2 Random Walks

These motions were such as to satisfy me ... that they arose neither from currents in the fluid, nor from its gradual evaporation, but belonged to the particle itself

(Robert Brown)

Consider a small particle suspended in a liquid. Due to the constant collisions with the surrounding liquid molecules, the path followed by the particle will be erratic, as was first noticed by Robert Brown in the nineteenth century in experiments where he was tracking the motion of pollen grains (Brown 1828). As a result, this is often known as Brownian motion – see also Pearle *et al*. (2010) for a modern take on Brown's experiments, and Fig. 2.1 for a later example of a quantitative study of particle diffusion by Jean Baptiste Perrin, which we will mention again in Chapter 3. This process was modeled by Albert Einstein, who derived the so-called diffusion equation, and understood the nature of the particle's dynamics. In this chapter, we will first study a simplified model for diffusion, where, following Einstein's original derivation from 1905, time will be discrete (i.e., at every time step the particle will move in some random direction). We will understand how far the particle typically gets after making N moves, and what its probability distribution is. These ideas are central in understanding numerous processes around us: from the dynamics of diffusing particles in liquids as well as in living cells, to modeling the dynamics of the stock market (which we will get to later in the book, in Chapter 3). In fact, the concept of "random walks," as this dynamics is often referred to, will also play an important role in our discussion of "Google PageRank," the algorithm at the heart of the search engine.

Asking the right question In science, it is often as important (and hard) to ask the right question than to come up with the right answer. The great statistician Karl Pearson (1905) sent a letter to *Nature* magazine posing, quite literally, the problem of the random walker: given that at every step a person chooses a random direction and walks a constant number of steps, what is the distribution of their position after N steps? It is remarkable that random walks were only introduced in the twentieth century, and a strange coincidence that they were almost simultaneously suggested by Pearson and Einstein.

---

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

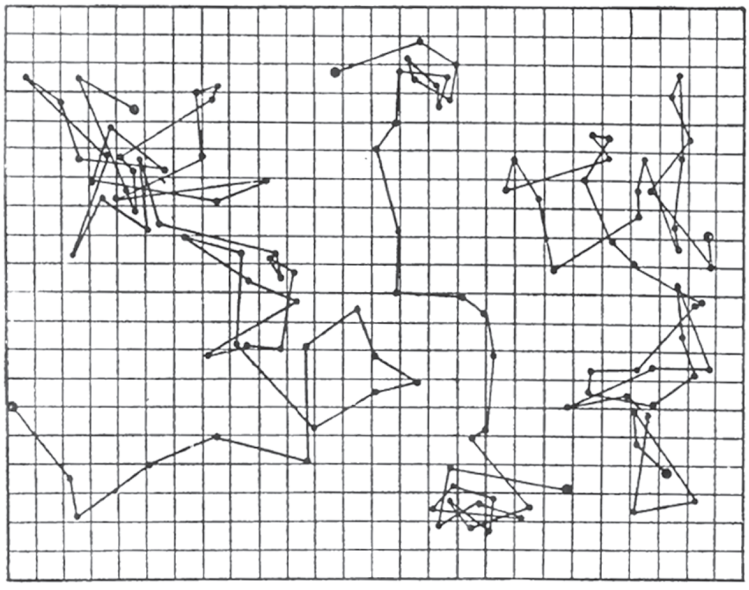

Figure 2.1 An experimental observation by Perrin of the diffusion of small (spherical) particles suspended in a liquid. From Perrin (2013).

2.1 Random Walks in 1D

Consider a "random walker" on a one-dimensional lattice. Time will be discrete, and at every time step the walker will take a step in one of the possible two directions, with equal probability. What is the mean value of, say, coordinate x, after N steps? From symmetry, it is clearly zero. But we certainly don't expect the walker to be precisely at the origin after a large number of steps. Its typical magnitude away from the origin may be found by calculating the RMS (root-mean-square) of the trajectories, i.e., the second moment of the distance from the origin.

Let's say the walker was at position x after N steps. After an additional step, it could be in one of two possible positions. Averaging the squared distance between these two possibilities, conditioned on being at position x at time N, gives us

$$\frac{1}{2}[(x+1)^2 + (x-1)^2] = x^2 + 1. \tag{2.1}$$

Upon averaging over the value of x^2 at time N, we find that

$$\langle x_{N+1}^2 \rangle = \langle x_N^2 \rangle + 1; \tag{2.2}$$

therefore, repeating the argument $N-1$ times, we find that

$$\langle x_N^2 \rangle = N. \tag{2.3}$$

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

This implies that the typical distance from the origin scales like $\sqrt{N}$. What about the position *distribution*? If N is even, then it is clear that the probability to be a distance M from the origin after N steps is zero for odd M, and for even M it is

$$p_M = \frac{1}{2^N}\binom{N}{R} = \frac{1}{2^N}\frac{N!}{R!\,(N-R)!}, \tag{2.4}$$

where R is the number of steps to the right, thus $R-(N-R) = 2R-N = M$. We can now evaluate this probability for $N \gg M$ using Stirling's formula, which provides an (excellent) approximation for $N!$, namely $N! \approx \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$. This leads to

$$\begin{aligned} p_M &\approx \frac{1}{\sqrt{2\pi}}\frac{1}{2^N}\frac{N^{N+1/2}}{R^{R+1/2}(N-R)^{N-R+1/2}} \\ &= \frac{1}{\sqrt{2\pi}}e^{-N\log(2)+(N+1/2)\log(N)-(R+1/2)\log(R)-(N-R+1/2)\log(N-R)}. \end{aligned} \tag{2.5}$$

We can proceed by using our assumption that $N \gg M$, implying that R is approximately equal to $N/2$.

Choosing the appropriate limit In physics and mathematics, it is often as important to analyze the results in the relevant limit as it is to derive the exact results. Based on the previous results we derived, namely, that the variance of the random walker increases linearly in time, it seems plausible that the limit $N \gg M$ will typically hold (i.e., for a large N the probability of M not obeying this condition will be small) – can you see why?

Under this approximation we find that

$$\log(R) = \log\left(\frac{N+M}{2}\right) = \log(N/2(1+M/N)) \approx \log(N/2) + M/N - \frac{1}{2}(M/N)^2, \tag{2.6}$$

where we used the Taylor series expansion $\log(1+x) \approx x - x^2/2 + x^3/3\ldots$, valid for small x. Similarly, we have

$$\begin{aligned} \log(N-R) = \log\left(\frac{N-M}{2}\right) &= \log(N/2(1-M/N)) \\ &\approx \log(N/2) - M/N - \frac{1}{2}(M/N)^2. \end{aligned} \tag{2.7}$$

Putting this together we find that

$$\begin{aligned} &(R+1/2)\log(R) + (N-R+1/2)\log(N-R) \\ &\quad\approx (R+1/2)\left[\log(N/2) + M/N - \frac{1}{2}(M/N)^2\right] \\ &\qquad + (N-R+1/2)\left[\log(N/2) - M/N - \frac{1}{2}(M/N)^2\right]. \end{aligned} \tag{2.8}$$

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

Simplifying leads to

$$(R+1/2)\log(R)+(N-R+1/2)\log(N-R)\approx(N+1)\log(N/2)+M^2/2N. \tag{2.9}$$

Finally,

$$p_M \approx \frac{2}{\sqrt{2\pi N}} e^{-\frac{M^2}{2N}}. \tag{2.10}$$

Hence the distribution is approximately Gaussian. In fact, this had to be the case since x is a sum of independent random variables, hence according to the central limit theorem (see Appendix B for a reminder) it should indeed converge to a Gaussian distribution! Note that the distribution indeed sums up to 1, since the support of this Gaussian is only on even sites.

