Reversible Markov Chains and Random Walks on Graphs

David Aldous and James Allen Fill

Unfinished monograph, 2002 (this is recompiled version, 2014)

Contents

Chapter 1

Introduction (July 20, 1999)

We start in section 1.1 with some "word problems", intended to provide some extrinsic motivation for the general field of this book. In section 1.2 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.

1.1 Word problems

1.1.1 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 numer of edges of the graph, giving the answer of 168 moves. See Chapter 3 yyy.

1.1.2 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 [336].*

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.

1.1.3 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 $(1, 2, \ldots, d)$. A *universal traversal sequence* $i_1, i_2, \ldots, i_u \in \{1, \ldots, d\}$ 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 $i_1, i_2, \ldots$ 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 [29].

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* τ_1, and it turns out that

$$\begin{aligned} \tau_1 &\sim \tfrac{3}{2}\log_2 d \quad \text{(riffle shuffle)} \\ \tau_1 &\sim d \log d \quad \text{(top-to-random shuffle)} \ . \end{aligned} \tag{1.1}$$

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

1.1.5 Sampling from high-dimensional distributions: Markov chain Monte Carlo

Suppose you have a function $f : R^d \to [0, \infty)$ 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 $t \to \infty$ 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.

1.1.6 Approximate counting of self-avoiding walks

A *self-avoiding walk (SAW)* of length l in the lattice Z^d is a walk $0 = v_0, v_1, v_2, \ldots, v_l$ for which the (v_i) are distinct and successive pairs (v_i, v_{i+1}) are adjacent. Understanding the $l \to \infty$ 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_l) of combinatorial sets of increasing size, provided one has some explicit connection between S_l and

S_{l+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.

1.1.7 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.

1.1.8 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.

1.1.9 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.

1.2 So what's in the book?

1.2.1 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 1.1.4) 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 (1.1) 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* ($\mathrm{P} \stackrel{?}{=} \mathrm{NP}$ and all that), which seeks to relate the time required to solve an algorithmic problem to the size of the problem.

1.2.2 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 [265] 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 [133] for this purpose.

1.2.3 Contents and alternate reading

Amongst the numerous introductory accounts of Markov chains, Norris [270] is closest to our style. That book, like the more concise treatment in Durrett [133] 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 [214] and by Hunter [186]. 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 [88] 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 [213] emphasizes stationary distributions of reversible

stochastic networks, Keilson [212] emphasizes structural properties such as complete monotonicity, and Doyle - Snell [131] 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 [93] 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 [112] 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 [231] 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 [332]. Chapter 13 xxx and mentions topics not treated in detail be-

cause 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 [237] (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.

Chapter 2

General Markov Chains (September 10, 1999)

The setting of this Chapter is a finite-state irreducible Markov chain (X_t), either in discrete time ($t = 0, 1, 2, \ldots$) or in continuous time ($0 \le t < \infty$). Highlights of the elementary theory of general (i.e. not-necessarily-reversible) Markov chains are readily available in several dedicated textbooks and in chapters of numerous texts on introductory probability or stochastic processes (see the Notes), so we just give a rapid review in sections 2.1 and 2.1.2. Subsequent sections emphasize several specific topics which are useful for our purposes but not easy to find in any one textbook: using the fundamental matrix in mean hitting times and the central limit theorem, metrics on distributions and submultiplicativity, Matthews' method for cover times, and martingale methods.

2.1 Notation and reminders of fundamental results

We recommend the textbook of Norris [270] for a clear treatment of the basic theory and a wide selection of applications.

Write $I = \{i, j, k, \ldots\}$ for a finite state space. Write $\mathbf{P} = p_{i,j}$ for the transition matrix of a discrete-time Markov chain $(X_t : t = 0, 1, 2, \ldots)$. To avoid trivialities let's exclude the one-state chain (*two*-state chains are useful, because surprisingly often general inequalities are sharp for two-state chains). The t-step transition probabilities are $P(X_t = j | X_0 = i) = p_{ij}^{(t)}$, where $\mathbf{P}^{(t)} = \mathbf{P}\mathbf{P}\ldots\mathbf{P}$ is the t-fold matrix product. Write $P_i(\cdot)$ and $E_i(\cdot)$

for probabilities and expectations for the chain started at state i and time 0. More generally, write $P_\rho(\cdot)$ and $E_\rho(\cdot)$ for probabilities and expectations for the chain started at time 0 with distribution ρ. Write

$$T_i = \min\{t \geq 0 : X_t = i\}$$

for the *first hitting time* on state i, and write

$$T_i^+ = \min\{t \geq 1 : X_t = i\}.$$

Of course $T_i^+ = T_i$ unless $X_0 = i$, in which case we call T_i^+ the *first return time* to state i. More generally, a subset A of states has first hitting time

$$T_A = \min\{t \geq 0 : X_t \in A\}.$$

We shall frequently use without comment "obvious" facts like the following.

> Start a chain at state i, wait until it first hits j, then wait until the time (S, say) at which it next hits k. Then $E_i S = E_i T_j + E_j T_k$.

The elementary proof sums over the possible values t of T_j. The sophisticated proof appeals to the *strong Markov* property ([270] section 1.4) of the *stopping time* T_j, which implies

$$E_i(S|X_t, t \leq T_j) = T_j + E_j T_k.$$

Recall that the symbol $|$ is the probabilist's shorthand for "conditional on".

2.1.1 Stationary distribution and asymptotics

Now assume the chain is *irreducible*. A fundamental result ([270] Theorems 1.7.7 and 1.5.6) is that there exists a unique *stationary distribution* $\pi = (\pi_i : i \in I)$, i.e. a unique probability distribution satisfying the *balance equations*

$$\pi_j = \sum_i \pi_i p_{ij} \text{ for all } j. \tag{2.1}$$

One way to prove this existence (liked by probabilists because it extends easily to the countable state setting) is to turn Lemma 2.6 below into a definition. That is, fix arbitrary i_0, define $\tilde{\pi}(i_0) = 1$, and define

$$\tilde{\pi}(j) = E_{i_0}(\text{number of visits to } j \text{ before time } T_{i_0}^+),\ j \neq i_0.$$

It can then be checked that $\pi_i := \tilde{\pi}(i)/\sum_j \tilde{\pi}(j)$ is a stationary distribution. The point of stationarity is that, if the initial position X_0 of the chain is random with the stationary distribution π, then the position X_t at any subsequent non-random time t has the same distribution π, and the process $(X_t, t = 0, 1, 2, \ldots)$ is then called the *stationary chain*.

A highlight of elementary theory is that the stationary distribution plays the main role in asymptotic results, as follows.

Theorem 2.1 (The ergodic theorem: [270] Theorem 1.10.2) *Let $N_i(t)$ be the number of visits to state i during times $0, 1, \ldots, t-1$. Then for any initial distribution,*

$$t^{-1}N_i(t) \to \pi_i \text{ a.s., as } t \to \infty.$$

Theorem 2.2 (The convergence theorem: [270] Theorem 1.8.3) *For any initial distribution,*

$$P(X_t = j) \to \pi_j \text{ as } t \to \infty, \text{ for all } j$$

provided the chain is aperiodic.

Theorem 2.1 is the simplest illustration of the *ergodic principle* "time averages equal space averages". Many general identities for Markov chains can be regarded as aspects of the ergodic principle – in particular, in section 2.2.1 we use it to derive expressions for mean hitting times. Such identities are important and useful.

The most classical topic in mathematical probability is time-asymptotics for i.i.d. (independent, identically distributed) random sequences. A vast number of results are known, and (broadly speaking) have simple analogs for Markov chains. Thus the analog of the strong law of large numbers is Theorem 2.1, and the analog of the central limit theorem is Theorem 2.17 below. As mentioned in Chapter 1 section 2.1 (yyy 7/20/99 version) this book has a different focus, on results which say something about the behavior of the chain over some specific finite time, rather than what happens in the indefinite future.

2.1.2 Continuous-time chains

The theory of continuous-time Markov chains closely parallels that of the discrete-time chains discussed above. To the reader with background in algorithms or discrete mathematics, the introduction of continuous time may at first seem artificial and unnecessary, but it turns out that certain

results are simpler in continuous time. See Norris [270] Chapters 2 and 3 for details on what follows.

A continuous-time chain is specified by *transition rates* ($q(i,j) = q_{ij}, j \neq i$) which are required to be non-negative but have no constraint on the sums. Given the transition rates, define

$$q_i := \sum_{j:j\neq i} q_{ij} \tag{2.2}$$

and extend (q_{ij}) to a matrix $\mathbf{Q}$ by putting $q_{ii} = -q_i$. The chain $(X_t : 0 \leq t < \infty)$ has two equivalent descriptions.

1. Infinitesimal description. Given that $X_t = i$, the chance that $X_{t+dt} = j$ is $q_{ij}dt$ for each $j \neq i$.

2. Jump-and-hold description. Define a transition matrix $\mathbf{J}$ by $J_{ii} = 0$ and

$$J_{ij} := q_{ij}/q_i, \ j \neq i. \tag{2.3}$$

Then the continuous-time chain may be constructed by the two-step procedure
(i) Run a discrete-time chain X^J with transition matrix $\mathbf{J}$.
(ii) Given the sequence of states $i_0, i_1, i_2, \ldots$ visited by X^J, the durations spent in states i_m are independent exponential random variables with rates q_{i_m}.
The discrete-time chain X^J is called the *jump chain* associated with X_t.

The results in the previous section go over to continuous-time chains with the following modifications.

(a) $P_i(X_t = j) = Q^{(t)}_{ij}$, where $\mathbf{Q}^{(t)} := \exp(\mathbf{Q}t)$.

(b) The definition of T_i^+ becomes

$$T_i^+ = \min\{t \geq T_{I\setminus i} : X_t = i\}.$$

(c) If the chain is irreducible then there exists a unique stationary distribution π characterized by

$$\sum_i \pi_i q_{ij} = 0 \text{ for all } j.$$

(d) In the ergodic theorem we interpret $N_i(t)$ as the total duration of time spent in state i during $[0,t]$:

$$N_i(t) := \int_0^t 1_{(X_s=i)}\, ds.$$

(e) In the convergence theorem the assumption of aperiodicity is unnecessary. [This fact is the one of the technical advantages of continuous time.]

(f) The evolution of $P(X_t = j)$ as a function of time is given by the *forwards equations*

$$\frac{d}{dt}P(X_t = j) = \sum_i P(X_t = i)q_{ij}. \tag{2.4}$$

Given a discrete-time chain X with some transition matrix $\mathbf{P}$, one can define the *continuized* chain $\widetilde{X}$ to have transition rates $q_{ij} = p_{ij}$, $j \neq i$. In other words, we replace the deterministic time-1 holds between jumps by holds with exponential(1) distribution. Many quantities are unchanged by the passage from the discrete time chain to the continuized chain. In particular the stationary distribution π and mean hitting times E_iT_A are unchanged. Therefore results stated in continuous time can often be immediately applied in discrete time, and vice versa.

In different parts of the book we shall be working with discrete or continuous time as a current convention, mentioning where appropriate how results change in the alternate setting. Chapter 4 (yyy section to be written) will give a survey of the differences between these two settings.

2.2 Identities for mean hitting times and occupation times

2.2.1 Occupation measures and stopping times

The purpose of this section is to give a systematic "probabilistic" treatment of a collection of general identities by deriving them from a single result, Proposition 2.3. We work in discrete time, but give the corresponding continuous-time results in section 2.2.3. Intuitively, a stopping time is a random time which can be specified by some on-line algorithm, together (perhaps) with external randomization.

Proposition 2.3 *Consider the chain started at state i. Let $0 < S < \infty$ be a stopping time such that $X_S = i$ and $E_iS < \infty$. Let j be an arbitrary state. Then*

$$E_i(\textit{number of visits to } j \textit{ before time } S) = \pi_j E_i S.$$

In the phrase "number of ... before time t", our convention is to include time 0 but exclude time t.

We shall give two different proofs. The first requires a widely-useful general theorem in stochastic processes.

Proof. Consider the renewal process whose inter-renewal time is distributed as S. The reward-renewal theorem (e.g. Ross [299] Thm. 3.6.1) says that the asymptotic proportion of time spent in state j equals

$$E_i(\text{number of visits to } j \text{ before time } S)/E_i S.$$

But this asymptotic average also equals π_j, by the ergodic theorem. □

We like that proof for philosophical reasons: a good way to think about general identities is that they show one quantity calculated in two different ways. Here is an alternative proof of a slightly more general assertion. We refer to Propositions 2.3 and 2.4 as *occupation measure identities*.

Proposition 2.4 *Let θ be a probability distribution on I. Let $0 < S < \infty$ be a stopping time such that $P_\theta(X_S \in \cdot) = \theta(\cdot)$ and $E_\theta S < \infty$. Let j be an arbitrary state. Then*

$$E_\theta(\textit{number of visits to } j \textit{ before time } S) = \pi_j E_\theta S.$$

Proof. Write $\rho_j = E_\theta(\text{number of visits to } j \text{ before time } S)$. We will show

$$\sum_j \rho_j p_{jk} = \rho_k \ \forall k. \tag{2.5}$$

Then by uniqueness of the stationary distribution, $\rho(\cdot) = c\pi(\cdot)$ for $c = \sum_k \rho_k = E_\theta S$.

Checking (2.5) is just a matter of careful notation.

$$\begin{aligned}
\rho_k &= \sum_{t=0}^{\infty} P_\theta(X_t = k, S > t) \\
&= \sum_{t=0}^{\infty} P_\theta(X_{t+1} = k, S > t) \text{ because } P_\theta(X_S = k) = P_\theta(X_0 = k) \\
&= \sum_{t=0}^{\infty} \sum_j P_\theta(X_t = j, S > t, X_{t+1} = k) \\
&= \sum_{t=0}^{\infty} \sum_j P_\theta(X_t = j, S > t)\ p_{jk} \text{ by the Markov property} \\
&= \sum_j \rho_j p_{jk}.
\end{aligned}$$

□

2.2.2 Mean hitting time and related formulas

The following series of formulas arise from particular choices of j and S in Proposition 2.3. For ease of later reference, we state them all together before starting the proofs. Some involve the quantity

$$Z_{ij} = \sum_{t=0}^{\infty} \left(p_{ij}^{(t)} - \pi_j \right) \tag{2.6}$$

In the periodic case the sum may oscillate, so we use the Cesaro limit or (equivalently, but more simply) the continuous-time limit (2.9). The matrix $\mathbf{Z}$ is called the *fundamental matrix* (see Notes for alternate standardizations). Note that from the definition

$$\sum_j Z_{ij} = 0 \text{ for all } i. \tag{2.7}$$

Lemma 2.5 $E_i T_i^+ = 1/\pi_i$.

Lemma 2.6

$$E_i(\textit{number of visits to } j \textit{ before time } T_i^+) = \pi_j/\pi_i.$$

Lemma 2.7 *For* $j \neq i$,

$$E_j(\textit{number of visits to } j \textit{ before time } T_i) = \pi_j(E_j T_i + E_i T_j).$$

Corollary 2.8 *For* $j \neq i$,

$$P_i(T_j < T_i^+) = \frac{1}{\pi_i(E_i T_j + E_j T_i)}.$$

Lemma 2.9 *For* $i \neq l$ *and arbitrary* j,

$$E_i(\textit{number of visits to } j \textit{ before time } T_l) = \pi_j(E_i T_l + E_l T_j - E_i T_j).$$

Corollary 2.10 *For* $i \neq l$ *and* $j \neq l$,

$$P_i(T_j < T_l) = \frac{E_i T_l + E_l T_j - E_i T_j}{E_j T_l + E_l T_j}.$$

Lemma 2.11 $\pi_i E_\pi T_i = Z_{ii}$.

Lemma 2.12 $\pi_j E_i T_j = Z_{jj} - Z_{ij}$.

Corollary 2.13 $\sum_j \pi_j E_i T_j = \sum_j Z_{jj}$ *for each* i.

Corollary 2.14 (The random target lemma) $\sum_j \pi_j E_i T_j$ *does not depend on* i.

Lemma 2.15

$$E_\pi(\textit{number of visits to } j \textit{ before time } T_i) = \frac{\pi_j}{\pi_i} Z_{ii} - Z_{ij}.$$

Lemmas 2.11 and 2.12, which will be used frequently throughout the book, will both be referred to as *the mean hitting time formula*. See the Remark following the proofs for a two-line heuristic derivation of Lemma 2.12. A consequence of the mean hitting time formula is that knowing the matrix $\mathbf{Z}$ is equivalent to knowing the matrix $(E_i T_j)$, since we can recover Z_{ij} as $\pi_j(E_\pi T_j - E_i T_j)$.

Proofs. The simplest choice of S in Proposition 2.3 is of course the first return time T_i^+. With this choice, the Proposition says

$$E_i(\text{number of visits to } j \text{ before time } T_i^+) = \pi_j E_i T_i^+.$$

Setting $j = i$ gives $1 = \pi_i E_i T_i^+$, which is Lemma 2.5, and then the case of general j gives Lemma 2.6.

Another choice of S is "the first return to i after the first visit to j". Then $E_i S = E_i T_j + E_j T_i$ and the Proposition becomes Lemma 2.7, because there are no visits to j before time T_j. For the chain started at i, the number of visits to i (including time 0) before hitting j has geometric distribution, and so

$$E_i(\text{number of visits to } i \text{ before time } T_j) = 1/P_i(T_j < T_i^+).$$

So Corollary 2.8 follows from Lemma 2.7 (with i and j interchanged).

Another choice of S is "the first return to i after the first visit to j after the first visit to l", where i, j, l are distinct. The Proposition says

$$\pi_j(E_i T_l + E_l T_j + E_j T_i) = E_i(\text{number of visits to } j \text{ before time } T_l)$$

$$+E_j(\text{number of visits to } j \text{ before time } T_i).$$

Lemma 2.7 gives an expression for the final expectation, and we deduce that (for distinct i, j, l)

$$E_i(\text{number of visits to } j \text{ before time } T_l) = \pi_j(E_i T_l + E_l T_j - E_i T_j).$$

This is the assertion of Lemma 2.9, and the identity remains true if $j = i$ (where it becomes Lemma 2.7) or if $j = l$ (where it reduces to $0 = 0$). We deduce Corollary 2.10 by writing

$$E_i(\text{number of visits to } j \text{ before time } T_l) =$$

$$P_i(T_j < T_l)E_j(\text{number of visits to } j \text{ before time } T_l)$$

and using Lemma 2.7 to evaluate the final expectation.

We now get slightly more ingenious. Fix a time $t_0 \geq 1$ and define S as the time taken by the following 2-stage procedure (for the chain started at i).

(i) wait time t_0

(ii) then wait (if necessary) until the chain next hits i.

Then the Proposition (with $j = i$) says

$$\sum_{t=0}^{t_0-1} p_{ii}^{(t)} = \pi_i(t_0 + E_\rho T_i) \tag{2.8}$$

where $\rho(\cdot) = P_i(X_{t_0} = \cdot)$. Rearranging,

$$\sum_{t=0}^{t_0-1} (p_{ii}^{(t)} - \pi_i) = \pi_i E_\rho T_i.$$

Letting $t_0 \to \infty$ we have $\rho \to \pi$ by the convergence theorem (strictly, we should give a separate argument for the periodic case, but it's simpler to translate the argument to continuous time where the periodicity issue doesn't arise) and we obtain Lemma 2.11.

For Lemma 2.12, where we may take $j \neq i$, we combine the previous ideas. Again fix t_0 and define S as the time taken by the following 3-stage procedure (for the chain started at i).

(i) wait until the chain hits k.

(ii) then wait a further time t_0.

(iii) then wait (if necessary) until the chain next hits i.

Applying Proposition 2.3 with this S and with $j = i$ gives

$$E_i(\text{number of visits to } i \text{ before time } T_k) + \sum_{t=0}^{t_0-1} p_{ki}^{(t)} = \pi_i(E_i T_k + t_0 + E_\rho T_i),$$

where $\rho(\cdot) = P_k(X_{t_0} = \cdot)$. Subtracting the equality of Lemma 2.7 and rearranging, we get

$$\sum_{t=0}^{t_0-1} (p_{ki}^{(t)} - \pi_i) = \pi_i(E_\rho T_i - E_k T_i).$$

Letting $t_0 \to \infty$, we have (as above) $\rho \to \pi$, giving

$$Z_{ki} = \pi_i(E_\pi T_i - E_k T_i).$$

Appealing to Lemma 2.11 we get Lemma 2.12. Corollary 2.13 follows from Lemma 2.12 by using (2.7).

To prove Lemma 2.15, consider again the argument for (2.8), but now apply the Proposition with $j \neq i$. This gives

$$\sum_{t=0}^{t_0-1} p_{ij}^{(t)} + E_\rho(\text{number of visits to } j \text{ before time } T_i) = \pi_j(t_0 + E_\rho T_i)$$

where $\rho(\cdot) = P_i(X_{t_0} = \cdot)$. Rearranging,

$$\sum_{t=0}^{t_0-1} (p_{ij}^{(t)} - \pi_j) + E_\rho(\text{number of visits to } j \text{ before time } T_i) = \pi_j E_\rho T_i.$$

Letting $t_0 \to \infty$ gives

$$Z_{ij} + E_\pi(\text{number of visits to } j \text{ before time } T_i) = \pi_j E_\pi T_i.$$

Applying Lemma 2.11 gives Lemma 2.15.

Remark. We promised a two-line heuristic derivation of the mean hitting time formula, and here it is. Write

$$\sum_{t=0}^{\infty} \left(1_{(X_t=j)} - \pi_j\right) = \sum_{t=0}^{T_j-1} \left(1_{(X_t=j)} - \pi_j\right) + \sum_{t=T_j}^{\infty} \left(1_{(X_t=j)} - \pi_j\right).$$

Take $E_i(\cdot)$ of each term to get $Z_{ij} = -\pi_j E_i T_j + Z_{jj}$. Of course this argument doesn't make sense because the sums do not converge. Implicit in our (honest) proof is a justification of this argument by a limiting procedure.

Example 2.16 *Patterns in coin-tossing.*

This is a classical example for which $\mathbf{Z}$ is easy to calculate. Fix n. Toss a fair coin repeatedly, and let $X_0, X_1, X_2, \ldots$ be the successive overlapping n-tuples. For example (with $n = 4$)

tosses	H	T	H	H	T	T
$X_0 =$	H	T	H	H		
$X_1 =$		T	H	H	T	
$X_2 =$			H	H	T	T

So $\mathbf{X}$ is a Markov chain on the set $I = \{H, T\}^n$ of n-tuples $i = (i_1, \ldots, i_n)$, and the stationary distribution π is uniform on I. For $0 \le d \le n-1$ write $I(i, j, d)$ for the indicator of the set "pattern j, shifted right by d places, agrees with pattern i where they overlap": formally, of the set

$$j_u = i_{u+d},\ 1 \le u \le n-d.$$

For example, with $n = 4$, $i = HHTH$ and $j = HTHH$,

d	0	1	2	3
$I(i,j,d)$	0	1	0	1

Then write

$$c(i, j) = \sum_{d=0}^{n-1} 2^{-d} I(i, j, d).$$

From the definition of $\mathbf{Z}$, and the fact that X_0 and X_t are independent for $t \ge n$,

$$Z_{ij} = c(i, j) - n2^{-n}.$$

So we can read off many facts about patterns in coin-tossing from the general results of this section. For instance, the mean hitting time formula (Lemma 2.11) says $E_\pi T_i = 2^n c(i, i) - n$. Note that "time 0" for the chain is the n'th toss, at which point the chain is in its stationary distribution. So the mean number of tosses until first seeing pattern i equals $2^n c(i, i)$. For $n = 5$ and $i = HHTHH$, the reader may check this mean number is 38. We leave the interested reader to explore further — in particular, find three patterns i, j, k such that

$$P(\text{pattern } i \text{ occurs before pattern } j) > 1/2$$

$$P(\text{pattern } j \text{ occurs before pattern } k) > 1/2$$

$$P(\text{pattern } k \text{ occurs before pattern } i) > 1/2.$$

Further results. One can of course obtain expressions in the spirit of Lemmas 2.5–2.15 for more complicated quantities. The reader may care to find expressions for

$$E_i \min(T_k, T_l)$$

$$E_i(\text{number of visits to } j \text{ before time } \min(T_k, T_l))$$

$$P_i(\text{hit } j \text{ before time } \min(T_k, T_l)).$$

Warning. Hitting times T_A on subsets will be studied later (e.g. Chapter 3 section 5.3) (yyy 9/2/94 version) in the reversible setting. It is important

to note that results often do not extend simply from singletons to subsets. For instance, one might guess that Lemma 2.11 could be extended to

$$E_\pi T_A = \frac{Z_{AA}}{\pi(A)}, \quad Z_{AA} := \sum_{t=0}^{\infty}(P_\pi(X_t \in A | X_0 \in A) - \pi(A)),$$

but it is easy to make examples where this is false.

2.2.3 Continuous-time versions

Here we record the continuous-time versions of the results of the previous section. Write

$$Z_{ij} = \int_0^\infty (P_i(X_t = j) - \pi_j)dt \tag{2.9}$$

This is consistent with (2.6) in that $\mathbf{Z}$ is the same for a discrete-time chain and its continuized chain. Recall from section 2.1.2 the redefinition (b) of T_i^+ in continuous time. In place of "number of visits to i" we use "total duration of time spent in i". With this substitution, Proposition 2.3 and the other results of the previous section extend to continuous time with only the following changes, which occur because the mean sojourn time in a state i is $1/q_i$ in continuous time, rather than 1 as in discrete time.

Lemma 2.5. $E_i T_i^+ = \frac{1}{q_i \pi_i}$.

Lemma 2.6.

$$E_i(\text{duration of time spent in } j \text{ before time } T_i^+) = \frac{\pi_j}{q_i \pi_i}.$$

Corollary 2.8. For $j \neq i$,

$$P_i(T_j < T_i^+) = \frac{1}{q_i \pi_i (E_i T_j + E_j T_i)}.$$

2.3 Variances of sums

In discrete time, consider the number $N_i(t)$ of visits to state i before time t. (Recall our convention is to count a visit at time 0 but not at time t.) For the *stationary* chain, we have (trivially)

$$E_\pi N_i(t) = t\pi_i.$$

It's not hard to calculate the variance:

$$\text{var}_{\ \pi} N_i(t) \quad = \quad \sum_{r=0}^{t-1} \sum_{s=0}^{t-1} (P_\pi(X_r = i, X_s = i) - \pi_i^2)$$

$$= \quad \pi_i \left(\sum_{u=0}^{t-1} 2(t-u)(p_{ii}^{(u)} - \pi_i) - t(1-\pi_i) \right)$$

setting $u = |s - r|$. This leads to the asymptotic result

$$\frac{\text{var }_\pi N_i(t)}{t} \to \pi_i(2Z_{ii} - 1 + \pi_i). \tag{2.10}$$

The fundamental matrix $\mathbf{Z}$ of (2.6) reappears in an apparently different context. Here is the more general result underlying (2.10). Take arbitrary functions $f : I \to R$ and $g : I \to R$ and center so that $E_\pi f(X_0) := \sum_i \pi_i f(i) = 0$ and $E_\pi g(X_0) = 0$. Write

$$S_t^f = \sum_{s=0}^{t-1} f(X_s)$$

and similarly for S_t^g. Then

$$E_\pi S_t^f S_t^g = \sum_i \sum_j f(i)g(j) \sum_{r=0}^{t-1} \sum_{s=0}^{t-1} (P_\pi(X_r = i, X_s = j) - \pi_i \pi_j).$$

The contribution to the latter double sum from terms $r \le s$ equals, putting $u = s - r$,

$$\pi_i \sum_{u=0}^{t-1} (t-u)(p_{ij}^{(u)} - \pi_j) \sim t\pi_i Z_{i,j}.$$

Collecting the other term and subtracting the twice-counted diagonal leads to the following result.

$$\frac{E_\pi S_t^f S_t^g}{t} \to f\Gamma g := \sum_i \sum_j f(i)\Gamma_{ij} g(j) \tag{2.11}$$

where Γ is the symmetric positive-definite matrix

$$\Gamma_{ij} := \pi_i Z_{ij} + \pi_j Z_{ji} + \pi_i \pi_j - \pi_i \delta_{ij}. \tag{2.12}$$

As often happens, the formulas simplify in continuous time. The asymptotic result (2.10) becomes

$$\frac{\text{var }_\pi N_i(t)}{t} \to 2\pi_i Z_{ii}$$

and the matrix Γ occurring in (2.11) becomes

$$\Gamma_{ij} := \pi_i Z_{ij} + \pi_j Z_{ji}.$$

Of course these asymptotic variances appear in the central limit theorem for Markov chains.

Theorem 2.17 *For centered f,*

$$t^{-1/2} S_t^f \quad \stackrel{d}{\to} \quad \text{Normal}(0, f\Gamma f) \text{ as } t \to \infty.$$

The standard proofs (e.g. [133] p. 378) don't yield any useful finite-time results, so we won't present a proof. We return to this subject in Chapter 4 section 4.1 (yyy 10/11/94 version) in the context of reversible chains. In that context, getting finite-time bounds on the approximation (2.10) for variances is not hard, but getting informative finite-time bounds on the Normal approximation remains quite hard.

Remark. Here's another way of seeing why asymptotic variances should relate (via $\mathbf{Z}$) to mean hitting times. Regard $N_i(t)$ as counts in a renewal process; in the central limit theorem for renewal counts ([133] Exercise 2.4.13) the variance involves the variance var $_i(T_i^+)$ of the inter-renewal time, and by (2.22) below this in turn relates to $E_\pi T_i$.

2.4 Two metrics on distributions

A major theme of this book is quantifying the convergence theorem (Theorem 2.2) to give estimates of how close the distribution of a chain is to the stationary distribution at *finite* times. Such quantifications require some explicit choice of "distance" between distributions, and two of the simplest choices are explained in this section. We illustrate with a trivial

Example 2.18 *Rain or shine?*

Suppose the true probability of rain tomorrow is 80% whereas we think the probability is 70%. How far off are we? In other words, what is the "distance" between π and θ, where

$$\pi(\text{rain}) = 0.8, \ \pi(\text{shine}) = 0.2$$

$$\theta(\text{rain}) = 0.7, \ \theta(\text{shine}) = 0.3.$$

Different notions of distance will give different numerical answers. Our first notion abstracts the idea that the "additive error" in this example is $0.8 - 0.7 = 0.1$.

2.4.1 Variation distance

Perhaps the simplest notion of distance between probability distributions is *variation distance*, defined as

$$||\theta_1 - \theta_2|| = \max_{A \subseteq I} |\theta_1(A) - \theta_2(A)|.$$

So variation distance is just the maximum *additive* error one can make, in using the "wrong" distribution to evaluate the probability of an event. In example 2.18, variation distance is 0.1. Several equivalent definitions are provided by

Lemma 2.19 *For probability distributions θ_1, θ_2 on a finite state space I,*

$$\begin{aligned}
\frac{1}{2}\sum_i |\theta_1(i) - \theta_2(i)| &= \sum_i (\theta_1(i) - \theta_2(i))^+ \\
&= \sum_i (\theta_1(i) - \theta_2(i))^- \\
&= 1 - \sum_i \min(\theta_1(i), \theta_2(i)) \\
&= \max_{A \subseteq I} |\theta_1(A) - \theta_2(A)| \\
&= \min P(V_1 \neq V_2)
\end{aligned}$$

the minimum taken over random pairs (V_1, V_2) such that V_m has distribution θ_m $(m = 1, 2)$. So each of these quantities equals the variation distance $||\theta_1 - \theta_2||$.

Proof. The first three equalities are clear. For the fourth, set $B = \{i : \theta_1(i) > \theta_2(i)\}$. Then

$$\begin{aligned}
\theta_1(A) - \theta_2(A) &= \sum_{i \in A} (\theta_1(i) - \theta_2(i)) \\
&\leq \sum_{i \in A \cap B} (\theta_1(i) - \theta_2(i)) \\
&\leq \sum_{i \in B} (\theta_1(i) - \theta_2(i)) \\
&= \sum_i (\theta_1(i) - \theta_2(i))^+
\end{aligned}$$

with equality when $A = B$. This, and the symmetric form, establish the fourth equality. In the final equality, the "$\leq$" follows from

$$|\theta_1(A) - \theta_2(A)| = |P(V_1 \in A) - P(V_2 \in A)| \leq P(V_2 \neq V_1).$$

And equality is attained by the following joint distribution. Let $\theta(i) = \min(\theta_1(i), \theta_2(i))$ and let

$$P(V_1 = i, V_2 = i) = \theta(i)$$

$$P(V_1 = i, V_2 = j) = \frac{(\theta_1(i) - \theta(i))(\theta_2(j) - \theta(j))}{1 - \sum_k \theta(k)},\ i \neq j.$$

(If the denominator is zero, then $\theta_1 = \theta_2$ and the result is trivial.) □

In the context of Markov chains we may use

$$d_i(t) := ||P_i(X_t = \cdot) - \pi(\cdot)|| \tag{2.13}$$

as a measure of deviation from stationarity at time t, for the chain started at state i. Also define

$$d(t) := \max_i d_i(t) \tag{2.14}$$

as the worst-case deviation from stationarity. Finally, it is technically convenient to introduce also

$$\bar{d}(t) = \max_{ij} ||P_i(X_t = \cdot) - P_j(X_t = \cdot)||. \tag{2.15}$$

In Chapter 4 we discuss, for reversible chains, relations between these "variation distance" notions and other measures of closeness-to-stationarity, and discuss parameters τ measuring "time until $d(t)$ becomes small" and their relation to other parameters of the chain. For now, let's just introduce a fundamental technical fact, the submultiplicativity property.

Lemma 2.20
(a) $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t),\ s, t \geq 0$ **[the submultiplicativity property]**.
(b) $d(s+t) \leq 2d(s)d(t),\ s, t \geq 0$.
(c) $d(t) \leq \bar{d}(t) \leq 2d(t),\ t \geq 0$.
(d) $d(t)$ *and* $\bar{d}(t)$ *decrease as* t *increases.*

Proof. We use the characterization of variation distance as

$$||\theta_1 - \theta_2|| = \min P(V_1 \neq V_2), \tag{2.16}$$

the minimum taken over random pairs (V_1, V_2) such that V_m has distribution θ_m $(m = 1, 2)$.

Fix states i_1, i_2 and times s, t, and let $\mathbf{Y}^1, \mathbf{Y}^2$ denote the chains started at i_1, i_2 respectively. By (2.16) we can construct a joint distribution for (Y^1_s, Y^2_s) such that

$$\begin{aligned} P(Y^1_s \neq Y^2_s) &= ||P_{i_1}(X_s = \cdot) - P_{i_2}(X_s = \cdot)|| \\ &\leq \bar{d}(s). \end{aligned}$$

Now for each pair (j_1, j_2), we can use (2.16) to construct a joint distribution for (Y^1_{s+t}, Y^2_{s+t}) given $(Y^1_s = j_1, Y^2_s = j_2)$ with the property that

$$P(Y^1_{s+t} \neq Y^2_{s+t} | Y^1_s = j_1, Y^2_s = j_2) = ||P_{j_1}(X_t = \cdot) - P_{j_2}(X_t = \cdot)||.$$

The right side is 0 if $j_1 = j_2$, and otherwise is at most $\bar{d}(t)$. So unconditionally

$$P(Y^1_{s+t} \neq Y^2_{s+t}) \leq \bar{d}(s)\bar{d}(t)$$

and (2.16) establishes part (a) of the lemma. For part (b), the same argument (with $\mathbf{Y}^2$ now being the stationary chain) shows

$$d(s+t) \leq d(s)\bar{d}(t) \tag{2.17}$$

so that (b) will follow from the upper bound $\bar{d}(t) \leq 2d(t)$ in (c). But this upper bound is clear from the triangle inequality for variation distance. And the lower bound in (c) follows from the fact that $\mu \to ||\theta - \mu||$ is a convex function, so that averaging over j with respect to π in (2.15) can only decrease distance. Finally, the "decreasing" property for $\bar{d}(t)$ follows from (a), and for $d(t)$ follows from (2.17). □

The assertions of this section hold in either discrete or continuous time. But note that the numerical value of $d(t)$ changes when we switch from a discrete-time chain to the continuized chain. In particular, for a discrete-time chain with period q we have $d(t) \to (q-1)/q$ as $t \to \infty$ (which incidently implies, taking $q = 2$, that the factor 2 in Lemma 2.20(b) cannot be reduced) whereas for the continuized chain $d(t) \to 0$.

One often sees slightly disguised corollaries of the submultiplicativity property in the literature. The following is a typical one.

Corollary 2.21 *Suppose there exists a probability measure μ, a real $\delta > 0$ and a time t such that*

$$p^{(t)}_{ij} \geq \delta \mu_j \ \forall i, j.$$

Then

$$d(s) \leq (1-\delta)^{\lfloor s/t \rfloor}, \ s \geq 0.$$

Proof. The hypothesis implies $\bar{d}(t) \leq 1 - \delta$, by the third equality in Lemma 2.19, and then the conclusion follows by submultiplicativity.

2.4.2 L^2 distance

Another notion of distance, which is less intuitively natural but often more mathematically tractable, is L^2 distance. This is defined with respect to

some fixed reference probability distribution π on I, which for our purposes will be the stationary distribution of some irreducible chain under consideration (and so $\pi_i > 0 \ \forall i$). The L^2 norm of a *function* $f : I \to R$ is

$$||f||_2 = \sqrt{\sum_i \pi_i f^2(i)}. \tag{2.18}$$

We define the L^2 norm of a *signed measure* ν on I by

$$||\nu||_2 = \sqrt{\sum_i \nu_i^2/\pi_i}. \tag{2.19}$$

This may look confusing, because a signed measure ν and a function f are in a sense "the same thing", being determined by values $(f(i); i \in I)$ or $(\nu_i; i \in I)$ which can be chosen arbitrarily. But the measure ν can also be determined by its density function $f(i) = \nu_i/\pi_i$, and so (2.18) and (2.19) say that the L^2 norm of a signed measure is defined to be the L^2 norm of its density function.

So $||\theta - \mu||_2$ is the "L^2" measure of distance between probability distributions θ, μ. In particular, the distance between θ and the reference distribution π is

$$||\theta - \pi||_2 = \sqrt{\sum_i \frac{(\theta_i - \pi_i)^2}{\pi_i}} = \sqrt{\sum_i \frac{\theta_i^2}{\pi_i} \quad - 1}.$$

In Example 2.18 we find $||\theta - \pi||_2 = 1/4$.

Writing $\theta(t)$ for the distribution at time t of a chain with stationary distribution π, it is true (cf. Lemma 2.20(d) for variation distance) that $||\theta(t) - \pi||_2$ is decreasing with t. Since there is a more instructive proof in the reversible case (Chapter 3 Lemma 23) (yyy 9/2/94 version) we won't prove the general case (see Notes).

Analogous to the L^2 norms are the L^1 norms

$$||f||_1 = \sum_i \pi_i |f(i)|$$

$$||\nu||_1 = \sum_i |\nu_i|.$$

The Cauchy-Schwarz inequality gives $|| \cdot ||_1 \le || \cdot ||_2$. Note that in the definition of $||\nu||_1$ the reference measure π has "cancelled out". Lemma 2.19

shows that for probability measures θ_1, θ_2 the L^1 distance is the same as variation distance, up to a factor of 2:

$$||\theta_1 - \theta_2|| = \tfrac{1}{2}||\theta_1 - \theta_2||_1.$$

As a trivial example in the Markov chain setting, consider

Example 2.22 *Take $I = \{0, 1, \ldots, n-1\}$, fix a parameter $0 < a < 1$ and define a transition matrix*

$$p_{ij} = a 1_{(j=i+1 \mod n)} + \frac{1-a}{n}.$$

In this example the t-step transition probabilities are

$$p_{ij}^{(t)} = a^t 1_{(j=i+t \mod n)} + \frac{1-a^t}{n}$$

and the stationary distribution π is uniform. We calculate (for arbitrary $j \neq i$)

$$d(t) = ||P_i(X_t \in \cdot) - \pi|| = (1 - n^{-1})a^t$$
$$\bar{d}(t) = ||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)|| = a^t$$
$$||P_i(X_t \in \cdot) - \pi||_2 = (n-1)^{1/2} a^t.$$

2.4.3 Exponential tails of hitting times

The submultiplicative property of $\bar{d}(t)$ is one instance of a general principle:

> because our state space is *finite*, many quantities which converge to zero as $t \to \infty$ must converge exponentially fast, by iterating over worst-case initial states.

Here's another instance, tails of hitting time distributions.

Consider the first hitting time T_A on a subset A. Define $t_A^* := \max_i E_i T_A$. For any initial distribution μ, any time $s > 0$ and any integer $m \geq 1$,

$$\begin{aligned} P_\mu(T_A > ms | T_A > (m-1)s) &= P_\theta(T_A > s) \text{ for some dist. } \theta \\ &\leq \max_i P_i(T_A > s) \\ &\leq t_A^*/s. \end{aligned}$$

So by induction on m

$$P_\mu(T_A > js) \leq (t_A^*/s)^j$$

implying

$$P_\mu(T_A > t) \le (t^*_A/s)^{\lfloor t/s \rfloor}, \quad t > 0.$$

In continuous time, a good (asymptotically optimal) choice of s is $s = et^*_A$, giving the exponential tail bound

$$\sup_\mu P_\mu(T_A > t) \le \exp\left(-\left\lfloor \frac{t}{et^*_A} \right\rfloor\right), \quad 0 < t < \infty. \tag{2.20}$$

A messier bound holds in discrete time, where we have to choose s to be an integer.

2.5 Distributional identities

It is much harder to get useful information about distributions (rather than mere expectations). Here are a few general results.

2.5.1 Stationarity consequences

A few useful facts about stationary Markov chains are, to experts, just specializations of facts about arbitrary (i.e. not-necessarily-Markov) stationary processes. Here we give a bare-hands proof of one such fact, the relation between the distribution of return time to a subset A and the distribution of first hitting time to A from a stationary start. We start in discrete time.

Lemma 2.23 *For $t = 1, 2, \ldots$,*

$$P_\pi(T_A = t - 1) = P_\pi(T^+_A = t) = \pi(A) P_{\pi_A}(T^+_A \ge t)$$

where $\pi_A(i) := \pi_i/\pi(A),\ i \in A$.

Proof. The first equality is obvious. Now let (X_t) be the chain started with its stationary distribution π. Then

$$\begin{aligned}
P_\pi(T^+_A = t) &= P(X_1 \notin A, \ldots, X_{t-1} \notin A, X_t \in A) \\
&= P(X_1 \notin A, \ldots, X_{t-1} \notin A) - P(X_1 \notin A, \ldots, X_t \notin A) \\
&= P(X_1 \notin A, \ldots, X_{t-1} \notin A) - P(X_0 \notin A, \ldots, X_{t-1} \notin A) \\
&= P(X_0 \in A, X_1 \notin A, \ldots, X_{t-1} \notin A) \\
&= \pi(A) P_{\pi_A}(T^+_A \ge t),
\end{aligned}$$

establishing the Lemma.

