

# Chapter 1

## Introduction

Aldous - Fill

July 20, 1999

We start in section 1 with some “word problems”, intended to provide some extrinsic motivation for the general field of this book. In section 2 we talk about the broad conceptual themes of the book, partly illustrated by the word problems, and then outline the actual contents and give advice to readers.

### 1 Word problems

#### 1.1 Random knight moves

Imagine a knight on a corner square of an otherwise empty chessboard. Move the knight by choosing at random from the legal knight-moves. What is the mean time until the knight first returns to the starting square?

At first sight this looks like a messy problem which will require numerical calculations. But the knight is moving as random walk on a finite graph (rather than just some more general Markov chain), and elementary theory reduces the problem to counting the number of edges of the graph, giving the answer of 168 moves. See Chapter 3 yyy.

#### 1.2 The white screen problem

Around 1980 I wrote a little Basic program that would display a random walk on the screen of my home computer. First, a pixel in the middle of the screen was lit up. Then one of the four directions N,E,W,S was selected uniformly at random and the walk proceeded one step in the chosen direction. That new pixel was lit up on the screen, and the process was repeated

from the new point, etc. For a while, the walk is almost always quickly visiting pixels it hasn't visited before, so one sees an irregular pattern that grows in the center of the screen. After quite a long while, when the screen is perhaps 95% illuminated, the growth process will have slowed down tremendously, and the viewer can safely go read *War and Peace* without missing any action. After a minor eternity every cell will have been visited. Any mathematician will want to know how long, on the average, it takes until each pixel has been visited. *Edited from Wilf [16]*.

Taking the screen to be  $m \times m$  pixels, we have a random walk on the discrete two-dimensional torus  $Z_d^2$ , and the problem asks for the mean *cover time*, that is the time to visit every vertex of the graph. Such questions have been studied for general graphs (see Chapter 6), though ironically this particular case of the two-dimensional torus is the hardest special graph. It is known that mean cover time is asymptotically at most  $4\pi^{-1}m^2 \log^2 m$ , and conjectured this is asymptotically correct (see Chapter 7 yyy). For  $m = 512$  this works out to be about 13 million. Of course, in accordance with Moore's Law what took a minor eternity in 1980 takes just a few seconds today.

### 1.3 Universal traversal sequences

Let  $S(n, d)$  be the set of all  $d$ -regular graphs  $G$  with  $n$  vertices and with the edges at each vertex labeled  $(1, 2, \dots, d)$ . A *universal traversal sequence*  $i_1, i_2, \dots, i_u \in \{1, \dots, d\}$  is a sequence that satisfies

for each  $G \in S(n, d)$  and each initial vertex of  $G$  the deterministic walk "at step  $t$  choose edge  $i_t$ " visits every vertex.

What is the shortest length  $u = u(n, d)$  of such a sequence?

To get a partial answer, instead of trying to be clever about picking the sequence, consider what happens if we just choose  $i_1, i_2, \dots$  uniformly at random. Then the walk on a graph  $G$  is just simple random walk on  $G$ . Using a result that the mean cover time on a regular graph is  $O(n^2)$  one can show (see Chapter 6 yyy) that *most* sequences of length  $O(dn^3 \log n)$  are universal traversal sequences.

Paradoxically, no explicit example of a universal traversal sequence this short is known. The argument above fits a general theme that probabilistic methods can be useful in combinatorics to establish the existence of objects which are hard to exhibit constructively: numerous examples are in the monograph by Alon and Spencer [1].

## 1.4 How long does it take to shuffle a deck of cards?

Repeated random shuffles of a  $d$ -card deck may be modeled as a Markov chain on the space of all  $d!$  possible configurations of the deck. Different physical methods of shuffling correspond to different chains. The model for the most common method, *riffle shuffle*, is described carefully in Chapter 9 (xxx section to be written). A mathematically simpler method is *top-to-random*, in which the top card is reinserted at one of the  $d$  possible positions, chosen uniformly at random (Chapter 9 section yyy). Giving a precise mathematical interpretation to the question

how many steps of the chain (corresponding to a specified physical shuffle) are needed until the distribution of the deck is approximately uniform (over all  $d!$  configurations)?

is quite subtle; we shall formalize different interpretations as different *mixing times*, and relations between mixing times are discussed in Chapter 4 for reversible chains and in Chapter 8 (xxx section to be written) for general chains. Our favorite formalization is via the *variation threshold time*  $\tau_1$ , and it turns out that

$$\begin{aligned}\tau_1 &\sim \frac{3}{2} \log_2 d && \text{(riffle shuffle)} \\ \tau_1 &\sim d \log d && \text{(top-to-random shuffle)} .\end{aligned}\tag{1}$$

For the usual deck with  $d = 52$  these suggest 8 and 205 shuffles respectively.