In summary, we learnt that

1. A diffusing particle would get to a distance $\sim \sqrt{t}$ from its starting position after a time t.
2. The probability distribution describing the particle's position is approximately Gaussian.

Random walks in 1D are strange and interesting! The problem set at the end of the chapter will expose you to some of the peculiarities associated with random walks in one dimension. Problems 2.6 and 2.7 deal with recurrent vs. transient random walks. It turns out that in one- and two-dimensional random walks a walk will *always* return to the origin sometime in the future ("recurrent random walks") – but that this is *not* the case for higher dimensions, where the random walk is called "transient." This is known as Pólya's theorem. Given that 1D random walks are recurrent, we may ask what the "first return time" distribution is – the distribution of the time to return to the origin for the first time. Solving Problems 2.1, 2.3 or 2.9 will show you that this is a power-law distribution, which, remarkably, has diverging mean – so while you *always* return to the origin, the mean time to return is *infinite*! This property is discussed in detail in Krapivsky, Redner, and Ben-Naim (2010). In fact, mean first passage times of random walkers (in any dimension) have a beautiful analogy with resistor networks (see also Doyle and Snell 1984) and relations with harmonic functions. Additional peculiarities arise if we consider the distribution of the *last* time a random walker returns to the origin within a given time interval. This is studied in Problem 2.2 (see also Kostinski and Amir 2016).

2.2 Derivation of the Diffusion Equation for Random Walks in Arbitrary Spatial Dimension

We shall now approach the problem with more generality, following nearly precisely the derivation by Einstein. We will work in 3D but the approach would be the same

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

in any dimension. The approach will have discrete time steps, but the step direction will be a continuous random variable, described by a probability distribution $g(\vec{\Delta})$ (here $\vec{\Delta}$ is a vector describing the step in the 3D space – not to be confused with the Laplacian operator!). We will not limit the random walker to a lattice, though it is possible to implement such a scenario by taking $g(\vec{\Delta})$ to be a sum of δ-functions (can you see how?).

We will seek to obtain the probability distribution $p(\vec{r})$, i.e., $p(\vec{r})dV$ will be the probability to find the particle in a volume dV around the point $\vec{r}$ (at some point in time corresponding to a a given number of steps). If the original problem is cast on a lattice, this distribution will be relevant to describe the *coarse grained* problem, when we shall zoom-out far enough such that we will not care about the details at the level of the lattice constant.

If at time t the probability distribution is described by $p(\vec{r},t)$, let us consider what it will be a time τ later, where τ denotes the duration of each step. As you can guess, in a realistic scenario the time of a step is non-constant, and τ would be the mean step time. Thus, we haven't lost too much in making time discrete – but we did make an assumption that a mean time *exists*. In Chapter 7 we will revisit this point, and show that in certain situations when the mean time diverges, we can get *subdiffusion* (slower spread of the probability distribution over time compared with diffusion).

To find $p(\vec{r},t+\tau)$, we need to integrate over all space, and consider the probability to have the "right" jump size to bring us to $\vec{r}$. This leads to

$$p(\vec{r},t+\tau) = \int p(\vec{r}-\vec{\Delta},t)d^3\vec{\Delta}g(\vec{\Delta}). \tag{2.11}$$

To proceed, we will Taylor expand p, assuming that the probability distribution is smooth on the scale of the typical jump. If we expand it to first order, we will find that

$$p(\vec{r}-\vec{\Delta}) \approx p(\vec{r}) - (\nabla p)\cdot\vec{\Delta}. \tag{2.12}$$

If the diffusion process is isotropic in space, there is no difference between a jump in the $\pm\vec{\Delta}$ direction, so the integral associated with the second term trivially vanishes:

$$\int ((\nabla p)\cdot\vec{\Delta})g(\vec{\Delta})d^3\vec{\Delta} = 0. \tag{2.13}$$

This means we have to expand to second order:

$$p(\vec{r}-\vec{\Delta}) \approx p(\vec{r}) - (\nabla p)\cdot\vec{\Delta} + \frac{1}{2}\sum_{i,j}\left.\frac{\partial^2 p}{\partial x_i \partial x_j}\right|_{\vec{r}}\Delta_i\Delta_j. \tag{2.14}$$

To which order should we expand? In deriving both Eqs. 2.6 and 2.14 we had to decide to which order we should Taylor expand a function. Were we to only retain the first-order term in the expansion, the results would have been nonsensical. In principle, we should make sure that the next order in the expansion is negligible compared with the terms we have kept, but we will often omit this step and rely on our physical intuition instead in deciding the "correct" order of expansion.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

Plugging this into Eq. (2.11), we find that

$$p(\vec{r},t+\tau)-p(\vec{r},t)\approx\int\frac{1}{2}\sum_{i,j}\frac{\partial^2 p}{\partial x_i\partial x_j}\Delta_i\Delta_j g(\vec{\Delta})d^3\vec{\Delta}. \tag{2.15}$$

Once again we may make further progress if we make assumptions regarding the symmetries associated with g: If we assume isotropic diffusion then $g(x,y,z) = g(-x,y,z)$, etc., implying that the only terms that would survive in the integration are the "diagonal" ones, and they would all be equal. Hence

$$p(\vec{r},t+\tau)-p(\vec{r},t)\approx\int\frac{1}{2}\sum_{i}\frac{\partial^2 p}{\partial x_i^2}\Delta_i^2 g(\vec{\Delta})d^3\vec{\Delta}. \tag{2.16}$$

From symmetry, we can replace $\int\Delta_i^2 g(\vec{\Delta})d^3\vec{\Delta} = \frac{1}{3}\int||\vec{\Delta}||^2 g(\vec{\Delta})d^3\vec{\Delta} = \frac{1}{3}\langle\Delta^2\rangle$, where $\langle\Delta^2\rangle$ is the second moment of the jump length distribution. Therefore, we find that

$$p(\vec{r},t+\tau)-p(\vec{r},t)\approx\frac{1}{6}\nabla^2 p\langle\Delta^2\rangle. \tag{2.17}$$

Replacing $p(\vec{r},t+\tau)-p(\vec{r},t)\approx\frac{\partial p}{dt}\tau$, we finally reach the famous diffusion equation

$$\frac{\partial p}{dt}=D\nabla^2 p, \tag{2.18}$$

with $D\equiv\frac{1}{6\tau}\langle\Delta^2\rangle$.

In the case of a random walker in dimension d, repeating the same derivation will reproduce Eq. (2.18) albeit with a diffusion constant $D\equiv\frac{1}{2d\tau}\langle\Delta^2\rangle$.