We'll give two consequences of Lemma 2.23. Summing over t gives

Corollary 2.24 (Kac's formula) $\pi(A)E_{\pi_A}T_A^+ = 1$

which extends the familiar fact $E_iT_i^+ = 1/\pi_i$. Multiplying the identity of Lemma 2.23 by t and summing gives

$$\begin{aligned} E_\pi T_A + 1 &= \sum_{t\geq 1} tP_{\pi_A}(T_A = t-1) \\ &= \pi(A)\sum_{t\geq 1} tP_{\pi_A}(T_A^+ \geq t) \\ &= \pi(A)\sum_{m\geq 1}\frac{1}{2}m(m+1)P_{\pi_A}(T_A^+ = m) \\ &= \frac{\pi(A)}{2}\left(E_{\pi_A}T_A^+ + E_{\pi_A}(T_A^+)^2\right). \end{aligned}$$

Appealing to Kac's formula and rearranging,

$$E_{\pi_A}(T_A^+)^2 = \frac{2E_\pi T_A + 1}{\pi(A)}, \tag{2.21}$$

$$\mathrm{var}_{\pi_A}(T_A^+) = \frac{2E_\pi T_A + 1}{\pi(A)} - \frac{1}{\pi^2(A)}. \tag{2.22}$$

More generally, there is a relation between $E_{\pi_A}(T_A^+)^p$ and $E_\pi(T_A^+)^{p-1}$.

In continuous time, the analog of Lemma 2.23 is

$$P_\pi(T_A \in (t, t+dt)) = Q(A, A^c)P_{\rho_A}(T_A > t)dt,\ t > 0 \tag{2.23}$$

where

$$Q(A, A^c) := \sum_{i\in A}\sum_{j\in A^c} q_{ij}, \quad \rho_A(j) := \sum_{i\in A} q_{ij}/Q(A, A^c), j \in A^c.$$

Integrating over $t > 0$ gives the analog of Kac's formula

$$Q(A, A^c)E_{\rho_A}T_A = \pi(A^c). \tag{2.24}$$

2.5.2 A generating function identity

Transform methods are useful in analyzing special examples, though that is not the main focus of this book. We record below just the simplest "transform fact". We work in discrete time and use generating functions – the corresponding result in continuous time can be stated using Laplace transforms.

Lemma 2.25 *Define*

$$G_{ij}(z) = \sum_{t\geq 0} P_i(X_t = j)z^t,\ F_{ij}(z) = \sum_{t\geq 0} P_i(T_j = t)z^t.$$

Then $F_{ij} = G_{ij}/G_{jj}$.

Analysis proof. Conditioning on T_j gives

$$p_{ij}^{(t)} = \sum_{l=0}^{t} P_i(T_j = l)p_{jj}^{(t-l)}$$

and so

$$\sum_{t\geq 0} p_{ij}^{(t)} z^t = \sum_{l\geq 0} \sum_{t-l\geq 0} P_i(T_j = l)z^l p_{jj}^{(t-l)} z^{t-l}$$

Thus $G_{ij}(z) = F_{ij}(z)G_{jj}(z)$, and the lemma follows. □

Probability proof. Let ζ have geometric(z) law $P(\zeta > t) = z^t$, independent of the chain. Then

$$\begin{aligned} G_{ij}(z) &= E_i(\text{number of visits to } j \text{ before time } \zeta) \\ &= P_i(T_j < \zeta)\ E_j(\text{number of visits to } j \text{ before time } \zeta) \\ &= F_{ij}(z)G_{jj}(z). \end{aligned}$$

□

Note that, differentiating term by term,

$$E_i T_j = \tfrac{d}{dz} F_{ij}(z)\Big|_{z=1}.$$

This and Lemma 2.25 can be used to give an alternative derivation of the mean hitting time formula, Lemma 2.12.

2.5.3 Distributions and continuization

The distribution at time t of the continuization $\hat{X}$ of a discrete-time chain X is most simply viewed as a Poisson mixture of the distributions (X_s). That is, $\hat{X}_t \stackrel{d}{=} X_{N_t}$ where N_t has Poisson(t) distribution independent of X. At greater length,

$$P_i(\hat{X}_t = j) = \sum_{s=0}^{\infty} \frac{e^{-t}t^s}{s!} P_i(X_s = j).$$

This holds because we can construct $\hat{X}$ from X by replacing the deterministic "time 1" holds by random, exponential(1), holds (ξ_j) between jumps,

and then the number N_t of jumps before time t has Poisson(t) distribution. Now write $S_n = \sum_{j=1}^{n} \xi_j$ for the time of the n'th jump. Then the hitting time $\hat{T}_A$ for the continuized chain is related to the hitting time T_A of the discrete-time chain by $\hat{T}_A = S_{T_A}$. Though these two hitting time distributions are different, their expectations are the same, and their variances are related in a simple way. To see this, the conditional distribution of $\hat{T}_A$ given T_A is the distribution of the sum of T_A independent ξ's, so (using the notion of conditional expectation given a random variable)

$$E(\hat{T}_A|T_A) = T_A, \text{ var } (\hat{T}_A|T_A) = T_A.$$

Thus (for any initial distribution)

$$E\hat{T}_A = EE(\hat{T}_A|T_A) = ET_A.$$

And the *conditional variance formula* ([133] p. 198)

$$\text{var } Z = E \text{ var } (Z|Y) + \text{var } E(Z|Y)$$

tells us that

$$\begin{aligned} \text{var } \hat{T}_A &= E\text{var } (\hat{T}_A|T_A) + \text{var } E(\hat{T}_A|T_A) \\ &= ET_A + \text{var } T_A. \end{aligned} \tag{2.25}$$

2.6 Matthews' method for cover times

Theorem 2.26 below is the only non-classical result in this Chapter. We make extensive use of this *Matthews' method* in Chapter 6 to analyze cover times for random walks on graphs.

Consider the *cover time* $C := \max_j T_j$ of the chain, i.e. the time required to visit every state. How can we bound E_iC in terms of the mean hitting times E_iT_j? To appreciate the cleverness of Theorem 2.26 let us first consider a more routine argument. Write $t^* := \max_{i,j} E_iT_j$. Since E_iC is unaffected by continuization, we may work in continuous time. By (2.20)

$$P_i(T_j > ket^*) \le e^{-k}, \ k = 1, 2, 3, \ldots.$$

By Boole's inequality, for an n-state chain

$$P_i(C > ket^*) \le ne^{-k}, \ k = 1, 2, 3, \ldots.$$

One can rewrite this successively as

$$\begin{aligned} P_i\left(\frac{C}{et^*} > x\right) &\leq ne \cdot e^{-x}, \quad 0 \leq x < \infty \\ P_i\left(\frac{C}{et^*} - \log(en) > x\right) &\leq e^{-x}, \quad 0 \leq x < \infty. \end{aligned}$$

In words, this says that the distribution of $\frac{C}{et^*} - \log(en)$ is stochastically smaller that the exponential(1) distribution, implying $E_i\left(\frac{C}{et^*} - \log(en)\right) \leq 1$ and hence

$$\max_i E_i C \leq (2 + \log n) et^*.$$

This argument does lead to a bound, but one suspects the factors 2 and e are artifacts of the proof; also, it seems hard to obtain a lower bound this way. The following result both "cleans up" the upper bound and gives a lower bound.

Theorem 2.26 (Matthews [256]) *For any n-state Markov chain,*

$$\max_v E_v C \leq h_{n-1} \max_{i,j} E_i T_j$$

$$\min_v E_v C \geq h_{n-1} \min_{i \neq j} E_i T_j$$

where $h_{n-1} := \sum_{m=1}^{n-1} \frac{1}{m}$.

Proof. We'll prove the lower bound — the upper bound proof is identical. Let $J_1, J_2, \ldots, J_n$ be a uniform random ordering of the states, independent of the chain. Define $C_m := \max_{i \leq m} T_{J_i}$ to be the time until all of $\{J_1, J_2, \ldots, J_m\}$ have been visited, in some order. The key identity is

$$E(C_m - C_{m-1} | J_1, \ldots, J_m; X_t, t \leq C_{m-1}) = t(L_{m-1}, J_m) 1_{(L_m = J_m)} \quad (2.26)$$

where $t(i,j) := E_i T_j$ and

$$L_m \text{ is the state amongst } \{J_1, J_2, \ldots, J_m\} \text{ hit last.}$$

To understand what this says, suppose we are told which are the states $\{J_1, J_2, \ldots, J_m\}$ and told the path of the chain up through time C_{m-1}. Then we know whether or not $L_m = J_m$: if not, then $C_m = C_{m-1}$, and if so, then the conditional distribution of $C_m - C_{m-1}$ is the distribution of the time to hit J_m from the state at time C_{m-1}, which we are told is state L_{m-1}.

Writing $t_* := \min_{i \neq j} t(i,j)$, the right side of (2.26) is $\geq t_* 1_{(L_m = J_m)}$, and so taking expectations

$$E(C_m - C_{m-1}) \geq t_* P(L_m = J_m).$$

But obviously $P(L_m = J_m) = 1/m$ by symmetry. So

$$E_v C = E_v C_1 + \sum_{m=2}^{n} E_v(C_m - C_{m-1}) \geq E_v C_1 + t_* \sum_{m=2}^{n} \frac{1}{m}.$$

Allowing for the possibility $J_1 = v$ we see $E_v C_1 \geq (1 - \frac{1}{n}) t_*$, and the lower bound follows.

2.7 New chains from old

Consider a chain (X_t) on state-space I, and fix $A \subseteq I$. There are many different constructions of new chains whose state space is (exactly or roughly) just A, and it's important not to confuse them. Three elementary constructions are described here. Anticipating the definition of *reversible* from Chapter 3, it is easy to check that if the original chain is reversible then each new chain is reversible.

2.7.1 The chain watched only on A

This is the chain (Y_n) defined by

$$S_0 = T_A = \min\{t \geq 0 : X_t \in A\}$$

$$S_n = \min\{t > S_{n-1} : X_t \in A\}$$

$$Y_n = X_{S_n}.$$

The chain (Y_n) has state space A and transition matrix

$$\bar{P}_A(i,j) = P_i(X_{T_A} = j),\ i,j \in A.$$

From the ergodic theorem (Theorem 2.1) it is clear that the stationary distribution π_A of (Y_t) is just π conditioned on A, that is

$$\pi_A(i) = \pi_i / \pi(A),\ i \in A. \tag{2.27}$$

2.7.2 The chain restricted to A

This is the chain with state space A and transition matrix $\hat{P}_A$ defined by

$$\begin{aligned} \hat{P}_A(i,j) &= P(i,j),\ i,j \in A, i \neq j \\ \hat{P}_A(i,i) &= 1 - \sum_{j \in A, j \neq i} P(i,j),\ i \in A. \end{aligned}$$

In general there is little connection between this chain and the original chain (X_t), and in general it is not true that the stationary distribution is given by (2.27). However, when the original chain is reversible, it is easy to check that the restricted chain does have the stationary distribution (2.27).

2.7.3 The collapsed chain

This chain has state space $I^* = A \cup \{a\}$ where a is a new state. We interpret the new chain as "the original chain with states A^c collapsed to a single state a". *Warning. In later applications we switch the roles of A and A^c, i.e. we collapse A to a single state a and use the collapsed chain on states $I^* = A^c \cup \{a\}$.* The collapsed chain has transition matrix

$$\begin{aligned} p^*_{ij} &= p_{ij},\ i,j \in A \\ p^*_{ia} &= \sum_{k \in A^c} p_{ik},\ i \in A \\ p^*_{ai} &= \frac{1}{\pi(A^c)} \sum_{k \in A^c} \pi_k p_{ki},\ i \in A \\ p^*_{aa} &= \frac{1}{\pi(A^c)} \sum_{k \in A^c} \sum_{l \in A^c} \pi_k p_{kl}. \end{aligned}$$

The collapsed chain has stationary distribution π^* given by

$$\pi^*_i = \pi_i, i \in A; \quad \pi^*_a = \pi(A^c).$$

Obviously the $\mathbf{P}$-chain started at i and run until T_{A^c} is the same as the $\mathbf{P}^*$-chain started at i and run until T_a. This leads to the general *collapsing principle*

> To prove a result which involves the behavior of the chain only up to time T_{A^c}, we may assume A^c is a singleton.

For we may apply the singleton result to the $\mathbf{P}^*$-chain run until time T_a, and the same result will hold for the $\mathbf{P}$-chain run until time T_{A^c}.

It is important to realize that typically (even for reversible chains) all three constructions give different processes. Loosely, the chain restricted to A "rebounds off the boundary of A^c where the boundary is hit", the collapsed chain "exits A^c at a random place independent of the hitting place", and the chain watched only on A "rebounds at a random place *dependent* on the hitting place".

2.8 Miscellaneous methods

2.8.1 Martingale methods

Modern probabilists regard the martingale optional stopping theorem as one of the most important results in their subject. As propaganda for martingales we give below four quick applications of that theorem, and a few more will appear later. All of these results could be proved in alternative, elementary ways. For the reader unfamiliar with martingales, Durrett [133] Chapter 4 contains much more than you need to know: Karlin and Taylor [208] Chapter 6 is a gentler introduction.

Lemma 2.27 *Given a non-empty subset $A \subset I$ and a function $f(i)$ defined for $i \in A$, there exists a unique extension of f to all I satisfying*

$$f(i) = \sum_j p_{ij} f(j), \ i \notin A.$$

Proof. If f satisfies the equations above then for any initial distribution the process $M_t := f(X_{\min(t,T_A)})$ is a martingale. So by the optional stopping theorem

$$f(i) = E_i f(X_{T_A}) \text{ for all } i. \tag{2.28}$$

This establishes uniqueness. Conversely, if we define f by (2.28) then the desired equations hold by conditioning on the first step.

Corollary 2.28 *If h is* harmonic, *i.e. if*

$$h(i) = \sum_j p_{ij} h(j) \text{ for all } i$$

then h is constant.

Proof. Clearly a constant function is harmonic. So the Corollary follows from the uniqueness assertion of Lemma 2.27, taking A to be some singleton.

Lemma 2.29 (The random target lemma) *The sum $\sum_j E_i T_j \pi_j$ does not depend on i.*

Proof. This repeats Corollary 2.14 with a different argument. The first-step recurrence for $g_j(i) := E_i T_j$ is

$$g_j(i) = 1_{(i \neq j)} + 1_{(i \neq j)} \sum_k p_{ik} g_j(k).$$

By Corollary 2.28 it is enough to show that $h(i) := \sum_j \pi_j g_j(i)$ is a harmonic function. We calculate

$$\begin{aligned} h(i) &= 1 - \pi_i + \sum_{j,k} \pi_j p_{ik} g_j(k) 1_{(i \neq j)} \\ &= 1 - \pi_i + \sum_k p_{ik}(h(k) - \pi_i g_i(k)) \text{ by definition of } h(k) \\ &= \sum_k p_{ik} h(k) \quad + 1 - \pi_i \left(1 + \sum_k p_{ik} g_i(k)\right). \end{aligned}$$

But $1/\pi_i = E_i T_i^+ = 1 + \sum_k p_{ik} g_i(k)$, so h is indeed harmonic.

Lemma 2.30 *For any stopping time S and any states i, j, k,*

$$E_i(\textit{number of transitions } j \to k \textit{ starting before time } S)$$

$$= p_{jk} E_i(\textit{number of visits to } j \textit{ before time } S).$$

Proof. Recall that "before" means strictly before. The assertion of the lemma is intuitively obvious, because each time the chain visits j it has chance p_{jk} to make a transition $j \to k$, and one can formalize this as in the proof of Proposition 2.4. A more sophisticated proof is to observe that $M(t)$ is a martingale, where

$$M(t) := N_{jk}(t) - p_{jk} N_j(t).$$

$$N_j(t) := \text{ number of visits to } j \text{ before time } t$$

$$N_{jk}(t) := \text{ number of transitions } j \to k \text{ starting before time } t \ .$$

And the assertion of the lemma is just the optional stopping theorem applied to the martingale M and the stopping time S.

Lemma 2.31 *Let A be a non-empty subset of I and let $h : I \to R$ satisfy*
(i) $h(i) \geq 0,\ i \in A$
(ii) $h(i) \geq 1 + \sum_j p_{ij} h(j),\ i \in A^c$.
Then $E_i T_A \leq h(i),\ i \in I$.

Proof. For arbitrary h, define g by

$$h(i) = 1 + \sum_j p_{ij} h(j) + g(i)$$

and then define

$$M_t = t + h(X_t) + \sum_{s=0}^{t-1} g(X_s).$$

Then $M_{\min(t,T_A)}$ is a martingale, so the optional sampling theorem says

$$E_i M_{T_A} = E_i M_0 = h(i).$$

But the hypotheses on h imply $M_{T_A} \geq T_A$.

2.8.2 A comparison argument

A theme running throughout the book is the idea of getting inequalities for a "hard" chain by making a comparison with some "easier" chain for which we can do exact calculations. Here is a simple example.

Lemma 2.32 *Let X be a discrete-time chain on states $\{0, 1, 2, \ldots, n\}$ such that $p_{ij} = 0$ whenever $j > i$. Write $m(i) = i - E_i X_1$, and suppose $0 < m(1) \leq m(2) \leq \ldots \leq m(n)$. Then $E_n T_0 \leq \sum_{j=1}^n \frac{1}{m(j)}$.*

Proof. The proof implicitly compares the given chain to the continuous-time chain with $q_{i,i-1} = m(i)$. Write $h(i) = \sum_{j=1}^i 1/m(j)$, and extend h by linear interpolation to real $0 \leq x \leq n$. Then h is concave and for $i \geq 1$

$$\begin{aligned} E_i h(X_1) &\leq h(E_i X_1) \text{ by concavity} \\ &= h(i - m(i)) \\ &\leq h(i) - m(i) h'(i) \text{ by concavity} \\ &= h(i) - 1 \end{aligned}$$

where h' is the left derivative. The result now follows from Lemma 2.31.

2.8.3 Wald equations

As mentioned previously, the results above don't really require martingales. Next we record a genuine martingale result, not directly involving Markov chains but ultimately useful in their analysis. Part (c) is *Wald's equation* and part (b) is *Wald's equation for martingales*. The result is a standard consequence of the optional sampling theorem: see [133] (3.1.6) for (c) and [133] Theorem 4.7.5 for (a).

Lemma 2.33 *(a) Let* $0 = Y_0 \le Y_1 \le Y_2 \ldots$ *be such that*

$$E(Y_{i+1} - Y_i | Y_j, j \le i) \le c, \ i \ge 0$$

for a constant c*. Then for any stopping time* T*,*

$$EY_T \le cET.$$

(b) If in the hypothesis we replace "$\le c$*" by "*$= c$*", then* $EY_T = cET$*.*
(c) In particular, if $Y_n = \sum_{i=1}^{n} \xi_i$ *for i.i.d. nonnegative* (ξ_i) *then* $EY_T = (E\xi_i)(ET)$*.*

2.9 Notes on Chapter 2.

Textbooks on Markov chains.

> It is easy to write books on ... or finite Markov chains, or on any of the other well-understood topics for which no further expositions are needed. *G.-C. Rota*
>
> `Your search for the Subject: MARKOV PROCESSES retrieved 273 records`. *U.C. Berkeley Library book catalog, September 1999.*

Almost every introductory textbook on stochastic processes has a chapter or two about Markov chains: among the best are Karlin-Taylor [208, 209], Grimmett-Stirzaker [177] and, slightly more advanced, Asmussen [34]. In addition to Norris [270] there are several other undergraduate-level textbooks entirely or mostly devoted to Markov chains: Adke-Manjanuth [1], Hunter [186], Iosifescu [189], Isaacson-Madsen [191], Kemeny-Snell [214], Romanovsky [297]. At the graduate level, Durrett [133] has a concise chapter on the modern approach to the basic limit theory. Several more advanced texts which overlap our material were mentioned in Chapter 1 section 2.3 (yyy 7/20/99 version); other texts are Freedman [154], Anderson [31], and

the treatize of Syski [318] on hitting times. Most textbooks leave an exaggerated impression of the difference between discrete- and continuous-time chains.

Section 2.1.2. Continuized is an ugly neologism, but no-one has collected my $5 prize for suggesting a better name!

Section 2.2. Elementary matrix treatments of results like those in section 2.2.2 for finite state space can be found in [186, 214]. On more general spaces, this is part of *recurrent potential* theory: see [96, 215] for the countable-state setting and Revuz [289] for continuous space. Our treatment, somewhat novel at the textbook level, Pitman [283] studied occupation measure identities more general than thos in section 2.2.1 and their applications to hitting time formulas, and we follow his approach in sectionMHTF. We are being slightly dishonest in treating Lemmas 2.5 and 2.6 this way, because these facts figure in the "right" proof of the ergodic theorems we use. We made a special effort *not* to abbreviate "number of visits to j before time S" as $N_j(S)$, which forces the reader to decode formulas.

Kemeny and Snell [214] call $\mathbf{Z} + \Pi$ the fundamental matrix, and use $(E_i T_j^+)$ rather than $(E_i T_j)$ as the matrix of mean hitting times. Our set-up seems a little smoother – cf. Meyer [202] who calls $\mathbf{Z}$ the *group inverse* of $\mathbf{I} - \mathbf{P}$.

The name "random target lemma" for Corollary 2.14 was coined by Lovász and Winkler [241]; the result itself is classical ([214] Theorem 4.4.10).

Open Problem 2.34 *Portmanteau theorem for occupation times.*

Can the results of section 2.2.2 be formulated as a single theorem? To explain the goal by analogy, consider the use [194] of *Feynman diagrams* to calculate quantities such as $E(A^3BC^2)$ for dependent mean-zero Gaussian (A, B, C). One rewrites the expectation as $E\prod_{i=1}^6 \xi_i$ for $\xi_1 = \xi_2 = \xi_3 = A; \xi_4 = B, \xi_5 = \xi_6 = C$, and then applies the formula

$$E\prod_{i=1}^{6} \xi_i = \sum_M \nu(M)$$

where the sum is over *matchings* $M = \{\{u_1, v_1\}, \{u_2, v_2\}, \{u_3, v_3\}\}$ of $\{1, 2, 3, 4, 5, 6\}$ and where

$$\nu(M) = \prod_{j=1}^{3} E(\xi_{u_j}\xi_{v_j}).$$

By analogy, we seek a general rule which associates an expression like

$$E_i(\text{number of visits to } j \text{ before time } \min(T_k, T_l))$$

with a combinatorial structure involving $\{i, j, k, l\}$; then associates with the combinatorial structure some function of variables $\{p_v, z_{vw},\ v, w \in \{i, j, k, l\}\}$; then shows that the value of the expression applied to a finite Markov chain equals the function of $\{\pi_v, Z_{vw},\ v, w \in \{i, j, k, l\}\}$.

Section 2.4.1. Corollary 2.21 and variants are the basis for the theory of positive-recurrent chains on continuous spaces: see [133] section 5.6 and Meyn and Tweedie [263].

Section 2.4.2. The fact that $||\theta(t) - \pi||_2$ is decreasing is a special case ($H(u) = u^2$) of the following result (e.g. [213] Theorem 1.6).

Lemma 2.35 *Let $H : [0, \infty) \to [0, \infty)$ be concave [convex]. Let $\theta(t)$ be the distribution of an irreducible chain with stationary distribution π. Then $\sum_i \pi_i H(\theta_i(t)/\pi_i)$ is increasing [decreasing].*

Section 2.6. Matthews [256, 257] introduced his method (Theorem 2.26) to study some highly symmetric walks (cf. Chapter 7) and to study some continuous-space Brownian motion covering problems.

Section 2.7. A more sophisticated notion is "the chain conditioned never to hit A", which can be formalized using Perron-Frobenius theory.

Section 2.8.1. Applying the optional stopping theorem involves checking side conditions (involving integrability of the martingale or the stopping time), but these are trivially satisfied in our applications.

Numerical methods. In many applications of non-reversible chains, e.g. to queueing-type processes, one must resort to numerical computations of the stationary distribution: see Stewart [314]. We don't discuss such issues because in the reversible case we have conceptually simple expressions for the stationary distribution,

Matrix methods. There is a curious dichotomy between textbooks on Markov chains which use matrix theory almost everywhere and textbooks which use matrix theory almost nowhere. Our style is close to the latter; matrix formalism obscures more than it reveals. For our purposes, the one piece of matrix theory which is really essential is the spectral decomposition of reversible transition matrices in Chapter 3. Secondarily useful is the theory surrounding the Perron-Frobenius theorem, quoted for reversible chains in Chapter 3 section 6.5. (yyy 9/2/94 version)

yyy move both subsections to Chapter 8 "A Second Look ...".

2.10 Move to other chapters

2.10.1 Attaining distributions at stopping times

We quote a result, Theorem 2.36, which may look superficially like the identities in section 2.2.1 but which in fact is deeper, in that it cannot be proved by mere matrix manipulations or by Proposition 2.3. The result goes back to Baxter and Chacon [44] (and is implicit in Rost [301]) in the more general continuous-space setting: a proof tailored to the finite state space case has recently been given by Lovász and Winkler [241].

Given distributions σ, μ, consider a stopping time T such that

$$P_\sigma(X_T \in \cdot) = \mu(\cdot). \tag{2.29}$$

Clearly, for any state j we have $E_\sigma T_j \le E_\sigma T + E_\mu T_j$, which rearranges to $E_\sigma T \ge E_\sigma T_j - E_\mu T_j$. So if we define

$$\bar{t}(\sigma, \mu) = \inf\{E_\sigma T : \ T \text{ a stopping time satisfying (2.29)}\}$$

then we have shown that $\bar{t}(\sigma, \mu) \ge \max_j(E_\sigma T_j - E_\mu T_j)$. Surprisingly, this inequality turns out to be an equality.

Theorem 2.36 $\bar{t}(\sigma, \mu) = \max_j(E_\sigma T_j - E_\mu T_j)$.

2.10.2 Differentiating stationary distributions

From the definition (2.6) of the fundamental matrix Z we can write, in matrix notation,

$$(\mathbf{I} - \mathbf{P})\mathbf{Z} = \mathbf{Z}(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \Pi \tag{2.30}$$

where Π is the matrix with (i, j)-entry π_j. The matrix $\mathbf{I} - \mathbf{P}$ is not invertible but (2.30) expresses $\mathbf{Z}$ as a "generalized inverse" of $\mathbf{I} - \mathbf{P}$, and one can use matrix methods to verify general identities in the spirit of section 2.2.1. See e.g. [186, 214]. Here is a setting where such matrix methods work well.

Lemma 2.37 *Suppose* $\mathbf{P}$ *(and hence* π *and* $\mathbf{Z}$*) depend on a real parameter* α*, and suppose* $\mathbf{R} = \frac{d}{d\alpha}\mathbf{P}$ *exists. Then, at* α *such that* $\mathbf{P}$ *is irreducible,*

$$\frac{d}{d\alpha}\pi = \pi\mathbf{R}\mathbf{Z}.$$

Proof. Write $\eta = \frac{d}{d\alpha}\pi$. Differentiating the balance equations $\pi = \pi\mathbf{P}$ gives $\eta = \eta\mathbf{P} + \pi\mathbf{R}$, in other words $\eta(\mathbf{I} - \mathbf{P}) = \pi\mathbf{R}$. Right-multiply by $\mathbf{Z}$ to get

$$\pi\mathbf{R}\mathbf{Z} = \eta(\mathbf{I} - \mathbf{P})\mathbf{Z} = \eta(\mathbf{I} - \Pi) = \eta - \eta\Pi.$$

But $\eta\Pi = 0$ because $\sum_i \eta_i = \frac{d}{d\alpha}(\sum_i \pi_i) = 0$.

Chapter 3

Reversible Markov Chains (September 10, 2002)

Chapter 2 reviewed some aspects of the elementary theory of general finite irreducible Markov chains. In this chapter we specialize to reversible chains, treating the discrete-time and continuous-time cases in parallel. After Section 3.3 we shall assume that we are dealing with reversible chains without continually repeating this assumption, and shall instead explicitly say "general" to mean not necessarily reversible. 9/10/99 version

3.1 Introduction

Recall $\mathbf{P}$ denotes the transition matrix and π the stationary distribution of a finite irreducible discrete-time chain (X_t). Call the chain *reversible* if

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all} \quad i, j. \tag{3.1}$$

Equivalently, suppose (for given irreducible $\mathbf{P}$) that π is a probability distribution satisfying (3.1). Then π is the unique stationary distribution and the chain is reversible. This is true because (3.1), sometimes called the *detailed balance equations*, implies

$$\sum_i \pi_i p_{ij} = \pi_j \sum_i p_{ji} = \pi_j \quad \text{for all } j$$

and therefore π satisfies the *balance equations* of (1) in Chapter 2. 9/10/99 version

The name *reversible* comes from the following fact. If (X_t) is the stationary chain, that is, if X_0 has distribution π, then

$$(X_0, X_1, \ldots, X_t) \stackrel{d}{=} (X_t, X_{t-1}, \ldots, X_0).$$

More vividly, given a movie of the chain run forwards and the same movie run backwards, you cannot tell which is which.

It is elementary that the same symmetry property (3.1) holds for the t-step transition matrix $\mathbf{P}^t$:

$$\pi_i p_{i,j}^{(t)} = \pi_j p_{ji}^{(t)}$$

9/10/99 version

and thence for the matrix $\mathbf{Z}$ of (6) in Chapter 2:

$$\pi_i Z_{ij} = \pi_j Z_{ji}. \tag{3.2}$$

But beware that the symmetry property does not work for mean hitting times: the assertion

$$\pi_i E_i T_j = \pi_j E_j T_i$$

is definitely *false* in general (see the Notes for one intuitive explanation). See Chapter 7 for further discussion. The following general lemma will be useful there.

1/31/94 version

The lemma has been copied here from Section 1.2 of Chapter 7 (1/31/94 version); reminder: it still needs to be deleted there!

Lemma 3.1 *For an irreducible reversible chain, the following are equivalent.*

(a) $P_i(X_t = i) = P_j(X_t = j), \quad i, j \in I, \quad t \geq 1$

(b) $P_i(T_j = t) = P_j(T_i = t), \quad i, j \in I, \quad t \geq 1.$

Proof. In either case the stationary distribution is uniform—under (a) by letting $t \to \infty$, and under (b) by taking $t = 1$, implying $p_{ij} \equiv p_{ji}$. So by reversibility $P_i(X_t = j) = P_j(X_t = i)$ for $i \neq j$ and $t \geq 1$. But recall from

9/10/99 version

Chapter 2 Lemma 25 that the generating functions

$$G_{ij}(z) := \sum_t P_i(X_t = j) z^j, \qquad F_{ij}(z) := \sum_t P_i(T_t = j) z^j$$

satisfy

$$F_{ij} = G_{ij}/G_{jj}. \tag{3.3}$$

For $i \neq j$ we have seen that $G_{ij} = G_{ji}$, and hence by (3.3)

$$F_{ij} = F_{ji} \text{ iff } G_{jj} = G_{ii},$$

which is the assertion of Lemma 3.1. ∎

The discussion above extends to continuous time with only notational changes, e.g., the detailed balance equation (3.1) becomes

$$\pi_i q_{ij} = \pi_j q_{ji} \text{ for all } i, j. \tag{3.4}$$

3.1.1 Time-reversals and cat-and-mouse games

For a general chain we can define the *time-reversed* chain to have transition matrix $\mathbf{P}^*$ where

$$\pi_i p_{ij} = \pi_j p^*_{ji}$$

so that the chain is reversible iff $\mathbf{P}^* = \mathbf{P}$. One can check [cf. (3.2)]

$$\pi_i Z_{ij} = \pi_j Z^*_{ji}. \tag{3.5}$$

The stationary $\mathbf{P}^*$-chain is just the stationary $\mathbf{P}$-chain run backwards in time. Consider Examples 16 and 22 from Chapter 2. In Example 16 (patterns in coin tossing) the time-reversal $\mathbf{P}^*$ just "shifts left" instead of shifting right, i.e., from $HTTTT$ the possible transitions are to $HHTTT$ and $THTTT$. In Example 22 the time-reversal just reverses the direction of motion around the n-cycle:

9/10/99 version

$$p^*_{ij} = a 1_{(j=i-1)} \ + \ \frac{1-a}{n}.$$

Warning. These examples are simple because the stationary distributions are uniform. If the stationary distribution has no simple form then typically $\mathbf{P}^*$ will have no simple form.

A few facts about reversible chains are really specializations of facts about general chains which involve both $\mathbf{P}$ and $\mathbf{P}^*$. Here is a simple instance.

Lemma 3.2 (The cyclic tour property) *For states $i_0, i_1, \ldots, i_m$ of a reversible chain,*

$$E_{i_0}T_{i_1} + E_{i_1}T_{i_2} + \cdots + E_{i_m}T_{i_0} = E_{i_0}T_{i_m} + E_{i_m}T_{i_{m-1}} + \cdots + E_{i_1}T_{i_0}.$$

The explanation is that in a general chain we have

$$E_{i_0}T_{i_1} + E_{i_1}T_{i_2} + \cdots + E_{i_m}T_{i_0} = E^*_{i_0}T_{i_m} + E^*_{i_m}T_{i_{m-1}} + \cdots + E^*_{i_1}T_{i_0} \tag{3.6}$$

where E^* refers to the time-reversed chain $\mathbf{P}^*$. Equality (3.6) is intuitively obvious when we visualize running a movie backwards. But a precise argument requires a little sophistication (see Notes). It is however straightforward to *verify* (3.6) using (3.5) and the mean hitting time formula $E_iT_j = (Z_{jj} - Z_{ij})/\pi_j$.

We shall encounter several results which have amusing interpretations as cat-and-mouse games. The common feature of these games is that the cat moves according to a transition matrix $\mathbf{P}$ and the mouse moves according to the time-reversed transition matrix $\mathbf{P}^*$.

Cat-and-mouse game 1. Both animals are placed at the same state, chosen according to the stationary distribution. The mouse makes a jump according to $\mathbf{P}^*$, and then stops. The cat starts moving according to $\mathbf{P}$ and continues until it finds the mouse, after M steps.

The notable feature of this game is the simple formula for EM:

$$EM = n - 1, \text{ where } n \text{ is the number of states.} \tag{3.7}$$

This is simple once you see the right picture. Consider the stationary $\mathbf{P}$-chain $(X_0, X_1, X_2, \ldots)$. We can specify the game in terms of that chain by taking the initial state to be X_1, and the mouse's jump to be to X_0, and the cat's moves to be to $X_2, X_3, \ldots$. So $M = T^+ - 1$ with

$$T^+ := \min\{t \geq 1 : X_t = X_0\}.$$

And $ET^+ = \sum_i \pi_i E_i T_i^+ = \sum_i \pi_i \frac{1}{\pi_i} = n$.

Cat-and-mouse game 2. This game, and Proposition 3.3, are rephrasings of results of Coppersmith et al [101]. Think of the cat and the mouse as pieces in a solitaire board game. The player sees their positions and chooses which one to move: if the cat is chosen, it makes a move according to $\mathbf{P}$, and if the mouse is chosen, it makes a move according to $\mathbf{P}^*$. Let M denote the number of moves until the cat and mouse meet. Then one expects the mean $E_{(x,y)}M$ to depend on the initial positions (x, y) of (cat, mouse) and on the player's strategy. But consider the example of asymmetric random walk on a n-cycle, with (say) chance 2/3 of moving clockwise and chance 1/3 of moving counterclockwise. A moment's thought reveals that the distance (measured clockwise from cat to mouse) between the animals does not depend on the player's strategy, and hence neither does $E_{(x,y)}M$. In general EM does depend on the strategy, but the following result implies that the size of the effect of strategy changes can be bounded in terms of a measure of non-symmetry of the chain.

Proposition 3.3 *Regardless of strategy,*

$$\min_z E_\pi T_z \leq E_{(x,y)}M - (E_x T_y - E_\pi T_y) \leq \max_z E_\pi T_z$$

where hitting times T refer to the $\mathbf{P}$-chain.

Symbol used here is "defined identically to be".

Proof. Consider the functions

$$f(x, y) :\equiv E_x T_y - E_\pi T_y$$

$$f^*(y, x) :\equiv E_y^* T_x - E_\pi^* T_x.$$

The first-step recurrences for $x \mapsto E_x T_y$ and $y \mapsto E^*_y T_x$ give

$$f(x,y) = 1 + \sum_z p_{xz} f(z,y), \quad y \neq x \tag{3.8}$$

$$f^*(y,x) = 1 + \sum_z p^*_{yz} f^*(z,x), \quad y \neq x. \tag{3.9}$$

By the mean hitting time formula

$$f(x,y) = \frac{-Z_{xy}}{\pi_y} = \frac{-Z^*_{yx}}{\pi_x} = f^*(y,x)$$

so we may rewrite (3.9) as

$$f(x,y) = 1 + \sum_z p^*_{yz} f(x,z), \quad y \neq x. \tag{3.10}$$

Now let $(\hat{X}_t, \hat{Y}_t)$ be the positions of (cat, mouse) after t moves according to some strategy. Consider

$$W_t \equiv t + f(\hat{X}_t, \hat{Y}_t).$$

Equalities (3.8) and (3.10) are exactly what is needed to verify

$$(W_t; 0 \leq t \leq M) \text{ is a martingale.}$$

So the optional stopping theorem says $E_{(x,y)} W_0 = E_{(x,y)} W_M$, that is,

$$f(x,y) = E_{(x,y)} M + E_{(x,y)} f(\hat{X}_M, \hat{Y}_M). \tag{3.11}$$

But $\hat{X}_M = \hat{Y}_M$ and $-f(z,z) = E_\pi T_z$, so

$$\min_z E_\pi T_z \leq -f(\hat{X}_M, \hat{Y}_M) \leq \max_z E_\pi T_z$$

and the result follows from (3.11). ∎

Remarks. Symmetry conditions in the reversible setting are discussed in Chapter 7. Vertex-transitivity forces $E_\pi T_z$ to be independent of z, and hence in the present setting implies $E_{(x,y)} M = E_x T_y$ regardless of strategy. For a reversible chain without this symmetry condition, consider (x_0, y_0) attaining the *min* and *max* of $E_\pi T_z$. The Proposition then implies $E_{y_0} T_{x_0} \leq E_{(x_0,y_0)} M \leq E_{x_0} T_{y_0}$ and the bounds are attained by keeping one animal fixed. But for general initial states the bounds of the Proposition are not attained. Indeed, the proof shows that to attain the bounds we need a strategy which forces the animals to meet at states attaining the extrema of $E_\pi T_z$. Finally, in the setting of random walk on a n-vertex graph we can combine Proposition 3.3 with mean hitting time bounds from Chapter 6 to show that EM is at worst $O(n^3)$.

1/31/94 version

10/31/94 version

3.1.2 Entrywise ordered transition matrices

9/10/99 version

This subsection adapted from Section 8.1 of Chapter MCMC (1/8/01 version); reminder: it still needs to be deleted there!

Recall from Chapter 2 Section 3 that for a function $f : S \to R$ with $\sum_i \pi_i f_i = 0$, the asymptotic variance rate is

$$\sigma^2(\mathbf{P}, f) := \lim_t t^{-1}\text{var} \sum_{s=1}^{t} f(X_s) = f\Gamma f \tag{3.12}$$

where $\Gamma_{ij} = \pi_i Z_{ij} + \pi_j Z_{ji} + \pi_i\pi_j - \pi_i\delta_{ij}$. These individual-function variance rates can be compared between chains with the same stationary distribution, under a very strong "(off-diagonal) entrywise ordering" of reversible transition matrices.

Lemma 3.4 (Peskun's Lemma [280]) *Let $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ be reversible with the same stationary distribution π. Suppose $p^{(1)}_{ij} \le p^{(2)}_{ij}$ for all $j \ne i$. Then $\sigma^2(\mathbf{P}^{(1)}, f) \ge \sigma^2(\mathbf{P}^{(2)}, f)$ for all f with $\sum_i \pi_i f_i = 0$.*

Proof. Introduce a parameter $0 \le \alpha \le 1$ and write

$$\mathbf{P} = \mathbf{P}(\alpha) := (1-\alpha)\mathbf{P}^{(1)} + \alpha\mathbf{P}^{(2)}.$$

Write $(\cdot)'$ for $\frac{d}{d\alpha}(\cdot)$. It is enough to show

$$(\sigma^2(\mathbf{P}, f))' \le 0.$$

By (3.12)

$$(\sigma^2(\mathbf{P}, f))' = f\Gamma' f = 2\sum_i \sum_j f_i \pi_i Z'_{ij} f_j.$$

Need to decide where to put statement and proof of that lemma
1/8/01 version

By Chapter MCMC Lemma 4, $\mathbf{Z}' = \mathbf{Z}\mathbf{P}'\mathbf{Z}$. By setting

$$g_i := \pi_i f_i; \quad a_{ij} := Z_{ij}/\pi_j; \quad w_{ij} := \pi_i p_{ij}$$

we find $\mathbf{A}' = \mathbf{A}\mathbf{W}'\mathbf{A}$ and can rewrite the equality above as

$$(\sigma^2(\mathbf{P}, f))' = 2\ g\mathbf{A}\mathbf{W}'\mathbf{A}g.$$

agreed: use negative semidefinite rather than nonpositive definite throughout

Since $\mathbf{A}$ is symmetric, it is enough to show that $\mathbf{W}'$ is negative semidefinite. By hypothesis $\mathbf{W}'$ is symmetric with zero row-sums and $w'_{ij} \ge 0$ for $j \ne i$. Ordering states arbitrarily, we may write

$$\mathbf{W}' = \sum_{i,j:\, i<j} w'_{ij}\mathbf{M}^{ij}$$

where $\mathbf{M}^{ij}$ is the matrix whose only nonzero entries are $m(i,i) = m(j,j) = -1$ and $m(i,j) = m(j,i) = 1$. Plainly $\mathbf{M}^{ij}$ is negative semidefinite, hence so is $\mathbf{W}'$. ∎

3.2 Reversible chains and weighted graphs

Our convention is that a *graph* has finite vertex-set $\mathcal{V} = \{v, x, y, \ldots\}$ and edge-set $\mathcal{E} = \{e_1, e_2, \ldots\}$, is connected and undirected, has no multiple edges, and has no self-loops. In a *weighted graph*, each edge (v, w) also has a weight $0 < w_{v,x} = w_{x,v} < \infty$, and we allow a weighted graph to have self-loops.