## 1.5 Sampling from high-dimensional distributions: Markov chain Monte Carlo

Suppose you have a function  $f : R^d \rightarrow [0, \infty)$  with  $\kappa := \int_{R^d} f(x) dx < \infty$ , where  $f$  is given by some explicit but maybe complicated formula. How can you devise a scheme to sample a random point in  $R^d$  with the normalized probability density  $f(x)/\kappa$ ?

For  $d = 1$  the elementary “inverse distribution function” trick is available, and for small  $d$  simple acceptance/rejection methods are often practical. For large  $d$  the most popular method is some form of *Markov chain Monte Carlo (MCMC)* method, and this specific  $d$ -dimensional sampling problem is a prototype problem for MCMC methods. The scheme is to design a chain to have stationary distribution  $f(x)/\kappa$ . A simple such chain is as follows. From a point  $x$ , the next point  $X_1$  is chosen by a two-step procedure. First choose  $Y$  from some reference distribution (e.g. multivariate

Normal with specified variance, or uniform of a sphere of specified radius) on  $R^d$ ; then set  $X_1 = x + Y$  with probability  $\min(1, f(x + Y)/f(x))$  and set  $X_1 = x$  with the remaining probability.

Routine theory says that the stationary density is indeed  $f(x)/\kappa$  and that as  $t \rightarrow \infty$  the distribution of the chain after  $t$  steps converges to this stationary distribution. So a *heuristic* algorithm for the sampling problem is

Choose a starting point, a reference distribution and a number  $t$  of steps, simulate the chain for  $t$  steps, and output the state of the chain after  $t$  steps.

To make a rigorous algorithm one needs to know how many steps are needed to guarantee closeness to stationarity; this is a *mixing time* question. The conceptual issues here are discussed in Chapter 11. Despite a huge literature on methodology and applications of MCMC in many different settings, rigorous results are rather scarce. A notable exception is in the sampling setting above where  $\log f$  is a concave function, where there exist complicated results (outlined in Chapter 11 xxx to be written) proving that a polynomial (in  $d$ ) number of steps suffice.

## 1.6 Approximate counting of self-avoiding walks

A *self-avoiding walk (SAW)* of length  $l$  in the lattice  $Z^d$  is a walk  $0 = v_0, v_1, v_2, \dots, v_l$  for which the  $(v_i)$  are distinct and successive pairs  $(v_i, v_{i+1})$  are adjacent. Understanding the  $l \rightarrow \infty$  asymptotics of the cardinality  $|S_l|$  of the set  $S_l$  of SAWs of length  $l$  (dimension  $d = 3$  is the most interesting case) is a famous open problem. A conceptual insight is that, for large  $l$ , the problem

find an algorithm which counts  $|S_l|$  approximately

can be reduced to the problem

find an algorithm which gives an approximately uniform random sample from  $S_l$ .

To explain, note that each walk in  $S_{l+1}$  is a one-step extension of some walk in  $S_l$ . So the ratio  $|S_{l+1}|/|S_l|$  equals the mean number of extensions of a uniform random SAW from  $S_l$ , which of course can be estimated from the empirical average of the number of extensions of a large sample of SAWs from  $S_l$ .

Similar schemes work for various other families ( $S_i$ ) of combinatorial sets of increasing size, provided one has some explicit connection between  $S_i$  and  $S_{i+1}$ . As in the previous word problem, one can get an approximately uniform random sample by MCMC, i.e. by designing a chain whose stationary distribution is uniform, and simulating a sufficiently large number of steps of the chain: in making a rigorous algorithm, the issue again reduces to bounding the mixing time of the chain. The case of SAWs is outlined in Chapter 11 section yyy.

### 1.7 Simulating a uniform random spanning tree

The last two word problems hinted at large classes of algorithmic problems; here is a different, more specific problem. A finite connected graph  $G$  has a finite number of *spanning trees*, and so it makes sense to consider a uniform random spanning tree of  $G$ . How can one simulate this random tree?

It turns out there is an exact method, which involves running random walk on  $G$  until every vertex  $v$  has been visited; then for each  $v$  (other than the initial vertex) let the tree include the edge by which the walk first visited  $v$ . This gives some kind of random spanning tree; it seems non-obvious that the distribution is uniform, but that is indeed true. See Chapter 8 section yyy.

### 1.8 Voter model on a finite graph

Consider a graph where each vertex is colored, initially with different colors. Each vertex from time to time (precisely, at times of independent Poisson processes of rate 1) picks an adjacent vertex at random and changes its color to the color of the picked neighbor. Eventually, on a finite graph, all vertices will have the same color: how long does this take?