To which order should we expand? Again! One might argue, correctly, that to better approximate $p(\vec{r},t+\tau)-p(\vec{r},t)$ we should evaluate the partial derivative with respect to time in the middle of the interval $(t,t+\tau)$. Can you see why in the above derivation it suffices to evaluate it at time t?

Notice that on the way we made another "hidden" assumption: that the second moment of the jump distribution *exists*. If it doesn't, the $\sqrt{t}$ scaling that we are familiar with will break down, and this time we will get *superdiffusion* (faster spread of the probability distribution over time compared with diffusion) – this scenario is known as a Lévy-flight. An interesting case arises when the variance of the step size diverges as well as the mean time between steps. Should we get sub or super diffusion in this case? As one may anticipate, the two effects compete with each other, and both options can occur, depending on the details. We will study this in detail later on in the book, in Chapter 7.

Returning to Eq. (2.18), we will now find the probability distribution as a function of space and time for a particle found at the origin at time $t = 0$. To proceed, let us Fourier transform the equation, denoting by $\hat{p}$ the F.T. of the distribution (with respect to space), to find that

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

$$\frac{\partial \hat{p}}{dt} = -Dk^2 \hat{p} \tag{2.19}$$

(see Appendix C for a refresher on Fourier transforms). Here we used the fact that the F.T. of a derivative gives $i \cdot k$:

$$\hat{f}' \equiv \int_{-\infty}^{\infty} \frac{df(x)}{dx} e^{-ik_x \cdot x} dx = f e^{-ik_x \cdot x}|_{-\infty}^{\infty} + ik_x \int_{-\infty}^{\infty} f e^{-ik_x \cdot x} dx = ik_x \hat{f}. \tag{2.20}$$

Note: the F.T. is sometimes defined with an additional $\frac{1}{\sqrt{2\pi}}$ – here this would not have made any difference, but later on in the book we will actually have to be quite mindful of these subtleties.

In this way, we decoupled the different Fourier components, and clearly each of them will decay exponentially, with a rate Dk^2: The long wavelength modes, corresponding to small k, will have diverging timescales. What is the intuition for this? Consider an initial condition where the probability distribution is sinusoidal, with a wavelength λ, as illustrated in Fig. 2.2. Eventually, particles will diffuse around to give a uniform probability distribution, but to do that, they have to cross a characteristic length λ. How long does this take? From the previous analysis we know that $\lambda \sim \sqrt{Dt}$, hence this time will scale as λ^2/D, consistent with the above expression for the decay rate constant of this Fourier mode.

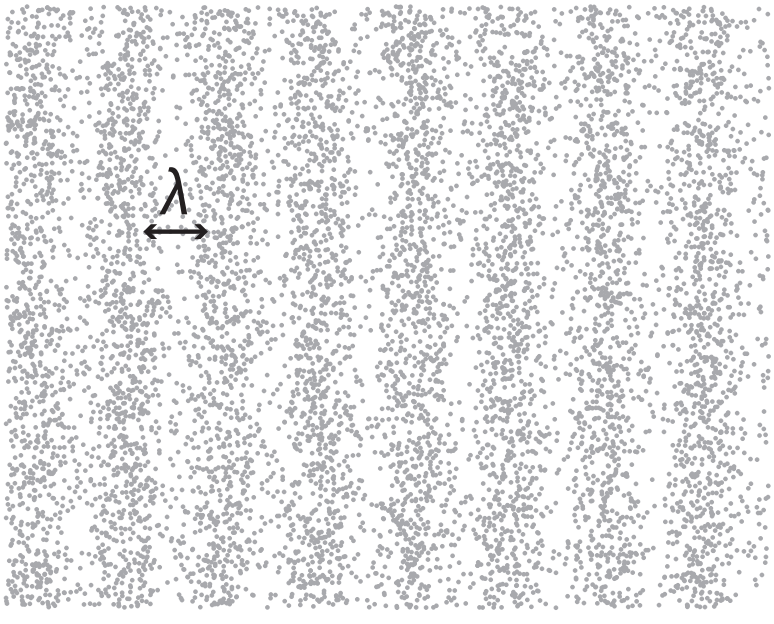

Figure 2.2 An initial configuration in which the density is modulated sinusoidally with a wavelength λ.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

Fourier analysis It is appropriate that we are using Fourier transforms to solve the diffusion equation – back in 1822, Joseph Fourier invented the notion of a Fourier series in order to solve the heat equation – the equation governing the flow of heat in a solid – mathematically identical to the diffusion equation we are studying here.

Next, we assume that our initial condition (at time $t = 0$) has a probability distribution concentrated around the origin (i.e., a δ-function at the origin), and we ask what the probability distribution is at some later time t. In fact, the solution of this problem is known as *Green's function*, and knowledge of it will allow us to cast the solution of *any* initial condition in terms of a convolution with Green's function – see Problem 2.4. Since the F.T. of a δ-function at the origin is a constant (i.e., all Fourier modes are equally excited) we have

$$\hat{p}(\vec{k}, t = 0) = 1. \tag{2.21}$$

According to Eq. (2.19), we have

$$\hat{p}(\vec{k}, t) = e^{-Dk^2 t}. \tag{2.22}$$

Scaling approach to solving the equation In Problem 2.11 a different route to solution of the diffusion equation in 1D is presented, which relies on guessing a scaling solution of a particular form. This is a commonly used method to solve PDE's of this sort, and is useful also for nonlinear PDE's where the Fourier analysis approach presented above fails.

The last step would be to inverse-Fourier this function to find what the probability distribution looks like in real space:

$$p(\vec{r}, t) = \frac{1}{(2\pi)^3} \int e^{-Dk^2 t} e^{i(k_x \cdot x + k_y \cdot y + k_z \cdot z)} dk_x dk_y dk_z. \tag{2.23}$$

Clearly, this separates into a product of three integrals (which means we can actually know the answer at this stage without further calculations if we draw on the 1D case!). Let us evaluate one of them:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-Dk_x^2 t} e^{ik_x \cdot x} dk_x = ?. \tag{2.24}$$

To proceed, we will use a trick that is often used in the context of Gaussian integrals, "completing the square":

$$I = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-Dt(k_x - \frac{ix}{2Dt})^2} e^{-\frac{x^2}{4Dt}} dk_x. \tag{2.25}$$

The last term can be taken out of the integral, since it does not depend on k_x, hence

$$I = \frac{1}{2\pi} e^{-\frac{x^2}{4Dt}} \int_{-\infty}^{\infty} e^{-Dt(k_x - \frac{ix}{2Dt})^2} dk_x. \tag{2.26}$$

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

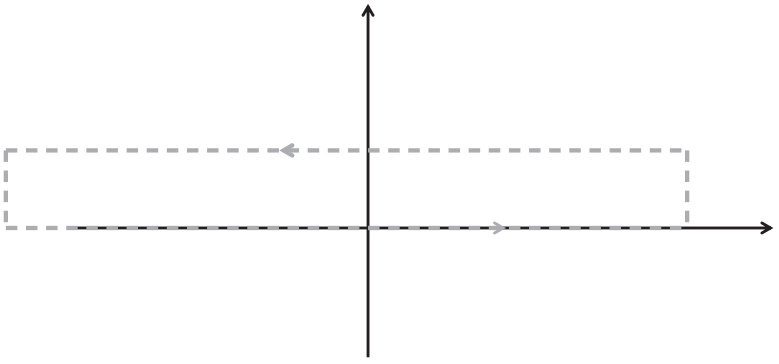

Figure 2.3 Consider the contour integral of the analytic function $f(z) = e^{-Az^2}$.