Given a weighted graph, there is a natural definition of a Markov chain on the vertices. This requires an arbitrary choice of convention: do we want to regard an absent edge as having weight 0 or weight $+\infty$? In terms of electrical networks (Section 3.3) the question is whether to regard weights as conductances or as resistances of wires. Conceptually one can make good arguments for either choice, but formulas look simpler with the conductance convention (absent edges have weight 0), so we'll adopt that convention. Define discrete-time *random walk* on a weighted graph to be the Markov chain with transition matrix

$$p_{vx} := w_{vx}/w_v, \ x \neq v \tag{3.13}$$

where

$$w_v := \sum_x w_{vx}, \quad w := \sum_v w_v.$$

Note that w is the total edge-weight, when each edge is counted *twice*, i.e., once in each direction. The fundamental fact is that this chain is automatically reversible with stationary distribution

$$\pi_v \equiv w_v/w \tag{3.14}$$

because (3.1) is obviously satisfied by $\pi_v p_{vx} = \pi_x p_{xv} = w_{vx}/w$. Our standing convention that graphs be connected implies that the chain is irreducible. Conversely, with our standing convention that chains be irreducible, any reversible chain can be regarded as as random walk on the weighted graph with edge-weights $w_{vx} := \pi_v p_{vx}$. Note also that the "aperiodic" condition for a Markov chain (occurring in the convergence theorem Chapter 2 Theorem 2) is just the condition that the graph be not bipartite. 9/10/99 version

An unweighted graph can be fitted into this setup by simply assigning weight 1 to each edge. Since we'll be talking a lot about this case, let's write out the specialization explicitly. The transition matrix becomes

$$p_{vx} = \begin{cases} 1/d_v & \text{if } (v,x) \text{ is an edge} \\ 0 & \text{if not} \end{cases}$$

where d_v is the degree of vertex v. The stationary distribution becomes

$$\pi_v = \frac{d_v}{2|\mathcal{E}|} \tag{3.15}$$

where $|\mathcal{E}|$ is the number of edges of the graph. In particular, on an unweighted *regular* graph the stationary distribution is uniform.

In continuous time there are two *different* ways to associate a walk with a weighted or unweighted graph. One way (*and we use this way unless otherwise mentioned*) is just to use (3.13) as the definition of the transition rates q_{vx}. In the language of Chapter 2 this is the continuization of the discrete-time walk, and has the same stationary distribution and mean hitting times as the discrete-time walk. The alternative definition, which we call the *fluid model*, uses the weights directly as transition rates:

9/10/99 version

$$q_{vx} := w_{vx}, \quad x \neq v. \tag{3.16}$$

In this model the stationary distribution is always uniform (cf. Section 3.2.1). In the case of an unweighted *regular* graph the two models are identical up to a deterministic time rescaling, but for non-regular graphs there are typically no exact relations between numerical quantities for the two continuous-time models. Note that, given an arbitrary continuous-time reversible chain, we can define edge-weights (w_{ij}) via

$$\pi_i q_{ij} = \pi_j q_{ji} = w_{ij}, \text{ say}$$

but the weights (w_{ij}) do not completely determine the chain: we can specify the π_i independently and then solve for the q's.

Though there's no point in writing out all the specializations of the general theory of Chapter 2, let us emphasize the simple expressions for mean return times of discrete-time walk obtained from Chapter 2 Lemma 5 and the expressions (3.14)–(3.15) for the stationary distribution.

9/10/99 version

9/10/99 version

Lemma 3.5 *For random walk on an n-vertex graph,*

$$\begin{aligned} E_v T_v^+ &= \frac{w}{w_v} \quad (\textit{weighted}) \\ &= \frac{2|\mathcal{E}|}{d_v} \quad (\textit{unweighted}) \\ &= n \quad (\textit{unweighted regular}). \end{aligned}$$

The example has been copied here from the start of Example 18 of Chapter 5 (4/22/96 version); reminder: that example needs to be modified accordingly!

Example 3.6 *Chess moves.*

Here is a classic homework problem for an undergraduate Markov chains course.

> Start a knight at a corner square of an otherwise-empty chessboard. Move the knight at random, by choosing uniformly from the legal knight-moves at each step. What is the mean number of moves until the knight returns to the starting square?

It's a good question, because if you don't know Markov chain theory it looks too messy to do by hand, whereas using Markov chain theory it becomes very simple. The knight is performing random walk on a graph (the 64 squares are the vertices, and the possible knight-moves are the edges). It is not hard to check that the graph is connected, so by the elementary Lemma 3.5, for a corner square v the mean return time is

$$E_v T_v^+ = \frac{1}{\pi_v} = \frac{2|\mathcal{E}|}{d_v} = |\mathcal{E}|,$$

and by drawing a sketch in the margin the reader can count the number of edges $|\mathcal{E}|$ to be 168.

The following cute variation of Lemma 3.5 is sometimes useful. Given the discrete-time random walk (X_t), consider the process

$$Z_t = (X_{t-1}, X_t)$$

recording the present position at time t and also the previous position. Clearly (Z_t) is a Markov chain whose state-space is the set $\vec{\mathcal{E}}$ of *directed* edges, and its stationary distribution (ρ, say) is

$$\rho(v,x) = \frac{w_{vx}}{w}$$

in the general weighted case, and hence

$$\rho(v,x) = \frac{1}{|\vec{\mathcal{E}}|}, \quad (x,v) \in \vec{\mathcal{E}}$$

in the unweighted case. Now given an edge (x,v), we can apply Chapter 2 Lemma 5 to (Z_t) and the state (x,v) to deduce the following. 9/10/99 version

Lemma 3.7 *Given an edge (v,x) define*

$$U := \min\{t \geq 1 : X_t = v,\ X_{t-1} = x\}.$$

Then

$$\begin{aligned} E_v U &= \frac{w}{w_{vx}} \quad (\textit{weighted}) \\ &= 2|\mathcal{E}| \quad (\textit{unweighted}). \end{aligned}$$

Corollary 3.8 (The edge-commute inequality) *For an edge (v, x),*

$$E_v T_x + E_x T_v \le \begin{array}{ll} \frac{w}{w_{vx}} & (\textit{weighted}) \\ 2|\mathcal{E}| & (\textit{unweighted}) \end{array} .$$

We shall soon see (Section 3.3.3) this inequality has a natural interpretation in terms of electrical resistance, but it is worth remembering that the result is more elementary than that.

Here is another variant of Lemma 3.5.

Lemma 3.9 *For random walk on a weighted n-vertex graph,*

$$\sum_{e=(v,x)} w_e (E_v T_x + E_x T_v) = w(n-1)$$

where the sum is over undirected edges.

Proof. Writing $\sum_v \sum_x$ for the sum over directed edges (v, x), the left side equals

$$\begin{aligned} & \frac{1}{2} \sum_v \sum_x w_{vx} (E_v T_x + E_x T_v) \\ = & \sum_v \sum_x w_{vx} E_x T_v \ \text{ by symmetry} \\ = & w \sum_v \pi_v \sum_x p_{vx} E_x T_v \\ = & w \sum_v \pi_v (E_v T_v^+ - 1) \\ = & w \sum_v \pi_v (\tfrac{1}{\pi_v} - 1) \\ = & w(n-1). \end{aligned}$$

■

3.2.1 The fluid model

Imagine a finite number of identical buckets which can hold unit quantity of fluid. Some pairs of buckets are connected by tubes through their bottoms. If a tube connects buckets i and j then, when the quantities of fluid in

buckets i and j are p_i and p_j, the flow rate through the tube should be proportional to the pressure difference and hence should be $w_{ij}(p_i - p_j)$ in the direction $i \to j$, where $w_{ij} = w_{ji}$ is a parameter. Neglecting the fluid in the tubes, the quantities of fluid ($p_i(t)$) at time t will evolve according to the differential equations

$$\frac{dp_j(t)}{dt} = \sum_{i \neq j} w_{ij}(p_i(t) - p_j(t)).$$

These of course are the same equations as the *forward equations* [(4) of Chapter 2] for $p_i(t)$ (the probability of being in state i at time t) for the continuous-time chain with transition rates $q_{ij} = w_{ij},\ j \neq i$. Hence we call this particular way of defining a continuous-time chain in terms of a weighted graph the *fluid model*. Our main purpose in mentioning this notion is to distinguish it from the electrical network analogy in the next section. Our intuition about fluids says that as $t \to \infty$ the fluid will distribute itself uniformly amongst buckets, which corresponds to the elementary fact that the stationary distribution of the "fluid model" chain is always uniform. Our intuition also says that increasing a "specific flow rate" parameter w_{ij} will make the fluid settle faster, and this corresponds to a true fact about the "fluid model" Markov chain (in terms of the eigenvalue interpretation of asymptotic convergence rate—see Corollary 3.28). On the other hand the same assertion for the usual discrete-time chain or its continuization is simply false.

9/10/99 version

3.3 Electrical networks

3.3.1 Flows

This is a convenient place to record some definitions. A *flow* $\mathbf{f} = (f_{ij})$ on a graph is required only to satisfy the conditions

$$f_{ij} = \begin{cases} -f_{ji} & \text{if } (i,j) \text{ is an edge} \\ 0 & \text{if not.} \end{cases}$$

So the net flow out of i is $f_{(i)} := \sum_{j \neq i} f_{ij}$, and by symmetry $\sum_i f_{(i)} = 0$. We will be concerned with flows satisfying extra conditions. Given disjoint non-empty subsets A, B of vertices, a *unit flow from B to A* is a flow satisfying

$$\sum_{i \in B} f_{(i)} = 1, \qquad f_{(j)} = 0 \ \text{ for all } j \notin A \cup B \tag{3.17}$$

which implies $\sum_{i \in A} f_{(i)} = -1$. Given a Markov chain X (in particular, given a weighted graph we can use the random walk) we can define a special flow as follows. Given $v_0 \notin A$, define $\mathbf{f}^{v_0 \to A}$ by

$$f_{ij} := E_{v_0} \sum_{t=1}^{T_A} \left(1_{(X_{t-1}=i, X_t=j)} - 1_{(X_{t-1}=j, X_t=i)} \right). \tag{3.18}$$

So f_{ij} is the mean number of transitions $i \to j$ minus the mean number of transitions $j \to i$, for the chain started at v_0 and run until hitting A. Clearly $\mathbf{f}^{v_0 \to A}$ is a unit flow from v_0 to A. Note that the mean net transitions definition of f_{ij} works equally well in continuous time to provide a unit flow $\mathbf{f}^{v_0 \to A}$ from v_0 to A.

In Section 3.7.2 we will define the notion of "a unit flow from v_0 to a probability distribution ρ" and utilize a special unit flow from v_0 to the stationary distribution.

3.3.2 The analogy

Given a weighted graph, consider the graph as an electrical network, where a wire linking v and x has conductance w_{vx}, i.e., resistance $1/w_{vx}$. Fix a vertex v_0 and a subset A of vertices not containing v_0. Apply voltage 1 at v_0 and ground (i.e., set at voltage 0) the set A of vertices. As we shall see, this determines the voltage $g(v)$ at each vertex v; in particular,

$$g(v_0) = 1; \quad g(\cdot) = 0 \text{ on } A. \tag{3.19}$$

Physically, according to *Ohm's law*,

$$\text{Current} = \frac{\text{Potential difference}}{\text{Resistance}}$$

for each wire; that is, the current I_{vx} along each wire (v, x) satisfies

$$I_{vx} = (g(v) - g(x)) w_{vx}. \tag{3.20}$$

Clearly, I is a flow, and according to *Kirchoff's node law*

$$I_{(v)} = 0, \quad v \notin \{v_0\} \cup A. \tag{3.21}$$

Regarding the above as intuition arising from the study of physical electrical networks, we can define an *electrical network* mathematically as a weighted graph together with a function g and a flow I, called *voltage* and *current*, respectively, satisfying (3.20)–(3.21) and the normalization (3.19).

As it turns out, these three conditions specify g [and hence also I, by (3.20)] uniquely since (3.20)–(3.21) imply

$$g(v) = \sum_x p_{vx} g(x), \quad v \notin \{v_0\} \cup A \tag{3.22}$$

with p_{vx} defined at (3.13), and Chapter 2 Lemma 27 shows that this equation, together with the boundary conditions (3.19), has a unique solution. Conversely, if g is the unique function satisfying (3.22) and (3.19), then I defined by (3.20) satisfies (3.21), as required. Thus a weighted graph uniquely determines both a random walk and an electrical network. 9/10/99 version

The point of this subsection is that the voltage and current functions can be identified in terms of the random walk. Recall the flow $\mathbf{f}^{v_0 \to A}$ defined at (3.18).

Proposition 3.10 *Consider a weighted graph as an electrical network, where a wire linking v and x has conductance w_{vx}. Suppose that the voltage function g satisfies* (3.19)*. Then the voltage at any vertex v is given in terms of the associated random walk by*

$$g(v) = P_v(T_{v_0} < T_A) \in [0, 1] \tag{3.23}$$

and the current I_{vx} along each wire (v, x) is f_{vx}/r, where $\mathbf{f} = \mathbf{f}^{v_0 \to A}$ and

$$r = \frac{1}{w_{v_0} P_{v_0}(T_A < T_{v_0}^+)} \in (0, \infty). \tag{3.24}$$

Since $\mathbf{f}$ is a unit flow from v_0 to A and $g(v_0) = 1$, we find $I_{(v_0)} = g(v_0)/r$. Since $g = 0$ on A, it is thus natural in light of Ohm's law to regard the entire network as effectively a single conductor from v_0 to A with resistance r; for this reason r is called the *effective resistance* between v_0 and A. Since (3.19) and (3.21) are clearly satisfied, to establish Proposition 3.10 it suffices by our previous comments to prove (3.20), i.e.,

$$\frac{f_{vx}}{r} = (g(v) - g(x)) w_{vx}. \tag{3.25}$$

Proof of (3.25). Here is a "brute force" proof by writing everything in terms of mean hitting times. First, there is no less of generality in assuming that A is a singleton a, by the collapsing principle (Chapter 2 Section 7.3). Now by the Markov property 9/10/99 version

$$\begin{aligned} f_{vx} \;=\; & E_{v_0}(\text{number of visits to } v \text{ before time } T_a)\; p_{vx} \\ & - E_{v_0}(\text{number of visits to } x \text{ before time } T_a)\; p_{xv}. \end{aligned}$$

Chapter 2 Lemma 9 gives a formula for the expectations above, and using 9/10/99 version
$\pi_v p_{vx} = \pi_x p_{xv} = w_{vx}/w$ we get

$$\frac{w}{w_{vx}} f_{vx} = E_a T_v - E_{v_0} T_v - E_a T_x + E_{v_0} T_x. \tag{3.26}$$

9/10/99 version And Chapter 2 Corollaries 8 and 10 give a formula for g:

$$g(v) = (E_v T_a + E_a T_{v_0} - E_v T_{v_0}) \pi_{v_0} P_{v_0}(T_a < T^+_{v_0})$$

which leads to

$$\frac{g(v) - g(x)}{\pi_{v_0} P_{v_0}(T_a < T^+_{v_0})} = E_v T_a - E_x T_a - E_v T_{v_0} + E_x T_{v_0}. \tag{3.27}$$

But the right sides of (3.27) and (3.26) are equal, by the cyclic tour property (Lemma 3.2) applied to the tour v_0, x, a, v, v_0, and the result (3.25) follows after rearrangement, using $\pi_{v_0} = w_{v_0}/w$. ■

Remark. Note that, when identifying a reversible chain with an electrical network, the procedure of collapsing the set A of states of the chain to a singleton corresponds to the procedure of shorting together the vertices A of the electrical network.

3.3.3 Mean commute times

The classical use of the electrical network analogy in the mathematical literature is in the study of the recurrence or transience of infinite-state reversible 6/23/01 version chains by comparison arguments (Chapter 13). As discussed in Doyle and Snell [131], the comparisons involve "cutting or shorting". Cutting an edge, or more generally decreasing an edge's conductance, can only increase an effective resistance. Shorting two vertices together (i.e., linking them with an edge of infinite conductance), or more generally increasing an edge's conductance, can only decrease an effective resistance. These ideas can be formalized via the extremal characterizations of Section 3.7 without explicitly relying on the electrical analogy.

In our context of finite-state chains the key observation is the follow- 9/10/99 version ing. For not-necessarily-reversible discrete-time chains we have (Chapter 2 Corollary 8)

$$\frac{1}{\pi_v P_v(T_a < T^+_v)} = E_v T_a + E_a T_v, \quad v \neq a, \tag{3.28}$$

where we may call the right side the *mean commute time* between v and a. [For continuous-time chains, π_v is replaced by $q_v \pi_v$ in (3.28).] Comparing with (3.24) and using $\pi_v = w_v/w$ gives

Corollary 3.11 (commute interpretation of resistance) *Given two vertices v, a in a weighted graph, the effective resistance r_{va} between v and a is related to the mean commute time of the associated random walk by*

$$E_v T_a + E_a T_v = w r_{va}.$$

Note that the Corollary takes a simple form in the case of unweighted graphs:

$$E_v T_a + E_a T_v = 2|\mathcal{E}| r_{va}. \tag{3.29}$$

Note also that the Corollary does not hold so simply if a and v are both replaced by subsets—see Corollary 3.37.

Corollary 3.11 apparently was not stated explicitly or exploited until a 1989 paper of Chandra et al [85], but then rapidly became popular in the "randomized algorithms" community. The point is that "cutting or shorting" arguments can be used to bound mean commute times. As the simplest example, it is obvious that the effective resistance r_{vx} across an edge (v, x) is at most the resistance $1/w_{vx}$ of the edge itself, and so Corollary 3.11 implies the edge-commute inequality (Corollary 3.8). Finally, we can use Corollary 3.11 to get simple exact expressions for mean commute times in some special cases, in particular for birth-and-death processes (i.e., weighted linear graphs) discussed in Chapter 5.

add pointers to uses of cutting and shorting later in book

4/22/96 version

As with the infinite-space results, the electrical analogy provides a vivid language for comparison arguments, but the arguments themselves can be justified via the extremal characterizations of Section 3.7 without explicit use of the analogy.

3.3.4 Foster's theorem

The commute interpretation of resistance allows us to rephrase Lemma 3.9 as the following result about electrical networks, due to Foster [153].

Corollary 3.12 (Foster's Theorem) *In a weighted n-vertex graph, let r_e be the effective resistance between the ends (a, b) of an edge e. Then*

$$\sum_e r_e w_e = n - 1.$$

* *

CONVENTION.

For the rest of the chapter we make the convention that we are dealing with a finite-state, irreducible, reversible chain, and we will not repeat the "reversible" hypothesis in each result. Instead we will say "general chain" to mean not-necessarily-reversible chain.

* *

3.4 The spectral representation

Use the transition matrix $\mathbf{P}$ to define

$$s_{ij} = \pi_i^{1/2} p_{ij} \pi_j^{-1/2}.$$

From definition (3.1), $\mathbf{S}$ is a *symmetric* matrix. So we can apply the elementary diagonalization theorem. The authors find it helpful to distinguish between the state space $I = \{i, j, \ldots\}$, of size n say, and the index set of integers $[n] = \{1, 2, \ldots, n\}$, the point being that the state space may have a lot of extra structure, whereas the index set has no obvious structure. The spectral theorem ([183] Theorem 4.1.5) gives a representation

$$\mathbf{S} = \mathbf{U} \Lambda \mathbf{U}^T$$

where $\mathbf{U} = (u_{im})_{i \in I, m \in [n]}$ is an orthonormal matrix, and $\Lambda = (\lambda_{m,m'})_{m,m' \in [n]}$ is a diagonal real matrix. We can write the diagonal entries of Λ as (λ_m), and arrange them in decreasing order. Then

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1. \tag{3.30}$$

The classical fact that $|\lambda_i| \leq 1$ follows easily from the fact that the entries of $\mathbf{S}^{(t)}$ are bounded as $t \to \infty$ by (3.31) below. These λ's are the eigenvalues of $\mathbf{P}$, as well as of $\mathbf{S}$. That is, the solutions $(\lambda; x)$ with $x_i \not\equiv 0$ of

$$\sum_i x_i p_{ij} = \lambda x_j \quad \text{for all } j$$

are exactly the pairs

$$(\lambda = \lambda_m; x_i = c_m \pi_i^{1/2} u_{im}, i = 1, \ldots, n)$$

for $m = 1, \ldots, n$, where $c_m \neq 0$ is arbitrary. And the solutions of

$$\sum_j p_{ij} y_j = \lambda y_i \quad \text{for all } i$$

are exactly the pairs

$$(\lambda = \lambda_m;\ \ y_i = c_m \pi_i^{-1/2} u_{im},\ i = 1, \ldots, n).$$

Note that an eigenvector (u_{i1}) of $\mathbf{S}$ corresponding to the eigenvalue $\lambda_1 = 1$ is

$$u_{i1} = \pi_i^{1/2}.$$

Uniqueness of the stationary distribution now implies $\lambda_2 < 1$.

Now consider matrix powers. We have

$$\mathbf{S}^{(t)} = \mathbf{U}\Lambda^{(t)}\mathbf{U}^T$$

and

$$p_{ij}^{(t)} = \pi_i^{-1/2} s_{ij}^{(t)} \pi_j^{1/2}, \tag{3.31}$$

so

$$P_i(X_t = j) = \pi_i^{-1/2} \pi_j^{1/2} \sum_{m=1}^{n} \lambda_m^t u_{im} u_{jm}. \tag{3.32}$$

This is the *spectral representation formula*. In continuous time, the analogous formula is

$$P_i(X_t = j) = \pi_i^{-1/2} \pi_j^{1/2} \sum_{m=1}^{n} \exp(-\lambda_m t) u_{im} u_{jm}. \tag{3.33}$$

As before, $\mathbf{U}$ is an orthonormal matrix and $u_{i1} = \pi_i^{1/2}$, and now the λ's are the eigenvalues of $-\mathbf{Q}$. In the continuous-time setting, the eigenvalues satisfy

$$0 = \lambda_1 < \lambda_2 \le \cdots \le \lambda_n. \tag{3.34}$$

Rather than give the general proof, let us consider the effect of continuizing the discrete-time chain (3.32). The continuized chain (Y_t) can be represented as $Y_t = X_{N(t)}$ where $N(t)$ has Poisson(t) distribution, so by conditioning on $N(t) = \nu$,

$$\begin{aligned} P_i(Y_t = j) &= \pi_i^{-1/2} \pi_j^{1/2} \sum_{m=1}^{n} u_{im} u_{jm} \sum_{\nu=0}^{\infty} \lambda_m^\nu \frac{e^{-t} t^\nu}{\nu!} \\ &= \pi_i^{-1/2} \pi_j^{1/2} \sum_{m=1}^{n} u_{im} u_{jm} \exp(-(1 - \lambda_m)t). \end{aligned}$$

So when we compare the spectral representations (3.32),(3.33) for a discrete-time chain and its continuization, the orthonormal matrices are identical, and the eigenvalues are related by

$$\lambda_m^{(c)} = 1 - \lambda_m^{(d)} \tag{3.35}$$

superscripts (c) and (d) indicating continuous or discrete time. In particular, this relation holds for the basic discrete and continuous time random walks on a graph.

Let us point out some interesting simple consequences of the spectral representation. For these purposes continuous time is simpler. First,

$$P_i(X_t = j) - \pi_j = c_{ij} e^{-\lambda_2 t} + o(e^{-\lambda_2 t}) \text{ as } t \to \infty \tag{3.36}$$

where $c_{ij} = \pi_i^{-1/2} \pi_j^{1/2} \sum_{m:\lambda_m = \lambda_2} u_{im} u_{jm}$ and where "typically" $c_{ij} \neq 0$. (A precise statement is this: there exists i such that

$$P_i(X_t = i) - \pi_i \sim c_{ii} e^{-\lambda_2 t}, \quad c_{ii} > 0, \tag{3.37}$$

by considering i such that $u_{i2} \neq 0$.) Thus λ_2 has the interpretation of "asymptotic rate of convergence to the stationary distribution". The authors find it simpler to interpret parameters measuring "time" rather than "1/time", and so prefer to work with the *relaxation time* τ_2 defined by

$$\tau_2 := 1/\lambda_2 \text{ for a continuous-time chain} \tag{3.38}$$

$$\tau_2 := 1/(1 - \lambda_2) \text{ for a discrete-time chain.} \tag{3.39}$$

Note that by (3.35) the value of τ_2 is unchanged by continuizing a discrete-time chain.

Still in continuous time, the spectral representation gives

$$P_i(X_t = i) = \pi_i + \sum_{m \geq 2} u_{im}^2 \exp(-\lambda_m t) \tag{3.40}$$

so the right side is decreasing with t, and in fact is *completely monotone*, a
9/10/99 version subject pursued in Section 3.5. Thus Z_{ii} defined in Chapter 2 Section 2.3 satisfies

$$\begin{aligned} Z_{ii} &= \int_0^\infty (P_i(X_t = i) - \pi_i)\, dt \\ &= \sum_{m \geq 2} u_{im}^2 \lambda_m^{-1} \text{ by (3.40).} \end{aligned} \tag{3.41}$$

Using the orthonormal property of $\mathbf{U}$,

$$\sum_i Z_{ii} = \sum_{m \geq 2} \lambda_m^{-1}.$$

99 version

Applying Corollary 13 of Chapter 2, we obtain a fundamental result relating average hitting times to eigenvalues.

Proposition 3.13 (The eigentime identity) *For each i,*

$$\begin{aligned} \sum_j \pi_j E_i T_j &= \sum_{m \geq 2} \lambda_m^{-1} \qquad (\textit{continuous time}) \\ \sum_j \pi_j E_i T_j &= \sum_{m \geq 2} (1 - \lambda_m)^{-1} \quad (\textit{discrete time}). \end{aligned}$$

[The discrete-time version follows from (3.35).] Proposition 3.13 expands upon the *random target lemma*, which said that (even for non-reversible chains) $\sum_j \pi_j E_i T_j$ does not depend on i.

3.4.1 Mean hitting times and reversible chains

In Chapter 2 Section 2.2 we listed identities for general chains such as the mean hitting time formulas 9/10/99 version

$$E_i T_j = (Z_{jj} - Z_{ij})/\pi_j; \quad E_\pi T_j = Z_{jj}/\pi_j.$$

There are a number of more complicated identities for general chains in which one side becomes zero for any reversible chain (by the symmetry property $\pi_i Z_{ij} = \pi_j Z_{ji}$) and which therefore simplify to give identities for reversible chains. We have already seen one example, the cyclic tour lemma, and the following result may be considered an extension of that lemma. [Indeed, sum the following equation over successive pairs (i, j) along a cycle to recapture the cyclic tour lemma.]

Corollary 3.14 $E_\pi T_j - E_\pi T_i = E_i T_j - E_j T_i$.

This identity follows immediately from the mean hitting time formulas and the symmetry property. Note the following interpretation of the corollary. Define an ordering $i \preceq j$ on the states by

$$i \preceq j \text{ iff } E_\pi T_i \leq E_\pi T_j.$$

Then Corollary 3.14 implies

$$E_i T_j \geq E_j T_i \text{ iff } i \preceq j.$$

Warning. Corollary 3.14 does not imply

$$\max_{i,j} E_i T_j \text{ is attained by some pair } (i_*, j_*) \text{ such that}$$
$$i_* \text{ attains } \min_i E_\pi T_i \text{ and } j_* \text{ attains } \max_j E_\pi T_j.$$

I haven't tried to find counterexamples with more than three states.

Here is a counterexample. Choose $0 < \varepsilon < 1/2$ arbitrarily and let

$$\mathbf{P} := \begin{bmatrix} 2\varepsilon & 1-2\varepsilon & 0 \\ \varepsilon & 1-2\varepsilon & \varepsilon \\ 0 & 1-2\varepsilon & 2\varepsilon \end{bmatrix}.$$

We invite the reader to perform the computations necessary to verify that $\mathbf{P}$ is reversible with $\pi = [\varepsilon, 1-2\varepsilon, \varepsilon]$ and

$$(E_i T_j) = \varepsilon^{-1}(1-2\varepsilon)^{-1} \begin{bmatrix} 0 & \varepsilon & 1 \\ 1-\varepsilon & 0 & 1-\varepsilon \\ 1 & \varepsilon & 0 \end{bmatrix},$$

so that $(E_\pi T_i) = \varepsilon^{-1}(1-2\varepsilon)^{-1}[1-2\varepsilon+2\varepsilon^2, 2\varepsilon^2, 1-2\varepsilon+2\varepsilon^2]$. Thus $E_\pi T_i$ is minimized uniquely by $i^* = 2$, while $\max_{i,j} E_i T_j$ is attained only by the pairs $(1,3)$ and $(3,1)$.

As a second instance of what reversibility implies, note, from (3.33) and the definition of Z_{ij}, that

$$Z_{ij} = \pi_i^{-1/2} \pi_j^{1/2} \sum_{m \geq 2} \lambda_m^{-1} u_{im} u_{jm}.$$

This implies

$$\text{the symmetrized matrix } \pi_i^{1/2} Z_{ij} \pi_j^{-1/2} \text{ is positive semidefinite.} \tag{3.42}$$

Note that a symmetric positive semidefinite matrix (M_{ij}) has the property $M_{ij}^2 \leq M_{ii} M_{jj}$. This gives

$$Z_{ij}^2 \leq Z_{ii} Z_{jj} \pi_j / \pi_i, \tag{3.43}$$

which enables us to upper-bound mean hitting times from arbitrary starts in terms of mean hitting times from stationary starts.

Lemma 3.15 $\max_{ij} E_i T_j \leq 2 \max_k E_\pi T_k$.

Proof. Using (3.43),

$$(Z_{ij}/\pi_j)^2 \le (Z_{ii}/\pi_i)\ (Z_{jj}/\pi_j)$$

and so

$$-Z_{ij}/\pi_j \le \max_k Z_{kk}/\pi_k.$$

So the mean hitting time formula gives the two equalities in

$$E_iT_j = \frac{Z_{jj}}{\pi_j} - \frac{Z_{ij}}{\pi_j} \le 2\max_k \frac{Z_{kk}}{\pi_k} = 2\max_k E_\pi T_k. \qquad \blacksquare$$

3.5 Complete monotonicity

One advantage of working in continuous time is to exploit complete monotonicity properties. Abstractly, call $f : [0,\infty) \to [0,\infty)$ *completely monotone* (CM) if there is a nonnegative measure μ on $[0,\infty)$ such that

$$f(t) = \int_{[0,\infty)} e^{-\theta t}\mu(d\theta), \quad 0 \le t < \infty. \tag{3.44}$$

Our applications will use only the special case of a finite sum

$$f(t) = \sum_m a_m e^{-\theta_m t}, \quad \text{for some } a_m > 0,\ \theta_m \ge 0, \tag{3.45}$$

but finiteness plays no essential role. If f is CM then (provided they exist) so are

$$-f'(t),$$
$$\bar{F}(t) := \int_t^\infty f(s)\,ds \tag{3.46}$$

A *probability distribution* ν on $[0,\infty)$ is called CM if its tail distribution function $\bar{F}(t) :\equiv \nu(t,\infty)$ is CM; equivalently, if its density function f is CM (except that here we must in the general case allow the possibility $f(0) = \infty$). In more probabilistic language, ν is CM iff it can be expressed as the distribution of ξ/Λ, where ξ and Λ are independent random variables such that

$$\xi \text{ has Exponential(1) distribution;}\ \ \Lambda > 0. \tag{3.47}$$

Given a CM function or distribution, the *spectral gap* $\lambda \ge 0$ can be defined consistently by

$$\begin{array}{lcll} \lambda & := & \inf\{t > 0 : \mu[0,t] > 0\} & \text{in setting (3.44)} \\ \lambda & := & \min\{\theta_m\} & \text{in setting (3.45)} \\ \lambda & := & \operatorname{ess\,inf} \Lambda & \text{in setting (3.47).} \end{array}$$

This λ controls the behavior of $f(t)$ as $t \to \infty$. A key property of CM functions is that their value at a general time t can be bounded in terms of their behavior at 0 and at ∞, as follows.

Lemma 3.16 *Let f be CM with $0 < f(0) < \infty$. Then*

$$\exp\left(\frac{f'(0)\,t}{f(0)}\right) \le \frac{f(t)}{f(0)} \le \frac{\bar{F}(t)}{\bar{F}(0)} \le \exp(-\lambda t),\ 0 \le t < \infty$$

where λ is the spectral gap.

We might have $\bar{F}(0) = \infty$, but then $\bar{F}(t) = \infty$ and $\lambda = 0$ so the convention $\infty/\infty = 1$ works.

Proof. By scaling we may suppose $f(0) = 1$. So we can rewrite (3.44) as

$$f(t) = Ee^{-\Theta t} \tag{3.48}$$

where Θ has distribution μ. Then $f'(t) = -E(\Theta e^{-\Theta t})$. Because $\theta \mapsto e^{-\theta t}$ is decreasing, the random variables Θ and $e^{-\Theta t}$ are negatively correlated (this fact is sometimes called "Chebyshev's *other* inequality", and makes a nice exercise [HINT: Symmetrize!]) and so $E(\Theta e^{-\Theta t}) \le (E\Theta)(Ee^{-\Theta t})$. This says $-f'(t) \le -f'(0)\,f(t)$, or in other words $\frac{d}{dt}\log f(t) \ge f'(0)$. Integrating gives $\log f(t) \ge t f'(0)$, which is the leftmost inequality. (Recall we scaled to make $f(0) = 1$.) For the second inequality,

$$\begin{aligned}
\bar{F}(t) &= E(\Theta^{-1}e^{-\Theta t}) \text{ by integrating (3.48)} \\
&\ge (E\Theta^{-1})(Ee^{-\Theta t}) \text{ by positive correlation} \\
&= \bar{F}(0)\,f(t).
\end{aligned}$$

Finally, from the definition of the spectral gap λ it is clear that $f(t)/f(0) \le e^{-\lambda t}$. But $\bar{F}$ has the same spectral gap as f. ∎

Returning to the study of continuous-time reversible chains, the spectral representation (3.40) says that $P_i(X_t = i)$ is a CM function. It is often convenient to subtract the limit and say

$$P_i(X_t = i) - \pi_i \text{ is a CM function.} \tag{3.49}$$

More generally, given any function $g : I \to R$ the function

$$\rho(t) :\equiv E[g(X_t)g(X_0)] \tag{3.50}$$

is CM for the *stationary* chain, because by (3.33)

$$\begin{aligned}\rho(t) &= \sum_{m=1}^{n}\left(\sum_i \pi_i^{1/2} g(i) u_{im}\right)\left(\sum_j \pi_j^{1/2} g(j) u_{jm}\right)\exp(-\lambda_m t) \\ &= \sum_{m=1}^{n}\left(\sum_i \pi_i^{1/2} g(i) u_{im}\right)^2 \exp(-\lambda_m t). \qquad (3.51)\end{aligned}$$

Specializing to the case $g = 1_A$ and conditioning,

$$P(X_t \in A \,|\, X_0 \in A) \text{ is a CM function} \qquad (3.52)$$

again assuming the stationary chain. When A is a singleton, this is (3.49).

Remark. To study directly discrete-time reversible chains, one would replace CM functions by sequences (f_n) of the form

$$f_n = \int_{-1}^{1} \theta^n \ \mu(d\theta).$$

But analogs of Lemma 3.16 and subsequent results (e.g., Proposition 3.22) become messier—so we prefer to derive discrete-time results by continuization.

3.5.1 Lower bounds on mean hitting times

As a quick application, we give bounds on mean hitting times to a single state from a stationary start. Recall $q_i = \sum_{j \neq i} q_{ij}$ is the exit rate from i, and τ_2 is the relaxation time of the chain.

Lemma 3.17 *For any state i in a continuous-time chain,*

$$\frac{(1-\pi_i)^2}{q_i \pi_i} \leq E_\pi T_i \leq \frac{\tau_2(1-\pi_i)}{\pi_i}.$$

By continuization, the Lemma holds in discrete time, replacing q_i by $1 - p_{ii}$.

Proof. The mean hitting time formula is

$$\pi_i E_\pi T_i = Z_{ii} = \int_0^\infty (P_i(X_t = i) - \pi_i)\, dt.$$

Write $f(t)$ for the integrand. We know f is CM, and here $\lambda \geq \lambda_2$ by (3.40), and $f'(0) = -q_i$, so the extreme bounds of Lemma 3.16 become, after multiplying by $f(0) = 1 - \pi_i$,

$$(1-\pi_i)\exp(-q_i t/(1-\pi_i)) \leq f(t) \leq (1-\pi_i)e^{-\lambda_2 t}.$$

Integrating these bounds gives the result. ■

We can now give general lower bounds on some basic parameters we will study in Chapter 4.

Proposition 3.18 *For a discrete-time chain on n states,*

$$\sum_j \pi_j E_\pi T_j \geq \frac{(n-1)^2}{n} \tag{3.53}$$

$$\max_{i,j}(E_i T_j + E_j T_i) \geq 2(n-1) \tag{3.54}$$

$$\max_{i,j} E_i T_j \geq n-1 \tag{3.55}$$

$$\tau_2 \geq \frac{n-1}{n}. \tag{3.56}$$

Remark. These inequalities become equalities for random walk on the complete graph (Chapter 5 Example 9). By examining the proof, it can be shown that this is the only chain where an equality holds.

4/22/96 version

Proof. We go to the continuized chain, which has $q_i = 1 - p_{ii} \leq 1$. Then

$$\begin{aligned}
\sum_j \pi_j E_\pi T_j &\geq \sum_j (1-\pi_j)^2 \text{ by Lemma 3.17} \\
&= n - 2 + \sum_j \pi_j^2 \\
&\geq n - 2 + \tfrac{1}{n} \\
&= (n-1)^2/n,
\end{aligned}$$

giving (3.53). By the eigentime identity,

$$\sum_j \pi_j E_\pi T_j = \sum_{m \geq 2} \lambda_m^{-1} \leq (n-1)\tau_2$$

and so (3.56) follows from (3.53).

Now fix i and write $\tau_0 = \sum_j \pi_j E_k T_j$, which (by the random target lemma) doesn't depend on k. Then

$$\sum_{j \neq i} \frac{\pi_j}{1-\pi_i}(E_i T_j + E_j T_i) = \frac{\tau_0 + E_\pi T_i}{1-\pi_i}. \tag{3.57}$$

If the right side were strictly less than $2(n-1)$ for all i, then

$$\sum_i \pi_i(\tau_0 + E_\pi T_i) < 2(n-1)\sum_i \pi_i(1-\pi_i),$$

which implies

$$2\tau_0 < 2(n-1)\left(1 - \sum_i \pi_i^2\right) \le 2(n-1)\left(1 - \frac{1}{n}\right) = \frac{2(n-1)^2}{n},$$

contradicting (3.53). Therefore there exists an i such that

$$\sum_{j \ne i} \frac{\pi_j}{1-\pi_i}(E_i T_j + E_j T_i) \ge 2(n-1)$$

and so there exists $j \ne i$ such that $E_i T_j + E_j T_i \ge 2(n-1)$. This is (3.54), and (3.55) follows immediately. ■

There are several other results in the spirit of Lemma 3.17 and Proposition 3.18. For instance, (22) in Chapter 2 says that for a general discrete-time chain, 9/10/99 version

$$\mathrm{var}_i\, T_i^+ = \frac{2E_\pi T_i + 1}{\pi_i} - \frac{1}{\pi_i^2}.$$

Appealing to Lemma 3.17 gives, after a little algebra,

Corollary 3.19 *For any state i in a discrete-time chain,*

$$\mathrm{var}_i T_i^+ \ge \frac{(1-\pi_i)(1-2\pi_i)}{\pi_i^2}.$$

Again, equality holds for random walk on the complete graph.

3.5.2 Smoothness of convergence

We're going to build some vague discussion around the following simple result.

Lemma 3.20

$$\sum_j \frac{p_{ij}^2(t)}{\pi_j} = \frac{p_{ii}(2t)}{\pi_i} \tag{3.58}$$

$$\left|\frac{p_{ik}(t+s)}{\pi_k} - 1\right| \le \sqrt{\left(\frac{p_{ii}(2t)}{\pi_i} - 1\right)\left(\frac{p_{kk}(2s)}{\pi_k} - 1\right)} \tag{3.59}$$

$$\frac{p_{ik}(t+s)}{\pi_k} \le \sqrt{\frac{p_{ii}(2t)}{\pi_i}\frac{p_{kk}(2s)}{\pi_k}} \quad \textit{and so } \max_{i,k} \tfrac{p_{ik}(2t)}{\pi_k} \le \max_i \tfrac{p_{ii}(2t)}{\pi_i}. \tag{3.60}$$

Proof.

$$\frac{p_{ik}(t+s)}{\pi_k} = \sum_j p_{ij}(t)\frac{p_{jk}(s)}{\pi_k} = \sum_j p_{ij}(t)\frac{p_{kj}(s)}{\pi_j}$$

by reversibility. Putting $k = i$, $s = t$ gives (3.58). Rewriting the above equality as

$$\frac{p_{ik}(t+s)}{\pi_k} - 1 = \sum_j \pi_j \frac{p_{ij}(t) - \pi_j}{\pi_j}\frac{p_{kj}(s) - \pi_j}{\pi_j}$$

and applying the Cauchy–Schwarz inequality, we get the bound $\sqrt{a_i(t)a_k(s)}$, where

$$a_i(t) = \sum_j \frac{(p_{ij}(t) - \pi_j)^2}{\pi_j} = \sum_j \frac{p_{ij}^2(t)}{\pi_j} - 1 = \frac{p_{ii}(2t)}{\pi_i} - 1.$$

This proves (3.59). The cruder bound (3.60) is sometimes easier to use than (3.59) and is proved similarly. ■

9/10/99 version

Discussion. Recalling from Chapter 2 Section 4.2 the definition of L^2 distance between distributions, (3.58) says

$$\|P_i(X_t \in \cdot) - \pi\|_2^2 = \frac{p_{ii}(2t)}{\pi_i} - 1. \tag{3.61}$$

In continuous time, we may regard the assertion "$\|P_i(X_t \in \cdot) - \pi\|_2$ is decreasing in t" as a consequence of the equality in (3.61) and the CM property of $p_{ii}(t)$. This assertion in fact holds for general chains, as pointed out in

9/10/99 version

Chapter 2 Lemma 35. Loosely, the general result of Chapter 2 Lemma 35

9/10/99 version

says that in a general chain the ratios $(P_\rho(X_t = j)/\pi_j,\ j \in I)$ considered as an *unordered* set tend to smooth out as t increases. For a reversible chain, much more seems to be true. There is some "intrinsic geometry" on the state space such that, for the chain started at i, the probability distribution as time increases from 0 "spreads out smoothly" with respect to the geometry. It's hard to formalize that idea convincingly. On the other hand, (3.61) does say convincingly that the rate of convergence of the single probability $p_{ii}(t)$ to $\pi(i)$ is connected to a rate of convergence of the entire distribution $P_i(X_t \in \cdot)$ to $\pi(\cdot)$. This intimate connection between the local and the global behavior of reversible chains underlies many of the technical

10/11/94 version

inequalities concerning mixing times in Chapter 4 and subsequent chapters.