This question turns out to be related (via a certain notion of *duality*) to the following question. Imagine particles, initially one at each vertex, which perform continuous-time random walk on the graph, but which coalesce when they meet. Eventually they will all coalesce into one particle: how long does this take? On the complete graph on  $n$  vertices, the mean time in each question is  $\sim n$ . See Chapter 10 section yyy.

### 1.9 Are you related to your ancestors?

You have two parents, four grandparents and eight great-grandparents. In other words, for small  $g \geq 1$

you have exactly  $2^g$   $g$ 'th-generation ancestors, and you are related to each of them.

But what about larger  $g$ ? Clearly you didn't have  $2^{120} \approx 10^{12}$  distinct 120'th-generation ancestors! Even taking  $g = 10$ , one can argue it's unlikely you had 1,024 different 10th-generation ancestors, though the number is likely only a bit smaller – say 1,000, in round numbers. Whether you are actually *related to* these people is a subtle question. At the level of grade-school genetics, you have 46 chromosomes, each a copy of one parental chromosome, and hence each a copy of some 10th-generation ancestor's chromosome. So you're genetically related to at most 46 of your 10th-generation ancestors. Taking account of *crossover* during chromosome duplication leads to a more interesting model, in which the issue is to estimate hitting probabilities in a certain continuous-time reversible Markov chain. It turns out (Chapter 13 yyy) that the number of 10th-generation ancestors who are genetically related to you is about 340. So you're unlikely to be related to a particular 10th-generation ancestor, a fact which presents a curious sidebar to the principle of hereditary monarchy.

## 2 So what's in the book?

### 2.1 Conceptual themes

Classical mathematical probability focuses on *time-asymptotics*, describing what happens if some random process runs for ever. In contrast, the word problems each ask “how long until a chain does something?”, and the focus of this book is on finite-time behavior. More precisely, the word problems ask about *hitting times*, the time until a state or a set of states is first visited, or until each state in a set is visited; or ask about *mixing times*, the number of steps until the distribution is approximately the stationary distribution. The card-shuffling problems (section 1.4) provide a very intuitive setting for such questions; how many shuffles are needed, as a function of the size of the deck, until the deck is well shuffled? Such *size-asymptotic* results, of which (1) is perhaps the best-known, are one of the themes of this book. Thus in one sense our work is in the spirit of the *birthday* and *coupon-collector's* problems in undergraduate probability; in another sense our goals are reminiscent of those of *computational complexity* ( $P \stackrel{?}{=} NP$  and all that), which seeks to relate the time required to solve an algorithmic problem to the size of the problem.

## 2.2 Prerequisites

The reader who has taken a first-year graduate course in mathematical probability will have no difficulty with the mathematical content of this book. Though if the phrase “randomized algorithm” means nothing to you, then it would be helpful to look at Motwani - Raghavan [13] to get some feeling for the algorithmic viewpoint.

We have tried to keep much of the book accessible to readers whose mathematical background emphasizes discrete math and algorithms rather than analysis and probability. The minimal background required is an undergraduate course in probability including classical limit theory for finite Markov chains. Graduate-level mathematical probability is usually presented within the framework of measure theory, which (with some justification) is often regarded as irrelevant “general abstract nonsense” by those interested in concrete mathematics. We will point out as we go the pieces of graduate-level probability that we use (e.g. martingale techniques, Wald’s identity, weak convergence). Advice: if your research involves probability then you should at some time see what’s taught in a good first-year-graduate course, and we strongly recommend Durrett [6] for this purpose.

## 2.3 Contents and alternate reading

Amongst the numerous introductory accounts of Markov chains, Norris [14] is closest to our style. That book, like the more concise treatment in Durrett [6] Chapter 5, emphasizes probabilistic methods designed to work in the countable-state setting. Matrix-based methods designed for the finite-state setting are emphasised by Kemeny - Snell [10] and by Hunter [7]. We start in Chapter 2 by briskly reviewing standard asymptotic theory of finite-state chains, and go on to a range of small topics less often emphasised: obtaining general identities from the reward-renewal theorem, and useful metrics on distributions, for instance. Chapter 3 starts our systematic treatment of reversible chains: their identification as random walks on weighted graphs, the analogy with electrical networks, the spectral representation and its consequences for the structure of hitting time distributions, the Dirichlet formalism, extremal characterization of eigenvalues and various mean hitting times. This material has not been brought together before. Chen [2] gives a somewhat more advanced treatment of some of the analytic techniques and their applications to infinite particle systems (also overlapping partly with our Chapters 10 and 11), but without our finite-time emphasis. Kelly [9]