To evaluate the integral over the first term, it is tempting to make a change of variables and get rid of the $\frac{ix}{2Dt}$ term. This is incorrect, of course, and more care has to be taken when dealing with complex variables (see Appendix C for a brief summary of contour integrals and the residue theorem). It is instructive to consider a contour integral in the complex plane, as is shown in Fig. 2.3. Since the function is analytic and the contributions of the vertical parts of the contour are negligible, we find that

$$\int_{-\infty}^{\infty} e^{-Dt(k_x - \frac{ix}{2Dt})^2} dk_x = \int_{-\infty}^{\infty} e^{-Dtk_x^2} dk_x = \sqrt{\frac{\pi}{Dt}} \tag{2.27}$$

(see Appendix A.1.2 for a detailed calculation of this one-dimensional Gaussian integral).

Putting everything together we conclude that

$$p(\vec{r},t) = \frac{1}{(\sqrt{4\pi Dt})^3} e^{-\frac{r^2}{4Dt}} \tag{2.28}$$

(as a sanity check, you can verify that the probability distribution is normalized).

Therefore, the probability distribution is Gaussian, and its variance is

$$\int r^2 p(\vec{r},t) d^3r = \int 3x^2 p(\vec{r},t) d^3r = 3 \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} dx. \tag{2.29}$$

This can be evaluated (how?) to give

$$\langle r^2 \rangle = 6Dt. \tag{2.30}$$

Random walks are useful Einstein's derivation was geared toward understanding diffusion of particles, but random walks have found applications in many other fields of science. Problem 2.10 explores the application of random walks to polymer physics. The next section will show why studying random walks on *networks* can lead to a useful algorithm that formed the beginning of a little search engine called Google.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

In future chapters, we will also see how building on the ideas of random walks in the context of economics will lead to the celebrated Black–Scholes equation (in fact, a few years earlier than Einstein the ideas of random walks were introduced by Louis Bachelier (1900) in the context of finance).

2.3 Markov Processes and Markov Chains

Let us approach our analysis of the random walker from a new perspective, which will be easier to generalize to other networks. Within our model for diffusion on a 1D lattice, the probability to go to site j does not depend on the history of the random walker, but only on its current state – this is an example of a *Markov process* (named after Andrey Markov). A familiar childhood example of a Markov process is the game Chutes and Ladders (as well as many other board games – see also Problem 2.15). But we emphasize that whether a process is Markovian depends on the space: Consider, for example, a game where we throw a die at each round, and keep track of the running sum. This is clearly a Markov process – knowing the current sum determines the probabilities to go to the next states. But what about a game in which we reset the sum to zero each time we get two identical numbers in a row? In this case, the process is memory dependent, so it is non-Markovian. But, if we work in the space where a state is defined by a *pair*, the running sum, and the result of the last throw – then the process becomes Markovian again. It is also worth noting that in cases where time steps are discrete, the process is referred to as a *Markov chain*. In what follows we will deal with Markov chains, though in the next chapter we will study Markov processes with *continuous* time. For an extended discussion of Markov chains, see Feller (2008).

Let us denote by $\boldsymbol{P}$ the matrix describing the transition probabilities, i.e., $\boldsymbol{P}_{ij}$ will be the probability to go from i to j. The key insight to note is that for a Markov chain if we know the current probabilities to be at every site, which we will denote by the vector $\vec{p}$, and we know the matrix P, we can easily find the vector of probabilities to be at every site after an additional move. By the definition of the matrix P, the probability to be in the ith site after an additional move is given by

$$p_i^{n+1} = \sum_j p_j^n \boldsymbol{P}_{ji}. \tag{2.31}$$

This can be written in more concise notation using the matrix formulation

$$\vec{p}_{n+1} = \boldsymbol{P}^T \vec{p}_n, \tag{2.32}$$

where ^{T} denotes the transpose operation, and $\vec{p}_n$ a column vector the entries of which are the probabilities to be at a given site after n steps.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

If we start at, say, the first site, then our initial vector of probabilities $\vec{p}_0$ will be

$$\vec{p}_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.33}$$

and if we would like to find it after N steps, we simply have to operate with the matrix P^T on this vector N times. In other words:

$$\vec{p}_N = [\boldsymbol{P}^T]^N \vec{p}_0. \tag{2.34}$$

This operation is computationally cheap, hence this is a much more efficient way of getting the answer than the simulation of many random walks and averaging over them.

2.4 Google PageRank: Random Walks on Networks as an Example of a Useful Markov Chain

We shall now describe a beautiful application of the concepts of random walks and Markov chains that contributed to one of the most influential algorithms of all time, as first described by Larry Page *et al*. (1999) in a seminal paper (which was the basis for Google).

The internet is a complex, directed graph. Imagine building a search engine, and attempting to respond to a particular search query. We would first have to find the relevant webpages (by searching the words in the text), and then choose the most appropriate one. It turns out that prior to 1998 the latter step was the more challenging, and that was the novelty which Google brought.

The idea behind PageRank is simple: We will model "surfing" on the WWW as a random walk process. We will assume that when a surfer reaches some webpage, with probability d (the "damping factor") he or she will continue to one of the pages that the current one links to, with equal probability. In practice, $d = 0.85$ is apparently used. With probability $1 - d$, the surfer is assumed to go to an arbitrary page in the web, with equal probability for all pages. The idea now is to associate the "merit" of a webpage with the (relative) number of times (i.e., frequency) that a surfer will visit the webpage starting from an arbitrary page. We will define the PageRank as the probability of being at a particular webpage in the limit of spending an infinite time surfing. Such a process only makes sense if an identical stationary distribution is reached starting from all initial conditions – which we will show is indeed the case. In other words, if the random walker surfs the web for a very long time, we will show that the probability to be in each page will approach a constant value (different for each page, usually), and that will be the "PageRank." Finally, if a page has no outgoing links, the random walker will choose a page at random from all pages, which is mathematically equivalent to having that page link to *all* other pages (including itself).

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

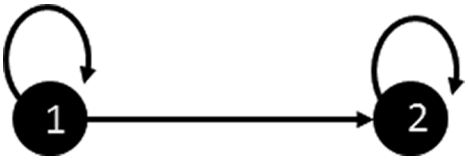

Figure 2.4 A simple network, with only two webpages.

Validating PageRank It is interesting to think how Page *et al*. knew that their newly developed search engine was indeed superior – how can that be done if there is no "benchmark" to compare against? The curious reader can look at their original paper for their "solution" to this.