3.5.3 Inequalities for hitting time distributions on subsets

99 version

We mentioned in Chapter 2 Section 2.2 that most of the simple identities there for mean hitting times E_iT_j on singletons have no simple analogs for hitting times T_A on subsets. One exception is Kac's formula (Chapter 2 Corollary 24), which says that for a general discrete-time chain 9/10/99 version

$$E_{\pi_A} T_A^+ = 1/\pi(A). \tag{3.62}$$

It turns out that for reversible chains there are useful inequalities relating the distributions of T_A under different initial distributions. These are simplest in continuous time as consequences of CM: as always, interesting consequences may be applied to discrete-time chains via continuization.

Recall π_A is the stationary distribution conditioned to A:

$$\pi_A(i) \equiv \pi(i)/\pi(A),\ i \in A.$$

Trivially

$$P_\pi(T_A > t) = \pi(A^c) P_{\pi_{A^c}}(T_A > t) \tag{3.63}$$

$$E_\pi T_A = \pi(A^c) E_{\pi_{A^c}} T_A. \tag{3.64}$$

Define the *ergodic exit* distribution ρ_A from A by

$$\rho_A(j) := \frac{\sum_{i \in A} \pi_i q_{ij}}{Q(A, A^c)},\quad j \in A^c, \tag{3.65}$$

where $Q(A, A^c)$ is the *ergodic flow rate* out of A:

$$Q(A, A^c) := \sum_{i \in A} \sum_{k \in A^c} \pi_i q_{ik}. \tag{3.66}$$

By stationarity, $Q(A, A^c) = Q(A^c, A)$.

Proposition 3.21 *Fix a subset A in a continuous-time chain.*

(i) *T_A has CM distribution when the initial distribution of the chain is any of the three distributions π or π_{A^c} or ρ_A.*

(ii) *The three hitting time distributions determine each other via* (3.63) *and*

$$P_{\pi_{A^c}}(T_A \in (t, t + dt)) = \frac{P_{\rho_A}(T_A > t)}{E_{\rho_A} T_A}\, dt. \tag{3.67}$$

(iii) *Write λ_A for the spectral gap associated with T_A (which is the same for each of the three initial distributions). Then*

$$P_{\rho_A}(T_A > t) \le P_{\pi_{A^c}}(T_A > t) = \frac{P_\pi(T_A > t)}{\pi(A^c)} \le \exp(-\lambda_A t),\quad t > 0 \tag{3.68}$$

and in particular

$$\frac{\pi(A^c)}{Q(A,A^c)} = E_{\rho_A} T_A \le E_{\pi_{A^c}} T_A = \frac{E_\pi T_A}{\pi(A^c)} \le 1/\lambda_A. \tag{3.69}$$

(iv)

$$E_\pi T_A \le \frac{\tau_2 \pi(A^c)}{\pi(A)}. \tag{3.70}$$

Concerning (b): The results cited from later require that the chain restricted to A^c be irreducible, but I think that requirement can be dropped using a limiting argument.

Remarks. (a) In discrete time we can define ρ_A and $Q(A, A^c)$ by replacing q_{ij} by p_{ij} in (3.65)–(3.66), and then (3.69) holds in discrete time. The left equality of (3.69) is then a reformulation of Kac's formula (3.62), because

$$\begin{aligned} E_{\pi_A} T_A^+ &= 1 + P_{\pi_A}(X_1 \in A^c) E_{\pi_A}(T_A^+ - 1 | X_1 \in A^c) \\ &= 1 + \frac{Q(A,A^c)}{\pi(A)} E_{\rho_A} T_A. \end{aligned}$$

(b) Equation (3.83) and Corollary 3.34 [together with remark (b) following Theorem 3.33] later show that $1/\lambda_A \le \tau_2/\pi(A)$. So (3.70) can be regarded as a consequence of (3.69). Reverse inequalities will be studied in Chapter 4.

10/11/94 version

Proof of Proposition 3.21. First consider the case where A is a singleton $\{a\}$. Then (3.70) is an immediate consequence of Lemma 3.17. The equalities in (3.69) and in (3.67) are general identities for stationary processes [(24) and (23) in Chapter 2]. We shall prove below that T_A is CM under $P_{\pi_{I\setminus\{a\}}}$. Then by (3.63), (3.67), and (3.46), T_A is also CM under the other two initial distributions. Then the second inequality of (3.68) is the upper bound in Lemma 3.16, and the first is a consequence of (3.67) and Lemma 3.16. And (3.69) follows from (3.68) by integrating over t.

9/10/99 version

To prove that T_A is CM under $P_{\pi_{I\setminus\{a\}}}$, introduce a parameter $0 < \varepsilon < 1$ and consider the modified chain (X_t^ε) with transition rates

$$\begin{aligned} q_{ij}^\varepsilon &:= q_{ij}, \ i \ne a \\ q_{aj}^\varepsilon &:= \varepsilon q_{aj}. \end{aligned}$$

The modified chain remains reversible, and its stationary distribution is of the form

$$\pi_i^\varepsilon = b_1 \pi_i, \quad i \ne a; \qquad \pi_a^\varepsilon = b_2$$

where the weights b_1, b_2 depend only on ε and π_a. Now as $\varepsilon \to 0$ with t fixed,

$$P_{\pi_{I\setminus\{a\}}}(X_t^\varepsilon \in I \setminus \{a\}) \to P_{\pi_{I\setminus\{a\}}}(T_a > t) \tag{3.71}$$

because the chain gets "stuck" upon hitting a. But the left side is CM by (3.52), so the right side (which does not depend on ε) is CM, because the class of CM distributions is closed under pointwise limits. (The last assertion is in general the continuity theorem for Laplace transforms [133] p. 83, though for our purposes we need only the simpler fact that the set of functions of the form (3.45) with at most n summands is closed.)

This completes the proof when A is a singleton. We now claim that the case of general A follows from the collapsing principle (Chapter 2 Section 7.3), 9/10/99 version i.e., by applying the special case to the chain in which the subset A is collapsed into a single state. This is clear for all the assertions of Proposition 3.21 except for (3.70), for which we need the fact that the relaxation time τ_2^A of the collapsed chain is at most τ_2. This fact is proved as Corollary 3.27 below. ■

Remark. Note that the CM property implies a *super*multiplicitivity property for hitting times from stationarity in a continuous-time reversible chain:

$$P_\pi(T_A > s + t) \geq P_\pi(T_A > s)P_\pi(T_A > t).$$

Contrast with the general *sub*multiplicitivity property (Chapter 2 Section 4.3) 9/10/99 version which holds when P_π is replaced by $\max_i P_i$.

3.5.4 Approximate exponentiality of hitting times

In many circumstances, the distribution of the first hitting time T_A on a subset A of states with $\pi(A)$ small (equivalently, with ET_A large) can be approximated by the exponential distribution with the same mean. As with the issue of convergence to the stationary distribution, such approximations can be proved for general chains (see Notes), but it is easier to get explicit bounds in the reversible setting. If T has a CM distribution, then [as at (3.47), but replacing $1/\Lambda$ by Θ] we may suppose $T \stackrel{d}{=} \Theta\xi$. We calculate

$$ET = (E\Theta)(E\xi) = E\Theta; \quad ET^2 = (E\Theta^2)(E\xi^2) = 2E\Theta^2$$

and so

$$\frac{ET^2}{2(ET)^2} = \frac{E\Theta^2}{(E\Theta)^2} \geq 1$$

with equality iff Θ is constant, i.e., iff T has exponential distribution. This suggests that the difference $\frac{ET^2}{2(ET)^2} - 1$ can be used as a measure of "deviation from exponentiality". Let us quote a result of Mark Brown ([72] Theorem 4.1(iii)) which quantifies this idea in a very simple way.

Proposition 3.22 *Let T have CM distribution. Then*

$$\sup_t |P(T > t) - e^{-t/ET}| \le \frac{ET^2}{2(ET)^2} - 1.$$

So we can use this bound for hitting times T_A in a stationary reversible chain. At first sight the bound seems useful only if we can estimate $E_\pi T_A^2$ and $E_\pi T_A$ accurately. But the following remarkable variation shows that for the hitting time distribution to be approximately exponential it is sufficient that the mean hitting time be large compared to the relaxation time τ_2.

Proposition 3.23 *For a subset A of a continuous-time chain,*

$$\sup_t |P_\pi(T_A > t) - \exp(-t/E_\pi T_A)| \le \tau_2 / E_\pi T_A.$$

9/10/99 version

Proof. By the collapsing principle (Chapter 2 Section 7.3) we may suppose A is a singleton $\{j\}$, because (Corollary 3.27 below) collapsing cannot increase the relaxation time. Combining the mean hitting time formula with the expression (3.41) for Z_{jj} in terms of the spectral representation (3.33),

$$E_\pi T_j = \pi_j^{-1} Z_{jj} = \pi_j^{-1} \sum_{m \ge 2} u_{jm}^2 \lambda_m^{-1}. \tag{3.72}$$

A similar calculation, exhibited below, shows

$$\frac{E_\pi T_j^2 - 2(E_\pi T_j)^2}{2} = \pi_j^{-1} \sum_{m \ge 2} u_{jm}^2 \lambda_m^{-2}. \tag{3.73}$$

But $\lambda_m^{-2} \le \lambda_2^{-1} \lambda_m^{-1} = \tau_2 \lambda_m^{-1}$ for $m \ge 2$, so the right side of (3.73) is bounded by $\pi_j^{-1} \tau_2 \sum_{m \ge 2} u_{jm}^2 \lambda_m^{-1}$, which by (3.72) equals $\tau_2 E_\pi T_j$. Applying Proposition 3.22 gives Proposition 3.23.

We give a straightforward but tedious verification of (3.73) (see also Notes). The identity $x^2/2 = \int_0^\infty (x-t)^+ \, dt, \ x \ge 0$ starts the calculation

Chapter 2 reference in following display is to 9/10/99 version.

$$\begin{aligned}
\frac{1}{2} E_\pi T_j^2 &= \int_0^\infty E_\pi (T_j - t)^+ \, dt \\
&= \int_0^\infty \sum_i P_\pi(X_t = i, T_j > t) E_i T_j \, dt \\
&= \sum_i E_i T_j \, E_\pi(\text{time spent at } i \text{ before } T_j) \\
&= \sum_i \frac{Z_{jj} - Z_{ij}}{\pi_j} \, \frac{Z_{jj}\pi_i - Z_{ji}\pi_j}{\pi_j} \\
&\qquad \text{by Chapter 2 Lemmas 12 and 15 (continuous-time version)} \\
&= \sum_i \pi_j^{-2} \pi_i (Z_{jj} - Z_{ij})^2.
\end{aligned}$$

Expanding the square, the cross-term vanishes and the first term becomes $(Z_{jj}/\pi_j)^2 = (E_\pi T_j)^2$, so

$$\tfrac{1}{2} E_\pi T_j^2 - (E_\pi T_j)^2 = \pi_j^{-2} \sum_i \pi_i Z_{ij}^2.$$

To finish the calculation,

$$\begin{aligned}
&\pi_j^{-1} \sum_i \pi_i Z_{ij}^2 \\
&\quad = \pi_j^{-1} \sum_i \pi_i \left(\int (p_{ij}(s) - \pi_j)\, ds \right) \left(\int (p_{ij}(t) - \pi_j)\, dt \right) \\
&\quad = \sum_i \left(\int (p_{ji}(s) - \pi_i)\, ds \right) \left(\int (p_{ij}(t) - \pi_j)\, dt \right) \\
&\quad = \int\int (p_{jj}(s+t) - \pi_j)\; ds\, dt \\
&\quad = \int t (p_{jj}(t) - \pi_j)\, dt \\
&\quad = \int \sum_{m \geq 2} u_{jm}^2 t e^{-\lambda_m t}\, dt \\
&\quad = \sum_{m \geq 2} u_{jm}^2 \lambda_m^{-2}. \qquad \blacksquare
\end{aligned}$$

See the Notes for a related result, Theorem 3.43.

3.6 Extremal characterizations of eigenvalues

3.6.1 The Dirichlet formalism

A reversible chain has an associated *Dirichlet form* $\mathcal{E}$, defined as follows. For functions $g : I \to R$ write

$$\mathcal{E}(g,g) := \tfrac{1}{2} \sum_i \sum_{j \neq i} \pi_i p_{ij} (g(j) - g(i))^2 \tag{3.74}$$

in discrete time, and substitute q_{ij} for p_{ij} in continuous time. One can immediately check the following equivalent definitions. In discrete time

$$\mathcal{E}(g,g) = \tfrac{1}{2} E_\pi (g(X_1) - g(X_0))^2 = E_\pi [g(X_0)(g(X_0) - g(X_1))]. \tag{3.75}$$

In continuous time

$$\begin{aligned}\mathcal{E}(g,g) &= \tfrac{1}{2}\lim_{t\to 0} t^{-1}E_\pi(g(X_t)-g(X_0))^2 \\ &= \lim_{t\to 0} t^{-1}E_\pi[g(X_0)(g(X_0)-g(X_t))] \\ &= -\sum_i\sum_j \pi_i g(i) q_{ij} g(j) \qquad (3.76)\end{aligned}$$

where the sum includes $j = i$. Note also that for random walk on a weighted graph, (3.74) becomes

$$\mathcal{E}(g,g) := \frac{1}{2}\sum_i\sum_{j\neq i}\frac{w_{ij}}{w}(g(j)-g(i))^2. \qquad (3.77)$$

9/10/99 version

Recall from Chapter 2 Section 6.2 the discussion of L^2 norms for functions and measures. In particular

$$\begin{aligned}||g||_2^2 &= \sum_i \pi_i g^2(i) = E_\pi g^2(X_0) \\ \|\mu-\pi\|_2^2 &= \sum_i \frac{\mu_i^2}{\pi_i} - 1 \text{ for a probability distribution } \mu.\end{aligned}$$

The relevance of $\mathcal{E}$ can be seen in the following lemma.

Lemma 3.24 *Write $\rho(t) = (\rho_j(t))$ for the distribution at time t of a continuous-time chain, with arbitrary initial distribution. Write $f_j(t) = \rho_j(t)/\pi_j$. Then*

$$\frac{d}{dt}\|\rho(t)-\pi\|_2^2 = -2\mathcal{E}(f(t),f(t)).$$

Proof. $\|\rho(t)-\pi\|_2^2 = \sum_j \pi_j^{-1}\rho_j^2(t) - 1$, so using the forward equations

$$\frac{d}{dt}\rho_j(t) = \sum_i \rho_i(t) q_{ij}$$

we get

$$\begin{aligned}\frac{d}{dt}\|\rho(t)-\pi\|_2^2 &= \sum_j\sum_i 2\pi_j^{-1}\rho_j(t)\rho_i(t)q_{ij} \\ &= 2\sum_j\sum_i f_j(t) f_i(t)\pi_i q_{ij}\end{aligned}$$

and the result follows from (3.76). ∎

3.6.2 Summary of extremal characterizations

For ease of comparison we state below three results which will be proved in subsequent sections. These results are commonly presented "the other way up" using *infs* rather than *sups*, but our presentation is forced by our convention of consistently defining parameters to have dimensions of "time" rather than "1/time". The *sups* are over functions $g : I \to R$ satisfying specified constraints, and excluding $g \equiv 0$. The results below are the same in continuous and discrete time—that is, continuization doesn't change the numerical values of the quantities we consider. We shall give the proofs in discrete time.

Extremal characterization of relaxation time. *The relaxation time τ_2 satisfies*

$$\tau_2 = \sup\{||g||_2^2/\mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0\}.$$

Extremal characterization of quasistationary mean hitting time. *Given a subset A, let α_A be the quasistationary distribution on A^c defined at* (3.82)*. Then the quasistationary mean exit time is*

$$E_{\alpha_A} T_A = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \geq 0,\ g = 0 \text{ on } A\}.$$

Extremal characterization of mean commute times. *For distinct states i, j the mean commute time satisfies*

$$E_i T_j + E_j T_i = \sup\{1/\mathcal{E}(g,g) : 0 \leq g \leq 1,\ g(i) = 1,\ g(j) = 0\}.$$

Because the state space is finite, the *sups* are attained, and there are theoretical descriptions of the g attaining the extrema in all three cases. An immediate practical use of these characterizations in concrete examples is to obtain lower bounds on the parameters by inspired guesswork, that is by choosing some simple explicit "test function" g which seems qualitatively right and computing the right-hand quantity. See Chapter 14 Example 32 for a typical example. Of course we cannot obtain *upper* bounds this way, but extremal characterizations can be used as a starting point for further theoretical work (see in particular the bounds on τ_2 in Chapter 4 Section 4). 3/10/94 version 10/11/94 version

3.6.3 The extremal characterization of relaxation time

The first two extremal characterizations are in fact just reformulations of the classical *Rayleigh–Ritz* extremal characterization of eigenvalues, which

goes as follows ([183] Theorem 4.2.2 and eq. 4.2.7). Let S be a symmetric matrix with eigenvalues $\mu_1 \geq \mu_2 \geq \cdots$. Then

$$\mu_1 = \sup_{\mathbf{x}} \frac{\sum_i \sum_j x_i s_{ij} x_j}{\sum_i x_i^2} \tag{3.78}$$

and an $\mathbf{x}$ attaining the *sup* is an eigenvalue corresponding to μ_1 (of course *sups* are over $\mathbf{x} \not\equiv 0$). And

$$\mu_2 = \sup_{\mathbf{y}: \sum_i y_i x_i = 0} \frac{\sum_i \sum_j y_i s_{ij} y_j}{\sum_i y_i^2} \tag{3.79}$$

and a $\mathbf{y}$ attaining the *sup* is an eigenvalue corresponding to μ_2.

As observed in Section 3.4, given a discrete-time chain with transition matrix P, the symmetric matrix ($s_{ij} = \pi_i^{1/2} p_{ij} \pi_j^{-1/2}$) has maximal eigenvalue 1 with corresponding eigenvector ($\pi_i^{1/2}$). So applying (3.79) and writing $y_i = \pi_i^{1/2} g(i)$, the second-largest eigenvalue (of S and hence of P) is given by

$$\lambda_2 = \sup_{g: \sum_i \pi_i g(i) = 0} \frac{\sum_i \sum_j \pi_i g(i) p_{ij} g(j)}{\sum_i \pi_i g^2(i)}.$$

In probabilistic notation the fraction is

$$\frac{E_\pi[g(X_0) g(X_1)]}{E_\pi g^2(X_0)} = 1 - \frac{E_\pi[g(X_0)(g(X_1) - g(X_0))]}{E_\pi g^2(X_0)} = 1 - \frac{\mathcal{E}(g,g)}{\|g\|_2^2}.$$

Since $\tau_2 = 1/(1-\lambda_2)$ in discrete time we have proved the first of our extremal characterizations.

Theorem 3.25 (Extremal characterization of relaxation time) *The relaxation time τ_2 satisfies*

$$\tau_2 = \sup\{||g||_2^2 / \mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0\}.$$

A function g_0, say, attaining the *sup* in the extremal characterization is, by examining the argument above, a right eigenvector associated with λ_2:

$$\sum_j p_{ij} g_0(j) = \lambda_2 g_0(i).$$

(From this point on in the discussion, we assume g_0 is normalized so that $\|g_0\|_2 = 1$.) The corresponding left eigenvector θ :

$$\sum_i \theta_i p_{ij} = \lambda_2 \theta_j \text{ for all } j$$

is the signed measure θ such that $\theta_i = \pi_i g_0(i)$. To continue a somewhat informal discussion of the interpretation of g_0, it is convenient to switch to continuous time (to avoid issues of negative eigenvalues) and to assume λ_2 has multiplicity 1. The equation which relates distribution at time t to initial distribution,

$$\rho_j(t) = \sum_i \rho_i(0) p_{ij}(t),$$

can also be used to define signed measures evolving from an initial signed measure. For the initial measure θ we have

$$\theta(t) = e^{-t/\tau_2}\theta.$$

For any signed measure $\nu = \nu(0)$ with $\sum_i \nu_i(0) = 0$ we have

$$\nu(t) \sim ce^{-t/\tau_2}\theta; \quad c = \sum_i \nu_i(0)\theta_i/\pi_i = \sum_i \nu_i(0) g_0(i).$$

So θ can be regarded as "the signed measure which relaxes to 0 most slowly". For a probability measure $\rho(0)$, considering $\rho(0) - \pi$ gives

$$\rho(t) - \pi \sim ce^{-t/\tau_2}\theta, \quad c = \sum_i (\rho_i(0) - \pi_i) g_0(i) = \sum_i \rho_i(0) g_0(i). \tag{3.80}$$

So θ has the interpretation of "the asymptotic normalized difference between the true distribution at time t and the stationary distribution". Finally, from (3.80) with $\rho(0)$ concentrated at i (or from the spectral representation)

$$P_i(X_t \in \cdot) - \pi \sim g_0(i) e^{-t/\tau_2}\theta.$$

So g_0 has the interpretation of "the asymptotic normalized size of deviation from stationarity, as a function of the starting state". When the state space has some geometric structure – jumps go to nearby states – one expects g_0 to be a "smooth" function, exemplified by the cosine function arising in the n-cycle (Chapter 5 Example 7). 4/22/96 version

3.6.4 Simple applications

Here is a fundamental "finite-time" result.

Good name for Lemma 3.26 – looks good to DA !

Lemma 3.26 (L^2 contraction lemma) *Write $\rho(t) = (\rho_j(t))$ for the distribution at time t of a continuous-time chain, with arbitrary initial distribution. Then*

$$\|\rho(t) - \pi\|_2 \le e^{-t/\tau_2} \|\rho(0) - \pi\|_2.$$

Proof. Write $f_j(t) = \rho_j(t)/\pi_j$. Then

$$\begin{aligned}
\frac{d}{dt}\|\rho(t) - \pi\|_2^2 &= -2\mathcal{E}(f(t), f(t)) \text{ by Lemma 3.24} \\
&= -2\mathcal{E}(f(t) - 1, f(t) - 1) \\
&\leq -2\,\frac{\|f(t) - 1\|_2^2}{\tau_2} \text{ by the extremal characterization of } \tau_2 \\
&= \frac{-2}{\tau_2}\,\|\rho(t) - \pi\|_2^2.
\end{aligned}$$

Integrating, $\|\rho(t) - \pi\|_2^2 \leq e^{-2t/\tau_2}\|\rho(0) - \pi\|_2^2$, and the result follows. ∎

Alternatively, Lemma 3.26 follows by observing that

$$\|\rho(t) - \pi\|_2^2 \text{ is CM with spectral gap at least } 2\lambda_2 = 2/\tau_2 \tag{3.81}$$

and applying Lemma 3.16. The fact (3.81) can be established directly from the spectral representation, but we will instead apply the observation at (3.50)–(3.51). Indeed, with $g(i) :\equiv \rho_i(0)/\pi_i$, we have

$$\begin{aligned}
E_\pi\left[g(X_{2t})g(X_0)\right] &= \sum_i \rho_i(0) \sum_j p_{ij}(2t)\frac{\rho_j(0)}{\pi_j} \\
&= \sum_i \rho_i(0) \sum_j \sum_k p_{ik}(t)p_{kj}(t)\frac{\rho_j(0)}{\pi_j} \\
&= \sum_k \frac{1}{\pi_k}\left[\sum_i \rho_i(0)p_{ik}(t)\right]\left[\sum_j \rho_j(0)p_{jk}(t)\right] \\
&= \sum_k \frac{1}{\pi_k}\rho_k^2(t) = \|\rho(t) - \pi\|_2^2 + 1.
\end{aligned}$$

Thus by (3.51)

$$\|\rho(t) - \pi\|_2^2 = \sum_{m=2}^{n}\left(\sum_i \pi_i^{1/2} g(i) u_{im}\right)^2 \exp(-\lambda_m t).$$

Our main use of the extremal characterization is to compare relaxation times of different chains on the same (or essentially the same) state space. Here are three instances. The first is a result we have already exploited in Section 3.5.

Corollary 3.27 *Given a chain with relaxation time τ_2, let τ_2^A be the relaxation time of the chain with subset A collapsed to a singleton $\{a\}$ (Chapter 2 Section 7.3). Then $\tau_2^A \leq \tau_2$.*

Proof. Any function g on the states of the collapsed chain can be extended to the original state space by setting $g = g(a)$ on A, and $\mathcal{E}(g,g)$ and $\sum_i \pi_i g(i)$ and $||g||_2^2$ are unchanged. So consider a g attaining the *sup* in the extremal characterization of τ_2^A and use this as a test function in the extremal characterization of τ_2. ■

Remark. An extension of Corollary 3.27 will be provided by the contraction principle (Chapter 4 Proposition 44). 10/11/94 version

Corollary 3.28 *Let τ_2 be the relaxation time for a "fluid model" continuous-time chain associated with a graph with weights (w_e) [recall (3.16)] and let τ_2^* be the relaxation time when the weights are (w_e^*). If $w_e^* \geq w_e$ for all edges e then $\tau_2^* \leq \tau_2$.*

Proof. Each stationary distribution is uniform, so $\|g\|_2^2 = \|g\|_2^{*2}$ while $\mathcal{E}^*(g,g) \geq \mathcal{E}(g,g)$. So the result is immediate from the extremal characterization. ■

The next result is a prototype for more complicated "indirect comparison" arguments later. It is convenient to state it in terms of random walk on a weighted graph. Recall (Section 3.2) that a reversible chain specifies a weighted graph with edge-weights $w_{ij} = \pi_i p_{ij}$, vertex-weights $w_i = \pi_i$, and total weight $w = 1$. refer to not-net-written Poincare chapter

Lemma 3.29 (the direct comparison lemma) *Let (w_e) and (w_e^*) be edge-weights on a graph, let (w_i) and (w_i^*) be the vertex-weights, and let τ_2 and τ_2^* be the relaxation times for the associated random walks. Then*

$$\frac{\min_e(w_e/w_e^*)}{\max_i(w_i/w_i^*)} \leq \frac{\tau_2}{\tau_2^*} \leq \frac{\max_i(w_i/w_i^*)}{\min_e(w_e/w_e^*)}$$

where in $\min_e$ we don't count loops $e = (v,v)$.

Proof. For any g, by (3.77)

$$w^* \mathcal{E}^*(g,g) \geq w \mathcal{E}(g,g) \min_e(w_e^*/w_e).$$

And since $w\|g\|_2^2 = \sum_i w_i g^2(i)$,

$$w^* \|g\|_2^{*2} \leq w\|g\|_2^2 \max_i(w_i^*/w_i).$$

So if g has π^*-mean 0 and π-mean b then

$$\frac{\|g\|_2^{*2}}{\mathcal{E}^*(g,g)} \leq \frac{\|g-b\|_2^{*2}}{\mathcal{E}^*(g-b,g-b)} \leq \frac{\|g-b\|_2^2}{\mathcal{E}(g-b,g-b)} \frac{\max_i(w_i^*/w_i)}{\min_e(w_e^*/w_e)}.$$

By considering the g attaining the extremal characterization of τ_2^*,

$$\tau_2^* \le \tau_2 \, \frac{\max_i(w_i^*/w_i)}{\min_e(w_e^*/w_e)}.$$

This is the lower bound in the lemma, and the upper bound follows by reversing the roles of w_e and w_e^*. ■

Remarks. Sometimes τ_2 is very sensitive to apparently-small changes in the chain. Consider random walk on an unweighted graph. If we add extra edges, but keeping the total number of added edges small relative to the number of original edges, then we might guess that τ_2 could not increase or decrease much. But the examples outlined below show that τ_2 may in fact change substantially in either direction.

Example 3.30 Take two complete graphs on n vertices and join with a single edge. Then $w = 2n(n-1) + 2$ and $\tau_2 \sim n^2/2$. But if we extend the single join-edge to an n-edge matching of the vertices in the original two complete graphs, then $w^* = 2n(n-1) + 2n \sim w$ but $\tau_2^* \sim n/2$.

Example 3.31 Take a complete graph on n vertices. Take $k = o(n^{1/2})$ new vertices and attach each to distinct vertices of the original complete graph. Then $w = n(n-1) + 2k$ and τ_2 is bounded. But if we now add all edges within the new k vertices, $w^* = n(n-1) + 2k + k(k-1) \sim w$ but $\tau_2^* \sim k$ provided $k \to \infty$.

As these examples suggest, comparison arguments are most effective when the stationary distributions coincide. Specializing Lemma 3.29 to this case, and rephrasing in terms of (reversible) chains, gives

Lemma 3.32 (the direct comparison lemma) *For transition matrices* $\mathbf{P}$ *and* $\mathbf{P}^*$ *with the same stationary distribution* π*, if*

$$p_{ij} \ge \delta p_{ij}^* \quad \textit{for all } j \ne i$$

then $\tau_2 \le \delta^{-1}\tau_2^*$.

Remarks. The hypothesis can be rephrased as $\mathbf{P} = \delta\mathbf{P}^* + (1-\delta)\mathbf{Q}$, where $\mathbf{Q}$ is a (maybe not irreducible) reversible transition matrix with stationary distribution π. When $\mathbf{Q} = \mathbf{I}$ we have $\tau_2 = \delta^{-1}\tau_2^*$, so an interpretation of the lemma is that "combining transitions of $\mathbf{P}^*$ with noise can't increase mixing time any more than combining transitions with holds".

3.6.5 Quasistationarity

Given a subset A of states in a discrete-time chain, let $\mathbf{P}^A$ be $\mathbf{P}$ restricted to A^c. Then $\mathbf{P}^A$ will be a substochastic matrix, i.e., the row-sums are at most 1, and some row-sum is strictly less than 1. Suppose $\mathbf{P}^A$ is irreducible. As a consequence of the Perron–Frobenius theorem (e.g., [183] Theorem 8.4.4) for the nonnegative matrix $\mathbf{P}^A$, there is a unique $0 < \lambda < 1$ (specifically, the largest eigenvalue of $\mathbf{P}^A$) such that there is a probability distribution α satisfying

$$\alpha = 0 \text{ on } A, \qquad \sum_i \alpha_i p_{ij} = \lambda \alpha_j, \ j \in A^c. \tag{3.82}$$

Writing α_A and λ_A to emphasize dependence on A, (3.82) implies that under P_{α_A} the hitting time T_A has geometric distribution

$$P_{\alpha_A}(T_A \geq m) = \lambda_A^m, \quad m \geq 0,$$

whence

$$E_{\alpha_A} T_A = \frac{1}{1 - \lambda_A}.$$

Call α_A the *quasistationary distribution* and $E_{\alpha_A} T_A$ the *quasistationary mean exit time*.

Similarly, for a continuous-time chain let $\mathbf{Q}^A$ be $\mathbf{Q}$ restricted to A^c. Assuming irreducibility of the substochastic chain with generator $\mathbf{Q}^A$, there is a unique $\lambda \equiv \lambda_A > 0$ such that there is a probability distribution $\alpha \equiv \alpha_A$ (called the *quasistationary distribution*) satisfying

$$\alpha = 0 \text{ on } A, \qquad \sum_i \alpha_i q_{ij} = -\lambda \alpha_j, \ j \in A^c.$$

This implies that under P_{α_A} the hitting time T_A has exponential distribution

$$P_{\alpha_A}(T_A > t) = \exp(-\lambda_A t), \quad t > 0,$$

whence the *quasistationary mean exit time* is

$$E_{\alpha_A} T_A = 1/\lambda_A. \tag{3.83}$$

Note that both α_A and $E_{\alpha_A} T_A$ are unaffected by continuization of a discrete-time chain.

The facts above do not depend on reversibility, but invoking now our standing assumption that chains are reversible we will show in remark (c) following Theorem 3.33 that, for continuous-time chains, λ_A here agrees with the spectral gap λ_A discussed in Proposition 3.21, and we can also now prove our second extremal characterization.

Theorem 3.33 (Extremal characterization of quasistationary mean hitting time) *The quasistationary mean exit time satisfies*

$$E_{\alpha_A} T_A = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \geq 0,\ g = 0 \text{ on } A\}. \tag{3.84}$$

Proof. As usual, we give the proof in discrete time. The matrix $(s_{ij}^A = \pi_i^{1/2} p_{ij}^A \pi_j^{-1/2})$ is symmetric with largest eigenvalue λ_A. Putting $x_i = \pi_i^{1/2} g(i)$ in the characterization (3.78) gives

$$\lambda_A = \sup_g \frac{\sum_i \sum_j \pi_i g(i) p_{ij}^A g(j)}{\sum_i \pi_i g^2(i)}.$$

Clearly the *sup* is attained by nonnegative g, and though the sums above are technically over A^c we can sum over all I by setting $g = 0$ on A. So

$$\lambda_A = \sup \left\{ \frac{\sum_i \sum_j \pi_i g(i) p_{ij}^A g(j)}{\sum_i \pi_i g^2(i)} : g \geq 0,\ g = 0 \text{ on } A \right\}.$$

As in the proof of Theorem 3.25 this rearranges to

$$\frac{1}{1-\lambda_A} = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \geq 0,\ g = 0 \text{ on } A\},$$

establishing Theorem 3.33. ∎

Remarks. (a) These remarks closely parallel the remarks at the end of Section 3.6.3. The *sup* in Theorem 3.33 is attained by the function g_0 which is the right eigenvector associated with λ_A, and by reversibility this is

$$g_0(i) = \alpha_A(i)/\pi_i. \tag{3.85}$$

It easily follows from (3.82) that

$$P_{\alpha_A}(X_t = j \,|\, T_A > t) = \alpha_A(j) \text{ for all } j \text{ and } t,$$

which explains the name *quasistationary distribution* for α_A. A related interpretation of α_A is as the distribution of the Markov chain conditioned on having been in A^c for the infinite past. More precisely, one can use Perron–Frobenius theory to prove that

$$P(X_t = j \,|\, T_A > t) \to \alpha_A(j) \text{ as } t \to \infty \tag{3.86}$$

provided P^A is aperiodic as well as irreducible.

(b) Relation (3.86) holds in continuous time as well (assuming irreducibility of the chain restricted to A^c), yielding

$$\begin{aligned}\exp(-\lambda_A t) &= P_{\alpha_A}(T > t) \\ &= \lim_{s\to\infty} P_\pi(T_A > t+s \,|\, T_A > s) = \lim_{s\to\infty} \frac{P_\pi(T_A > t+s)}{P_\pi(T_A > s)}.\end{aligned}$$

Since by Proposition 3.21 the distribution of T_A for the stationary chain is CM with spectral gap (say) σ_A, the limit here is $\exp(-\sigma_A t)$. Thus $\lambda_A = \sigma_A$, that is, our two uses of λ_A refer to the same quantity.

(c) We conclude from remark (b), (3.83), and the final inequality in (3.69) that, in either continuous or discrete time,

$$E_{\alpha_A} T_A \geq \frac{E_\pi T_A}{\pi(A^c)} \geq E_\pi T_A. \tag{3.87}$$

This is needed at the bottom of page 27 (9/22/96 version) in Chapter 5.

Our fundamental use of quasistationarity is the following.

Corollary 3.34 *For any subset A, the quasistationary mean hitting time satisfies*

$$E_{\alpha_A} T_A \leq \tau_2/\pi(A).$$

Proof. As at (3.85) set $g(i) = \alpha_A(i)/\pi_i$, so

$$E_{\alpha_A} T_A = ||g||_2^2/\mathcal{E}(g,g). \tag{3.88}$$

Now $E_\pi g(X_0) = 1$, so applying the extremal characterization of relaxation time to $g - 1$,

$$\tau_2 \geq \frac{\|g-1\|_2^2}{\mathcal{E}(g-1,g-1)} = \frac{\|g\|_2^2 - 1}{\mathcal{E}(g,g)} = (E_{\alpha_A} T_A)\left(1 - \frac{1}{\|g\|_2^2}\right), \tag{3.89}$$

the last equality using (3.88). Since α_A is a probability distribution on A^c we have

$$1 = E_\pi \left[1_{A^c}(X_0) g(X_0)\right]$$

and so by Cauchy–Schwarz

$$1^2 \leq (E_\pi 1_{A^c}(X_0)) \times ||g||_2^2 = (1 - \pi(A))||g||_2^2.$$

Rearranging,

$$1 - \frac{1}{||g||_2^2} \geq \pi(A)$$

and substituting into (3.89) gives the desired bound. ∎

Combining Corollary 3.34 with (3.68) and (3.83) gives the result below.

DA has deleted discrete-time claim in previous version. It seems true but not worth sweating over.

Lemma 3.35

$$(\textit{continuous time}) \quad P_\pi(T_A > t) \;\leq\; \exp(-t\pi(A)/\tau_2), \qquad t \geq 0$$

3.7 Extremal characterizations and mean hitting times

Theorem 3.36 (Extremal characterization of mean commute times) *For distinct states i and a, the mean commute time satisfies*

$$E_iT_a + E_aT_i = \sup\{1/\mathcal{E}(g,g) : 0 \leq g \leq 1,\ g(i) = 1,\ g(a) = 0\} \tag{3.90}$$

and the sup *is attained by $g(j) = P_j(T_i < T_a)$. In discrete time, for a subset A and a state $i \notin A$,*

$$\pi_i P_i(T_A < T_i^+) = \inf\{\mathcal{E}(g,g) : 0 \leq g \leq 1,\ g(i) = 1,\ g(\cdot) = 0 \text{ on } A\} \tag{3.91}$$

and the inf *is attained by $g(j) = P_j(T_i < T_A)$. Equation* (3.91) *remains true in continuous time, with π_i replaced by $q_i\pi_i$ on the left.*

Proof. As noted at (3.28), form (3.90) follows (in either discrete or continuous time) from form (3.91) with $A = \{a\}$. To prove (3.91), consider g satisfying the specified boundary conditions. Inspecting (3.74), the contribution to $\mathcal{E}(g,g)$ involving a fixed state j is

$$\sum_{k \neq j} \pi_j p_{jk}(g(k) - g(j))^2. \tag{3.92}$$

As a function of $g(j)$ this is minimized by

$$g(j) = \sum_k p_{jk} g(k). \tag{3.93}$$

Thus the g which minimizes $\mathcal{E}$ subject to the prescribed boundary conditions on $A \cup \{i\}$ must satisfy (3.93) for all $j \notin A \cup \{i\}$, and by Chapter 2 Lemma 27 the unique solution of these equations is $g(j) = P_j(T_i < T_A)$. Now apply to this g the general expression (3.75):

$$\mathcal{E}(g,g) = \sum_j \pi_j g(j) \left(g(j) - \sum_k p_{jk} g(k) \right).$$

For $j \notin A \cup \{i\}$ the factor $(g(j) - \sum_k p_{jk} g(k))$ equals zero, and for $j \in A$ we have $g(j) = 0$, so only the $j = i$ term contributes. Thus

$$\begin{aligned} \mathcal{E}(g,g) &= \pi_i \left(1 - \sum_k p_{ik} g(k)\right) \\ &= \pi_i (1 - P_i(T_i^+ < T_A)) \\ &= \pi_i P_i(T_A < T_i^+), \end{aligned} \tag{3.94}$$

giving (3.91). ■

The analogous result for two disjoint subsets A and B is a little complicated to state. The argument above shows that

$$\inf\{\mathcal{E}(g,g) : g(\cdot) = 0 \text{ on } A,\ g(\cdot) = 1 \text{ on } B\}$$

is attained by $g_0(j) = P_j(T_B < T_A)$ and that this g_0 satisfies

$$\mathcal{E}(g_0, g_0) = \sum_{i \in B} \pi_i P_i(T_A < T_B^+). \tag{3.95}$$

We want to interpret the reciprocal of this quantity as a mean time for the chain to commute from A to B and back. Consider the stationary chain $(X_t; -\infty < t < \infty)$. We can define what is technically called a "marked point process" which records the times at which A is first visited after a visit to B and vice versa. Precisely, define Z_t taking values in $\{\alpha, \beta, \delta\}$ by

$$Z_t := \begin{cases} \beta & \text{if } \exists s < t \text{ such that } X_s \in A,\ X_t \in B,\ X_u \notin A \cup B\ \forall s < u < t \\ \alpha & \text{if } \exists s < t \text{ such that } X_s \in B,\ X_t \in A,\ X_u \notin B \cup A\ \forall s < u < t \\ \delta & \text{otherwise.} \end{cases}$$

So the times t when $Z_t = \beta$ are the times of first return to B after visiting A, and the times t when $Z_t = \alpha$ are the times of first return to A after visiting B. Now (Z_t) is a stationary process. By considering the time-reversal of X, we see that for $i \in B$

$$P(X_0 = i,\ Z_0 = \beta) = P(X_0 = i,\ T_A < T_B^+) = \pi_i P_i(T_A < T_B^+).$$

So (3.95) shows $P(Z_0 = \beta) = \mathcal{E}(g_0, g_0)$. If we define T_{BAB}, "the typical time to go from B to A and back to B", to have the conditional distribution of $\min\{t \geq 1 : Z_t = \beta\}$ given $Z_0 = \beta$, then Kac's formula for the (non-Markov) stationary process Z (see e.g. [133] Theorem 6.3.3) says that $ET_{BAB} = 1/P(Z_0 = \beta)$. So we have proved (discrete time chain)

Corollary 3.37

$$ET_{BAB} = \sup\{1/\mathcal{E}(g,g) : 0 \le g \le 1,\ g(\cdot) = 0 \text{ on } A,\ g(\cdot) = 1 \text{ on } B\}$$

and the sup *is attained by* $g(i) = P_i(T_B < T_A)$.

As another interpretation of this quantity, define

$$\rho_B(\cdot) = P(X_0 \in \cdot \,|\, Z_0 = \beta), \quad \rho_A(\cdot) = P(X_0 \in \cdot \,|\, Z_0 = \alpha).$$

Interpret ρ_B and ρ_A as the distribution of hitting places on B and on A in the commute process. It is intuitively clear, and not hard to verify, that

$$P_{\rho_A}(X(T_B) \in \cdot) = \rho_B(\cdot), \quad P_{\rho_B}(X(T_A) \in \cdot) = \rho_A(\cdot)$$

$$ET_{BAB} = E_{\rho_B} T_A + E_{\rho_A} T_B.$$

In particular

$$\min_{i \in B} E_i T_A + \min_{i \in A} E_i T_B \le ET_{BAB} \le \max_{i \in B} E_i T_A + \max_{i \in A} E_i T_B.$$

3.7.1 Thompson's principle and leveling networks

Theorem 3.36 was stated in terms of (reversible) Markov chains. Rephrasing in terms of discrete-time random walk on a weighted graph gives the usual "electrical network" formulation of the Dirichlet principle stated below, using (3.77),(3.91) and (3.94). Recall from Proposition 3.10 that the effective resistance r between v_0 and A is, in terms of the random walk,

$$r = \frac{1}{w_{v_0} P_{v_0}(T_A < T^+_{v_0})}. \tag{3.96}$$

Proposition 3.38 (The Dirichlet principle) *Take a weighted graph and fix a vertex* v_0 *and a subset* A *of vertices not containing* v_0. *Then the quantity* $\frac{1}{2}\sum_i \sum_j w_{ij}(g(j) - g(i))^2$ *is minimized, over all functions* $g : I \to [0,1]$ *with* $g(v_0) = 1$ *and* $g(\cdot) = 0$ *on* A, *by the function* $g(i) :\equiv P_i(T_{v_0} < T_A)$ *(where probabilities refer to random walk on the weighted graph), and the minimum value equals* $1/r$, *where* r *is the effective resistance* (3.96).