emphasizes stationary distributions of reversible stochastic networks, Keilson [8] emphasizes structural properties such as complete monotonicity, and Doyle - Snell [5] give a delightful elementary treatment of the electrical network connection. Chapter 4 is the centerpiece of our attempt to create coherent intermediate-level theory. We give a detailed analysis of different mixing times: the relaxation time (1/spectral gap), the variation threshold (where variation distance becomes small, uniformly in initial state) and the Cheeger time constant (related to weighted connectivity). We discuss relations between these times and their surprising connection with mean hitting times; the *distinguished paths* method for bounding relaxation time, Cheeger-type inequalities, and how these parameters behave under operations on chains (watching only on a subset, taking product chains). Little of this exists in textbooks, though Chung [3] gives a more graph-theoretic treatment of Cheeger inequalities and of the advanced analytic techniques in Chapter 12.

The rather technical Chapter 4 may seem tough going, but the payoff is that subsequent chapters tend to “branch out” without developing further theoretical edifices. Chapter 5 gives bare-hands treatments of numerous examples of random walks on special graphs, and of two classes of chains with special structure: birth-and-death chains, and random walks on trees. Chapter 6 treats cover times (times to visit every vertex), which feature in several of our word problems, and for which a fairly complete theory exists. Chapter 7 discusses a hierarchy of symmetry conditions for random walks on graphs and groups, emphasising structural properties. A conspicuous gap is that we do not discuss how analytic techniques (e.g. group representation theory, orthogonal polynomials) can be systematically used to derive exact formulas for  $t$ -step transition probabilities or hitting time distributions in the presence of enough symmetry. Diaconis [4] has material on this topic, but an updated account would be valuable. Chapter 8 returns to not-necessarily reversible chains, treating topics such as certain optimal stopping times, the Markov chain tree theorem, and coupling from the past. Chapter 9 xxx. Chapter 10 describes the coupling method of bounding the variation threshold mixing time, and then discusses several interacting particle systems on finite graphs related to random walks. As background, Liggett [11] is the standard reference for interacting particle systems on infinite lattices. Chapter 11 xxx. Chapter 12 recounts work of work of Diaconis and Saloff-Coste, who bring the techniques of Nash inequalities, log-Sobolev inequalities and local Poincaré inequalities to bear to obtain sharper estimates for reversible Markov chains. These techniques were originally developed by analysts in



the study of heat kernels, cf. the sophisticated treatment in Varopoulos et al [15]. Chapter 13 xxx and mentions topics not treated in detail because of mathematical depth or requirements for extraneous mathematical techniques or the authors' exhaustion.

As previously mentioned, our purpose is to provide systematic intermediate-level discussion of reversible Markov chains and random walks on graphs, built around the central theme of mixing times and hitting times developed in Chapter 4. Various topics could be tackled in a more bare-hands way; an opposite approach by Lovász [12] (N.B. second edition) is to lead the reader through half a chapter of problems concerning random walk on graphs. Our approach is to treat random walk on an unweighted graph as a specialization of reversible chain, which makes it clear where non-trivial graph theory is being used (basically, not until Chapter 6).

We have not included exercises, though filling in omitted details will provide ample exercise for a conscientious reader. Of the open problems, some seem genuinely difficult while others have just not been thought about before.

## References

- [1] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley, 1992.
- [2] M. F. Chen. *From Markov Chains to Non-Equilibrium Particle Systems*. World Scientific, Singapore, 1992.
- [3] F.R.K. Chung. *Spectral Graph Theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Amer. Math. Soc., 1997.
- [4] P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward CA, 1988.
- [5] P.G. Doyle and J.L. Snell. *Random Walks and Electrical Networks*. Mathematical Association of America, Washington DC, 1984.
- [6] R. Durrett. *Probability: Theory and Examples*. Wadsworth, Pacific Grove CA, 1991.
- [7] J.J. Hunter. *Mathematical Techniques of Applied Probability*. Academic Press, 1983.
- [8] J. Keilson. *Markov Chain Models - Rarity and Exponentiality*. Springer-Verlag, 1979.
- [9] F.P. Kelly. *Reversibility and Stochastic Networks*. Wiley, 1979.
- [10] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Van Nostrand, 1960.
- [11] T.M. Liggett. *Interacting Particle Systems*. Springer-Verlag, 1985.
- [12] L. Lovász. *Combinatorial Problems and Exercises*. North-Holland, 1993. Second Edition.
- [13] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
- [14] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.
- [15] N. Th. Varopoulos, L. Saloff-Coste, and T. Coulhon. *Analysis and Geometry on Groups*. Cambridge University Press, 1992.
- [16] H.S. Wilf. The editor's corner: The white screen problem. *Amer. Math. Monthly*, 96:704–707, 1989.