2.4.1 Simple Example of the Dynamics of a Random Walker on a Network

Let's first illustrate how PageRank works on a simple example. Consider the page architecture shown in Fig. 2.4. (Note that it is also important to specify whether the webpages link to themselves or not.) As before, we denote by $\boldsymbol{P}$ the matrix describing the transition probabilities (i.e., $\boldsymbol{P}_{ij}$ will be the probability to go from i to j). For a damping factor d the matrix will be

$$\boldsymbol{P} = (1-d)\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} + d\begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ \frac{1-d}{2} & \frac{1+d}{2} \end{pmatrix}. \tag{2.35}$$

Assuming that we start in page (1), what is the probability to be in page (1) after N steps? What happens if we ask the same question starting with page (2)?

As Eq. (2.34) shows, we have to take $\boldsymbol{P}^t$ to the Nth power.

This can be done more economically by finding the eigenvalues and eigenvectors of the matrix (see Appendix B if a refresher is needed).

Its characteristic polynomial is

$$|\lambda \boldsymbol{I} - \boldsymbol{P}| = (\lambda - 1/2)(\lambda - (1+d)/2) - \frac{1-d}{4}. \tag{2.36}$$

The eigenvalues are the roots of this polynomial. It is easy to check that $\lambda = 1$ is a root – we will shortly see why this has to be the case in general. Since the trace of the matrix is equal to the sum of eigenvalues, the second one must be such that they sum up to $1 + d/2$, hence it is $d/2$. We can now write the initial probability vector as a sum of the two eigenmodes, i.e.,

$$\vec{v} = a_1\vec{v}_1 + a_2\vec{v}_2. \tag{2.37}$$

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

The advantage of working with the eigenmodes is that after operating with P^T we will get $\lambda_j \vec{v_j}$, hence after operating with $(P^T)^N$ we will get $\lambda_j^N \vec{v_j}$. Therefore, the probability to be at (1) after N operations has the form:

$$p_1^N = a_1 \lambda_1^N (\vec{v_1})_1 + a_2 \lambda_2^N (\vec{v_2})_1. \tag{2.38}$$

We know that $\lambda_1 = 1$ and $\lambda_2 = d/2$, hence

$$p_1^N = c_1 + c_2 (d/2)^N. \tag{2.39}$$

It remains to find c_1 and c_2. One way to do it would be to find the eigenmodes, and then find a_1 and a_2. But a shortcut would be to use the fact that $p_1^0 = 1$ and $p_1^1 = 1/2$, hence $c_1 + c_2 = 1; c_1 + c_2(d/2) = 1/2 \Rightarrow c_1 = \frac{1-d}{2-d}, c_2 = \frac{1}{2-d}$.

Note that, in this example, the probability to be at (1) decays exponentially to its asymptotic value $\frac{1-d}{2-d}$, with a characteristic timescale which is associated with the second eigenvalue. In general, this exponential decay can be oscillatory – corresponding to complex eigenvalues (the initial matrix is typically non-Hermitian, so the eigenvalues are not guaranteed to be real – see Appendix B for a concise summary of these linear algebra facts). As a sanity check, note that for $d = 1$ the asymptotic value to be at (1) vanishes, while for $d = 0$ it is 1/2.

Repeating the same analysis when starting with page (2), Eq. (2.39) is still intact, but now $c_1 + c_2(d/2) = \frac{1+d}{2}$, leading to $c_1 = \frac{1}{2-d}$, $c_2 = \frac{1-d}{2-d}$. Thus, at long times the probability to be at (2) would approach $\frac{1}{2-d}$, and the probability for (1) would approach $\frac{1-d}{2-d}$ – just as they have for the other initial condition. We will now generalize this example for arbitrary networks.

2.4.2 Existence and Uniqueness of the Stationary Distribution in the Context of PageRank

It is natural to ask whether a stationary distribution exists – one that is invariant under the operation of the matrix P^T – and whether we always converge to it starting from any initial conditions. We shall see that for the PageRank setup with damping factor $d < 1$ such a distribution indeed exists and we necessarily converge to it. For $d = 1$ we can construct examples where the probability vector would never converge to a constant one. For example, consider a network with two pages, and

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.40}$$

If we start in page (1), then the probability to be at (1) will oscillate forever between the values of 0 and 1, and hence does not converge to a constant. This non-convergent behavior will not occur once we put in the "damping factor," as we shall shortly prove.

First, let us prove that there is an eigenvector with eigenvalue of precisely 1. For the matrix P, this is easy to see, by considering the product of this matrix by a vector $\vec{v}$ the entries of which are all equal to 1. We have

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

$$\sum_j \boldsymbol{P}_{ij} v_j = \sum_j \boldsymbol{P}_{ij} = 1 = v_i. \tag{2.41}$$

Therefore, there must also be an eigenvalue 1 for the matrix P^T (since they share the same characteristic polynomial). The corresponding eigenvector will therefore not decay in time, and is a proof for the existence of a stationary distribution. What about its uniqueness? To establish that, we will show that all other eigenvalues will have magnitude smaller than 1. Consider an eigenvector with eigenvalue λ and entries v_j. Let us assume that the entry with the largest magnitude is the mth one. Without loss of generality, we can choose it to be real and positive (can you see why?). Consider now the matrix P operating on the vector, and in particular the summation corresponding to the mth element. It reads:

$$\sum_j \boldsymbol{P}_{mj} v_j = \lambda v_m. \tag{2.42}$$

Additionally, we have $\sum_j \boldsymbol{P}_{mj} = 1$. Considering that the entries of $\boldsymbol{P}$ are positive, this implies that

$$\left| \sum_j \boldsymbol{P}_{mj} v_j \right| \leq v_m \sum_j \boldsymbol{P}_{mj} = v_m. \tag{2.43}$$

Hence all eigenvalues have a magnitude smaller than or equal to 1. When can the equality occur? In order for it to occur, we need every v_j that multiplies a non-vanishing matrix element $\boldsymbol{P}_{mj}$ to be equal to v_m. But for the PageRank matrix, for $d < 1$, all matrix entries are nonzero, hence *all* the elements of $\vec{v}$ have to be equal – proving that the only eigenvalue with magnitude 1 is the eigenvector we guessed previously. Think about the oscillatory example shown above, and explain why the eigenvalue -1 of that matrix does not violate this proof! Note that while this proof was for the eigenvalues of $\boldsymbol{P}$, the same holds for its transpose (since they share the same characteristic polynomial). Also, note that for symmetric matrices the unique eigenmode the entries of which are equal will be proportional (up to trivial normalization) to the stationary distribution. The fact that the matrix corresponding to the WWW is asymmetric is essential for this algorithm to be useful!

What we just proved is part of the Perron–Frobenius theorem (the full theorem being slightly more general (Feller 2008)).

2.4.3 Explicit Formula for PageRank

By now we understand that a unique eigenvector with eigenvalue 1 exists, and understand why it is a potentially good way of ranking the web. For the case of PageRank we can write an explicit formula for it. Consider the product of $\boldsymbol{P}^T$ on the PageRank vector. This, as we now know, results in precisely the PageRank vector. We also note that $\boldsymbol{P}^T_{A,B}$ corresponds to the transition rate from page B to page A, which for the PageRank algorithm equals $\frac{1-d}{N}$ if the page *doesn't* link to A, and $\frac{1-d}{N} + \frac{d}{deg(B)}$ if it

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

does, where $deg(B)$ is the *outgoing* degree of the node B (i.e., the number of pages that it links to). Calculating explicitly the Ath entry of the product of $\boldsymbol{P}^T$ and the PageRank vector then leads to

$$PR(A) = \frac{1-d}{N} + d \sum_{B \to A} \frac{PR(B)}{deg(B)}, \tag{2.44}$$

where the sum is over pages which have a link to A.