There is a dual form of the Dirichlet principle, which following Doyle and Snell [131] we call

Proposition 3.39 (Thompson's principle) *Take a weighted graph and fix a vertex v_0 and a subset A of vertices not containing v_0. Let $\mathbf{f} = f_{ij}$ denote a unit flow from v_0 to A. Then $\frac{1}{2}\sum_i \sum_j (f_{ij}^2/w_{ij})$ is minimized, over all such flows, by the flow $\mathbf{f}^{v_0 \to A}$ [defined at (3.18)] associated with the random walk from v_0 to A, and the minimum value equals the effective resistance r appearing in (3.96).*

Recall that a flow is required to have $f_{ij} = 0$ whenever $w_{ij} = 0$, and interpret sums $\sum_i \sum_j$ as sums over ordered pairs (i, j) with $w_{ij} > 0$.

Proof. Write $\psi(\mathbf{f}) := \frac{1}{2}\sum_i \sum_j (f_{ij}^2/w_{ij})$. By formula (3.25) relating the random walk notions of "flow" and "potential", the fact that $\psi(\mathbf{f}^{v_0 \to A}) = r$ is immediate from the corresponding equality in the Dirichlet principle. So the issue is to prove that for a unit flow $\mathbf{f}^*$, say, attaining the minimum of $\psi(\mathbf{f})$, we have $\psi(\mathbf{f}^*) = \psi(\mathbf{f}^{v_0 \to A})$. To prove this, consider two arbitrary paths (y_i) and (z_j) from v_0 to A, and let $\mathbf{f}^\varepsilon$ denote the flow $\mathbf{f}^*$ modified by adding flow rates $+\varepsilon$ along the edges (y_i, y_{i+1}) and by adding flow rates $-\varepsilon$ along the edges (z_i, z_{i+1}). Then $\mathbf{f}^\varepsilon$ is still a unit flow from v_0 to A. So the function $\varepsilon \to \psi(\mathbf{f}^\varepsilon)$ must have derivative zero at $\varepsilon = 0$, and this becomes the condition that

$$\sum_i (f^*_{y_i,y_{i+1}}/w_{y_i,y_{i+1}}) = \sum_i (f^*_{z_i,z_{i+1}}/w_{z_i,z_{i+1}}).$$

So the sum is the same for all paths from v_0 to A. Fixing x, the sum must be the same for all paths from x to A, because two paths from x to A could be extended to paths from v_0 to A by appending a common path from v_0 to x. It follows that we can define $g^*(x)$ as the sum $\sum_i (f^*_{x_i,x_{i+1}}/w_{x_i,x_{i+1}})$ over some path (x_i) from x to A, and the sum does not depend on the path chosen. So

(This is essentially the same argument used in Section 3.3.2.)

$$g^*(x) - g^*(z) = \frac{f^*_{xz}}{w_{xz}} \text{ for each edge } (x, z) \text{ not contained within } A. \quad (3.97)$$

The fact that $\mathbf{f}^*$ is a flow means that, for $x \notin A \cup \{v_0\}$,

$$0 = \sum_{z: w_{xz} > 0} f^*_{xz} = \sum_z w_{xz}(g^*(x) - g^*(z)).$$

So g^* is a harmonic function outside $A \cup \{v_0\}$, and $g^* = 0$ on A. So by the uniqueness result (Chapter 2 Lemma 27) we have that g^* must be proportional to g, the minimizing function in Proposition 3.38. So $\mathbf{f}^*$ is proportional to $\mathbf{f}^{v_0 \to A}$, because the relationship (3.97) holds for both, and then $\mathbf{f}^* = \mathbf{f}^{v_0 \to A}$ because both are *unit* flows. ∎

9/10/99 version

A remarkable statistical interpretation was discussed in a monograph of Borre and Meissl [57]. Imagine a finite set of locations such as hilltops. For each pair of locations (i, j) with a clear line-of-sight, measure the elevation difference D_{ij} = (height of j minus height of i). Consider the associated graph [whose edges are such pairs (i, j)], and suppose it is connected. Take one location v_0 as a benchmark "height 0". If our measurements were exact we could determine the height of location x by adding the D's along a path from v_0 to x, and the sum would not depend on the path chosen. But suppose our measurements contain random errors. Precisely, suppose $D_{i,j}$ equals the true height difference $h(j) - h(i)$ plus an error Y_{ij} which has mean 0 and variance $1/w_{ij}$ and is independent for different measurements. Then it seems natural to estimate the height of x by taking some average $\hat{h}(x)$ over paths from v_0 to x, and it turns out that the "best" way to average is to use the random walk from v_0 to x and average (over realizations of the walk) the net height climbed by the walk.

In mathematical terms, the problem is to choose weights f_{ij}, not depending on the function h, such that

$$\hat{h}(x) := \frac{1}{2} \sum_i \sum_j f_{ij} D_{ij}$$

has $E\hat{h}(x) = h(x)$ and minimal variance. It is not hard to see that the former "unbiased" property holds iff f is a unit flow from v_0 to x. Then

$$\text{var } \hat{h}(x) = \frac{1}{4} \sum_i \sum_j f_{ij}^2 \text{var}(D_{ij}) = \frac{1}{4} \sum_i \sum_j \frac{f_{ij}^2}{w_{ij}}$$

and Proposition 3.39 says this is minimized when we use the flow from v_0 to x obtained from the random walk on the weighted graph. But then

$$\hat{h}(x) = E_{v_0} \sum_{t=1}^{T_x} D_{X_{t-1} X_t},$$

the expectation referring to the random walk.

3.7.2 Hitting times and Thompson's principle

Using the commute interpretation of resistance (Corollary 3.11) to translate Thompson's principle into an assertion about mean commute times gives the following.

Corollary 3.40 *For random walk on a weighted graph and distinct vertices v and a,*

$$E_vT_a + E_aT_v = w \inf \left\{ \frac{1}{2} \sum_i \sum_j (f_{ij}^2/w_{ij}) : \mathbf{f} \text{ is a unit flow from } a \text{ to } v \right\}$$

and the min *is attained by the flow $\mathbf{f}^{a\to v}$ associated with the random walk.*

Comparing with Theorem 3.36 we have two different extremal characterizations of mean commute times, as a *sup* over potential functions and as an *inf* over flows. In practice this "flow" form is less easy to use than the "potential" form, because writing down a flow $\mathbf{f}$ is harder than writing down a function g. But, when we can write down and calculate with some plausible flow, it gives upper bounds on mean commute times.

One-sided mean hitting times E_iT_j don't have simple extremal characterizations of the same kind, with the exception of hitting times from stationarity. To state the result, we need two definitions. First, given a probability distribution ρ on vertices, a *unit flow from a to ρ* is a flow f satisfying

$$f_{(i)} = 1_{(i=a)} - \rho_i \quad \text{for all } i; \tag{3.98}$$

more generally, a *unit flow from a set A to ρ* is defined to satisfy

$$\sum_{i\in A} f_{(i)} = 1 - \rho(A) \quad \text{and} \quad f_{(i)} = -\rho_i \text{ for all } i \in A^c.$$

Now fix a state a and define the special flow $\mathbf{f}^{a\to\pi}$ by

$$f_{ij} := \lim_{t_0\to\infty} E_a \sum_{t=1}^{t_0} \left(1_{(X_{t-1}=i,\, X_t=j)} - 1_{(X_{t-1}=j,\, X_t=i)} \right) \tag{3.99}$$

with the usual convention in the periodic case. So f_{ij} is the mean excess of transitions $i \to j$ compared to transitions $j \to i$, for the chain started at a and run forever. This is a unit flow from a to π, in the above sense. Equation (6) (the definition of $\mathbf{Z}$) in Chapter 2 and reversibility give the first equality, and Chapter 2 Lemma 12 gives the last equality, in

9/10/99 version

9/10/99 version

$$\begin{aligned} f_{ij} &= Z_{ai}p_{ij} - Z_{aj}p_{ji} \\ &= \frac{Z_{ia}\pi_i p_{ij}}{\pi_a} - \frac{Z_{ja}\pi_j p_{ji}}{\pi_a} \\ &= \frac{(Z_{ia} - Z_{ja})\pi_i p_{ij}}{\pi_a} \qquad (3.100) \\ &= \frac{(E_jT_a - E_iT_a)w_{ij}}{w}, \qquad (3.101) \end{aligned}$$

switching to "weighted graphs" notation. Note also that the first-step recurrence for the function $i \mapsto Z_{ia}$ is

$$Z_{ia} = \left(1_{(i=a)} - \pi_a\right) + \sum_j p_{ij} Z_{ja}. \tag{3.102}$$

Proposition 3.41 *For random walk on a weighted graph and a subset A of vertices,*

$$\begin{aligned} E_\pi T_A &= w \inf \left\{ \frac{1}{2} \sum_i \sum_j (f_{ij}^2 / w_{ij}) : \mathbf{f} \text{ is a unit flow from } A \text{ to } \pi \right\} \\ &= \sup \left\{ 1/\mathcal{E}(g,g) : -\infty < g < \infty,\ g(\cdot) = 1 \text{ on } A,\ \sum_i \pi_i g(i) = 0 \right\}. \end{aligned}$$

When A is a singleton $\{a\}$, the minimizing flow is the flow $\mathbf{f}^{a \to \pi}$ defined above, and the maximizing function g is $g(i) = Z_{ia}/Z_{aa}$. For general A the maximizing function g is $g(i) = 1 - \frac{E_i T_A}{E_\pi T_A}$.

Added the final observation about general A

Proof. Suppose first that $A = \{a\}$. We start by showing that the extremizing flow, $\mathbf{f}^*$ say, is the asserted $\mathbf{f}^{a \to \pi}$. By considering adding to $\mathbf{f}^*$ a flow of size ε along a directed cycle, and copying the argument for (3.97) in the proof of Proposition 3.39, there must exist a function g^* such that

$$g^*(x) - g^*(z) = \frac{f^*_{xz}}{w_{xz}} \text{ for each edge } (x,z). \tag{3.103}$$

The fact that $\mathbf{f}^*$ is a unit flow from a to π says that

$$1_{(x=a)} - \pi_x = \sum_z f^*_{xz} = \sum_z w_{xz}(g^*(x) - g^*(z))$$

which implies

$$\frac{1_{(x=a)} - \pi_x}{w_x} = \sum_z p_{xz}(g^*(x) - g^*(z)) = g^*(x) - \sum_z p_{xz} g^*(z).$$

Since $w_x = w\pi_x$ and $1/\pi_a = E_a T_a^+$, this becomes

$$g^*(x) = \sum_z p_{xz} g^*(z) - w^{-1} \left(1 - (E_a T_a^+) 1_{(x=a)}\right).$$

Now these equations have a unique solution g^*, up to an additive constant, because the difference between two solutions is a harmonic function. (Recall

99 version

Chapter 2 Corollary 28.) On the other hand, a solution is $g^*(x) = -\frac{E_x T_a}{w}$, by considering the first-step recurrence for $E_x T_a^+$. So by (3.103) $f^*_{xz} = (E_z T_a - E_x T_a) w_{xz}/w$, and so $\mathbf{f}^* = \mathbf{f}^{a\to\pi}$ by (3.101).

Now consider the function g which minimizes $\mathcal{E}(g,g)$ under the constraints $\sum_i \pi_i g(i) = 0$ and $g(a) = 1$. By introducing a Lagrange multiplier γ we may consider g as minimizing $\mathcal{E}(g,g) + \gamma \sum_i \pi_i g(i)$ subject to $g(a) = 1$. Repeating the argument at (3.92), the minimizing g satisfies

$$-2 \sum_{k \neq j} \pi_j p_{jk} (g(k) - g(j)) \ + \gamma \pi_j = 0, \quad j \neq a.$$

Rearranging, and introducing a term $\beta 1_{(j=a)}$ to cover the case $j = a$, we have

$$g(j) = \sum_k p_{jk} g(k) \ - (\gamma/2) + \beta 1_{(j=a)} \text{ for all } j,$$

for some γ, β. Because $\sum_j \pi_j g(j) = 0$ we have

$$0 = 0 - (\gamma/2) + \beta \pi_a,$$

allowing us to rewrite the equation as

$$g(j) = \sum_k p_{jk} g(k) \ + \beta (1_{(j=a)} - \pi_a).$$

By the familiar "harmonic function" argument this has a unique solution, and (3.102) shows the solution is $g(j) = \beta Z_{ja}$. Then the constraint $g(a) = 1$ gives $g(j) = Z_{ja}/Z_{aa}$.

Next consider the relationship between the flow $\mathbf{f} = \mathbf{f}^{a\to\pi}$ and the function $g(i) \equiv Z_{ia}/Z_{aa}$. We have

Chapter 2 reference in following display is to 9/10/99 version.

$$\begin{aligned}
w \ \frac{f_{ij}^2}{w_{ij}} &= (E_j T_a - E_i T_a)^2 \frac{w_{ij}}{w} \ \text{ by (3.101)} \\
&= (E_\pi T_a)^2 \frac{w_{ij}}{w} \left(\frac{E_j T_a - E_i T_a}{E_\pi T_a} \right)^2 \\
&= (E_\pi T_a)^2 \frac{w_{ij}}{w} \left(\frac{Z_{ia} - Z_{ja}}{Z_{aa}} \right)^2 \ \text{ by Chapter 2 Lemmas 11, 12} \\
&= (E_\pi T_a)^2 \frac{w_{ij}}{w} \ (g(i) - g(j))^2.
\end{aligned}$$

Thus it is enough to prove

$$w \frac{1}{2} \sum_i \sum_j \frac{f_{ij}^2}{w_{ij}} = E_\pi T_a \tag{3.104}$$

and it will then follow that

$$1/\mathcal{E}(g,g) = E_\pi T_a.$$

To prove (3.104), introduce a parameter ε (which will later go to 0) and a new vertex z and edge-weights $w_{iz} = \varepsilon w_i$. Writing superscripts $^\varepsilon$ to refer to this new graph and its random walk, the equation becomes $w^\varepsilon_{iz} := \varepsilon w_i$ and Corollary 3.40 says

$$E^\varepsilon_a T_z + E^\varepsilon_z T_a = w^\varepsilon \frac{1}{2} \sum_{i \in I^\varepsilon} \sum_{j \in I^\varepsilon} \frac{(f^\varepsilon_{ij})^2}{w^\varepsilon_{ij}} \tag{3.105}$$

where $\mathbf{f}^\varepsilon$ is the special unit flow from a to z associated with the new graph (which has vertex set $I^\varepsilon := I \cup \{z\}$). We want to interpret the ingredients to (3.105) in terms of the original graph. Clearly $w^\varepsilon = w(1+2\varepsilon)$. The new walk has chance $\varepsilon/(1+\varepsilon)$ to jump to z from each other vertex, so $E^\varepsilon_a T_z = (1+\varepsilon)/\varepsilon$. Starting from z, after one step the new walk has the stationary distribution π on the original graph, and it follows easily that $E^\varepsilon_z T_a = 1 + E_\pi T_a(1 + O(\varepsilon))$. We can regard the new walk up to time T_z as the old walk sent to z at a random time U^ε with Geometric$(\varepsilon/(1+\varepsilon))$ distribution, so for $i \neq z$ and $j \neq z$ the flow f^ε_{ij} is the expected net number of transitions $i \to j$ by the old walk up to time U^ε. From the spectral representation it follows easily that $f^\varepsilon_{ij} = f_{ij} + O(\varepsilon)$. Similarly, for $i \neq z$ we have $-f^\varepsilon_{zi} = f^\varepsilon_{iz} = P_a(X(U^\varepsilon - 1) = i) = \pi_i + O(\varepsilon)$; noting that $\sum_{i \in I} f^\varepsilon_{iz} = 1$, the total contribution of such terms to the double sum in (3.105) is

$$\begin{aligned} 2 \sum_{i \in I} \frac{(\pi_i + f^\varepsilon_{iz} - \pi_i)^2}{\varepsilon w_i} &= \frac{2}{w\varepsilon} \sum_{i \in I} \frac{(\pi_i + f^\varepsilon_{iz} - \pi_i)^2}{\pi_i} \\ &= \frac{2}{w\varepsilon} \left(1 + \sum_{i \in I} \frac{(f^\varepsilon_{iz} - \pi_i)^2}{\pi_i} \right) = \frac{2}{w\varepsilon} + O(\varepsilon). \end{aligned}$$

So (3.105) becomes

$$\frac{1+\varepsilon}{\varepsilon} + 1 + E_\pi T_a + O(\varepsilon) = w(1+2\varepsilon) \left(\frac{1}{2} \sum_{i \in I} \sum_{j \in I} \frac{f^2_{ij}}{w_{ij}} + \frac{1}{w\varepsilon} \right) + O(\varepsilon)$$

Subtracting $(1+2\varepsilon)/\varepsilon$ from both sides and letting $\varepsilon \to 0$ gives the desired (3.104). This concludes the proof for the case $A = \{a\}$, once we use the mean hitting time formula to verify

$$g(i) :\equiv \tfrac{Z_{ia}}{Z_{aa}} = 1 - \tfrac{E_i T_a}{E_\pi T_a}.$$

Finally, the extension to general A is an exercise in use of the chain, say X^*, in which the set A is collapsed to a single state a. Recall Chapter 2 Section 7.3. In particular, 9/10/99 version

$$w^*_{ij} = w_{ij},\ i,j \in A^c; \quad w^*_{ia} = \sum_{k\in A} w_{ik},\ i \in A^c; \quad w^*_{aa} = \sum_{k\in A}\sum_{l\in A} w_{kl}; \quad w^* = w.$$

We now sketch the extension. First, $E_\pi T_A = E^*_{\pi^*} T_a$. Then the natural one-to-one correspondence between functions g on I with $g(\cdot) = 1$ on A and $\sum_{i\in I} \pi_i g(i) = 0$ and functions g^* on I^* with $g^*(a) = 0$ and $\sum_{i\in I^*} \pi^*_i g^*(i) = 0$ gives a trivial proof of

$$E_\pi T_A = \sup\left\{1/\mathcal{E}(g,g) : -\infty < g < \infty,\ g(\cdot) = 1 \text{ on } A,\ \sum_i \pi_i g(i) = 0\right\}.$$

It remains to show that

$$\begin{aligned} &\inf\{\Psi(\mathbf{f}) : \mathbf{f} \text{ is a unit flow from } A \text{ to } \pi \text{ (for } X)\} \\ &\quad = \inf\{\Psi^*(\mathbf{f}^*) : \mathbf{f}^* \text{ is a unit flow from } a \text{ to } \pi^* \text{ (for } X^*)\} \end{aligned} \quad (3.106)$$

where

$$\Psi(\mathbf{f}) := \frac{1}{2}\sum_{i\in I}\sum_{j\in I}(f^2_{ij}/w_{ij}), \quad \Psi^*(\mathbf{f}^*) := \frac{1}{2}\sum_{i\in I^*}\sum_{j\in I^*}((f^*_{ij})^2/w^*_{ij}).$$

Indeed, given a unit flow $\mathbf{f}$ from A to π, define f^*_{ij} as f_{ij} if $i,j \in A^c$ and as $\sum_{k\in A} f_{ik}$ if $i \in A^c$ and $j = a$. One can check that $\mathbf{f}^*$ is a unit flow from a to π^* (the key observation being that $\sum_{i\in A}\sum_{j\in A} f_{ij} = 0$) and, using the Cauchy–Schwarz inequality, that $\Psi^*(\mathbf{f}^*) \le \Psi(\mathbf{f})$. Conversely, given a unit flow $\mathbf{f}^*$ from a to π^*, define f_{ij} as f^*_{ij} if $i,j \in A^c$, as $f^*_{ia} w_{ij} / \sum_{k\in A} w_{ik}$ if $i \in A^c$ and $j \in A$, and as 0 if $i,j \in A$. One can check that $\mathbf{f}$ is a unit flow from A to π and that $\Psi(\mathbf{f}) = \Psi(\mathbf{f}^*)$. We have thus established (3.106), completing the proof. ■

Corollary 3.42 *For chains with transition matrices $\mathbf{P}, \tilde{\mathbf{P}}$ and the same stationary distribution π,*

$$\min_{i\ne j}\frac{p_{ij}}{\tilde{p}_{ij}} \le \frac{\tilde{E}_\pi T_a}{E_\pi T_a} \le \max_{i\ne j}\frac{p_{ij}}{\tilde{p}_{ij}}.$$

Proof. Plug the minimizing flow $\mathbf{f}^{a\to\pi}$ for the $\mathbf{P}$-chain into Proposition 3.41 for the $\tilde{\mathbf{P}}$-chain to get the second inequality. The first follows by reversing the roles of $\mathbf{P}$ and $\tilde{\mathbf{P}}$. ■

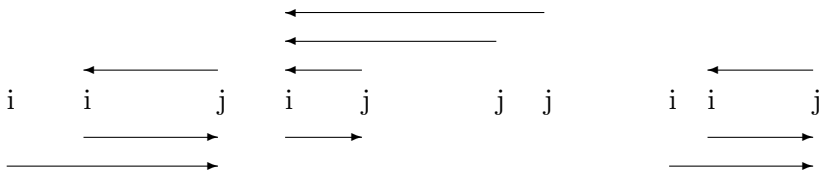
3.8 Notes on Chapter 3

Textbooks. Almost all the results have long been known to (different groups of) experts, but it has not been easy to find accessible textbook treatments. Of the three books on reversible chains at roughly the same level of sophistication as ours, Kelly [213] emphasizes stationary distributions of stochastic networks; Keilson [212] emphasizes mathematical properties such as complete monotonicity; and Chen [88] discusses those aspects useful in the study of interacting particle systems.

Section 3.1. In abstract settings reversible chains are called *symmetrizable*, but that's a much less evocative term. Elementary textbooks often give *Kolmogorov's criterion* ([213] Thm 1.7) for reversibility, but we have never found it to be useful.

I put a pointer from the first section to here - DA

The following figure may be helpful in seeing why $\pi_i E_i T_j \neq \pi_j E_j T_i$ for a general reversible chain, even if π is uniform. Run such a chain (X_t) for $-\infty < t < \infty$ and record only the times when the chain is in state i or state j. Then $E_i T_j$ is the long-run empirical average of the passage times from i to j, indicated by arrows $\longrightarrow$; and $E_j T_i$ is the long-run empirical average of the passage times from j to i in the reverse time direction, indicated by arrows $\longleftarrow$. One might think these two quantities were averaging the same empirical intervals, but a glance at the figure shows they are not.



Section 3.1.1. Though probabilists would regard the "cyclic tour" Lemma 3.2 as obvious, László Lovász pointed out a complication, that with a careful definition of starts and ends of tours these times are not invariant under time-reversal. The sophisticated fix is to use doubly-infinite stationary chains and observe that tours in reversed time just interleave tours in forward time, so by ergodicity their asymptotic rates are equal. Tetali [326] shows that the cyclic tour property implies reversibility. Tanushev and Arratia [321] show that the *distributions* of forward and reverse tour times are equal.

9/1/99 version

Cat-and-mouse game 1 is treated more opaquely in Coppersmith et al [99], whose deeper results are discussed in Chapter 9 Section 4.4. Underlying the use of the optional sampling theorem in game 2 is a general result about optimal stopping, but it's much easier to prove what we want here than to

appeal to general theory. Several algorithmic variations on Proposition 3.3 are discussed in Coppersmith et al [101] and Tetali and Winkler [327].

Section 3.2. Many textbooks on Markov chains note the simple explicit form of the stationary distribution for random walks on graphs. An historical note (taken from [107]) is that the first explicit treatment of random walk on a general finite graph was apparently given in 1935 by Bottema [58], who proved the convergence theorem. Amongst subsequent papers specializing Markov theory to random walks on graphs let us mention Gobel and Jagers [168], which contains a variety of the more elementary facts given in this book, for instance the unweighted version of Lemma 3.9. Another observation from [168] is that for a reversible chain the quantity

$$\beta_{ijl} \equiv \pi_j^{-1} E_i(\text{number of visits to } j \text{ before time } T_l)$$

satisfies $\beta_{ijl} = \beta_{jil}$. Indeed, by Chapter 2 Lemma 9 we have 9/10/99 version

$$\beta_{ijl} = (E_i T_l + E_l T_j + E_j T_i) - (E_j T_i + E_i T_j)$$

and so the result follows from the cyclic tour property.

Just as random walks on undirected graphs are as general as reversible Markov chains, so random walks on *directed* graphs are as general as general Markov chains. In particular, one usually has no simple expression like (3.15) for the stationary distribution. The one tractable case is a *balanced* directed graph, where the in-degree d_v of each vertex v equals its out-degree.

Section 3.2.1. Yet another way to associate a continuous-time reversible chain with a weighted graph is to set $q_{ij} = w_{ij}/\sqrt{w_i w_j}$. This construction was used by Chung and Yau [95] as the simplest way to set up discrete analogs of certain results from differential geometry.

Another interpretation of continuous-time random walk on a weighted graph is to write $w_{ij} = 1/L_{ij}$ and interpret L_{ij} as edge-length. Then run Brownian motion on the edges of the graph. Starting from vertex i, the chance that j is the first vertex other than i visited is $w_{ij}/\sum_k w_{ik}$, so the embedded discrete chain is the usual discrete random walk. This construction could be used as an intermediate step in the context of approximating Brownian motion on a manifold by random walk on a graph embedded in the manifold.

Section 3.3. Doyle and Snell [131] gave a detailed elementary textbook exposition of Proposition 3.10 and the whole random walk / electrical network connection. Previous brief textbook accounts were given by Kemeny et al [215] and Kelly [213]. Our development follows closely that of Chapter 2 in Lyons and Peres [250]. As mentioned in the text, the first explicit use

(known to us) of the mean commute interpretation was given by Chandra et al [85]. One can combine the commute formula with the general identities of Chapter 2 to obtain numerous identities relating mean hitting times and resistances, some of which are given (using bare-hands proofs instead) in Tetali [324]. The connection between Foster's theorem and Lemma 3.9 was noted in [99].

Section 3.4. The spectral theory is of course classical. In devising a symmetric matrix one could use $\pi_i p_{ij}$ or $p_{ij}\pi_j^{-1}$ instead of $\pi_i^{1/2} p_{ij} \pi_j^{-1/2}$— there doesn't seem any systematic advantage to a particular choice. We learned the eigentime identity from Andrei Broder who used it in [66], and Lemma 3.15 from David Zuckerman who used it in [341]. Apparently no-one has studied whether Lemma 3.15 holds for general chains. Mark Brown (personal communication) has noted several variations on the theme of Lemma 3.15, for example that the unweighted average of $(E_i T_j; i, j \in A)$ is bounded by the unweighted average of $(E_\pi T_j; j \in A)$. The name *eigentime identity* is our own coinage: once we call $1/\lambda_2$ the relaxation time it is natural to start thinking of the other $1/\lambda_m$ as "eigentimes".

Section 3.5. We regard complete monotonicity as a name for "mixtures of exponentials", and have not used the analytic characterization via derivatives of alternating signs. Of course the CM property is implicit in much analysis of reversible Markov processes, but we find it helpful to exhibit explicitly its use in obtaining inequalities. This idea in general, and in particular the "stochastic ordering of exit times" result (Proposition 3.21), were first emphasized by Keilson [212] in the context of reliability and queueing models. Brown [73] gives other interesting consequences of monotonicity.

Section 3.5.1. Parts of Proposition 3.18 have been given by several authors, e.g., Broder and Karlin [66] Corollary 18 give (3.53). One can invent many variations. Consider for instance $\min_i \max_j E_i T_j$. On the complete graph this equals $n-1$, but this is not the minimum value, as observed by Erik Ordentlich in a homework exercise. If we take the complete graph, distinguish a vertex i_0, let the edges involving i_0 have weight ε and the other edges have weight 1, then as $\varepsilon \to 0$ we have (for $j \neq i_0$)

$$E_{i_0} T_j \to \tfrac{1}{n-1} + \tfrac{n-2}{n-1}(1 + (n-2)) = n - 2 + \tfrac{1}{n-1}.$$

By the random target lemma and (3.53), the quantity under consideration is at least $\tau_0 \geq n - 2 + \frac{1}{n}$, so the example is close to optimal.

Section 3.5.4. The simple result quoted as Proposition 3.22 is actually weaker than the result proved in Brown [72]. The ideas in the proof of Proposition 3.23 are in Aldous [12] and in Brown [73], the latter containing

a shorter Laplace transform argument for (3.73). Aldous and Brown [20] give a more detailed account of the exponential approximation, including the following result which is useful in precisely the situation where Proposition 3.23 is applicable, that is, when $E_\pi T_A$ is large compared to τ_2.

Theorem 3.43 *Let α_A be the quasistationary distribution on A^c defined at* (3.82)*. Then*

$$\begin{aligned} P_\pi(T_A > t) &\geq \left(1 - \frac{\tau_2}{E_{\alpha_A} T_A}\right) \exp\left(\frac{-t}{E_{\alpha_A} T_A}\right), \quad t > 0 \\ E_\pi T_A &\geq E_{\alpha_A} T_A - \tau_2. \end{aligned}$$

Using this requires only a *lower* bound on $E_\alpha T_A$, which can often be obtained using the extremal characterization (3.84). Connections with "interleaving of eigenvalues" results are discussed in Brown [74].

For general chains, explicit bounds on exponential approximation are much messier: see Aldous [5] for a bound based upon total variation mixing and Iscoe and McDonald [192] for a bound involving spectral gaps.

Section 3.6.1. Dirichlet forms were developed for use with continuous-space continuous-time Markov processes, where existence and uniqueness questions can be technically difficult—see, e.g., Fukushima [159]. Their use subsequently trickled down to the discrete world, influenced, e.g., by the paper of Diaconis and Stroock [124]. Chen [88] is the most accessible introduction.

Section 3.6.2. Since mean commute times have two dual extremal characterizations, as *sups* over potential functions and as *infs* over flows, it is natural to ask

Open Problem 3.44 Does there exist a characterization of the relaxation time as exactly an *inf* over flows?

We will see in Chapter 4 Theorem 32 an inequality giving an upper bound on the relaxation time in terms of an *inf* over flows, but it would be more elegant to derive such inequalities from some exact characterization. 10/11/94 version

Section 3.6.4. Lemma 3.32 is sometimes used to show, by comparison with the i.i.d. chain,

$$\text{if } \min_{i,j:i\neq j} (p_{ij}/\pi_j) = \delta > 0 \text{ then } \tau_2 \leq \delta^{-1}.$$

But this is inefficient: direct use of submultiplicitivity of variation distance gives a stronger conclusion.

Section 3.6.5. Quasistationary distributions for general chains have long been studied in applied probability, but the topic lacks a good survey article. Corollary 3.34 is a good example of a repeatedly-rediscovered simple-yet-useful result which defies attempts at attribution.

Kahale [203] Corollary 6.1 gives a discrete time variant of Lemma 3.35, that is to say an upper bound on $P_\pi(T_A \geq t)$, and Alon et al [27] Proposition 2.4 give a lower bound in terms of the *smallest* eigenvalue (both results are phrased in the context of random walk on an undirected graph). In studying bounds on T_A such as Lemma 3.35 we usually have in mind that $\pi(A)$ is small. One is sometimes interested in exit times from a set A with $\pi(A)$ small, i.e., hitting times on A^c where $\pi(A^c)$ is near 1. In this setting one can replace inequalities using τ_2 or τ_c (parameters which involve the whole chain) by inequalities involving analogous parameters for the chain restricted to A and its boundary. See Babai [36] for uses of such bounds.

details in previous version now deleted

Section 3.7.1. Use of Thompson's principle and the Dirichlet principle to study transience / recurrence of countably infinite state space chains is given an elementary treatment in Doyle and Snell [131] and more technical treatments in papers of Nash-Williams [266], Griffeath and Liggett [173] and Lyons [251]. Some reformulations of Thompson's principle are discussed by Berman and Konsowa [45].

We learned about the work of Borre and Meissl [57] on leveling networks from Persi Diaconis. Here is another characterization of effective resistance that is of a similar spirit. Given a weighted graph, assign independent Normal$(0, w_{ij})$ random variables X_{ij} to the edges (i, j), with $X_{ji} = -X_{ij}$. Then condition on the event that the sum around any cycle vanishes. The conditional process (Y_{ij}) is still Gaussian. Fix a reference vertex v_* and for each vertex v let S_v be the sum of the Y-values along a path from v_* to v. (The choice of path doesn't matter, because of the conditioning event.) Then (obviously) S_v is mean-zero Normal but (not obviously) its variance is the effective resistance between v_* and v. This is discussed and proved in Janson [194] Section 9.4.

Section 3.7.2. We have never seen in the literature an explicit statement of the extremal characterizations for mean hitting times from a stationary start (Proposition 3.41), but these are undoubtedly folklore, at least in the "potential" form. Iscoe et al [193] implicitly contains the analogous characterization of $E_\pi \exp(-\theta T_A)$. Steve Evans once showed us an argument for Corollary 3.42 based on the usual Dirichlet principle, and that motivated us to present the "natural explanation" given by Proposition 3.41.

Initial distribution is indeed π – yes (DA)

Chapter 4

Hitting and Convergence Time, and Flow Rate, Parameters for Reversible Markov Chains (October 11, 1994)

The elementary theory of general finite Markov chains (cf. Chapter 2) focuses on exact formulas and limit theorems. My view is that, to the extent there is any intermediate-level mathematical theory of reversible chains, it is a theory of *inequalities*. Some of these were already seen in Chapter 3. This chapter is my attempt to impose some order on the subject of inequalities. We will study the following five parameters of a chain. Recall our standing assumption that chains are finite, irreducible and reversible, with stationary distribution π.

(i) The maximal mean commute time

$$\tau^* = \max_{ij}(E_iT_j + E_jT_i)$$

(ii) The average hitting time

$$\tau_0 = \sum_i \sum_j \pi_j \pi_i E_i T_j.$$

(iii) The variation threshold time

$$\tau_1 = \inf\{t > 0 : \bar{d}(t) \leq e^{-1}\}$$

where as in Chapter 2 section yyy

$$\bar{d}(t) = \max_{ij} ||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)||$$

(iv) The relaxation time τ_2, i.e. the time constant in the asymptotic rate of convergence to the stationary distribution.

(v) A "flow" parameter

$$\tau_c = \sup_A \frac{\pi(A)\pi(A^c)}{\sum_{i\in A}\sum_{j\in A^c}\pi_i p_{ij}} = \sup_A \frac{\pi(A^c)}{P_\pi(X_1 \in A^c | X_0 \in A)}$$

in discrete time, and

$$\tau_c = \sup_A \frac{\pi(A)\pi(A^c)}{\sum_{i\in A}\sum_{j\in A^c}\pi_i q_{ij}} = \sup_A \frac{\pi(A^c)\ dt}{P_\pi(X_{dt} \in A^c | X_0 \in A)}$$

in continuous time.

The following table may be helpful. "Average-case" is intended to indicate essential use of the stationary distribution.

	worst-case	average-case
hitting times	τ^*	τ_0
mixing times	τ_1	τ_2
flow		τ_c

The table suggests there should be a sixth parameter, but I don't have a candidate.

The ultimate point of this study, as will seen in following chapters, is

- For many questions about reversible Markov chains, the way in which the answer depends on the chain is related to one of these parameters
- so it is useful to have methods for estimating these parameters for particular chains.

This Chapter deals with relationships between these parameters, simple illustrations of properties of chains which are closely connected to the parameters, and methods of bounding the parameters. To give a preview, it turns out that these parameters are essentially decreasing in the order $(\tau^*, \tau_0, \tau_1, \tau_2, \tau_c)$: precisely,

$$\frac{1}{2}\tau^* \geq \tau_0 \geq \tau_2 \geq \tau_c$$

$$66\tau_0 \geq \tau_1 \geq \tau_2$$

and perhaps the constant 66 can be reduced to 1. There are no general reverse inequalities, but reverse bounds involving extra quantities provide a rich and sometimes challenging source of problems.

The reader may find it helpful to read this chapter in parallel with the list of examples of random walks on unweighted graphs in Chapter 5. As another preview, we point out that on regular n-vertex graphs each parameter may be as large as $\Theta(n^2)$ but no larger; and τ^*, τ_0 may be as small as $\Theta(n)$ and the other parameters as small as $\Theta(1)$, but no smaller. The property (for a sequence of chains) "$\tau_0 = O(n)$" is an analog of the property "transience" for a single infinite-state chain, and the property "$\tau_2 = O(\text{poly}(\log n))$" is an analog of the "non-trivial boundary" property for a single infinite-state chain. These analogies are pursued in Chapter yyy.

The next five sections discuss the parameters in turn, the relationship between two different parameters being discussed in the latter's section. Except for τ_1, the numerical values of the parameters are unchanged by continuizing a discrete-time chain. And the results of this Chapter not involving τ_1 hold for either discrete or continuous-time chains.

4.1 The maximal mean commute time τ^*

We start by repeating the definition

$$\tau^* \equiv \max_{ij}(E_iT_j + E_jT_i) \tag{4.1}$$

and recalling what we already know. Obviously

$$\max_{ij} E_iT_j \le \tau^* \le 2 \max_{ij} E_iT_j$$

and by Chapter 3 Lemma yyy

$$\max_{j} E_\pi T_j \le \tau^* \le 4 \max_{j} E_\pi T_j. \tag{4.2}$$

Arguably we could have used $\max_{ij} E_iT_j$ as the "named" parameter, but the virtue of τ^* is the resistance interpretation of Chapter 3 Corollary yyy.

Lemma 4.1 *For random walk on a weighted graph,*

$$\tau^* = w \max_{ij} r_{ij}$$

where r_{ij} is the effective resistance between i and j.

In Chapter 3 Proposition yyy we proved lower bounds for any n-state discrete-time reversible chain:

$$\tau^* \geq 2(n-1)$$

$$\max_{ij} E_i T_j \geq n-1$$

which are attained by random walk on the complete graph. Upper bounds will be discussed extensively in Chapter 6, but let's mention two simple ideas here. Consider a path $i = i_0, i_1, \ldots, i_m = j$, and let's call this path γ_{ij} (because we've run out of symbols whose names begin with "p"!) This path, considered in isolation, has "resistance"

$$r(\gamma_{ij}) \equiv \sum_{e \in \gamma_{ij}} 1/w_e$$

which by the Monotonicity Law is at least the effective resistance r_{ij}. Thus trivially

$$\tau^* \leq w \max_{i,j} \min_{\text{paths } \gamma_{ij}} r(\gamma_{ij}). \tag{4.3}$$

A more interesting idea is to combine the max-flow min-cut theorem (see e.g. [86] sec. 5.4) with Thompson's principle (Chapter 3 Corollary yyy). Given a weighted graph, define

$$c \equiv \min_A \sum_{i \in A} \sum_{j \in A^c} w_{ij} \tag{4.4}$$

the *min* over proper subsets A. The max-flow min-cut theorem implies that for any pair a, b there exists a flow $\mathbf{f}$ from a to b of size c such that $|f_{ij}| \leq w_{ij}$ for all edges (i, j). So there is a *unit* flow from a to b such that $|f_e| \leq c^{-1} w_e$ for all edges e. It is clear that by deleting any flows around cycles we may assume that the flow through any vertex i is at most unity, and so

$$\sum_j |f_{ij}| \leq 2 \text{ for all } i, \text{ and } = 1 \text{ for } i = a, b. \tag{4.5}$$

So

$$\begin{aligned} E_a T_b + E_b T_a &\leq w \sum_e \frac{f_e^2}{w_e} \text{ by Thompson's principle} \\ &\leq \frac{w}{c} \sum_e |f_e| \\ &\leq \frac{w}{c} (n-1) \text{ by (4.5).} \end{aligned}$$

and we have proved

Proposition 4.2 *For random walk on an n-vertex weighted graph,*

$$\tau^* \le \frac{w(n-1)}{c}$$

for c defined at (4.4).

Lemma 4.1 and the Monotonicity Law also make clear a one-sided bound on the effect of changing edge-weights monotonically.

Corollary 4.3 *Let $\tilde{w}_e \ge w_e$ be edge-weights and let $\tilde{\tau}^*$ and τ^* be the corresponding parameters for the random walks. Then*

$$\frac{E_i\tilde{T}_j + E_j\tilde{T}_i}{E_iT_j + E_jT_i} \le \frac{\tilde{w}}{w} \text{ for all } i, j$$

and so

$$\tilde{\tau}^*/\tau^* \le \tilde{w}/w.$$

In the case of unweighted graphs the bound in Corollary 4.3 is $|\tilde{\mathcal{E}}|/|\mathcal{E}|$. Example yyy of Chapter 3 shows there can be no lower bound of this type, since in that example $\tilde{w}/w = 1 + O(1/n)$ but (by straightforward calculations) $\tilde{\tau}^*/\tau^* = O(1/n)$.

4.2 The average hitting time τ_0

As usual we start by repeating the definition

$$\tau_0 \equiv \sum_i \sum_j \pi_j \pi_i E_i T_j \tag{4.6}$$

and recalling what we already know. We know (a result not using reversibility: Chapter 2 Corollary yyy) the *random target lemma*

$$\sum_j \pi_j E_i T_j = \tau_0 \text{ for all } i \tag{4.7}$$

and we know the *eigentime identity* (Chapter 3 yyy)

$$\tau_0 = \sum_{m \ge 2} (1 - \lambda_m)^{-1} \text{ in discrete time} \tag{4.8}$$

$$\tau_0 = \sum_{m \ge 2} \lambda_m^{-1} \text{ in continuous time} \tag{4.9}$$

In Chapter 3 yyy we proved a lower bound for n-state discrete-time chains:

$$\tau_0 \geq \frac{(n-1)^2}{n}$$

which is attained by random walk on the complete graph.

We can give a flow characterization by averaging over the characterization in Chapter 3 yyy. For each vertex a let $\mathbf{f}^{a\to\pi} = (f_{ij}^{a\to\pi})$ be a flow from a to π of volume π_a, that is a unit flow scaled by π_a. Then

$$\tau_0 = w \ \min \left\{ \frac{1}{2} \sum_i \sum_j \sum_a \frac{(f_{ij}^{a\to\pi})^2}{\pi_a w_{ij}} \right\}$$

the *min* being over families of flows $\mathbf{f}^{a\to\pi}$ described above.

