{\left(\frac{1}{2(t-s)}\right)}} \\ &= \frac{3(t-s)^2}{\sqrt{t-s}\sqrt{2\pi}} \cdot \sqrt{2(t-s)\pi} = 3(t-s)^2. \end{aligned}$$

Hence for $\alpha = 4, D = 3, \beta = 1$, Brownian motion satisfies the continuity condition $E [|B_t - B_s|^\alpha] \leq D|t - s|^{1+\beta}$.

Although the Brownian motion as we defined it above is not unique, we do know that for any $\omega \in \Omega$ the function $t \rightarrow B_t(\omega)$ is continuous almost surely. Thus, every $\omega \in \Omega$ gives rise to a continuous function from $[0, \infty)$ to \mathbb{R}^d . In this way, we can think of Brownian motion as a randomly chosen element in the set of continuous functions from $[0, \infty)$ to \mathbb{R}^d according to the probability measure P . Or, in our context: a random continuous path starting at the origin.

References

- [1] Probability with Martingales, D. Williams, Cambridge University Press, 1991.
- [2] Stochastic Differential Equations, B. Øksendal, Springer-Verlag, 1985.
- [3] Principals of Random Walk, F. Spitzer, Springer-Verlag, second edition, 1976.

- [4] Random Walks and Random Environments, B. Hughes, Oxford University Press, 1995.
- [5] Probability: Theory and Examples, R. Durrett, Duxbury Press, second edition, 1996.