Another way of thinking about it You can view Eq. (2.44) as basically saying that the importance of a page is the weighted sum of the importance of all the pages that link to it. How important are *they*? The equation applies in the same way to all pages, and that's why we get an eigenvalue problem.

In fact, there are many other contexts in which *ranking* a set of objects (not webpages) is useful. Perhaps not surprisingly given its simplicity, the PageRank algorithm, although most prominently utilized by Page *et al.* in the context of ranking webpages, was actually used earlier in various contexts. For example, Pinski and Narin (1976) used it in the context of ranking scientific journals and Keener (1993) in the context of ranking football teams. See also Ermann, Frahm, and Shepelyansky (2015) for various additional applications and an extended discussion of the algorithm.

2.4.4 Convergence to Stationary Distribution Starting from Arbitrary Initial Conditions

So far we have proved the uniqueness and existence of a stationary distribution, and saw one example where starting from some initial conditions we converge to it. But how do we know that this is true in general? (For example, perhaps the system can undergo oscillations between different states.)

Typically, the transition matrix P will be diagonalizable: It will be similar to a diagonal matrix, which has the eigenvalues on the diagonal. In the basis in which P is diagonal, taking the matrix to a large power would imply that all eigenvalues will decay to 0 (since their magnitude is smaller than 1) except for the one corresponding to the stationary distribution: Hence we will converge to it, no matter what the initial conditions are.

The situation is trickier in cases where the matrix is not diagonalizable (can you think of one that obeys our assumptions?) In that case, although the matrix is *not* similar to a diagonal matrix, it can be shown that it is similar to an *upper triangular* matrix (for example, by using the Schur decomposition or the Jordan form of the matrix). This matrix still has the N eigenvalues on the diagonal. It is then possible to show that taking that matrix to a large power leads to a matrix where all but one row vanish – implying the convergence to the stationary distribution; Can you see why? (see also Problem 2.14).

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

2.5 Relation between Markov Chains and the Diffusion Equation

Let us now relate the PageRank formulation to the diffusion equation which we derived before, focusing on the 1D case for concreteness. It will also be convenient to work with periodic boundary conditions – so you should envision a set of N sites organized at regular intervals on a circle, with the random walker jumping clockwise or anticlockwise at every time step. The matrix corresponding to this network for $N = 5$ is

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{pmatrix}. \tag{2.45}$$

In the following we will prove that for odd N there will be convergence to a unique stationary distribution for any initial conditions (note that the matrix does not satisfy the PageRank conditions, so this is not obvious a priori). Since the matrix is symmetric, the stationary distribution is uniform on all sites. But how long will it take the stationary distribution to be established, if we start at one of the sites? According to the previous analysis in terms of the eigenmodes, this time is set by the eigenvalue with the slowest decay rate, which is the one with the second to largest real part (since the stationary distribution corresponds to that with the largest real part, namely 1). In fact, this matrix is real and symmetric so all eigenvalues are real. Moreover, its particular structure allows for an exact solution of the eigenvalues and eigenmodes.

Let us "guess" a solution in the form of plane-waves, i.e.,

$$A_d = e^{ikd}. \tag{2.46}$$

Multiplying this vector by the matrix $\boldsymbol{P}$, we find the following equation for every entry c save for the first and the last:

$$\sum_d \boldsymbol{P}_{cd} A_d = \sum_d \boldsymbol{P}_{cd} e^{ikd} = \frac{1}{2}\left(e^{ik(c-1)} + e^{ik(c+1)}\right) = \lambda A_c = \lambda e^{ikc}. \tag{2.47}$$

This would indeed be a solution, as long as

$$\lambda = \cos(k). \tag{2.48}$$

It remains to consider the equations for the first and last entries. It is easy to see that they would also be satisfied if and only if

$$e^{ikN} = 1. \tag{2.49}$$

This implies that the allowed values for k are: $k = 2\pi M/N$, with M an integer. It is easy to check that taking $M = 0, 1..N-1$ leads to N eigenvectors which are all *orthogonal* to each other, and thus are independent and span the vector space.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

Relation to Fast Fourier Transform It is interesting to point out that the eigenmodes of this particular matrix play a pivotal role in a remarkable algorithm known as "Fast Fourier Transform" (FFT), which allows one, among other applications, to multiply two N digit number using $O(N \log N)$ operations rather than the naive N^2 you might expect (Knuth 1998).

Since the eigenvalues are equal to the cosine of k, it is clear that other than the unique $\lambda = 1$, all other eigenvalues are doubly degenerate (save the eigenvalue -1 for even N). The (doubly degenerate) eigenvalue closest to 1 (but distinct from it) corresponds to $M = 1$ or $M = N - 1$, and its value is

$$\lambda = \cos(2\pi/N) \approx 1 - \frac{1}{2}(2\pi/N)^2. \tag{2.50}$$

Therefore, after t steps the amplitude corresponding to this eigenmode would decay as

$$\lambda^t = \left(1 - \frac{1}{2}(2\pi/N)^2\right)^t. \tag{2.51}$$

We can now use the relation $(1 - x/t)^t \approx e^{-x}$, to approximate

$$\lambda^t \approx e^{-\frac{t}{2}(2\pi/N)^2}. \tag{2.52}$$

In fact, if one wants to quantify the correction to this approximation, it is helpful to write

$$(1 - x/t)^t = e^{t \log(1-x/t)} \approx e^{-x - x^2/2t}, \tag{2.53}$$

hence the correction is negligible when $x \ll \sqrt{t}$, which for our problem implies $N^2 \gg \sqrt{t}$.

From Eq. (2.52), we see that the relaxation time is $\frac{N^2}{2\pi^2}$. This is plausible since in order to homogenize the distribution on the scale of the ring, we need to diffuse over N sites, which demands $O(N^2)$ steps – according to our previous results on 1D diffusion. Note that for times of the order of the relaxation time, the condition $N^2 \gg \sqrt{t}$ reduces to $N \gg 1$, thus validating the approximation previously used.

Another nice way to see the relation between the "discrete" approach and the continuous one is by considering the *difference* equation for the discrete probabilities at the sites, p_i. It is given by

$$\delta \vec{p} = \tilde{\boldsymbol{P}} \vec{p}, \tag{2.54}$$

with $\tilde{\boldsymbol{P}} = \boldsymbol{P} - I$, I being the identity matrix.

In the continuum limit, the LHS will become a time derivative ($\delta p_i \approx \frac{\partial p}{\partial t}\tau$, with τ the duration of each step), while the RHS is proportional to the discrete version of the second derivative operator: Its ith component is given by

$$\frac{1}{2}(p_{i-1} + p_{i+1} - 2p_i) = \frac{1}{2}[(p_{i+1} - p_i) - (p_i - p_{i-1})] \approx \frac{1}{2}(p'|_{i+1} - p'|_i) \approx p''/2. \tag{2.55}$$

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

Hence in the continuum limit this 1D random walk leads to the diffusion equation, with $D = \frac{1}{2\tau}$.