By writing

$$\tau_0 = \frac{1}{2} \sum_i \sum_j \pi_i \pi_j (E_i T_j + E_j T_i) \leq \frac{1}{2} \max_{ij} (E_i T_j + E_j T_i)$$

we see that $\tau_0 \leq \frac{1}{2}\tau^*$. It may happen that τ^* is substantially larger than τ_0. A fundamental example is the $M/M/1/n$ queue (xxx) where τ_0 is linear in n but τ^* grows exponentially. A simple example is the two-state chain with

$$p_{01} = \varepsilon, p_{10} = 1-\varepsilon, \quad \pi_0 = 1-\varepsilon, \pi_1 = \varepsilon$$

for which $\tau_0 = 1$ but $\tau^* = \frac{1}{\varepsilon} + \frac{1}{1-\varepsilon}$. This example shows that (without extra assumptions) we can't improve much on the bound

$$\tau^* \leq \frac{2\tau_0}{\min_j \pi_j} \tag{4.10}$$

which follows from the observation $E_i T_j \leq \tau_0/\pi_j$.

One can invent examples of random walks on regular graphs in which also τ^* is substantially larger than τ_0. Under symmetry conditions (vertex-transitivity, Chapter 7) we know *a priori* that $E_\pi T_i$ is the same for all i and hence by (4.2) $\tau^* \leq 4\tau_0$. In practice we find that τ_0 and τ^* have the same order of magnitude in most "naturally-arising" graphs, but I don't know any satisfactory formalization of this idea.

The analog of Corollary 4.3 clearly holds, by averaging over i and j.

Corollary 4.4 *Let $\tilde{w}_e \geq w_e$ be edge-weights and let $\tilde{\tau}_0$ and τ_0 be the corresponding parameters for the random walks. Then*

$$\tilde{\tau}_0/\tau_0 \leq \tilde{w}/w.$$

In one sense this is mysterious, because in the eigentime identity the largest term in the sum is the first term, the relaxation time τ_2, and Example yyy of Chapter 3 shows that there is no such upper bound for τ_2.

4.3 The variation threshold τ_1.

4.3.1 Definitions

Recall from Chapter 2 yyy that $||\quad||$ denotes variation distance and

$$d(t) \equiv \max_i ||P_i(X_t \in \cdot) - \pi(\cdot)||$$

$$\bar{d}(t) \equiv \max_{ij} ||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)||$$

$$d(t) \le \bar{d}(t) \le 2d(t)$$

$$\bar{d}(s+t) \le \bar{d}(s)\bar{d}(t)$$

We define the parameter

$$\tau_1 \equiv \min\{t : \bar{d}(t) \le e^{-1}\}. \tag{4.11}$$

The choice of constant e^{-1}, and of using $\bar{d}(t)$ instead of $d(t)$, are rather arbitrary, but this choice makes the numerical constants work out nicely (in particular, makes $\tau_2 \le \tau_1$ – see section 4.4). Submultiplicativity gives

Lemma 4.5 $d(t) \le \bar{d}(t) \le \exp(-\lfloor t/\tau_1 \rfloor) \le \exp(1 - t/\tau_1),\ t \ge 0.$

The point of parameter τ_1 is to formalize the idea of "time to approach stationarity, from worst starting-place". The fact that variation distance is just one of several distances one could use may make τ_1 seem a very arbitrary choice, but Theorem 4.6 below says that three other possible quantifications of this idea are equivalent. Here *equivalent* has a technical meaning: parameters τ_a and τ_b are equivalent if their ratio is bounded above and below by numerical constants not depending on the chain. (Thus (4.2) says τ^* and $\max_j E_\pi T_j$ are equivalent parameters). More surprisingly, τ_1 is also equivalent to two more parameters involving mean hitting times. We now define all these parameters.

xxx *Warning.* Parameters $\tau_1^{(4)}, \tau_1^{(5)}$ in this draft were parameters $\tau_1^{(3)}, \tau_1^{(4)}$ in the previous draft.

The first idea is to measure distance from stationarity by using ratios of probabilities. Define *separation* from stationarity to be

$$s(t) \equiv \min\{s : p_{ij}(t) \ge (1-s)\pi_j \text{ for all } i, j\}.$$

Then $s(\cdot)$ is submultiplicative, so we naturally define the separation threshold time to be

$$\tau_1^{(1)} \equiv \min\{t : s(t) \le e^{-1}\}.$$

The second idea is to consider minimal *random* times at which the chain has *exactly* the stationary distribution. Let

$$\tau_1^{(2)} \equiv \max_i \min_{U_i} E_i U_i$$

where the *min* is over stopping times U_i such that $P_i(X(U_i) \in \cdot) = \pi(\cdot)$. As a variation on this idea, let us temporarily write, for a probability distribution μ on the state space,

$$\tau(\mu) \equiv \max_i \min_{U_i} E_i U_i$$

where the *min* is over stopping times U_i such that $P_i(X(U_i) \in \cdot) = \mu(\cdot)$. Then define

$$\tau_1^{(3)} = \min_\mu \tau(\mu).$$

Turning to the parameters involving mean hitting times, we define

$$\tau_1^{(4)} \equiv \max_{i,k} \sum_j \pi_j |E_i T_j - E_k T_j| = \max_{i,k} \sum_j |Z_{ij} - Z_{kj}| \tag{4.12}$$

where the equality involves the fundamental matrix $\mathbf{Z}$ and holds by the mean hitting time formula. Parameter $\tau_1^{(4)}$ measures variability of mean hitting times as the starting place varies. The final parameter is

$$\tau_1^{(5)} \equiv \max_{i,A} \pi(A) E_i T_A.$$

Here we can regard the right side as the ratio of $E_i T_A$, the Markov chain mean hitting time on A, to $1/\pi(A)$, the mean hitting time under independent sampling from the stationary distribution.

The definitions above make sense in either discrete or continuous time, but the following notational convention turns out to be convenient. For a discrete-time chain we define τ_1 to be the value obtained by applying the definition (4.11) to the *continuized* chain, and write τ_1^{disc} for the value obtained for the discrete-time chain itself. Define similarly $\tau_1^{(1)}$ and $\tau_1^{1,\text{disc}}$. But the other parameters $\tau_1^{(2)} - \tau_1^{(5)}$ are defined directly in terms of the discrete-time chain. We now state the equivalence theorem, from Aldous [6].

Theorem 4.6 *(a) In either discrete or continuous time, the parameters* $\tau_1, \tau_1^{(1)}, \tau_1^{(2)}, \tau_1^{(3)}, \tau_1^{(4)}$ *and* $\tau_1^{(5)}$ *are equivalent.*
(b) In discrete time, τ_1^{disc} *and* $\tau_1^{1,disc}$ *are equivalent, and* $\tau_1^{(2)} \le \frac{e}{e-1} \tau_1^{1,disc}$
.

This will be (partially) proved in section 4.3.2, but let us first give a few remarks and examples. The parameter τ_1 and total variation distance are closely related to the notion of *coupling* of Markov chains, discussed in Chapter 14. Analogously (see the Notes), the separation $s(t)$ and the parameter $\tau_1^{(1)}$ are closely related to the notion of *strong stationary times* V_i for which

$$P_i(X(V_i) \in \cdot \ | V_i = t) = \pi(\cdot) \text{ for all } t. \tag{4.13}$$

Under our standing assumption of reversibility there is a close connection between separation and variation distance, indicated by the next lemma.

Lemma 4.7 *(a)* $\bar{d}(t) \le s(t)$.
(b) $s(2t) \le 1 - (1 - \bar{d}(t))^2$.

Proof. Part (a) is immediate from the definitions. For (b),

$$\begin{aligned}
\frac{p_{ik}(2t)}{\pi_k} &= \sum_j \frac{p_{ij}(t)p_{jk}(t)}{\pi_k} \\
&= \sum_j \pi_j \frac{p_{ij}(t)p_{kj}(t)}{\pi_j^2} \text{ by reversibility} \\
&\ge \left(\sum_j \pi_j \frac{p_{ij}^{1/2}(t)p_{kj}^{1/2}(t)}{\pi_j} \right)^2 \text{ by } EZ \ge (EZ^{1/2})^2 \\
&\ge \left(\sum_j \min(p_{ij}(t), p_{kj}(t)) \right)^2 \\
&= (1 - ||P_i(X_t \in \cdot) - P_k(X_t \in \cdot)||)^2 \\
&\ge (1 - \bar{d}(t))^2. \quad \square
\end{aligned}$$

Note also that the definition of $s(t)$ involves *lower* bounds in the convergence $\frac{p_{ij}(t)}{\pi_j} \to 1$. One can make a definition involving *upper* bounds

$$\hat{d}(t) \equiv \max_{i,j} \frac{p_{ij}(t)}{\pi_j} - 1 = \max_i \frac{p_{ii}(t)}{\pi_i} - 1 \ge 0 \tag{4.14}$$

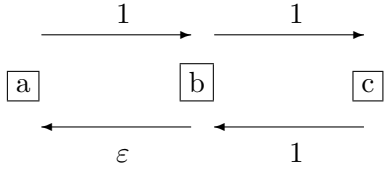
where the equality (Chapter 3 Lemma yyy) requires in discrete time that t be even. This yields the following one-sided inequalities, but Example 4.9 shows there can be no such reverse inequality.

Lemma 4.8 *(a)* $4||P_i(X_t \in \cdot) - \pi(\cdot)||^2 \le \frac{p_{ii}(2t)}{\pi_i} - 1, \ t \ge 0$.
(b) $d(t) \le \frac{1}{2}\sqrt{\hat{d}(2t)}, \ t \ge 0$

Proof. Part (b) follows from part (a) and the definitions. Part (a) is essentially just the "$||\ \ ||_1 \le ||\ \ ||_2$" inequality, but let's write it out bare-hands.

$$\begin{aligned}
4||P_i(X_t \in \cdot) - \pi(\cdot)||^2 &= \left(\sum_j |p_{ij}(t) - \pi_j|\right)^2 \\
&= \left(\sum_j \pi_j^{1/2} \left|\frac{p_{ij}(t) - \pi_j}{\pi_j^{1/2}}\right|\right)^2 \\
&\le \sum_j \frac{(p_{ij}(t) - \pi_j)^2}{\pi_j} \text{ by Cauchy-Schwarz} \\
&= -1 + \sum_j \frac{p_{ij}^2(t)}{\pi_j} \\
&= -1 + \frac{p_{ii}(2t)}{\pi_i} \text{ by Chapter 3 Lemma yyy.}
\end{aligned}$$

Example 4.9 Consider a continuous-time 3-state chain with transition rates



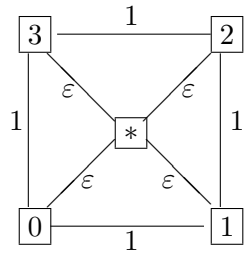
Here $\pi_a = \frac{\varepsilon}{2+\varepsilon},\ \pi_b = \pi_c = \frac{1}{2+\varepsilon}$. It is easy to check that τ_1 is bounded as $\varepsilon \to 0$. But $p_{aa}(t) \to e^{-t}$ as $\varepsilon \to 0$, and so by considering state a we have $\hat{d}(t) \to \infty$ as $\varepsilon \to 0$ for any fixed t.

Remark. In the nice examples discussed in Chapter 5 we can usually find a pair of states (i_0, j_0) such that

$$\bar{d}(t) = ||P_{i_0}(X_t \in \cdot) - P_{j_0}(X_t \in \cdot)|| \text{ for all } t.$$

The next example shows this is false in general.

Example 4.10 Consider random walk on the weighted graph



for suitably small ε. As $t \to 0$ we have $1 - \bar{d}(t) \sim c_\varepsilon t^2$, the *max* attained by pairs $(0,2)$ or $(1,3)$. But as $t \to \infty$ we have $\bar{d}(t) \sim a_\varepsilon \exp(-t/\tau_2(\varepsilon))$ where $\tau_2(\varepsilon) = \Theta(1/\varepsilon)$ and where the *max* is attained by pairs $(i, *)$. □

As a final comment, one might wonder whether the minimizing distribution μ in the definition of $\tau_1^{(3)}$ were always π, i.e. whether $\tau_1^{(3)} = \tau_1^{(2)}$ always. But a counter-example is provided by random walk on the n-star (Chapter 5 yyy) where $\tau_1^{(3)} = 1$ (by taking μ to be concentrated on the center vertex) but $\tau_1^{(2)} \to 3/2$.

4.3.2 Proof of Theorem 6

We will prove

Lemma 4.11 $\tau_1 \le \tau_1^{(1)} \le 4\tau_1$

Lemma 4.12 $\tau_1^{(3)} \le \tau_1^{(2)} \le \frac{e}{e-1}\tau_1^{(1)}$.

Lemma 4.13 $\tau_1^{(4)} \le 4\tau_1^{(3)}$

Lemma 4.14 $\tau_1^{(5)} \le \tau_1^{(4)}$

These lemmas hold in discrete and continuous time, interpreting $tau_1, \tau_1^{(1)}$ as $\tau_1^{\text{disc}}, \tau_1^{1,\text{disc}}$ is discrete time. Incidentally, Lemmas 4.12, 4.13 and 4.14 do not depend on reversibility. To complete the proof of Theorem 4.6 in continuous time we would need to show

$$\tau_1 \le K\tau_1^{(5)} \text{ in continuous time} \tag{4.15}$$

for some absolute constant K. The proof I know is too lengthy to repeat here – see [6]. Note that (from its definition) $\tau_1^{(2)} \le \tau_0$, so that (4.15) and the lemmas above imply $\tau_1 \le 2K\tau_0$ in continuous time. We shall instead give a direct proof of a result weaker than (4.15):

Lemma 4.15 $\tau_1 \le 66\tau_0$.

Turning to the assertions of Theorem 4.6 is discrete time, (b) is given by the discrete-time versions of Lemmas 4.11 and 4.12. To prove (a), it is enough to show that the numerical values of the parameters $\tau_1^{(2)} - -\tau_1^{(5)}$ are unchanged by continuizing the discrete-time chain. For $\tau_1^{(5)}$ and $\tau_1^{(4)}$ this is clear, because continuization doesn't affect mean hitting times. For $\tau_1^{(3)}$ and $\tau_1^{(2)}$ it reduces to the following lemma.

Lemma 4.16 *Let X_t be a discrete-time chain and Y_t be its continuization, both started with the same distribution. Let T be a randomized stopping time for Y. Then there exists a randomized stopping time $\hat{T}$ for X such that $P(X(\hat{T}) \in \cdot) = P(Y(T) \in \cdot)$ and $E\hat{T} = ET$.*

Proof of Lemma 4.11. The left inequality is immediate from Lemma 4.7(a), and the right inequality holds because

$$\begin{aligned} s(4\tau_1) &\le 1 - (1 - \bar{d}(2\tau_1))^2 \text{ by Lemma 4.7}(b) \\ &\le 1 - (1 - e^{-2})^2 \text{ by Lemma 4.5} \\ &\le e^{-1}. \end{aligned}$$

Proof of Lemma 4.12. The left inequality is immediate from the definitions. For the right inequality, fix i. Write $u = \tau_1^{(1)}$, so that

$$p_{jk}(u) \ge (1 - e^{-1})\pi_k \text{ for all } j, k.$$

We can construct a stopping time $U_i \in \{u, 2u, 3u, \ldots\}$ such that

$$P_i(X_{U_i} \in \cdot, U_i = u) = (1 - e^{-1})\pi(\cdot)$$

and then by induction on m such that

$$P_i(X_{U_i} \in \cdot, U_i = mu) = e^{-(m-1)}(1 - e^{-1})\pi(\cdot), \ m \ge 1.$$

Then $P_i(X_{U_i} \in \cdot) = \pi(\cdot)$ and $E_i U_i = u(1-e^{-1})^{-1}$. So $\tau_1^{(2)} \le (1-e^{-1})^{-1}\tau_1^{(1)}$.

Remark. What the argument shows is that we can construct a strong stationary time V_i (in the sense of (4.13)) such that

$$E_i V_i = (1 - e^{-1})^{-1}\tau_1^{(1)}. \tag{4.16}$$

Proof of Lemma 4.13. Consider the probability distribution μ attaining the *min* in the definition of $\tau_1^{(3)}$, and the associated stopping times U_i. Fix i. Since $P_i(X(U_i) \in \cdot) = \mu(\cdot)$,

$$E_i T_j \le E_i U_i + E_\mu T_j \le \tau_1^{(3)} + E_\mu T_j.$$

The random target lemma (4.7) says $\sum_j E_i T_j \pi_j = \sum_j E_\mu T_j \pi_j$ and so

$$\sum_j \pi_j |E_i T_j - E_\mu T_j| = 2 \sum_j \pi_j (E_i T_j - E_\mu T_j)^+ \le 2\tau_1^{(3)}.$$

Writing $b(i)$ for the left sum, the definition of $\tau_1^{(4)}$ and the triangle inequality give $\tau_1^{(4)} \le \max_{i,k}(b(i) + b(k))$, and the Lemma follows.

Proof of Lemma 4.14. Fix a subset A and a starting state $i \notin A$. Then for any $j \in A$,

$$E_i T_j = E_i T_A + E_\rho T_j$$

where ρ is the hitting place distribution $P_i(X_{T_A} \in \cdot)$. So

$$\pi(A) E_i T_A = \sum_{j \in A} \pi_j E_i T_A = \sum_{j \in A} \pi_j (E_i T_j - E_\rho T_j)$$

$$\le \max_k \sum_{j \in A} \pi_j (E_i T_j - E_k T_j) \le \tau_1^{(4)}.$$

Proof of Lemma 4.15. For small $\delta > 0$ to be specified later, define

$$A = \{j : E_\pi T_j \le \tau_0/\delta\}.$$

Note that Markov's inequality and the definition of τ_0 give

$$\pi(A^c) = \pi\{j : E_\pi T_j > \tau_0/\delta\} \le \frac{\sum_j \pi_j E_\pi T_j}{\tau_0/\delta} = \frac{\tau_0}{\tau_0/\delta} = \delta. \tag{4.17}$$

Next, for any j

$$\begin{aligned} E_\pi T_j &= \int_0^\infty \left(\frac{p_{jj}(s)}{\pi_j} - 1 \right) ds \text{ by Chapter 2 Lemma yyy} \\ &\ge t \left(\frac{p_{jj}(t)}{\pi_j} - 1 \right) \text{ for any } t \end{aligned}$$

by monotonicity of $p_{jj}(t)$. Thus for $j \in A$ we have

$$\frac{p_{jj}(t)}{\pi_j} - 1 \le \frac{E_\pi T_j}{t} \le \frac{\tau_0}{\delta t}$$

and applying Chapter 3 Lemma yyy (b)

$$\frac{p_{jk}(t)}{\pi_k} \ge 1 - \frac{\tau_0}{\delta t}, \ j, k \in A.$$

Now let i be arbitrary and let $k \in A$. For any $0 \le s \le u$,

$$\frac{P_i(X_{u+t} = k | T_A = s)}{\pi_k} \ge \min_{j \in A} \frac{P_j(X_{u+t-s} = k)}{\pi_k} \ge 1 - \frac{\tau_0}{\delta(u + t - s)} \ge 1 - \frac{\tau_0}{\delta t}$$

and so

$$\frac{p_{ik}(u+t)}{\pi_k} \geq \left(1 - \frac{\tau_0}{\delta t}\right)^+ P_i(T_A \leq u). \tag{4.18}$$

Now

$$P_i(T_A > u) \leq \frac{E_i T_A}{u} \leq \frac{\tau_1^{(5)}}{u\pi(A)}.$$

using Markov's inequality and the definition of $\tau_1^{(5)}$. And $\tau_1^{(5)} \leq \tau_1^{(4)} \leq 2\tau_0$, the first inequality being Lemma 4.14 and the second being an easy consequence of the definitions. Combining (4.18) and the subsequent inequalities shows that, for $k \in A$ and arbitrary i

$$\frac{p_{ik}(u+t)}{\pi_k} \geq \left(1 - \frac{\tau_0}{\delta t}\right)^+ \left(1 - \frac{2\tau_0}{u\pi(A)}\right)^+ \equiv \eta, \text{ say.}$$

Applying this to arbitrary i and j we get

$$\bar{d}(u+t) \leq 1 - \eta\pi(A) \leq 1 - \left(1 - \frac{\tau_0}{\delta t}\right)^+ \left(\pi(A) - \frac{2\tau_0}{u}\right)^+$$

$$\leq 1 - \left(1 - \frac{\tau_0}{\delta t}\right)^+ \left(1 - \delta - \frac{2\tau_0}{u}\right)^+ \text{ by (4.17).}$$

Putting $t = 49\tau_0, u = 17\tau_0, \delta = 1/7$ makes the bound $= \frac{305}{833} < e^{-1}$.

Remark. The ingredients of the proof above are complete monotonicity and conditioning on carefully chosen hitting times. The proof of (4.15) in [6] uses these ingredients, plus the minimal hitting time construction in the recurrent balayage theorem (Chapter 2 yyy).

Outline proof of Lemma 4.16. The observant reader will have noticed (Chapter 2 yyy) that we avoided writing down a careful definition of stopping times in the continuous setting. The definition involves measure-theoretic issues which I don't intend to engage, and giving a rigorous proof of the lemma is a challenging exercise in the measure-theoretic formulation of continuous-time chains. However, the underlying idea is very simple. Regard the chain Y_t as constructed from the chain $(X_0, X_1, X_2, \ldots)$ and exponential(1) holds (ξ_i). Define $\hat{T} = N(T)$, where $N(t)$ is the Poisson counting process $N(t) = \max\{m : \xi_1 + \ldots + \xi_m \leq t\}$. Then $X(\hat{T}) = Y(T)$ by construction and $E\hat{T} = ET$ by the optional sampling theorem for the martingale $N(t) - t$. □

4.3.3 τ_1 in discrete time, and algorithmic issues

Of course for period-2 chains we don't have convergence to stationarity in discrete time, so we regard $\tau_1^{\text{disc}} = \tau_1^{1,\text{disc}} = \infty$. Such chains – random

walks on bipartite weighted graphs – include several simple examples of unweighted graphs we will discuss in Chapter 5 (e.g. the n-path and n-cycle for even n, and the d-cube) and Chapter 7 (e.g. card-shuffling by random transpositions, if we insist on transposing distinct cards).

As mentioned in Chapter 1 xxx, a topic of much recent interest has been "Markov Chain Monte Carlo", where one constructs a discrete-time reversible chain with specified stationary distribution π and we wish to use the chain to sample from π. We defer systematic discussion to xxx, but a few comments are appropriate here. We have to start a simulation somewhere. In practice one might use as initial distribution some distribution which is feasible to simulate and which looks intuitively "close" to π, but this idea is hard to formalize and so in theoretical analysis we seek results which hold regardless of the initial distribution, i.e. "worst-case start" results. In this setting $\tau_1^{(2)}$ is, by definition, the minimum expected time to generate a sample with distribution π. But the definition of $\tau_1^{(2)}$ merely says a stopping time exists, and doesn't tell us how to implement it algorithmically. For algorithmic purposes we want rules which don't involve detailed structure of the chain. The most natural idea – stopping at a deterministic time – requires one to worry unnecessarily about near-periodicity. One way to avoid this worry is to introduce holds into the discrete-time chain, i.e. simulate $(\mathbf{P}+I)/2$ instead of $\mathbf{P}$. As an alternative, the distribution of the continuized chain at time t can be obtained by simply running the discrete-time chain for a Poisson(t) number of steps. "In practice" there is little difference between these alternatives. But the continuization method, as well as being mathematically less artificial, allows us to avoid the occasional messiness of discrete-time theory (see e.g. Proposition 4.29 below). In this sense our use of τ_1 for discrete-time chains as the value for continuous-time chains is indeed sensible: it measures the accuracy of a natural algorithmic procedure applied to a discrete-time chain.

Returning to technical matters, the fact that a periodic (reversible, by our standing assumption) chain can only have period 2 suggests that the discrete-time periodicity effect could be eliminated by averaging over times t and $t+1$ only, as follows.

Open Problem 4.17 *Show there exist $\psi(x) \downarrow 0$ as $x \downarrow 0$ and $\phi(t) \sim t$ as $t \to \infty$ such that, for any discrete-time chain,*

$$\max_i \left|\left| \frac{P_i(X_t \in \cdot) + P_i(X_{t+1} \in \cdot)}{2} - \pi(\cdot) \right|\right| \le \psi(d(\phi(t))), \quad t = 0, 1, 2, \ldots$$

where $d(\cdot)$ refers to the continuized chain.

See the Notes for some comments on this problem.

If one does wish to study distributions of discrete-time chains at deterministic times, then in place of τ_2 one needs to use

$$\beta \equiv \max(|\lambda_m| : 2 \le m \le n) = \max(\lambda_2, -\lambda_n). \tag{4.19}$$

The spectral representation then implies

$$|P_i(X_t = i) - \pi_i| \le \beta^t, \ t = 0, 1, 2, \ldots. \tag{4.20}$$

4.3.4 τ_1 and mean hitting times

In general τ_1 may be much smaller than τ^* or τ_0. For instance, random walk on the complete graph has $\tau_0 \sim n$ while $\tau_1 \to 1$. So we cannot (without extra assumptions) hope to improve much on the following result.

Lemma 4.18 *For an n-state chain, in discrete or continuous time,*

$$\tau_0 \le n\tau_1^{(2)}$$

$$\tau^* \le \frac{2\tau_1^{(2)}}{\min_j \pi_j}.$$

Lemmas 4.24 and 4.25 later are essentially stronger, giving corresponding upper bounds in terms of τ_2 instead of τ_1. But a proof of Lemma 4.18 is interesting for comparison with the cat-and-mouse game below.

Proof of Lemma 4.18. By definition of $\tau_1^{(2)}$, for the chain started at i_0 we can find stopping times $U_1, U_2, \ldots$ such that

$$E(U_{s+1} - U_s | X_u, u \le U_s) \le \tau_1^{(2)}$$

$$(X(U_s); s \ge 1) \text{ are independent with distribution } \pi.$$

So $S_j \equiv \min\{s : X(U_s) = j\}$ has $E_{i_0} S_j = 1/\pi_j$, and so

$$E_{i_0} T_j \le E_{i_0} U_{S_j} \le \frac{\tau_1^{(2)}}{\pi_j}$$

where the second inequality is justified below. The second assertion of the lemma is now clear, and the first holds by averaging over j.

The second inequality is justified by the following martingale result, which is a simple application of the optional sampling theorem. The "equality" assertion is sometimes called *Wald's equation* for martingales.

Lemma 4.19 *Let $0 = Y_0 \le Y_1 \le Y_2 \dots$ be such that*

$$E(Y_{i+1} - Y_i | Y_j, j \le i) \le c, \ i \ge 0$$

for a constant c. Then for any stopping time T,

$$EY_T \le cET.$$

If in the hypothesis we replace "$\le c$" by "$= c$", then $EY_T = cET$.

Cat-and-Mouse Game. Here is another variation on the type of game described in Chapter 3 section yyy. Fix a graph. The cat starts at some vertex v_c and follows a continuous-time simple random walk. The mouse starts at some vertex v_m and is allowed an arbitrary strategy. Recall the mouse can't see the cat, so it must use a deterministic strategy, or a random strategy independent of the cat's moves. The mouse seeks to maximize EM, the time until meeting. Write m^* for the *sup* of EM over all starting positions v_c, v_m and all strategies for the mouse. So m^* just depends on the graph. Clearly $m^* \ge \max_{i,j} E_i T_j$, since the mouse can just stand still.

Open Problem 4.20 *Does $m^* = \max_{i,j} E_i T_j$? In other words, is it never better to run than to hide?*

Here's a much weaker upper bound on m^*. Consider for simplicity a regular n-vertex graph. Then

$$m^* \le \frac{en\tau_1^{(1)}}{e-1}. \tag{4.21}$$

Because as remarked at (4.16), we can construct a strong stationary time V such that $EV = \frac{e\tau_1^{(1)}}{e-1} = c$, say. So we can construct $0 = V_0 < V_1 < V_2 \dots$ such that

$$E(V_{i+1} - V_i | V_j, j \le i) \le c, \ i \ge 0$$

$$(X(V_i), i \ge 1) \text{ are independent with the uniform distribution } \pi$$

$$(X(V_i), i \ge 1) \text{ are independent of } (V_i, i \ge 1).$$

So regardless of the mouse's strategy, the cat has chance $1/n$ to meet the mouse at time V_i, independently as i varies, so the meeting time M satisfies $M \le V_T$ where T is a stopping time with mean n, and (4.21) follows from Lemma 4.19. This topic will be pursued in Chapter 6 yyy.

4.3.5 τ_1 and flows

Since discrete-time chains can be identified with random walks on weighted graphs, relating properties of the chain to properties of "flows" on the graph is a recurring theme. Thompson's principle (Chapter 3 yyy) identified mean commute times and mean hitting times from stationarity as *infs* over flows of certain quantities. Sinclair [308] noticed that τ_1 could be related to "multicommodity flow" issues, and we give a streamlined version of his result (essentially Corollary 4.22) here. Recall from Chapter 3 section yyy the general notation of a unit flow from a to π, and the special flow $\mathbf{f}^{a\to\pi}$ induced by the Markov chain.

Lemma 4.21 *Consider a family $\mathbf{f} = (\mathbf{f}^{(a)})$, where, for each state a, $\mathbf{f}^{(a)}$ is a unit flow from a to the stationary distribution π. Define*

$$\psi(\mathbf{f}) = \max_{edges\ (i,j)} \sum_a \pi_a \frac{|f_{ij}^{(a)}|}{\pi_i p_{ij}}$$

in discrete time, and substitute q_{ij} for p_{ij} in continuous time. Let $\mathbf{f}^{a\to\pi}$ be the special flow induced by the chain. Then

$$\psi(\mathbf{f}^{a\to\pi}) \le \tau_1^{(4)} \le \Delta\psi(\mathbf{f}^{a\to\pi})$$

where Δ is the diameter of the transition graph.

Proof. We work in discrete time (the continuous case is similar). By Chapter 3 yyy

$$\frac{f_{ij}^{a\to\pi}}{\pi_i p_{ij}} = \frac{Z_{ia} - Z_{ja}}{\pi_a}$$

and so

$$\sum_a \pi_a \frac{|f_{ij}^{a\to\pi}|}{\pi_i p_{ij}} = \sum_a |Z_{ia} - Z_{ja}|.$$

Thus

$$\psi(\mathbf{f}^{a\to\pi}) = \max_{\text{edge}(i,j)} \sum_a |Z_{ia} - Z_{ja}|.$$

The result now follows because by (4.12)

$$\tau_1^{(4)} = \max_{i,k} \sum_a |Z_{ia} - Z_{ka}|$$

where i and k are not required to be neighbors. □

Using Lemmas 4.11 - 4.13 to relate $\tau_1^{(4)}$ to τ_1, we can deduce a lower bound on τ_1 in terms of flows.

Corollary 4.22 $\tau_1 \geq \frac{e-1}{16e} \inf_{\mathbf{f}} \psi(\mathbf{f})$.

Unfortunately it seems hard to get analogous upper bounds. In particular, it is not true that

$$\tau_1 = O\left(\Delta \inf_{\mathbf{f}} \psi(\mathbf{f})\right).$$

To see why, consider first random walk on the n-cycle (Chapter 5 Example yyy). Here $\tau_1 = \Theta(n^2)$ and $\psi(\mathbf{f}^{a\to\pi}) = \Theta(n)$, so the upper bound in Lemma 4.21 is the right order of magnitude, since $\Delta = \Theta(n)$. Now modify the chain by allowing transitions between arbitrary pairs (i, j) with equal chance $o(n^{-3})$. The new chain will still have $\tau_1 = \Theta(n^2)$, and by considering the special flow in the original chain we have $\inf_{\mathbf{f}} \psi(\mathbf{f}) = O(n)$, but now the diameter $\Delta = 1$.

4.4 The relaxation time τ_2

The parameter τ_2 is the relaxation time, defined in terms of the eigenvalue λ_2 (Chapter 3 section yyy) as

$$\begin{aligned} \tau_2 &= (1-\lambda_2)^{-1} \text{ in discrete time} \\ &= \lambda_2^{-1} \text{ in continuous time.} \end{aligned}$$

In Chapter 3 yyy we proved a lower bound for an n-state discrete-time chain:

$$\tau_2 \geq 1 - \frac{1}{n}$$

which is attained by random walk on the complete graph. We saw in Chapter 3 Theorem yyy the extremal characterization

$$\tau_2 = \sup\{||g||_2^2/\mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0\}. \tag{4.22}$$

The next three lemmas give inequalities between τ_2 and the parameters studied earlier in this chapter. Write $\pi_* \equiv \min_i \pi_i$.

Lemma 4.23 *In continuous time,*

$$\tau_2 \leq \tau_1 \leq \tau_2 \left(1 + \frac{1}{2}\log\frac{1}{\pi_*}\right).$$

In discrete time,

$$\tau_1^{(2)} \leq \frac{4e}{e-1}\ \tau_2 \left(1 + \frac{1}{2}\log\frac{1}{\pi_*}\right).$$

Lemma 4.24 $\tau_2 \le \tau_0 \le (n-1)\tau_2$.

Lemma 4.25 $\tau^* \le \frac{2\tau_2}{\pi_*}$.

Proof of Lemma 4.23. Consider first the continuous time case. By the spectral representation, as $t \to \infty$ we have $p_{ii}(t) - \pi_i \sim c_i \exp(-t/\tau_2)$ with some $c_i \neq 0$. But by Lemma 4.5 we have $|p_{ii}(t) - \pi_i| = O(\exp(-t/\tau_1))$. This shows $\tau_2 \le \tau_1$. For the right inequality, the spectral representation gives

$$p_{ii}(t) - \pi_i \le e^{-t/\tau_2}. \tag{4.23}$$

Recalling the definition (4.14) of $\hat{d}$,

$$\begin{aligned} \bar{d}(t) &\le 2d(t) \\ &\le \sqrt{\hat{d}(2t)} \text{ by Lemma 4.8}(b) \\ &\le \sqrt{\max_i \frac{e^{-2t/\tau_2}}{\pi_i}} \text{ by (4.14) and (4.23)} \\ &= \pi_*^{-1/2} e^{-t/\tau_2} \end{aligned}$$

and the result follows. The upper bound on $\tau_1^{(2)}$ holds in continuous time by Lemmas 4.11 and 4.12, and so holds in discrete time because $\tau_1^{(2)}$ and τ_2 are unaffected by continuization.

Proof of Lemma 4.24. $\tau_2 \le \tau_0$ because τ_2 is the first term in the eigentime identity for τ_0. For the other bound, Chapter 3 Lemma yyy gives the inequality in

$$\tau_0 = \sum_j \pi_j E_\pi T_j \le \sum_j (1-\pi_j)\tau_2 = (n-1)\tau_2.$$

Proof of Lemma 4.25. Fix states a, b such that $E_a T_b + E_b T_a = \tau^*$ and fix a function $0 \le g \le 1$ attaining the *sup* in the extremal characterization (Chapter 3 Theorem yyy), so that

$$\tau^* = \frac{1}{\mathcal{E}(g,g)}, \quad g(a) = 0, g(b) = 1.$$

Write $c = \sum_i \pi_i g(i)$. Applying the extremal characterization of τ_2 to the centered function $g - c$,

$$\tau_2 \ge \frac{||g-c, g-c||_2^2}{\mathcal{E}(g-c, g-c)} = \frac{\text{var }_\pi g(X_0)}{\mathcal{E}(g,g)} = \tau^* \text{ var }_\pi g(X_0).$$

But

$$\begin{aligned}
\text{var}_{\ \pi} g(X_0) &\geq \pi_a c^2 + \pi_b (1-c)^2 \\
&\geq \inf_{0 \leq y \leq 1} (\pi_a y^2 + \pi_b (1-y)^2) \\
&= \frac{\pi_a \pi_b}{\pi_a + \pi_b} \\
&\geq \frac{1}{2} \min(\pi_a, \pi_b) \\
&\geq \pi_*/2
\end{aligned}$$

establishing the lemma. □

Simple examples show that the bounds in these Lemmas cannot be much improved in general. Specifically
(a) on the complete graph (Chapter 5 Example yyy) $\tau_2 = (n-1)\tau_0$ and $\tau_2^* = \frac{2\tau_2}{\pi_*}$.
(b) On the barbell (Chapter 5 Example yyy), τ_2, τ_1 and τ_0 are asymptotic to each other.
(c) In the $M/M/1/n$ queue, $\tau_1/\tau_2 = \Theta(\log 1/\pi_*)$ as $n \to \infty$. □

In the context of Lemma 4.23, if we want to relate τ_1^{disc} itself to eigenvalues in discrete time we need to take almost-periodicity into account and use $\beta = \max(\lambda_2, -\lambda_n)$ in place of τ_2. Rephrasing the proof of Lemma 4.23 gives

Lemma 4.26 *In discrete time,*

$$\lceil \frac{1}{\log 1/\beta} \rceil \leq \tau_1^{disc} \leq \lceil \frac{1 + \frac{1}{2} \log 1/\pi_*}{\log 1/\beta} \rceil$$

Regarding a discrete-time chain as random walk on a weighted graph, let Δ be the diameter of the graph. By considering the definition of the variation distance $\bar{d}(t)$ and initial vertices i, j at distance Δ, it is obvious that $\bar{d}(t) = 1$ for $t < \Delta/2$, and hence $\tau_1^{\text{disc}} \geq \lceil \Delta/2 \rceil$. Combining with the upper bound in Lemma 4.26 leads to a relationship between the diameter and the eigenvalues of a weighted graph.

Corollary 4.27

$$\log \frac{1}{\beta} \leq \frac{2 + \log(1/\pi_*)}{\Delta}.$$

This topic will be discussed further in Chapter yyy.

4.4.1 Correlations and variances for the stationary chain

Perhaps the most natural probabilistic interpretation of τ_2 is as follows. Recall that the *correlation* between random variables Y, Z is

$$\mathrm{cor}(Y,Z) \equiv \frac{E(YZ) - (EY)(EZ)}{\sqrt{\mathrm{var}\ Y\ \mathrm{var}\ Z}}.$$

For a stationary Markov chain define the *maximal correlation* function

$$\rho(t) \equiv \max_{h,g} \mathrm{cor}(h(X_0), g(X_t))$$

This makes sense for general chains (see Notes for further comments), but under our standing assumption of reversibility we have

Lemma 4.28 *In continuous time,*

$$\rho(t) = \exp(-t/\tau_2),\ t \geq 0.$$

In discrete time,

$$\rho(t) = \beta^t,\ t \geq 0$$

where $\beta = \max(\lambda_2, -\lambda_n)$.

This is a translation of the Rayleigh-Ritz characterization of eigenvalues (Chapter 3 yyy) – we leave the details to the reader.

Now consider a function g with $E_\pi g(X_0) = 0$ and $||g||_2^2 \equiv E_\pi g^2(X_0) > 0$. Write

$$\begin{aligned} S_t \equiv &\quad \textstyle\int_0^t g(X_s)ds \quad \text{in continuous time} \\ S_t \equiv &\quad \textstyle\sum_{s=0}^{t-1} g(X_s) \quad \text{in discrete time.} \end{aligned}$$

Recall from Chapter 2 yyy that for general chains there is a limit variance $\sigma^2 = \lim_{t\to\infty} t^{-1}\mathrm{var}\ S_t$. Reversibility gives extra qualitative and quantitative information. The first result refers to the stationary chain.

Proposition 4.29 *In continuous time,* $t^{-1}\mathrm{var}\ {}_\pi S_t \uparrow \sigma^2$, *where*

$$0 < \sigma^2 \leq 2\tau_2 ||g||_2^2.$$

And $A(t/\tau_2)t\sigma^2 \leq \mathrm{var}\ {}_\pi S_t \leq t\sigma^2$, *where*

$$A(u) \equiv \int_0^u \left(1 - \frac{s}{u}\right) e^{-s} ds = 1 - u^{-1}(1 - e^{-u}) \uparrow 1 \ \textit{as}\ u \uparrow \infty.$$

In discrete time,

$$t^{-1}\text{var }_\pi S_t \to \sigma^2 \le 2\tau_2 ||g||_2^2$$

$$\sigma^2 t\left(1 - \frac{2\tau_2}{t}\right) \le \text{var }_\pi S_t \le \sigma^2 t + ||g||_2^2$$

and so in particular

$$\text{var }_\pi S_t \le t||g||_2^2(2\tau_2 + \frac{1}{t}). \tag{4.24}$$

Proof. Consider first the continuous time case. A brief calculation using the spectral representation (Chapter 3 yyy) gives

$$E_\pi g(X_0)g(X_t) = \sum_{m\ge 2} g_m^2 e^{-\lambda_m t} \tag{4.25}$$

where $g_m = \sum_i \pi_i^{1/2} u_{im} g(i)$. So

$$\begin{aligned} t^{-1}\text{var }_\pi S_t &= t^{-1}\int_0^t \int_0^t E_\pi g(X_u)g(X_s)\, dsdu \\ &= 2t^{-1}\int_0^t (t-s)E_\pi g(X_0)g(X_s)ds \\ &= 2\int_0^t \left(1 - \frac{s}{t}\right)\sum_{m\ge 2} g_m^2 e^{-\lambda_m s}\, ds \qquad (4.26) \\ &= 2\sum_{m\ge 2}\frac{g_m^2}{\lambda_m} A(\lambda_m t) \qquad (4.27) \end{aligned}$$

by change of variables in the integral defining $A(u)$. The right side increases with t to

$$\sigma^2 \equiv 2\sum_{m\ge 2} g_m^2/\lambda_m, \tag{4.28}$$

and the sum here is at most $\sum_{m\ge 2} g_m^2/\lambda_2 = ||g||_2^2\tau_2$. On the other hand, $A(\cdot)$ is increasing, so

$$t^{-1}\text{var }_\pi(S_t) \ge 2\sum_{m\ge 2}\frac{g_m^2}{\lambda_m} A(\lambda_2 t) = \sigma^2 A(t/\tau_2).$$

In discrete time the arguments are messier, and we will omit details of calculations. The analog of (4.26) becomes

$$t^{-1}\text{var }_\pi S_t = \sum_{s=-(t-1)}^{t-1}\left(1 - \frac{|i|}{t}\right)\sum_{m\ge 2} g_m^2 \lambda_m^s.$$

In place of the change of variables argument for (4.27), one needs an elementary calculation to get

$$t^{-1}\text{var }_\pi S_t = 2\sum_{m\geq 2}\frac{g_m^2}{1-\lambda_m}\ B(\lambda_m,t) \tag{4.29}$$

$$\text{where } B(\lambda,t) = \frac{1+\lambda}{2} - \frac{\lambda(1-\lambda^t)}{t(1-\lambda)}.$$

This shows

$$t^{-1}\text{var }_\pi S_t \to \sigma^2 \equiv \sum_{m\geq 2} g_m^2 \frac{1+\lambda_m}{1-\lambda_m}$$

and the sum is bounded above by

$$\frac{1+\lambda_2}{1-\lambda_2}\sum_{m\geq 2} g_m^2 = \frac{1+\lambda_2}{1-\lambda_2}||g||_2^2 \leq 2\tau_2||g||_2^2.$$

Next, rewrite (4.29) as

$$\text{var }_\pi S_t - \sigma^2 t = -2\sum_{m\geq 2} g_m^2 \frac{\lambda_m(1-\lambda_m^t)}{(1-\lambda_m)^2}.$$

Then the upper bound for var $_\pi S_t$ follows by checking

$$\inf_{-1\leq\lambda<1} \frac{\lambda(1-\lambda^t)}{(1-\lambda)^2} \geq -\frac{1}{2}.$$

For the lower bound, one has to verify

$$\sup_{-1\leq\lambda\leq\lambda_2} \frac{2\lambda(1-\lambda^t)}{(1-\lambda)(1+\lambda)} \text{ is attained at } \lambda_2 \text{ (and equals } C\text{, say)}$$

where in the sequel we may assume $\lambda_2 > 0$. Then

$$B(\lambda_m,t) \geq \frac{1+\lambda_m}{2}(1-C/t),\ m\geq 2$$

and so

$$t^{-1}\text{var }_\pi S_t \geq \sigma^2\ (1-C/t).$$

But

$$\frac{C}{t} = \frac{2\lambda_2(1-\lambda_2^t)}{t(1-\lambda_2)(1+\lambda_2)} \leq \frac{2}{t(1-\lambda_2)} = 2\tau_2/t$$

giving the lower bound. □

Note that even in discrete time it is τ_2 that matters in Proposition 4.29. Eigenvalues near -1 are irrelevant, except that for a periodic chain we have $\sigma = 0$ for one particular function g (which?).

Continuing the study of $S_t \equiv \int_0^t g(X_s)ds$ or its discrete analog for a stationary chain, standardize to the case where $E_\pi g(X_0) = 0, E_\pi g^2(X_0) = 1$. Proposition 4.29 provides finite-time bounds for the asymptotic approximation of variance. One would like a similar finite-time bound for the asymptotic Normal approximation of the distribution of S_t.

Open Problem 4.30 Is there some explicit function $\psi(b, s) \to 0$ as $s \to \infty$, not depending on the chain, such that for standardized g and continuous-time chains,

$$\sup_x \left| P_\pi \left(\frac{S_t}{\sigma t^{1/2}} \le x \right) - P(Z \le x) \right| \le \psi(||g||_\infty, t/\tau_2)$$

where $||g||_\infty \equiv \max_i |g(i)|$ and Z has Normal$(0, 1)$ distribution?

See the Notes for further comments. For the analogous result about large deviations see Chapter yyy.

4.4.2 Algorithmic issues

Suppose we want to estimate the average $\bar{g} \equiv \sum_i \pi_i g(i)$ of a function g defined on state space. If we could sample i.i.d. from π we would need order ε^{-2} samples to get an estimator with error about $\varepsilon\sqrt{\text{var}_\pi g}$. Now consider the setting where we cannot directly sample from π but instead use the "Markov Chain Monte Carlo" method of setting up a reversible chain with stationary distribution π. How many steps of the chain do we need to get the same accuracy? As in section 4.3.3, because we typically can't quantify the closeness to π of a feasible initial distribution, we consider bounds which hold for arbitrary initial states. In assessing the number of steps required, there are two opposite traps to avoid. The first is to say (cf. Proposition 4.29) that $\varepsilon^{-2}\tau_2$ steps suffice. This is wrong because the relaxation time bounds apply to the stationary chain and cannot be directly applied to a non-stationary chain. The second trap is to say that because it take $\Theta(\tau_1)$ steps to obtain one sample from the stationary distribution, we therefore need order $\varepsilon^{-2}\tau_1$ steps in order to get ε^{-2} independent samples. This is wrong because we don't need independent samples. The correct answer is order $(\tau_1 + \varepsilon^{-2}\tau_2)$ steps. The conceptual idea (cf. the definition of $\tau_1^{(2)}$) is to find a stopping time achieving distribution π and use it as an initial

state for simulating the stationary chain. More feasible to implement is the following algorithm.

Algorithm. For a specified real number $t_1 > 0$ and an integer $m_2 \geq 1$, generate $M(t_1)$ with Poisson(t_1) distribution. Simulate the chain X_t from arbitrary initial distribution for $M(t_1) + m_2 - 1$ steps and calculate

$$A(g, t_1, m_2) \equiv \frac{1}{m_2} \sum_{t=M(t_1)}^{M(t_1)+m_2-1} g(X_t).$$

Corollary 4.31

$$P\left(|A(g, t_1, m_2) - \bar{g}| > \varepsilon ||g||_2\right) \leq s(t_1) + \frac{2\tau_2 + 1/m_2}{\varepsilon^2 m_2}$$

where $s(t)$ is separation (recall section 4.3.1) for the continuized chain.

To make the right side approximately δ we may take

$$t_1 = \tau_1^{(1)} \lceil \log(2/\delta) \rceil; \quad m_2 = \lceil \frac{4\tau_2}{\varepsilon^2 \delta} \rceil.$$

Since the mean number of steps is $t_1 + m_2 - 1$, this formalizes the idea that we can estimate $\bar{g}$ to within $\varepsilon ||g||_2$ in order $(\tau_1 + \varepsilon^{-2}\tau_2)$ steps.

xxx if don't know tau's

Proof. We may suppose $\bar{g} = 0$. Since $X_{M(t_1)}$ has the distribution of the continuized chain at time t_1, we may use the definition of $s(t_1)$ to write

$$P(X_{M(t_1)} \in \cdot) = (1 - s(t_1))\pi \ + s(t_1)\rho$$

for some probability distribution ρ. It follows that

$$P\left(|A(g, t_1, m_2)| > \varepsilon ||g||_2\right) \leq s(t_1) + P_\pi \left(\left| \frac{1}{m_2} \sum_{t=0}^{m_2-1} g(X_t) \right| > \varepsilon ||g||_2 \right)$$

$$\leq s(t_1) + \frac{1}{m_2^2 \varepsilon^2 ||g||_2^2} \text{ var }_\pi \left(\sum_{t=0}^{m_2-1} g(X_t) \right).$$

Apply (4.24).

4.4.3 τ_2 and distinguished paths

The extremal characterization (4.22) can be used to get lower bounds on τ_2 by considering a tractable test function g. (xxx list examples). As mentioned in Chapter 3, it is an open problem to give an extremal characterization of τ_2 as exactly an *inf* over flows or similar constructs. As an alternative, Theorem 4.32 gives a non-exact upper bound on τ_2 involving quantities derived from arbitrary choices of paths between states. An elegant exposition of this idea, expressed by the first inequality in Theorem 4.32, was given by Diaconis and Stroock [124], and Sinclair [308] noted the second inequality. We copy their proofs.

We first state the result in the setting of random walk on a weighted graph. As in section 4.1, consider a path $x = i_0, i_1, \ldots, i_m = y$, and call this path γ_{xy}. This path has length $|\gamma_{xy}| = m$ and has "resistance"

$$r(\gamma_{xy}) \equiv \sum_{e \in \gamma_{xy}} 1/w_e$$

where here and below e denotes a directed edge.

Theorem 4.32 *For each ordered pair (x, y) of vertices in a weighted graph, let γ_{xy} be a path from x to y. Then for discrete-time random walk,*

$$\tau_2 \le w \max_e \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) 1_{(e \in \gamma_{xy})}$$

$$\tau_2 \le w \max_e \frac{1}{w_e} \sum_x \sum_y \pi_x \pi_y |\gamma_{xy}| 1_{(e \in \gamma_{xy})}.$$

Note that the two inequalities coincide on an *unweighted* graph.

Proof. For an edge $e = (i, j)$ write $\Delta g(e) = g(j) - g(i)$. The first equality below is the fact 2 var $(Y_1) = E(Y_1 - Y_2)^2$ for i.i.d. Y's, and the first inequality is Cauchy-Schwarz.

$$\begin{aligned} 2||g||_2^2 &= \sum_x \sum_y \pi_x \pi_y (g(y) - g(x))^2 \\ &= \sum_x \sum_y \pi_x \pi_y \left(\sum_{e \in \gamma_{xy}} \Delta g(e) \right)^2 \\ &= \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) \left(\sum_{e \in \gamma_{xy}} \frac{1}{\sqrt{w_e r(\gamma_{xy})}} \sqrt{w_e} \Delta g(e) \right)^2 \end{aligned} \quad (4.30)$$

$$\begin{aligned} &\leq \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) \sum_{e \in \gamma_{xy}} w_e (\Delta g(e))^2 && (4.31) \\ &= \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) \sum_e w_e (\Delta g(e))^2 1_{(e \in \gamma_{xy})} \\ &\leq \kappa \sum_e w_e (\Delta g(e))^2 = \kappa \; 2w\mathcal{E}(g,g) && (4.32) \end{aligned}$$

where κ is the *max* in the first inequality in the statement of the Theorem. The first inequality now follows from the extremal characterization (4.22). The second inequality makes a simpler use of the Cauchy-Schwarz inequality, in which we replace (4.30,4.31,4.32) by

$$\begin{aligned} &= \sum_x \sum_y \pi_x \pi_y \left(\sum_{e \in \gamma_{xy}} 1 \cdot \Delta g(e) \right)^2 \\ &\leq \sum_x \sum_y \pi_x \pi_y |\gamma_{xy}| \sum_{e \in \gamma_{xy}} (\Delta g(e))^2 && (4.33) \\ &\leq \kappa' \sum_e w_e (\Delta g(e))^2 = \kappa' \; 2w\mathcal{E}(g,g) \end{aligned}$$

where κ' is the *max* in the second inequality in the statement of the Theorem.

Remarks. (a) Theorem 4.32 applies to continuous-time (reversible) chains by setting $w_{ij} = \pi_i q_{ij}$.

(b) One can replace the deterministic choice of paths γ_{xy} by random paths Γ_{xy} of the form $x = V_0, V_1, \ldots, V_M = y$ of random length $M = |\Gamma_{xy}|$. The second inequality extends in the natural way, by taking expectations in (4.33) to give

$$\leq \sum_x \sum_y \pi_x \pi_y \; E \left(\; |\Gamma_{xy}| 1_{(e \in \Gamma_{xy})} \; \sum_e (\Delta g(e))^2 \right),$$

and the conclusion is

Corollary 4.33

$$\tau_2 \leq w \max_e \frac{1}{w_e} \sum_x \sum_y \pi_x \pi_y \; E \; |\Gamma_{xy}| 1_{(e \in \Gamma_{xy})}.$$

(c) Inequalities in the style of Theorem 4.32 are often called *Poincaré inequalities* because, to quote [124], they are "the discrete analog of the classical method of Poincaré for estimating the spectral gap of the Laplacian on a domain (see e.g. Bandle [39])". I prefer the descriptive name *the*

distinguished path method. This method has the same spirit as the *coupling method* for bounding τ_1 (see Chapter yyy), in that we get to use our skill and judgement in making wise choices of paths in specific examples. xxx list examples. Though its main utility is in studying hard examples, we give some simple illustrations of its use below.

Write the conclusion of Corollary 4.33 as $\tau_2 \le w \max_e \frac{1}{w_e} F(e)$. Consider a regular unweighted graph, and let $\Gamma_{x,y}$ be chosen uniformly from the set of minimum-length paths from x to y. Suppose that $F(e)$ takes the same value F for every directed edge e. A sufficient condition for this is that the graph be *arc-transitive* (see Chapter 8 yyy). Then, summing over edges in Corollary 4.33,

$$\tau_2 |\vec{\mathcal{E}}| \le w \sum_e \sum_x \sum_y \pi_x \pi_y E|\Gamma_{xy}| 1_{(e \in \Gamma_{xy})} = w \sum_x \sum_y \pi_x \pi_y E|\Gamma_{xy}|^2$$

where $|\vec{\mathcal{E}}|$ is the number of directed edges. Now $w = |\vec{\mathcal{E}}|$, so we may reinterpret this inequality as follows.

Corollary 4.34 *For random walk on an arc-transitive graph, $\tau_2 \le ED^2$, where $D = d(\xi_1, \xi_2)$ is the distance between independent uniform random vertices ξ_1, ξ_2.*

In the context of the d-dimensional torus Z_N^d, the upper bound is asymptotic (as $N \to \infty$) to $N^2 E \left(\sum_{i=1}^d U_i\right)^2$ where the U_i are independent uniform $[0, 1/2]$, This bound is asymptotic to $d(d + 1/3)N^2/16$. Here (Chapter 5 Example yyy) in fact $\tau_2 \sim dN^2/(2\pi^2)$, so for fixed d the bound is off by only a constant. On the d-cube (Chapter 5 Example yyy), D has Binomial$(d, 1/2)$ distribution and so the bound is $ED^2 = (d^2 + d)/4$, while in fact $\tau_2 = d/2$.

Intuitively one feels that the bound in Corollary 4.34 should hold for more general graphs, but the following example illustrates a difficulty.

Example 4.35 Consider the graph on $n = 2m$ vertices obtained from two complete graphs on m vertices by adding m edges comprising a matching of the two vertex-sets.

Here a straightforward implementation of Theorem 4.32 gives an upper bound of $2m$, while in fact $\tau_2 = m/2$. On the other hand the conclusion of Corollary 4.34 would give an $O(1)$ bound. Thus even though this example has a strong symmetry property (*vertex-transitivity*: Chapter 8 yyy) no bound like Corollary 4.34 can hold.

4.5 The flow parameter τ_c

In this section it's convenient to work in continuous time, but the numerical quantities involved here are unchanged by continuization.

4.5.1 Definition and easy inequalities

Define

$$\tau_c = \sup_A \frac{\pi(A)\pi(A^c)}{Q(A, A^c)} \tag{4.34}$$

where

$$Q(A, A^c) \equiv \sum_{i \in A} \sum_{j \in A^c} \pi_i q_{ij}$$

and where such *sups* are always over proper subsets A of states. This parameter can be calculated exactly in only very special cases, where the following lemma is helpful.

Lemma 4.36 *The* sup *in (4.34) is attained by some split $\{A, A^c\}$ in which both A and A^c are connected (as subsets of the graph of permissible transitions).*

Proof. Consider a split $\{A, A^c\}$ in which A is the union of $m \geq 2$ connected components (B_i). Write $\gamma = \min_i \frac{Q(B_i, B_i^c)}{\pi(B_i)\pi(B_i^c)}$. Then

$$\begin{aligned} Q(A, A^c) &= \sum_i Q(B_i, B_i^c) \\ &\geq \gamma \sum_i \pi(B_i)\pi(B_i^c) \\ &= \gamma \sum_i (\pi(B_i) - \pi^2(B_i)) \\ &= \gamma \left(\pi(A) - \sum_i \pi^2(B_i) \right) \end{aligned}$$

and so

$$\frac{Q(A, A^c)}{\pi(A)\pi(A^c)} \geq \gamma \, \frac{\pi(A) - \sum_i \pi^2(B_i)}{\pi(A) - \pi^2(A)}.$$

But for $m \geq 2$ we have $\sum_i \pi^2(B_i) \leq \left(\sum_i \pi(B_i)\right)^2 = \pi^2(A)$, which implies $\frac{Q(A,A^c)}{\pi(A)\pi(A^c)} > \gamma$. □

To see how τ_c arises, note that the extremal characterization of τ_2 (4.22) applied to $g = 1_A$ implies

$$\frac{\pi(A)\pi(A^c)}{Q(A, A^c)} \le \tau_2$$

for any subset A. But much more is true: Chapter 3 yyy may be rephrased as follows. For any subset A,

$$\frac{\pi(A)\pi(A^c)}{Q(A, A^c)} \le \frac{\pi(A)E_\pi T_A}{\pi(A^c)} \le \pi(A)E_{\alpha_A}T_A \le \tau_2$$

where α_A is the quasistationary distribution on A^c defined at Chapter 3 yyy. So taking *sups* gives

Corollary 4.37

$$\tau_c \le \sup_A \frac{\pi(A)E_\pi T_A}{\pi(A^c)} \le \sup_A \pi(A)E_{\alpha_A}T_A \le \tau_2.$$

In a two-state chain these inequalities all become equalities. This seems a good justification for our choice of definition of τ_c, instead of the alternative

$$\sup_{A:\pi(A)\le 1/2} \frac{\pi(A)}{Q(A, A^c)}$$

which has been used in the literature but which would introduce a spurious factor of 2 into the inequality $\tau_c \le \tau_2$.

Lemma 4.39 below shows that the final inequality of Corollary 4.37 can be reversed. In contrast, on the n-cycle $\tau_c = \Theta(n)$ whereas the other quantities in Corollary 4.37 are $\Theta(n^2)$. This shows that the "square" in Theorem 4.40 below cannot be omitted in general. It also suggests the following question (cf. τ_1 and $\tau_1^{(5)}$)

Open Problem 4.38 *Does there exist a constant K such that*

$$\tau_2 \le K \sup_A \frac{\pi(A)E_\pi T_A}{\pi(A^c)}$$

for every reversible chain?

A positive answer would provide, via Chapter 3 yyy, a correct order-of-magnitude extremal characterization of τ_2 in terms of flows.

Lemma 4.39

$$\tau_2 \le \sup_{A:\pi(A)\ge 1/2} E_{\alpha_A} T_A$$

and so in particular

$$\tau_2 \le 2 \sup_A \pi(A) E_{\alpha_A} T_A.$$

Proof. $\tau_2 = ||h||_2^2/\mathcal{E}(h,h)$ for the eigenvector h associated with λ_2. Put

$$A = \{x : h(x) \le 0\}$$

and assume $\pi(A) \ge 1/2$, by replacing h by $-h$ if necessary. Write $h_+ = \max(h, 0)$. We shall show

$$\tau_2 \le ||h_+||_2^2/\mathcal{E}(h_+, h_+) \tag{4.35}$$

and then the extremal characterization Chapter 3 yyy

$$E_{\alpha_A} T_A = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \ge 0, g = 0 \text{ on } A\} \tag{4.36}$$

implies $\tau_2 \le E_{\alpha_A} T_A$ for this specific A.

The proof of (4.35) requires us to delve slightly further into the calculus of Dirichlet forms. Write $\mathbf{P}_t f$ for the function $(\mathbf{P}_t f)(i) = E_i f(X_t)$ and write $\langle f, g\rangle$ for the inner product $\sum_i \pi_i f(i) g(i)$. Write $\partial(\cdot)$ for $\frac{d}{dt}(\cdot)_{t=0}$. Then

$$\partial\langle f, \mathbf{P}_t g\rangle = -\mathcal{E}(f,g)$$

where

$$\mathcal{E}(f,g) = \frac{1}{2} \sum_i \sum_j (f(j) - f(i))(g(j) - g(i)) q_{ij}.$$

Now consider $\partial\langle h_+, \mathbf{P}_t h\rangle$. On the one hand

$$\partial\langle h_+, \mathbf{P}_t h\rangle = -\mathcal{E}(h_+, h) \le -\mathcal{E}(h_+, h_+)$$

where the inequality follows from the inequality $(a^+ - b^+)^2 \le (a^+ - b^+)(a - b)$ for real a, b. On the other hand, $\langle h_+, h\rangle \le \langle h_+, h_+\rangle = ||h_+||_2^2$, and the eigenvector h satisfies $\partial(\mathbf{P}_t h) = -\lambda_2 h$, so

$$\partial\langle h_+, \mathbf{P}_t h\rangle \ge -\lambda_2 ||h_+||_2^2.$$

Combining these inequalities leads to (4.35).

4.5.2 Cheeger-type inequalities

A lot of attention has been paid to reverse inequalities which upper bound τ_2 in terms of τ_c or related "flow rate" parameters. Motivation for such results will be touched upon in Chapter yyy. The prototype for such results is

Theorem 4.40 (Cheeger's inequality) $\tau_2 \le 8q^*\tau_c^2$, *where* $q^* \equiv \max_i q_i$.

This result follows by combining Lemma 4.39 above with Lemma 4.41 below. In discrete time these inequalities hold with q^* deleted (i.e. replaced by 1), by continuization. Our treatment of Cheeger's inequality closely follows Diaconis and Stroock [124] – see Notes for more history.

Lemma 4.41 *For any subset* A,

$$E_{\alpha_A} T_A \le \frac{2q^*\tau_c^2}{\pi^2(A)}.$$

Proof. Fix A and g with $g \ge 0$ and $g = 0$ on A.

$$\begin{aligned}
&\left(\sum\sum_{x \ne y} |g^2(x) - g^2(y)| \pi_x q_{xy} \right)^2 \\
&\le \sum\sum_{x \ne y} (g(x) - g(y))^2 \pi_x q_{xy} \times \sum\sum_{x \ne y} (g(x) + g(y))^2 \pi_x q_{xy} \\
&\qquad \text{by the Cauchy-Schwarz inequality} \\
&= 2\mathcal{E}(g,g) \sum\sum_{x \ne y} (g(x) + g(y))^2 \pi_x q_{xy} \\
&\le 4\mathcal{E}(g,g) \sum\sum_{x \ne y} (g^2(x) + g^2(y)) \pi_x q_{xy} \\
&= 8\mathcal{E}(g,g) \sum_x \pi_x q_x g^2(x) \\
&\le 8q^* \mathcal{E}(g,g)\ ||g||_2^2.
\end{aligned}$$

On the other hand

$$\begin{aligned}
&\sum\sum_{x \ne y} |g^2(x) - g^2(y)| \pi_x q_{xy} \\
&\quad = 2 \sum\sum_{g(x) > g(y)} (g^2(x) - g^2(y)) \pi_x q_{xy}
\end{aligned}$$

$$
\begin{aligned}
&= 4 \sum \sum_{g(x)>g(y)} \left(\int_{g(y)}^{g(x)} t dt \right) \pi_x q_{xy} \\
&= 4 \int_0^\infty t \left(\sum \sum_{g(y) \le t < g(x)} \pi_x q_{xy} \right) dt \\
&= 4 \int_0^\infty t \; Q(B_t, B_t^c) \; dt \text{ where } B_t \equiv \{x : g(x) > t\} \\
&\ge 4 \int_0^\infty t \; \frac{\pi(B_t) \; \pi(B_t^c)}{\tau_c} \; dt \text{ by definition of } \tau_c \\
&\ge 4 \int_0^\infty t \; \frac{\pi(B_t) \; \pi(A)}{\tau_c} \; dt \text{ because } g = 0 \text{ on } A \\
&= \frac{2\pi(A)||g||_2^2}{\tau_c}.
\end{aligned}
$$

Rearranging,

$$\frac{||g||_2^2}{\mathcal{E}(g,g)} \le \frac{2q^* \tau_c^2}{\pi^2(A)}$$

and the first assertion of the Theorem follows from the extremal characterization (4.36) of $E_{\alpha_A} T_A$.

4.5.3 τ_c and hitting times

Lemma 4.25 and Theorem 4.40 imply a bound on τ^* in terms of τ_c. But a direct argument, using ideas similar to those in the proof of Lemma 4.41, does better.

Proposition 4.42

$$\tau^* \le \frac{4(1 + \log n)}{\min_j \pi_j} \; \tau_c.$$

Example 4.43 below will show that the *log* term cannot be omitted. Compare with graph-theoretic bounds in Chapter 6 section yyy.

Proof. Fix states a, b. We want to use the extremal characterization (Chapter 3 yyy). So fix a function $0 \le g \le 1$ with $g(a) = 0, g(b) = 1$. Order the states as $a = 1, 2, 3, \ldots, n = b$ so that $g(\cdot)$ is increasing.

$$
\begin{aligned}
\mathcal{E}(g,g) &= \sum \sum_{i<k} \pi_i q_{ik} (g(k) - g(i))^2 \\
&\ge \sum \sum \sum_{i \le j < k} \pi_i q_{ik} (g(j+1) - g(j))^2
\end{aligned}
$$

$$\begin{aligned} &= \sum_j (g(j+1)-g(j))^2 Q(A_j, A_j^c), \text{ where } A_j \equiv [1,j] \\ &\geq \sum_j (g(j+1)-g(j))^2 \frac{\pi(A_j)\pi(A_j^c)}{\tau_c} \end{aligned} \tag{4.37}$$

But

$$1 = \sum_j (g(j+1)-g(j)) = \sum_j (g(j+1)-g(j)) \, \frac{\pi^{1/2}(A_j)\pi^{1/2}(A_j^c)}{\tau_c^{1/2}} \, \frac{\tau_c^{1/2}}{\pi^{1/2}(A_j)\pi^{1/2}(A_j^c)}.$$

So by Cauchy-Schwarz and (4.37)

$$1 \leq \tau_c \mathcal{E}(g,g) \sum_j \frac{1}{\pi(A_j)\pi(A_j^c)}. \tag{4.38}$$

But $\pi(A_j) \geq j\pi_*$, where $\pi_* \equiv \min_i \pi_i$, so

$$\sum_{j:\pi(A_j)\leq 1/2} \frac{1}{\pi(A_j)\pi(A_j^c)} \leq \sum_j \frac{2}{j\pi_*} \leq \frac{2}{\pi_*}(1+\log n).$$

The same bound holds for the sum over $\{j : \pi(A_j) \geq 1/2\}$, so applying (4.38) we get

$$\frac{1}{\mathcal{E}(g,g)} \leq \tau_c \, \frac{4}{\pi_*}(1+\log n)$$

and the Proposition follows from the extremal characterization.

Example 4.43 Consider the weighted linear graph with loops on vertices $\{0,1,2,\ldots,n-1\}$, with edge-weights

$$w_{i-1,i} = i, \; 1 \leq i \leq n-1; \qquad w_{ii} = 2n - i1_{(i\neq 0)} - (i+1)1_{(i\neq n-1)}.$$

This gives vertex-weights $w_i = 2n$, and so the stationary distribution is uniform. By the commute interpretation of resistance,

$$\tau^* = E_0 T_{n-1} + E_{n-1} T_0 = w r_{0,n-1} = 2n^2 \sum_{i=1}^{n-1} 1/i \sim 2n^2 \log n.$$

Using Lemma 4.36, the value of τ_c is attained by a split of the form $\{[0,j],[j+1,n-1]\}$, and a brief calculation shows that the maximizing value is $j=0$ and gives

$$\tau_c = 2(n-1).$$

So in this example, the bound in Proposition 4.42 is sharp up to the numerical constant.

4.6 Induced and product chains

Here we record the behavior of our parameters under two natural operations on chains.

4.6.1 Induced chains

Given a Markov chain X_t on state space I and a function $f : I \to \hat{I}$, the process $f(X_t)$ is typically *not* a Markov chain. But we can invent a chain which substitutes. In discrete time (the continuous case is similar) define the *induced chain* $\hat{X}_t$ to have transition matrix

$$\hat{p}_{\hat{i},\hat{j}} = P_\pi(f(X_1) = \hat{j} | f(X_0) = \hat{i}) = \frac{\sum\sum_{i,j:f(i)=\hat{i},f(j)=\hat{j}} \pi_i p_{i,j}}{\sum_{i:f(i)=\hat{i}} \pi_i}. \tag{4.39}$$

More informatively, we are matching the stationary flow rates:

$$P_{\hat{\pi}}(\hat{X}_0 = \hat{i}, \hat{X}_1 = \hat{j}) = P_\pi(f(X_0) = \hat{i}, f(X_1) = \hat{j}). \tag{4.40}$$

The reader may check that (4.39) and (4.40) are equivalent. Under our standing assumption that X_t is reversible, the induced chain is also reversible (though the construction works for general chains as well). In the electrical network interpretation, we are shorting together vertices with the same f-values. It seems intuitively plausible that this "shorting" can only decrease our parameters describing convergence and mean hitting time behavior.

Proposition 4.44 (The contraction principle) *The values of τ^*, τ_0, τ_2 and τ_c in an induced chain are less than or equal to the corresponding values in the original chain.*

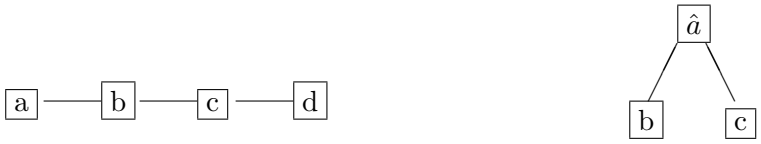
Proof. A function $\hat{g} : \hat{I} \to R$ pulls back to a function $g \equiv \hat{g}(f(\cdot)) : I \to R$. So the Dirichlet principle (Chapter 3 yyy) shows that mean commute times can only decrease when passing to an induced chain:

$$E_{f(i)}\hat{T}_{f(j)} + E_{f(j)}\hat{T}_{f(i)} \le E_i T_j + E_j T_i.$$

This establishes the assertion for τ^* and τ_0, and the extremal characterization of relaxation time works similarly for τ_2. The assertion about τ_c is immediate from the definition, since a partition of $\hat{I}$ pulls back to a partition of I. □

On the other hand, it is easy to see that shorting may *increase* a one-sided mean hitting time. For example, random walk on the unweighted

graph on the left has $E_a T_b = 1$, but when we short $\{a, d\}$ together to form vertex $\hat{a}$ in the graph on the right, $E_{\hat{a}} \hat{T}_b = 2$.



Finally, the behavior of the τ_1-family under shorting is unclear.

Open Problem 4.45 *Is the value of $\tau_1^{(2)}$ in an induced chain bounded by K times the value of $\tau_1^{(2)}$ in the original chain, for some absolute constant K? For $K = 1$?*

4.6.2 Product chains

Given Markov chains on state spaces $I^{(1)}$ and $I^{(2)}$, there is a natural concept of a "product chain" on state space $I^{(1)} \times I^{(2)}$. It is worth writing this concept out in detail for two reasons. First, to prevent confusion between several different possible definitions in discrete time. Second, because the behavior of relaxation times of product chains is relevant to simple examples and has a surprising application (section 4.6.3).

As usual, things are simplest in continuous time. Define the product chain to be

$$\mathbf{X}_t = (X_t^{(1)}, X_t^{(2)})$$

where the components $X_t^{(1)}$ and $X_t^{(2)}$ are independent versions of the given chains. So

$$P_{i_1,i_2}(\mathbf{X}_t = (j_1, j_2)) = P_{i_1}^{(1)}(X_t^{(1)} = j_1) P_{i_2}^{(2)}(X_t^{(2)} = j_2). \tag{4.41}$$

Using the interpretation of relaxation time as asymptotic rate of convergence of transition probabilities, (Chapter 3 yyy) it is immediate that $\mathbf{X}$ has relaxation time

$$\tau_2 = \max(\tau_2^{(1)}, \tau_2^{(2)}). \tag{4.42}$$

In discrete time there are two different general notions of "product chain". One could consider the chain $(X_t^{(1)}, X_t^{(2)})$ whose coordinates are the independent chains. This is the chain with transition probabilities

$$(i_1, i_2) \to (j_1, j_2): \quad \text{probability } P^{(1)}(i_1, j_1) P^{(2)}(i_2, j_2)$$

and has relaxation time

$$\tau_2 = \max(\tau_2^{(1)}, \tau_2^{(2)}).$$

But it is more natural to define the *product chain* $\mathbf{X}_t$ to be the chain with transition probabilities

$$(i_1, i_2) \to (j_1, i_2): \quad \text{probability } \frac{1}{2} P^{(1)}(i_1, j_1)$$

$$(i_1, i_2) \to (i_1, j_2): \quad \text{probability } \frac{1}{2} P^{(2)}(i_2, j_2).$$

This is the jump chain derived from the product of the continuized chains, and has relaxation time

$$\tau_2 = 2 \max(\tau_2^{(1)}, \tau_2^{(2)}). \tag{4.43}$$

Again, this can be seen without need for calculation: the continuized chain is just the continuous-time product chain *run at half speed*.

This definition and (4.43) extend to d-fold products in the obvious way. Random walk on Z^d is the product of d copies of random walk on Z^1, and random walk on the d-cube (Chapter 5 yyy) is the product of d copies of random walk on $\{0, 1\}$.

Just to make things more confusing, given graphs $G^{(1)}$ and $G^{(2)}$ the Cartesian product graph is defined to have edges

$$(v_1, w_1) \leftrightarrow (v_2, w_1) \text{ for } v_1 \leftrightarrow v_2$$

$$(v_1, w_1) \leftrightarrow (v_1, w_2) \text{ for } w_1 \leftrightarrow w_2.$$

If both $G^{(1)}$ and $G^{(2)}$ are r-regular then random walk on the product graph is the product of the random walks on the individual graphs. But in general, discrete-time random walk on the product graph is the jump chain of the product of the *fluid model* (Chapter 3 yyy) continuous-time random walks. So if the graphs are r_1- and r_2-regular then the discrete-time random walk on the product graph has the product distribution as its stationary distribution and has relaxation time

$$\tau_2 = (r_1 + r_2) \max(\tau_2^{(1)}/r_1, \tau_2^{(2)}/r_2).$$

But for non-regular graphs, neither assertion is true.

Let us briefly discuss the behavior of some other parameters under products. For the continuous-time product (4.41), the total variation distance $\bar{d}$ of section 4.3 satisfies

$$\bar{d}(t) = 1 - (1 - \bar{d}^{(1)}(t))(1 - \bar{d}^{(2)}(t))$$

and we deduce the crude bound

$$\tau_1 \le 2 \max(\tau_1^{(1)}, \tau_1^{(2)})$$

where superscripts refer to the graphs $G^{(1)}, G^{(2)}$ and not to the parameters in section 4.3.1. For the discrete-time chain, there is an extra factor of 2 from "slowing down" (cf. (4.42,4.43)), leading to

$$\tau_1 \le 4 \max(\tau_1^{(1)}, \tau_1^{(2)}).$$

Here our conventions are a bit confusing: this inequality refers to the discrete-time product chain, but as in section 4.3 we define τ_1 via the continuized chain – we leave the reader to figure out the analogous result for τ_1^{disc} discussed in section 4.3.3.

To state a result for τ_0, consider the continuous-time product $(X_t^{(1)}, X_t^{(2)})$ of independent copies of the same n-state chain. If the underlying chain has eigenvalues $(\lambda_i; 1 \le i \le n)$ then the product chain has eigenvalues $(\lambda_i + \lambda_j; 1 \le i, j \le n)$ and so by the eigentime identity

$$\begin{aligned}
\tau_0^{\text{product}} &= \sum_{i,j \ge 1; (i,j) \ne (1,1)} \frac{1}{\lambda_i + \lambda_j} \\
&= 2\tau_0 + \sum_{i,j \ge 2} \frac{1}{\lambda_i + \lambda_j} \\
&= 2\tau_0 + 2 \sum_{i=2}^{n} \sum_{j=i}^{n} \frac{1}{\lambda_i + \lambda_j} \\
&\le 2\tau_0 + \sum_{i=2}^{n} (n - i + 1) \frac{2}{\lambda_i} \\
&\le 2\tau_0 + (n-1) 2\tau_0 = 2n\tau_0.
\end{aligned}$$

Thus in discrete time

$$\tau_0^{\text{product}} \le 4n\tau_0. \tag{4.44}$$

4.6.3 Efron-Stein inequalities

The results above concerning relaxation times of product chains are essentially obvious using the interpretation of relaxation time as asymptotic rate of convergence of transition probabilities, but they are much less obvious using the extremal interpretation. Indeed, consider the n-fold product of a single chain X with itself. Write (X_0, X_1) for the distribution at times 0 and 1 of X, and τ_2 for the relaxation time of X. Combining (4.43) with the extremal characterization (4.22) of the relaxation time for the product chain, a brief calculation gives the following result.

Corollary 4.46 *Let $f : I^n \to R$ be arbitrary. Let $(X^{(i)}, Y^{(i)})$, $i = 1, \ldots, n$ be independent copies of (X_0, X_1). Let $Z = f(X^{(1)}, \ldots, X^{(n)})$ and let $Z^{(i)} = f(X^{(1)}, \ldots, X^{(i-1)}, Y^{(i)}, X^{(i+1)}, \ldots, X^{(n)})$. Then*

$$\frac{\text{var } (Z)}{\frac{1}{2n}\sum_{i=1}^n E(Z - Z^{(i)})^2} \le n\tau_2.$$

To appreciate this, consider the "trivial" case where the underlying Markov chain is just an i.i.d. sequence with distribution π on I. Then $\tau_2 = 1$ and the $2n$ random variables $(X^{(i)}, Y^{(i)}; 1 \le i \le n)$ are i.i.d. with distribution π. And this special case of Corollary 4.46 becomes (4.45) below, because for each i the distribution of $Z - Z^{(i)}$ is unchanged by substituting X_0 for $Y^{(i)}$.

Corollary 4.47 *Let $f : I^n \to R$ be arbitrary. Let $(X_0, X_1, \ldots, X_n)$ be i.i.d. with distribution π. Let $Z^{(i)} = f(X_1, \ldots, X_{i-1}, X_0, X_{i+1}, \ldots, X_n)$ and let $Z = f(X_1, \ldots, X_n)$. Then*

$$\text{var } (Z) \le \frac{1}{2}\sum_{i=1}^n E(Z - Z^{(i)})^2 \tag{4.45}$$

If f is symmetric then

$$\text{var } (Z) \le \sum_{i=0}^n E(Z^{(i)} - \bar{Z})^2 \tag{4.46}$$

where $Z^{(0)} = Z$ and $\bar{Z} = \frac{1}{n+1}\sum_{i=0}^n Z^{(i)}$.

Note that in the symmetric case we may rewrite

$$Z^{(i)} = f(X_0, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

This reveals (4.46) to be the celebrated *Efron-Stein* inequality in statistics, and in fact (4.45) is a known variant (see Notes).

Proof. As observed above, (4.45) is a special case of Corollary 4.46. So it is enough to show that for symmetric f the right sides of (4.45) and (4.46) are equal. Note that by symmetry $a = E(Z^{(i)})^2$ does not depend on i, and $b = EZ^{(i)}Z^{(j)}$ does not depend on (i,j), for $j \neq i$. So the right side of (4.45) is

$$\frac{1}{2}n(a - 2b + a) = n(a - b).$$

But it is easy to calculate

$$EZ^{(i)}\bar{Z} = E\bar{Z}^2 = \frac{a}{n+1} + \frac{nb}{n+1}$$

and then the right side of (4.46) equals

$$(n+1)(a - 2EZ^{(i)}\bar{Z} + E\bar{Z}^2) = na - nb.$$

4.6.4 Why these parameters?

The choice of parameters studied in this chapter is partly arbitrary, but our choice has been guided by two criteria, one philosophical and one technical. The philosophical criterion is

> when formalizing a vague idea, choose a definition which has several equivalent formulations.

This is why we used the maximal mean hitting time parameter $\max_{i,j}(E_iT_j + E_jT_i)$ instead of $\max_{i,j} E_iT_j$, because the former permits the equivalent "resistance" interpretation.

Here is the technical criterion. Given a continuous-time chain X_t and a state i, create a new chain X^*_t by splitting i into two states i_1, i_2 and setting

$$q^*_{i_1 j} = q^*_{i_2 j} = q_{ij};\ j \neq i$$

$$q^*_{j i_1} = q^*_{j i_2} = \frac{1}{2}q_{ji}\ j \neq i$$

$$q^*_{i_1 i_2} = q^*_{i_2 i_1} = \rho$$

with $q^* = q$ elsewhere. Then $\pi^*(i_1) = \pi^*(i_2) = \frac{1}{2}\pi(i)$, with $\pi^* = \pi$ elsewhere. As $\rho \to \infty$, we may regard the new chain as converging to the old chain in a certain sense. So our technical criterion for parameters τ is that the value of τ for X^* should converge, as $\rho \to \infty$, to the value for X. It is easy to check this holds for the τ's we have studied, but it does not hold for, say,

$$\tau \equiv \max_{ij} \pi_j E_i T_j$$

which at first sight might seem a natural parameter.

4.7 Notes on Chapter 4

Section 4.3.1. The definition of $\tau_1^{(2)}$ involves the idea of a stopping time U such that X_U has distribution π and is independent of the starting position. This idea is central to the standard modern theory of Harris-recurrent Markov chains, i.e. chains on continuous space which mimic the asymptotic behavior of discrete recurrent chains, and does not require reversibility. See [133] sec. 5.6 for an introduction, and [34, 263] for more comprehensive treatments. In that field, researchers have usually been content to obtain some finite bound on EU, and haven't faced up to our issue of the quantitative dependence of the bound on the underlying chain.

Separation and strong stationary times were introduced in Aldous and Diaconis [22], who gave some basic theory. These constructions can be used to bound convergence times in examples, but in practice are used in examples with much special structure, e.g. non-necessarily-symmetric random walks on groups. Examples can be found in [21, 22] and Matthews [258]. Development of theory, mostly for stochastically monotone chains on 1-dimensional state space, is in Diaconis and Fill [114, 115], Fill [150, 151] and Matthews [260].

The recurrent balayage theorem (Chapter 1 yyy) can be combined with the mean hitting time formula to get

$$\tau_1^{(2)} = \max_{ij} \frac{-Z_{ij}}{\pi_j}. \tag{4.47}$$

Curiously, this elegant result doesn't seem to help much with the inequalities in Theorem 4.6.

What happens with the τ_1-family of parameters for general chains remains rather obscure. Some counter-examples to equivalence, and weaker inequalities containing $\log 1/\pi_*$ factors, can be found in [6]. Recently, Lovasz and Winkler [241] initiated a detailed study of $\tau_1^{(2)}$ for general chains which promises to shed more light on this question.

Our choice of τ_1 as the "representative" of the family of $\tau_1^{(i)}$'s is somewhat arbitrary. One motivation was that it gives the constant "1" in the inequality $\tau_2 \leq \tau_1$. It would be interesting to know whether the constants in other basic inequalities relating the τ_1-family to other parameters could be made "1":

Open Problem 4.48 *(a) Is* $\tau_1 \leq \tau_0$*?*
(b) Is $\tau_2 \leq \tau_1^{(2)}$*?*

Much of recent sophisticated theory xxx refs bounds $d(t)$ by bounding

$\hat{d}(t)$ and appealing to Lemma 4.7(b). But it is not clear whether there is an analog of Theorem 4.6 relating the $\hat{d}$-threshold to other quantities.

Section 4.3.2. The parts of Theorem 4.6 involving $\tau_1^{(1)}$ and $\tau_1^{(3)}$ are implicit rather than explicit in [6]. That paper had an unnecessarily complicated proof of Lemma 4.13. The proof of (4.15) in [6] gives a constant $K \approx e^{13}$. It would be interesting to obtain a smaller constant! Failing this, a small constant in the inequality $\tau_1^{(1)} \leq K\tau_1^{(3)}$ would be desirable. As a weaker result, it is easy to show

$$\tau_1^{(1)} \leq 10 \min_j \max_i E_i T_j \tag{4.48}$$

which has some relevance to later examples (yyy).