2.6 Summary

In this chapter we introduced the important concept of a random walk – on which the next chapter will build. We used "brute-force" combinatorics to show why in the 1D case asymptotically the probability distribution of the random walker is Gaussian, with a variance that increases linearly in time – the hallmark of diffusion. We then followed Einstein's derivation to derive the diffusion equation in *any* dimension (under certain simplifying assumptions which we will relax in Chapter 7). Solving the diffusion equation showed that the insights gained from the 1D analysis remain intact. The problems at the end of this chapter further substantiate the various interesting properties and diverse applications of random walks in various spatial dimensions.

We next considered random walks on *networks*, which led us to a more general discussion of Markov chains and of the Perron–Frobenius theorem. This theorem forms the basis of the influential Google PageRank algorithm, which we also reviewed. Finally, we showed how one can relate the Markov chain formalism to the 1D diffusive behavior studied early on in this chapter, with an interesting mathematical relation to the FFT algorithm.

For further reading There are many excellent books dealing with random walks. Some useful resources are Feller (2008); Gardiner (2009); Schuss (2009); Paul and Baschnagel (2013); Hughes (1996); Krapivsky, Redner, and Ben-Naim (2010), all of which have extended discussions of many of the topics covered in this chapter.

2.7 Exercises

2.1 1D Random Walks and First Return Times*

Consider a 1D symmetric random walk, with step size a and time step T.

(a) Show that in the continuum limit (i.e., for times $\gg T$ and spatial resolution $\gg a$) the walker is described by the diffusion equation $\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}$, and find D.
(b) What is the probability that the walker is found precisely at the origin after the first N steps? In the continuum version, what is the probability *density* to be at the origin after time t?
(c) What is the probability that the walker is always to the right of the origin within the first N steps? *Hint*: Enumerate such paths by mapping this problem to a circular track on which are written N numbers, a fraction of which are $+1$ and the rest are -1. What is the probability that the sum of numbers along the track

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

(say, in the clockwise direction) starting from one of the N points is always positive? Which of the N points cannot be possible starting points? [See also Kostinski, S. and Amir, A. An elementary derivation of first and last return times of 1D random walks. *American Journal of Physics* 84 (1), 57–60 (2016) for an extended discussion].

(d) Use the previous results to prove that a 1D walk is recurrent, i.e., that in the course of time the walker will return to the origin with probability 1.

(e) In the continuum limit characterized by the diffusion constant D, what is the probability distribution of the time for the walker to first return to the origin, $F(t)$? What is the average time to return to the origin?

2.2 1D Random Walks and *Last* Return Times

Consider a discrete 1D random walk, with step size a, time step T, and number of steps N.

(a) Suppose that the walk ends at a time t_{tot}. What is the probability $p(t)$ to return to the origin for the last time at some time $t < t_{tot}$? *Hint:* consider the discrete case and the results of Problem 2.1, then go to the continuum limit.

(c) Assuming that the amount by which one sports team leads another is well described by a random walk, what does the result of (a) suggest about lead changes in sports games?

[See also: Clauset, A., Kogan, M., and Redner, S., Safe leads and lead changes in competitive team sports, *Physical Review E*, 91(6), p.062815 (2015), as well as the reference pointed out in Problem 2.1(c)].

2.3 First Passage Times in 1D

Consider a particle undergoing diffusion on a 1D lattice, starting at the origin. The objective of this problem would be to find the probability distribution of the time it takes the particle to reach a site a distance $N \gg 1$ lattice constants away to the right of the origin (this is known as the First Passage Time – FPT).

(a) Set up the appropriate PDE to describe $P(x,t)$, and define the relevant boundary conditions for finding the FPT distribution.

(b) Find an explicit solution of the PDE. *Hint:* consider superimposing the solution of the PDE on the infinite 1D line with $\delta(r)$ initial condition at $t = 0$ with the solution on the infinite line with an initial condition of $A\delta(r - r_0)$, and find the appropriate values for A and r_0.

(c) Using the results of part (b), what is the probability to hit the target at any time? What is the mean time to reach the target?

2.4 Green's Function Warmup

Consider a particle with uncertain initial position following the probability distribution $p_0(x)$. Show that the position distribution at time $t > 0$ is given by a convolution of the form

$$p(x,t) = \int_{-\infty}^{\infty} \varphi(x - y, t) p_0(y) dy.$$

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

2.5 Biased Random Walks

(a) Repeat Einstein's logic in deriving the diffusion equation in two dimensions, *without* assuming isotropy.
(b) Solve the resulting equation.

2.6 Recurrent and Nonrecurrent Random Walks

Consider a random walker on a d-dimensional lattice.

(a) Define u as the probability to return to the origin. What is the probability to visit the origin precisely k times?
(b) What is the expected number of times to visit the origin? Show that it is finite if and only if $u < 1$.
(c) In the continuum limit, what is the probability density to be at the origin after time t?
(d) Use the results of (b) and (c) to show that 2D walks are recurrent while for $d > 2$ they are not. This is known as Pólya's theorem, named after the mathematician George Pólya.

It has been suggested that this is why when moles look for their mates they stick to a given depth underground!

2.7 Recurrent and Nonrecurrent Random Walks Revisited

(a) Consider a possibly asymmetric random walk in one dimension. A random walker starts at $x = 0$, and at every time step, the walker takes a step to the right with probability p and a step to the left with probability $q = 1 - p$, as illustrated in Fig. 2.5.

Determine whether the random walk is recurrent. Recall that a random walk is said to be recurrent if it eventually returns to the origin with probability 1.
Hint: Prove first that a random walk is recurrent if and only if the sum

$$\sum_{n=1}^{\infty} p_0^n$$

is finite. Here p_0^n is the probability of the walker being at $x = 0$ at time n.
(b) Determine whether a symmetric random walk in d dimensions is recurrent.
Hint: Since the exact expression for p_0^n is quite complicated for $d > 1$, you may wish to use an approximation.

Problem credit: Pétur Rafn Bryde.

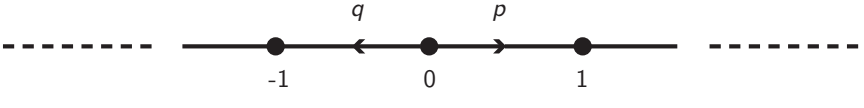

Figure 2.5 An asymmetric random walk.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

2.8 Gambler's Ruin*

Consider a discrete 1D random walk, with step size a and time step T. The walker starts at a point c with $0 < c < L$ ($L \gg a$ and $c \gg a$). In this problem we would like to find the probabilities for the walker to get to the origin or point L. Throughout the problem, it will be helpful to work in the continuum limit (see Problem 2.1(a)).

(a) Impose *absorbing* boundary conditions at 0 and at L. Once the walker reaches 0 or L, the walker cannot come back to the segment $[0, L]$. Find the eigenfunctions of the operator $\frac{\partial^2}{\partial x^2}$ on the line $[0, L]$.
(b) Write an expression for the initial probability distribution (a δ-function) as a sum of the eigenfunctions you got in part (b).
(c) Write out the probability distribution after time t ($p(x,t)$). (*Hint:* try using an ansatz where the space and time variables are separated $p(x,t) = f(x)g(t)$).
(d) Find expressions for the probability flux to exit the segment $[0, L]$ at 0 or L at time t (you may leave this as an infinite series).
(e) Find expressions for the overall probability to exit the segment at 0 or L (you may leave this as an infinite series).
(g) **Gambler's Ruin**: A gambler starts off with \$100. At every round the gambler flips a fair coin, and either wins or loses \$1 with equal probability. The game ends when they either have \$1000 or go bankrupt. What is, approximately, the probability for each of these two events to occur? (You may either sum up the series or approximate it numerically by keeping the first few terms).
(h) Sum up the infinite series. You will find a very simple expression for the exit probability.

2.9 First Passage Time – Again!

Consider a 1D symmetric random walk, with step size a, time step T, starting at the origin.

Based on the results from Problem 2.8 (Gambler's Ruin), find the probability distribution of the time it takes the random walker to reach a site a distance X away from the origin (the first passage time). What is the mean first passage time? *Hint:* Obtain a closed-form expression for the FPT distribution by replacing the series with an integral in the appropriate limit.

2.10 Modeling Polymers as Random Walks

In this problem, we will simulate a variant of a random walk on a 2D lattice that doesn't intersect itself. At each step, the random walker can go to one of the *empty* neighboring sites; note that it can get stuck after some time, in which case we should discard the simulation. Run the simulation for a total step number of 100. Repeat as many times as you need to achieve 1000 random walks (that did not terminate). Plot the average $\langle R^2 \rangle$ and N in a log-log scale. What is the empirical power-law?

This problem is related to the self-avoiding random walks introduced by P. Flory in 1953 to model polymers. For an extended discussion, see Hughes, B. D. *Random Walks and Random Environments*, Clarendon Press, Oxford (1996).

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

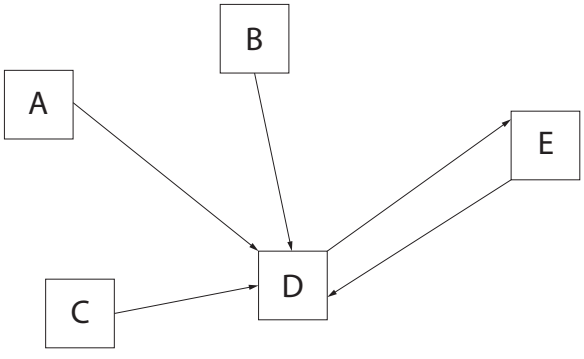

Figure 2.6 A simple network.

2.11 Scaling Approach to Solving the Diffusion Equation

In this problem we will use a different approach to solve the diffusion equation in one dimension, using scaling.

(a) Motivated by the result $\langle x^2 \rangle \propto t$, it might seem plausible to search for a solution of the form $p(x,t) = C(t)f(xt^{\alpha})$. What must the value of α be? Why must the time-dependent factor $c(t)$ be included, and what must be its form?
(b) Obtain an ODE (Ordinary Differential Equation) for the function $f(z)$ by plugging the ansatz of part (a) into the 1D diffusion equation.
(c) Solve the ODE. *Hint:* it might be useful to use the simple relation $[gz]' = g'z+g$.

2.12 PageRank

Consider the web shown in Fig. 2.6.

(a) How would Google's PageRank algorithm rank pages A–E for a "damping factor" of 0.85?
(b) If a walker starts at C at $t = 0$, what is the probability of being at site C after 2 moves? 5 moves? 20 moves? How is your answer related to part (a)?
(c) Using software of your choice, simulate a "random web surfer" on pages A–F, starting at page C, and compare the frequency of visits to page C to your answers in part (b).

2.13 Hacking PageRank

Suppose we have a database with N pages that are completely linked to each other (the corresponding graph is *complete*). A spammer adds their page, probability.com, to the database and wants probability.com to be ranked as high as possible in PageRank. All pages are also linked to themselves.

(a) Add k fake pages to the database and link all of them to probability.com. There are no links between the spammer's pages and the N pages. What is the PageRank of probability.com?
(b) Hack into k of the N pages in the database and link them to probability.com. What is the PageRank of probability.com?

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

(c) Which of (a) and (b) gives the higher PageRank for probability.com? As an example, take $N = 100, k = 10$, and $d = 0.85$. Then, consider the case $N \gg k$.

2.14 PageRank Analysis for Matrices That Cannot Be Diagonalized*

(a) Give an example of a stochastic matrix (i.e., for which the rows sum up to one), for which the eigenvectors do not span the vector space.
(b) Consider now the PageRank scenario. By utilizing the Jordan form or Schur decomposition, prove that starting from any initial condition the probabilities to be at each site would converge to those associated with the unique $\lambda = 1$ eigenvalue (without assuming that the matrix is diagonalizable).

2.15 Markov Chains in a Board Game

Consider the board game shown in Fig. 2.7. The game board has 3×3 squares arranged in order from $i = 1$ to $i = 9$. Each player rolls a three-sided die in turn with 3 equally likely outcomes from 1 to 3 denoted by a, which is the number of steps they can move. Players do not move if $i + a > 9$. Starting from 0 off the game board, the first player to hit the last square is the winner. There are two types of magic channels on the board, which are shown as dashed and solid lines. If the player stops at the foot of the solid-line channel, they will be transported to the top of that channel. Conversely, if the player stops at the top of the dashed-line channel, they will slide down to the bottom of that channel.

(a) Having a set of states $S = \{s_i\}, i = 0, 1, \ldots, 9$, set up the corresponding Markov chains for the game *with* and *without* magic channels. You need to build the transition matrix T whose ijth entry t_{ij} is the probability that the Markov chain starting in state s_i will be in state s_j after one step.
(b) Consider a Markov chain having r absorbing states (i.e., that once reached are never left) and t transient (non-absorbing) states from each of which an absorbing state can be reached. Assume that the first t states are transient and the last r states are absorbing. The transition matrix is then in a block form with a lower right block of $r \times r$ identity matrix:

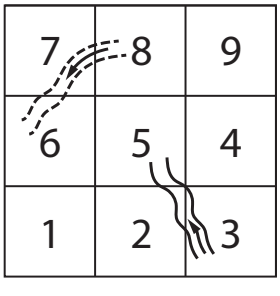

Figure 2.7 Game board.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press

$$T = \left(\begin{array}{c|c} Q & A \\ \hline 0 & I \end{array}\right).$$

Show that the matrix $I - Q$ is invertible and has the inverse matrix $N = I + Q + Q^2 + \cdots$. Also show that the ijth entry n_{ij} of N is the expected number of times the chain visits state s_j before getting absorbed given that it starts in s_i (where if $i = j$, the counting of visits starts from 1, i.e., n_{ii} will never be smaller than 1).

(c) Based on your results in (a) and (b), express analytically the average number of runs to complete the game *with* and *without* magic channels, and evaluate the expression numerically if needed.

Problem credit: Rui Fang.

https://doi.org/10.1017/9781108855259.003 Published online by Cambridge University Press