Section 4.3.3. The analog of Open Problem 4.17 in which we measure distance from stationarity by $\hat{d}$ instead of $d(t)$ is straightforward, using the "CM proxy" property of discrete time chains:

$$P_i(X_{2t} = i) + P_i(X_{2t+1} = i) \downarrow 0 \text{ as } t \to \infty.$$

Open Problem 4.17 itself seems deeper, though the weaker form in which we require only that $\phi(t) = O(t)$ can probably be proved by translating the proof of (4.15) into discrete time and using the CM proxy property.

Section 4.3.4. The cat-and-mouse game was treated briefly in Aleliunas et al [25], who gave a bare-hands proof of a result like (4.21). Variations in which the cat is also allowed an arbitrary strategy have been called "princess and monster" games – see Isaacs [190] for results in a different setting.

Section 4.3.5. Sinclair [308] points out that "hard" results of Leighton and Rao [223] on multicommodity flow imply

$$\inf_{\mathbf{f}} \psi(\mathbf{f}) \leq K\tau_2 \log n. \tag{4.49}$$

This follows from Corollary 4.22 and Lemma 4.23 when π is uniform, but Sinclair posed

Open Problem 4.49 (i) Is there a simple proof of (4.49) in general?
(ii) Does (4.49) hold with the diameter Δ in place of $\log n$?

Section 4.4. As an example of historical interest, before this topic became popular Fiedler [146] proved

Proposition 4.50 *For random walk on a n-vertex weighted graph where the stationary distribution is uniform,*

$$\tau_2 \leq \frac{w}{4nc \sin^2 \frac{\pi}{2n}} \sim \frac{wn}{\pi^2 c}$$

where c is the minimum cut defined at (4.4).

This upper bound is sharp. On the other hand, Proposition 4.2 gave the same upper bound (up to the numerical constant) for the *a priori* larger quantity τ^*, and so is essentially a stronger result.

Section 4.4.1. In the non-reversible case the definition of the maximal correlation $\rho(t)$ makes sense, and there is similar asymptotic behavior:

$$\rho(t) \sim c \exp(-\lambda t) \text{ as } t \to \infty$$

where λ is the "spectral gap". But we cannot pull back from asymptotia to the real world so easily: it is not true that $\rho(t)$ can be bounded by $K \exp(-\lambda t)$ for universal K. A dramatic example from Aldous [11] section 4 has for each n an n-state chain with spectral gap bounded away from 0 but with $\rho(n)$ also bounded away from 0, instead of being exponentially small. So implicit claims in the literature that estimates of the spectral gap for general chains have implications for finite-time behavior should be treated with extreme skepticism.

It is not surprising that the classical Berry-Esseen Theorem for i.i.d. sums ([133] Thm. 2.4.10) has an analog for chains. Write σ^2 for the asymptotic variance rate in Proposition 4.29 and write Z for a standard Normal r.v.

Proposition 4.51 *There is a constant K, depending on the chain, such that*

$$\sup_x |P_\pi(\frac{S_t}{\sigma t^{1/2}} \le x) - P(Z \le x)| \le K t^{-1/2}$$

for all $t \ge 1$ and all standardized g.

This result is usually stated for infinite-state chains satisfying various mixing conditions, which are automatically satisfied by finite chains. See e.g. Bolthausen [55]. At first sight the constant K depends on the function g as well as the chain, but a finiteness argument shows that the dependence on g can be removed. Unfortunately the usual proofs don't give any useful indications of how K depends on the chain, and so don't help with Open Problem 4.30.

The variance results in Proposition 4.29 are presumably classical, being straightforward consequences of the spectral representation. Their use in algorithmic settings such as Corollary 4.31 goes back at least to [10].

Section 4.4.3. Systematic study of the optimal choice of weights in the Cauchy-Schwarz argument for Theorem 4.32 may lead to improved bounds in examples. Alan Sokal has unpublished notes on this subject.

Section 4.5.1. The quantity $1/\tau_c$, or rather this quantity with the alternate definition of τ_c mentioned in the text, has been called *conductance*. I avoid that term, which invites unnecessary confusion with the electrical network terminology. However, the subscript c can be regarded as standing for "Cheeger" or "conductance".

In connection with Open Problem 4.38 we mention the following result. Suppose that in the definition (section 4.4.1) of the maximal correlation function $\rho(t)$ we considered only *events*, i.e. suppose we defined

$$\tilde{\rho}(t) \equiv \sup_{A,B} \mathrm{cor}(1_{(X_0 \in A)}, 1_{(X_t \in B)}).$$

Then $\tilde{\rho}(t) \leq \rho(t)$, but in fact the two definitions are equivalent in the sense that there is a universal function $\psi(x) \downarrow 0$ as $x \downarrow 0$ such that $\rho(t) \leq \psi(\tilde{\rho}(t))$. This is a result about "measures of dependence" which has nothing to do with Markovianness – see e.g. Bradley et al [59].

Section 4.5.2. The history of Cheeger-type inequalities up to 1987 is discussed in [222] section 6. Briefly, Cheeger [87] proved a lower bound for the eigenvalues of the Laplacian on a compact Riemannian manifold, and this idea was subsequently adapted to different settings – in particular, by Alon [26] to the relationship between eigenvalues and expansion properties of graphs. Lawler and Sokal [222], and independently Jerrum and Sinclair [307], were the first to discuss the relationship between τ_c and τ_2 at the level of reversible Markov chains. Their work was modified by Diaconis and Stroock [124], whose proof we followed for Lemmas 4.39 and 4.41. The only novelty in my presentation is talking explicitly about quasistationary distributions, which makes the relationships easier to follow.

xxx give forward pointer to results of [238, 158].

Section 4.6.2. See Efron-Stein [140] for the origin of their inequality. Inequality (4.45), or rather the variant mentioned above Corollary 4.47 involving the $2n$ i.i.d. variables

Chapter 5

Examples: Special Graphs and Trees (April 23 1996)

There are two main settings in which explicit calculations for random walks on large graphs can be done. One is where the graph is essentially just 1-dimensional, and the other is where the graph is highly symmetric. The main purpose of this chapter is to record some (mostly) bare-hands calculations for simple examples, in order to illuminate the general inequalities of Chapter 4. Our focus is on natural examples, but there are a few artificial examples devised to make a mathematical point. A second purpose is to set out some theory for birth-and-death chains and for trees.

Lemma 5.1 below is useful in various simple examples, so let's record it here. An edge (v, x) of a graph is *essential* (or a *bridge*) if its removal would disconnect the graph, into two components $A(v, x)$ and $A(x, v)$, say, containing v and x respectively. Recall that $\mathcal{E}$ is the set of (undirected) edges, and write $\mathcal{E}(v, x)$ for the set of edges of $A(v, x)$.

Lemma 5.1 (essential edge lemma) *For random walk on a weighted graph with essential edge (v, x),*

$$E_v T_x = \frac{2\sum_{(i,j)\in\mathcal{E}(v,x)} w_{ij}}{w_{vx}} + 1 \tag{5.1}$$

$$E_v T_x + E_x T_v = \frac{w}{w_{vx}}, \text{ where } w = \sum_i \sum_j w_{ij}. \tag{5.2}$$

Specializing to the unweighted case,

$$E_v T_x = 2|\mathcal{E}(v, x)| + 1 \tag{5.3}$$

$$E_v T_x + E_x T_v = 2|\mathcal{E}|. \tag{5.4}$$

Proof. It is enough to prove (5.1), since (5.2) follows by adding the two expressions of the form (5.1). Because (v,x) is essential, we may delete all vertices of $A(x,v)$ except x, and this does not affect the behavior of the chain up until time T_x, because x must be the first visited vertex of $A(x,v)$. After this deletion, $\pi_x^{-1} = E_x T_x^+ = 1 + E_v T_x$ by considering the first step from x, and $\pi_x = w_{vx}/(2w_{vx} + 2\sum_{(i,j)\in\mathcal{E}(v,x)} w_{ij})$, giving (5.1). ∎

Remarks. Of course Lemma 5.1 is closely related to the *edge-commute inequality* of Chapter 3 Lemma yyy. We can also regard (5.2), and hence (5.4), as consequences of the commute interpretation of resistance (Chapter 3 yyy), because the effective resistance across an essential edge (v,x) is obviously $1/w_{vx}$.

5.1 One-dimensional chains

5.1.1 Simple symmetric random walk on the integers

It is useful to record some elementary facts about simple symmetric random walk (X_t) on the (infinite) set of all integers. As we shall observe, these may be derived in several different ways.

A fundamental formula gives exit probabilities:

$$P_b(T_c < T_a) = \frac{b-a}{c-a},\ a < b < c. \tag{5.5}$$

An elementary argument is that $g(i) \equiv P_i(T_c < T_a)$ satisfies the 1-step recurrence

$$g(i) = \tfrac{1}{2}g(i+1) + \tfrac{1}{2}g(i-1),\ a < i < b$$
$$g(a) = 0,\quad g(b) = 1,$$

whose solution is $g(i) = (i-a)/(b-a)$. At a more sophisticated level, (5.5) is a martingale result. The quantity $p \equiv P_b(T_c < T_a)$ must satisfy

$$b = E_b\, X(T_a \wedge T_c) = pc + (1-p)a,$$

where the first equality is the optional sampling theorem for the martingale X, and solving this equation gives (5.5).

For $a < c$, note that $T_a \wedge T_c$ is the "exit time" from the open interval (a,c). We can use (5.5) to calculate the "exit before return" probability

$$\begin{aligned} P_b(T_b^+ > T_a \wedge T_c) &= \tfrac{1}{2}P_{b+1}(T_c < T_b) + \tfrac{1}{2}P_{b-1}(T_a < T_b) \\ &= \frac{1}{2}\frac{1}{c-b} + \frac{1}{2}\frac{1}{b-a} \\ &= \frac{c-a}{2(c-b)(b-a)}. \end{aligned} \tag{5.6}$$

For the walk started at b, let $m(b,x;a,c)$ be the mean number of visits to x before the exit time $T_a \wedge T_c$. (Recall from Chapter 2 our convention that "before time t" includes time 0 but excludes time t). The number of returns to b clearly has a Geometric distribution, so by (5.6)

$$m(b,b;a,c) = \frac{2(c-b)(b-a)}{c-a}, \quad a \le b \le c. \tag{5.7}$$

To get the analog for visits to x we consider whether or not x is hit at all before exiting; this gives

$$m(b,x;a,c) = P_b(T_x < T_a \wedge T_c)\ m(x,x;a,c).$$

Appealing to (5.5) and (5.7) gives the famous mean occupation time formula

$$m(b,x;a,c) \quad = \quad \begin{cases} \frac{2(x-a)(c-b)}{c-a}, & a \le x \le b \le c \\ \\ \frac{2(c-x)(b-a)}{c-a}, & a \le b \le x \le c. \end{cases} \tag{5.8}$$

Now the (random) time to exit must equal the sum of the (random) times spent at each state. So, taking expectations,

$$E_b(T_a \wedge T_c) = \sum_{x=a}^{c} m(b,x;a,c),$$

and after a little algebra we obtain

Lemma 5.2 $E_b(T_a \wedge T_c) = (b-a)(c-b),\ a < b < c.$

This derivation of Lemma 5.2 from (5.8) has the advantage of giving the mean occupation time formula (5.8) on the way. There are two alternative ways to prove Lemma 5.2. An elementary proof is to set up and solve the 1-step recurrence for $h(i) \equiv E_i(T_a \wedge T_c)$:

$$h(i) = 1 + \tfrac{1}{2}h(i+1) + \tfrac{1}{2}h(i-1),\ a < i < c$$
$$h(a) = h(c) = 0.$$

The more elegant proof uses a martingale argument. Taking $b = 0$ without loss of generality, the first equality below is the optional sampling theorem for the martingale $(X^2(t) - t)$:

$$\begin{aligned} E_0(T_a \wedge T_c) &= E_0 X^2(T_a \wedge T_c) \\ &= a^2 P_0(T_a < T_c) + c^2 P_0(T_c < T_a) \\ &= a^2 \frac{c}{c-a} + c^2 \frac{-a}{c-a} \text{ by (5.5)} \\ &= -ac. \end{aligned}$$

The preceding discussion works in discrete or continuous time. Exact distributions at time t will of course differ in the two cases. In discrete time we appeal to the Binomial distribution for the number of +1 steps, to get

$$P_0(X_{2t} = 2j) = \frac{(2t)!}{(t+j)!(t-j)!} 2^{-2t}, \ -t \le j \le t \tag{5.9}$$

and a similar expression for odd times t. In continuous time, the numbers of +1 and of −1 steps in time t are independent Poisson(t) variables, so

$$P_0(X_t = -j) = P_0(X_t = j) = e^{-2t} \sum_{i=0}^{\infty} \frac{t^{2i+j}}{i!(i+j)!}, \ j \ge 0. \tag{5.10}$$

5.1.2 Weighted linear graphs

Consider the n-vertex linear graph $0 - 1 - 2 - \cdots - (n-1)$ with arbitrary edge-weights $(w_1, \ldots, w_{n-1})$, where $w_i > 0$ is the weight on edge $(i-1, i)$. Set $w_0 = w_n = 0$ to make some later formulas cleaner. The corresponding discrete-time random walk has transition probabilities

$$p_{i,i+1} = \frac{w_{i+1}}{w_i + w_{i+1}}, \ p_{i,i-1} = \frac{w_i}{w_i + w_{i+1}}, \quad 0 \le i \le n-1$$

and stationary distribution

$$\pi_i = \frac{w_i + w_{i+1}}{w}, \quad 0 \le i \le n-1$$

where $w = 2 \sum_i w_i$. In probabilistic terminology, this is a *birth-and-death process*, meaning that a transition cannot alter the state by more than 1. It is elementary that such processes are automatically reversible (xxx spells out the more general result for trees), so as discussed in Chapter 3 yyy the set-up above with weighted graphs gives the general discrete-time birth-and-death process with $p_{ii} \equiv 0$. But note that the continuization does *not* give the general continuous-time birth-and-death process, which has $2(n-1)$ parameters $(q_{i,i-1}, q_{i,i+1})$ instead of just $n-1$ parameters (w_i). The formulas below could all be extended to this general case (the analog of Proposition 5.3 can be found in undergraduate textbooks, e.g., Karlin and Taylor [208] Chapter 4) but our focus is on the simplifications which occur in the "weighted graphs" case.

Proposition 5.3 *(a) For $a < b < c$,*

$$P_b(T_c < T_a) = \frac{\sum_{i=a+1}^{b} w_i^{-1}}{\sum_{i=a+1}^{c} w_i^{-1}}.$$

(b) For $b < c$,

$$E_b T_c = c - b + 2 \sum_{j=b+1}^{c} \sum_{i=1}^{j-1} w_i w_j^{-1}.$$

(c) For $b < c$,

$$E_b T_c + E_c T_b = w \sum_{i=b+1}^{c} w_i^{-1}.$$

Note that we can obtain an expression for $E_c T_b$, $b < c$, by reflecting the weighted graph about its center.

Proof. These are extensions of (5.5,5.1,5.2) and recycle some of the previous arguments. Writing $h(j) = \sum_{i=1}^{j} w_i^{-1}$, we have that $(h(X_t))$ is a martingale, so

$$h(b) = E_b h(X(T_a \wedge T_c)) = p h(c) + (1-p) h(a)$$

for $p \equiv P_b(T_c < T_a)$. Solving this equation gives $p = \frac{h(b)-h(a)}{h(c)-h(a)}$, which is (a).

The mean hitting time formula (b) has four different proofs! Two that we will *not* give are as described below Lemma 5.2: Set up and solve a recurrence equation, or use a well-chosen martingale. The slick argument is to use the essential edge lemma (Lemma 5.1) to show

$$E_{j-1} T_j = 1 + 2 \frac{\sum_{i=1}^{j-1} w_i}{w_j}.$$

Then

$$E_b T_c = \sum_{j=b+1}^{c} E_{j-1} T_j,$$

establishing (b). Let us also write out the non-slick argument, using mean occupation times. By considering mean time spent at i,

$$E_b T_c = \sum_{i=0}^{b-1} P_b(T_i < T_c) m(i,i,c) + \sum_{i=b}^{c-1} m(i,i,c), \tag{5.11}$$

where $m(i,i,c)$ is the expectation, starting at i, of the number of visits to i before T_c. But

$$\begin{aligned} m(i,i,c) &= \frac{1}{P_i(T_c < T_i^+)} \\ &= \frac{1}{p_{i.i+1} P_{i+1}(T_c < T_i)} \\ &= (w_i + w_{i+1}) \sum_{j=i+1}^{c} w_j^{-1} \text{ using (a).} \end{aligned}$$

Substituting this and (a) into (5.11) leads to the formula stated in (b).

Finally, (c) can be deduced from (b), but it is more elegant to use the essential edge lemma to get

$$E_{i-1}T_i + E_iT_{i-1} = w/w_i \tag{5.12}$$

and then use

$$E_bT_c + E_cT_b = \sum_{i=b+1}^{c} (E_{i-1}T_i + E_iT_{i-1}). \quad \blacksquare$$

We now start some little calculations relating to the parameters discussed in Chapter 4. Plainly, from Proposition 5.3

$$\tau^* = w \sum_{i=1}^{n-1} w_i^{-1}. \tag{5.13}$$

Next, consider calculating $E_\pi T_b$. We could use Proposition 5.3(b), but instead let us apply Theorem yyy of Chapter 3, giving $E_\pi T_b$ in terms of unit flows from b to π. In a linear graph there is only one such flow, which for $i \geq b$ has $f_{i,i+1} = \pi[i+1, n-1] = \sum_{j=i+1}^{n-1} \pi_j$, and for $i \leq b-1$ has $f_{i,i+1} = -\pi[0, i]$, and so the Proposition implies

$$E_\pi T_b = w \sum_{i=b+1}^{n-1} \frac{\pi^2[i, n-1]}{w_i} + w \sum_{i=1}^{b} \frac{\pi^2[0, i-1]}{w_i}. \tag{5.14}$$

There are several ways to use the preceding results to compute the average hitting time parameter τ_0. Perhaps the most elegant is

$$\begin{aligned}
\tau_0 &= \sum_i \sum_{j>i} \pi_i \pi_j (E_iT_j + E_jT_i) \\
&= \sum_{k=1}^{n-1} \pi[0, k-1]\pi[k, n-1](E_{k-1}T_k + E_kT_{k-1}) \\
&= \sum_{k=1}^{n-1} \pi[0, k-1]\pi[k, n-1]w/w_k \text{ by (5.12)} \\
&= w^{-1} \sum_{k=1}^{n-1} w_k^{-1} \left(w_k + 2\sum_{j=1}^{k-1} w_j \right) \left(w_k + 2\sum_{j=k+1}^{n-1} w_j \right).
\end{aligned} \tag{5.15}$$

There are sophisticated methods (see Notes) of studying τ_1, but let us just point out that Proposition 5.23 later (proved in the more general context

of trees) holds in the present setting, giving

$$\frac{1}{K_1} \min_x \max(E_0 T_x, E_{n-1} T_x) \le \tau_1 \le K_2 \min_x \max(E_0 T_x, E_{n-1} T_x). \quad (5.16)$$

We do not know an explicit formula for τ_2, but we can get an upper bound easily from the "distinguished paths" result Chapter 4 yyy. For $x < y$ the path γ_{xy} has $r(\gamma_{xy}) = \sum_{u=x+1}^{y} 1/w_u$ and hence the bound is

$$\tau_2 \le \frac{1}{w} \max_j \sum_{x=0}^{j-1} \sum_{y=j}^{n-1} \sum_{u=x+1}^{y} \frac{(w_x + w_{x+1})(w_y + w_{y+1})}{w_u}. \quad (5.17)$$

jjj This uses the Diaconis–Stroock version. The Sinclair version is

$$\tau_2 \le \frac{1}{w} \max_j \frac{1}{w_j} \sum_{x=0}^{j-1} \sum_{y=j}^{n-1} (w_x + w_{x+1})(w_y + w_{y+1})(y - x).$$

xxx literature on τ_2 (van Doorn, etc.)

jjj Also relevant is work of N. Kahale (and others) on how optimal choice of weights in use of Cauchy–Schwarz inequality for Diaconis–Stroock–Sinclair leads to equality in case of birth-and-death chains.

jjj See also Diaconis and Saloff-Coste Metropolis paper, which mentions work of Diaconis students on Metropolizing birth-and-death chains.

xxx examples of particular **w**. **jjj** might just bring up as needed?

xxx contraction principle and lower bounds on τ_2 (relating to current Section 6 of Chapter 4)

By Chapter 4 Lemma yyy,

$$\tau_c = \max_{1 \le i \le n-1} \frac{\pi[0, i-1]\pi[i, n-1]}{w_i}. \quad (5.18)$$

5.1.3 Useful examples of one-dimensional chains

Example 5.4 *The two-state chain.*

This is the birth-and-death chain on $\{0, 1\}$ with $p_{01} = 1 - p_{00} = p$ and $p_{10} = 1 - p_{11} = q$, where $0 < p < 1$ and $0 < q < 1$ are arbitrarily specified. Since p_{00} and p_{11} are positive, this does not quite fit into the framework of Section 5.1.2, but everything is nonetheless easy to calculate. The stationary distribution is given by

$$\pi_0 = q/(p+q), \quad \pi_1 = p/(p+q).$$

In discrete time, the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$, and in the notation of Chapter 3, Section yyy for the spectral representation, the matrix S has $s_{11} = 1 - p$, $s_{22} = 1 - q$, and $s_{12} = s_{21} = (pq)^{1/2}$ with normalized right eigenvectors

$$u_1 = [(q/(p+q))^{1/2}, (p/(p+q))^{1/2}]^T, \quad u_2 = [(p/(p+q))^{1/2}, -(q/(p+q))^{1/2}]^T.$$

The transition probabilities are given by

$$\begin{aligned} P_0(X_t = 1) &= 1 - P_0(X_t = 0) = \frac{p}{p+q}[1 - (1-p-q)^n], \\ P_1(X_t = 0) &= 1 - P_1(X_t = 1) = \frac{q}{p+q}[1 - (1-p-q)^n] \end{aligned}$$

in discrete time and by

$$\begin{aligned} P_0(X_t = 1) &= 1 - P_0(X_t = 0) = \frac{p}{p+q}[1 - e^{-(p+q)t}], \\ P_1(X_t = 0) &= 1 - P_1(X_t = 1) = \frac{q}{p+q}[1 - e^{-(p+q)t}] \end{aligned}$$

in continuous time. It is routine to calculate $E_0T_1 = 1/p$, $E_1T_0 = 1/q$, and

$$\bar{d}(t) = e^{-(p+q)t}, \quad d(t) = \max(p/(p+q), q/(p+q))\, e^{-(p+q)t},$$

and then

$$\max_{ij} E_iT_j = \max(E_0T_1, E_1T_0) = \frac{1}{\min(p,q)}, \quad \tau^* = E_0T_1 + E_1T_0 = \frac{p+q}{pq},$$

and

$$\tau_0 = \tau_1 = \tau_2 = \tau_c = 1/(p+q).$$

Example 5.5 *Biased random walk with reflecting barriers.*

We consider the chain on $\{0, 1, \ldots, n-1\}$ with reflecting barriers at 0 and $n-1$ that at each unit of time moves distance 1 rightward with probability p and distance 1 leftward with probability $q = 1 - p$. Formally, the setting is that of Section 5.1.2 with

$$w_i = \rho^{i-1}, \quad w = \frac{2(1-\rho^{n-1})}{1-\rho} \to \frac{2}{1-\rho},$$

where we assume $\rho \equiv p/q < 1$ and all asymptotics developed for this example are for fixed ρ and large n. If $\rho \neq 1$, there is by symmetry no loss of generality in assuming $\rho < 1$, and the case $\rho = 1$ will be treated later in Example 5.8.

Specializing the results of Section 5.1.2 to the present example, one can easily derive the asymptotic results

$$\max_{ij} E_i T_j \sim \tau^* \sim E_\pi T_{n-1} \sim 2\rho^{-(n-2)}/(1-\rho)^2 \tag{5.19}$$

and, by use of (5.15),

$$\tau_0 \sim \frac{1+\rho}{1-\rho} n. \tag{5.20}$$

For τ_c, the maximizing i in (5.18) equals $(1+o(1))n/2$, and this leads to

$$\tau_c \to (1+\rho)/(1-\rho). \tag{5.21}$$

The spectral representation can be obtained using the orthogonal polynomial techniques described in Karlin and Taylor [209] Chapter 10; see especially Section 5(b) there. The reader may verify that the eigenvalues of $\mathbf{P}$ in discrete time are 1, -1, and, for $m = 1, \ldots, n-2$,

$$\frac{2\rho^{1/2}}{1+\rho} \cos\theta_m, \text{ where } \theta_m \equiv \frac{m\pi}{n-1}$$

with (unnormalized) right eigenvector

$$\rho^{-i/2} \left[2\cos(i\theta_m) - (1-\rho)\frac{\sin((i+1)\theta_m)}{\sin(\theta_m)} \right], \quad i = 0, \ldots, n-1.$$

In particular,

$$\tau_2 = \left[1 - \frac{2\rho^{1/2}}{1+\rho} \cos\left(\frac{\pi}{n-1}\right) \right]^{-1} \to \frac{1+\rho}{(1-\rho^{1/2})^2}. \tag{5.22}$$

The random walk has drift $p-q = -(1-\rho)/(1+\rho) \equiv -\mu$. It is not hard to show for fixed $t > 0$ that the distances $\bar{d}_n(tn)$ and $d_n(tn)$ of Chapter 4 yyy converge to 1 if $t < \mu$ and to 0 if $t > \mu$.

jjj include details? In fact, the cutoff occurs at $\mu n + c_\rho n^{1/2}$: cf. (e.g.) Example 4.46 in [115]. Continue same paragraph:

In particular,

$$\tau_1 \sim \frac{1-\rho}{1+\rho} n \tag{5.23}$$

Example 5.6 *The $M/M/1$ queue.*

We consider the M/M/1/$(n-1)$ queue. Customers queue up at a facility to wait for a single server (hence the "1") and are handled according to a "first come, first served" queuing discipline. The first "M" specifies that the arrival point process is Markovian, i.e., a Poisson process with intensity parameter λ (say); likewise, the second "M" reflects our assumption that the service times are exponential with parameter μ (say). The parameter $n-1$ is the queue size limit; customers arriving when the queue is full are turned away.

We have described a continuous-time birth-and-death process with constant birth and death rates λ and μ, respectively. If $\lambda+\mu=1$, this is nearly the continuized biased random walk of Example 5.5, the only difference being in the boundary behavior. In particular, one can check that the asymptotics in (5.19)–(5.23) remain unchanged, where $\rho \equiv \lambda/\mu$, called the *traffic intensity*, remains fixed and n becomes large. For the M/M/1/$(n-1)$ queue, the stationary distribution is the conditional distribution of $G-1$ given $G \le n$, where G has the Geometric$(1-\rho)$ distribution. The eigenvalues are 1 and, for $m = 1, \ldots, n-1$,

$$\frac{2\rho^{1/2}}{1+\rho}\cos\theta_m, \text{ where now } \theta_m \equiv \frac{m\pi}{n}$$

with (unnormalized) right eigenvector

$$\frac{2\rho^{-i/2}}{1+\rho}\left[\cos(i\theta_m) + (\rho^{1/2}\cos\theta_m - 1)\frac{\sin((i+1)\theta_m)}{\sin(\theta_m)}\right], \quad i = 0, \ldots, n-1.$$

5.2 Special graphs

In this section we record results about some specific easy-to-analyze graphs. As in Section 5.1.3, we focus on the parameters $\tau^*, \tau_0, \tau_1, \tau_2, \tau_c$ discussed in Chapter 4; orders of magnitudes of these parameters (in the asymptotic setting discussed with each example) are summarized in terms of n, the number of vertices, in the following table. A minor theme is that some of the graphs are known or conjectured to be extremal for our parameters. In the context of extremality we ignore the parameter τ_1 since its exact definition is a little arbitrary.

jjj David: (1) Shall I add complete bipartite to table? (2) Please fill in missing entries for torus.

Orders of magnitude of parameters [$\tau = \Theta$(entry)] for special graphs.

Example	τ^*	τ_0	τ_1	τ_2	τ_c
5.7. cycle	n^2	n^2	n^2	n^2	n
5.8. path	n^2	n^2	n^2	n^2	n
5.9. complete graph	n	n	1	1	1
5.10. star	n	n	1	1	1
5.11. barbell	n^3	n^3	n^3	n^3	n^2
5.12. lollipop	n^3	n^2	n^2	n^2	n
5.13. necklace	n^2	n^2	n^2	n^2	n
5.14. balanced r-tree	$n \log n$	$n \log n$	n	n	n
5.15. d-cube ($d = \log_2 n$)	n	n	$d \log d$	d	d
5.16. dense regular graphs	n	n	1	1	1
5.17. d-dimensional torus					
$d = 2$	**jjj**?	$n \log n$	$n^{2/d}$	$n^{2/d}$	**jjj**?$n^{1/d}$
$d \geq 3$	**jjj**?	n	$n^{2/d}$	$n^{2/d}$	**jjj**?$n^{1/d}$
5.19. rook's walk	n	n	1	1	1

In simpler cases we also record the t-step transition probabilities $P_i(X_t = j)$ in discrete and continuous time. In fact one could write out exact expressions for $P_i(X_t = j)$ and indeed for hitting time distributions in almost all these examples, but complicated exact expressions are seldom very illuminating.

qqq names of graphs vary—suggestions for "standard names" from readers of drafts are welcome.

Example 5.7 *The n-cycle.*

This is just the graph $0 - 1 - 2 - \cdots - (n-1) - 0$ on n vertices. By rotational symmetry, it is enough to give formulas for random walk started at 0. If $(\hat{X}_t)$ is random walk on (all) the integers, then $X_t = \phi(\hat{X}_t)$ is random walk on the n-cycle, for

$$\phi(i) = i \bmod n.$$

Thus results for the n-cycle can be deduced from results for the integers. For instance,

$$E_0 T_i = i(n - i) \tag{5.24}$$

by Lemma 5.2, because this is the mean exit time from $(i-n, i)$ for random walk on the integers. We can now calculate

$$\max_{ij} E_i T_j \quad = \quad \lfloor n^2/4 \rfloor$$

$$\tau^* \equiv \max_{ij}(E_iT_j + E_jT_i) = 2\lfloor n^2/4 \rfloor \quad (5.25)$$

$$\tau_0 = n^{-1}\sum_j E_0T_j = (n^2-1)/6 \quad (5.26)$$

where for the final equality we used the formula

$$\sum_{m=1}^{n} m^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}.$$

As at (5.9) and (5.10) we can get an expression for the distribution at time t from the Binomial distribution (in discrete time) or the Poisson distribution (in continuous time). The former is

$$P_0(X_t = i) = \sum_{j:2j-t=i \bmod n} \frac{t!}{j!(t-j)!}2^{-t}.$$

A more useful expression is obtained from the spectral representation. The n eigenvalues of the transition matrix are $\cos(2\pi m/n)$, $0 \le m \le n-1$. That is, 1 and (if n is even) -1 are simple eigenvalues, with respective normalized eigenvectors $u_{i0} = 1/\sqrt{n}$ and $u_{i,n/2} = (-1)^i/\sqrt{n}$ $(0 \le i \le n-1)$. The multiplicity of $\cos(2\pi m/n)$ is 2 for $0 < m < n/2$; the corresponding normalized eigenvectors are $u_{im} = \sqrt{2/n}\cos(2\pi im/n)$ and $u_{i,-m} = \sqrt{2/n}\sin(2\pi im/n)$ $(0 \le i \le n-1)$. Thus

$$P_0(X_t = i) = \frac{1}{n}\sum_{m=0}^{n-1}(\cos(2\pi m/n))^t \cos(2\pi im/n),$$

a fact most easily derived using Fourier analysis.

jjj Cite Diaconis book [112]? Continue same paragraph:

So the relaxation time is

$$\tau_2 = \frac{1}{1-\cos(2\pi/n)} \sim \frac{n^2}{2\pi^2}.$$

As an aside, note that the eigentime identity (Chapter 3 yyy) gives the curious identity

$$\frac{n^2-1}{6} = \sum_{m=1}^{n-1}\frac{1}{1-\cos(2\pi m/n)}$$

whose $n \to \infty$ limit is the well-known formula $\sum_{m=1}^{\infty} m^{-2} = \pi^2/6$.

If n is even, the discrete-time random walk is periodic. This parity problem vanishes in continuous time, for which we have the formula

$$P_0(X(t) = i) = \frac{1}{n} \sum_{m=0}^{n-1} \exp(-t(1 - \cos(2\pi m/n))) \cos(2\pi i m/n). \qquad (5.27)$$

Turning to total variation convergence, we remain in continuous time and consider the distance functions $\bar{d}_n(t)$ and $d_n(t)$ of Chapter 4 yyy. The reader familiar with the notion of weak convergence of random walks to Brownian motion (on the circle, in this setting) will see immediately that

$$\bar{d}_n(tn^2) \to \bar{d}_\infty(t)$$

where the limit is "$\bar{d}$ for Brownian motion on the circle", which can be written as

$$\bar{d}_\infty(t) \equiv 1 - 2P((t^{1/2}Z) \bmod 1 \ \in (1/4, 3/4))$$

where Z has the standard Normal distribution. So

$$\tau_1 \sim cn^2$$

for the constant c such that $\bar{d}_\infty(c) = e^{-1}$, whose numerical value $c \doteq 0.063$ has no real significance.

jjj David: You got 0.054. Please check. Continue same paragraph:

Similarly

$$d_n(tn^2) \to d_\infty(t) \equiv \frac{1}{2} \int_0^1 |f_t(u) - 1| \, du,$$

where f_t is the density of $(t^{1/2}Z) \bmod 1$.

As for τ_c, the *sup* in its definition is attained by some A of the form $[0, i-1]$, so

$$\tau_c = \max_i \frac{\frac{i}{n}(1 - \frac{i}{n})}{1/n} = \frac{1}{n} \left\lfloor \frac{n^2}{4} \right\rfloor \sim \frac{n}{2}.$$

As remarked in Chapter 4 yyy, this provides a counter-example to reversing inequalities in Theorem yyy. But if we consider $\max_A(\pi(A)E_\pi T_A)$, the *max* is attained with $A = [\frac{n}{2} - \alpha n, \frac{n}{2} + \alpha n]$, say, where $0 \le \alpha < 1/2$. By Lemma 5.2, for $x \in (-\frac{1}{2} + \alpha, \frac{1}{2} - \alpha)$,

$$E_{\lfloor (x \bmod 1)n \rfloor} T_A \sim \left(\tfrac{1}{2} - \alpha - x\right)\left(\tfrac{1}{2} - \alpha + x\right) n^2,$$

and so

$$E_\pi T_A \sim n^2 \int_{-\frac{1}{2}+\alpha}^{\frac{1}{2}-\alpha} \left(\frac{1}{2} - \alpha - x\right)\left(\frac{1}{2} - \alpha + x\right) dx = \frac{4(\frac{1}{2} - \alpha)^3 n^2}{3}.$$

Thus

$$\max_A(\pi(A)E_\pi T_A) \sim n^2 \sup_{0<\alpha<1/2} \frac{4(\frac{1}{2}-\alpha)^3\ 2\alpha}{3} = \frac{9n^2}{512},$$

consistent with Chapter 4 Open Problem yyy.

xxx level of detail for $\bar{d}$ results, here and later.

Remark. Parameters τ^*, τ_0, τ_1, and τ_2 are all $\Theta(n^2)$ in this example, and in Chapter 6 we'll see that they are $O(n^2)$ over the class of regular graphs. However, the exact maximum values over all n-vertex regular graphs (or the constants c in the $\sim cn^2$ asymptotics) are not known. See Chapter 6 for the natural conjectures.

Example 5.8 *The n-path.*

This is just the graph $0 - 1 - 2 - \cdots - (n-1)$ on n vertices. If $(\hat{X}_t)$ is random walk on (all) the integers, then $X_t = \phi(\hat{X}_t)$ is random walk on the n-path, for the "concertina" map

$$\phi(i) = \begin{cases} i & \text{if } i \bmod 2(n-1) \le n-1 \\ 2(n-1) - (i \bmod 2(n-1)) & \text{otherwise.} \end{cases}$$

Of course the stationary distribution is not quite uniform:

$$\pi_i = \frac{1}{n-1},\ 1 \le i \le n-2; \quad \pi_0 = \pi_{n-1} = \frac{1}{2(n-1)}.$$

Again, results for the n-path can be deduced from results for the integers. Using Lemma 5.2,

$$E_iT_j = (j-i)(j+i),\ 0 \le i < j \le n-1. \tag{5.28}$$

¿From this, or from the more general results in Section 5.1.2, we obtain

$$\max_{ij} E_iT_j = (n-1)^2 \tag{5.29}$$

$$\tau^* \equiv \max_{ij}(E_iT_j + E_jT_i) = 2(n-1)^2 \tag{5.30}$$

$$\tau_0 = \sum_j \pi_j E_0T_j = \frac{1}{3}(n-1)^2 + \frac{1}{6} \tag{5.31}$$

We can also regard X_t as being derived from random walk $\tilde{X}_t$ on the $(2n-2)$-cycle via $X_t = \min(\tilde{X}_t, 2n-2-\tilde{X}_t)$. So we can deduce the spectral representation from that in the previous example:

$$P_i(X_t = j) = \sqrt{\pi_j/\pi_i} \sum_{m=0}^{n-1} \lambda_m^t u_{im}u_{jm}$$

where, for $0 \le m \le n-1$,

$$\lambda_m = \cos(\pi m/(n-1))$$

and

$$u_{0m} = \sqrt{\pi_m}; \quad u_{n-1,m} = \sqrt{\pi_m}(-1)^m;$$
$$u_{im} = \sqrt{2\pi_m}\cos(\pi im/(n-1)), \quad 1 \le i \le n-2.$$

In particular, the relaxation time is

$$\tau_2 = \frac{1}{1-\cos(\pi/(n-1))} \sim \frac{2n^2}{\pi^2}.$$

Furthermore, $\bar{d}_n(t) = \bar{\tilde{d}}_{2n-2}(t)$ and $d_n(t) = \tilde{d}_{2n-2}(t)$ for all t, so

$$\bar{d}_n(t(2n)^2) \to \bar{d}_\infty(t)$$

$$d_n(t(2n)^2) \to d_\infty(t)$$

where the limits are those in the previous example. Thus $\tau_1 \sim cn^2$, where $c \doteq 0.25$ is 4 times the corresponding constant for the n-cycle.

xxx explain: BM on $[0,1]$ and circle described in Chapter 16.

It is easy to see that

$$\tau_c = \begin{cases} \frac{n-1}{2} & \text{if } n \text{ is even} \\ \frac{n-1}{2} - \frac{1}{2(n-1)} & \text{if } n \text{ is odd} \end{cases}$$

In Section 5.3.2 we will see that the n-path attains the exact maximum values of our parameters over all n-vertex trees.

Example 5.9 *The complete graph.*

For the complete graph on n vertices, the t-step probabilities for the chain started at i can be calculated by considering the induced 2-state chain which indicates whether or not the walk is at i. This gives, in discrete time,

$$\begin{aligned} P_i(X_t = i) &= \frac{1}{n} + \left(1 - \frac{1}{n}\right)\left(-\frac{1}{n-1}\right)^t \\ P_i(X_t = j) &= \frac{1}{n} - \frac{1}{n}\left(-\frac{1}{n-1}\right)^t, \quad j \ne i \end{aligned} \tag{5.32}$$

and, in continuous time,

$$\begin{aligned} P_i(X_t = i) &= \frac{1}{n} + \left(1 - \frac{1}{n}\right) \exp\left(-\frac{nt}{n-1}\right) \\ P_i(X_t = j) &= \frac{1}{n} - \frac{1}{n} \exp\left(-\frac{nt}{n-1}\right), \quad j \neq i \end{aligned} \tag{5.33}$$

It is clear that the hitting time to $j \neq i$ has Geometric$(1/(n-1))$ distribution (in continuous time, Exponential$(1/(n-1))$ distribution), and so in particular

$$E_i T_j = n - 1, \ j \neq i. \tag{5.34}$$

Thus we can calculate the parameters

$$\tau^* \equiv \max_{ij}(E_j T_i + E_i T_j) = 2(n-1) \tag{5.35}$$

$$\max_{ij} E_i T_j = n - 1 \tag{5.36}$$

$$\tau_0 \equiv n^{-1} \sum_j E_i T_j = (n-1)^2/n. \tag{5.37}$$

From (5.32) the discrete-time eigenvalues are $\lambda_2 = \lambda_3 = \cdots = \lambda_n = -1/(n-1)$. So the relaxation time is

$$\tau_2 = (n-1)/n. \tag{5.38}$$

The total variation functions are

$$\bar{d}(t) = \exp\left(-\frac{nt}{n-1}\right), \quad d(t) = \frac{n-1}{n} \exp\left(-\frac{nt}{n-1}\right),$$

so

$$\tau_1 = (n-1)/n. \tag{5.39}$$

It is easy to check

$$\tau_c = (n-1)/n.$$

We have already proved (Chapter 3 yyy) that the complete graph attains the exact minimum of $\tau^*, \max_{ij} E_i T_j, \tau_0$, and τ_2 over all n-vertex graphs. The same holds for τ_c, by considering (in an arbitrary graph) a vertex of minimum degree.

Example 5.10 *The n-star.*

This is the graph on $n \geq 3$ vertices $\{0, 1, 2, \ldots, n-1\}$ with edges 0 – 1, 0 – 2, 0 – 3, …, 0 – $(n-1)$. The stationary distribution is

$$\pi_0 = 1/2,\ \pi_i = 1/(2(n-1)),\quad i \geq 1.$$

In discrete time the walk is periodic. Starting from the leaf 1, the walk at even times is simply independent and uniform on the leaves, so

$$P_1(X_{2t} = i) = 1/(n-1),\ i \geq 1$$

for $t \geq 1$. At odd times, the walk is at 0. Writing these t-step probabilities as

$$P_1(X_t = i) = \frac{1}{2(n-1)}(1 + (-1)^t)1_{(i\geq 1)} + \frac{1}{2}(1 + (-1)^{t+1})1_{(i=0)},\quad t \geq 1$$

we see that the discrete-time eigenvalues are $\lambda_2 = \cdots = \lambda_{n-1} = 0,\ \lambda_n = -1$ and hence the relaxation time is

$$\tau_2 = 1.$$

The mean hitting times are

$$E_1 T_0 = 1$$
$$E_1 T_j = 2(n-1),\quad j \geq 2,$$

where the latter comes from the fact that $T_j/2$ has Geometric$(1/(n-1))$ distribution, using the "independent uniform on leaves at even times" property. Then

$$E_0 T_1 = 2n - 3.$$

We can calculate the parameters

$$\tau^* \equiv \max_{ij}(E_i T_j + E_j T_i) = 4n - 4 \tag{5.40}$$
$$\max_{ij} E_i T_j = 2n - 2 \tag{5.41}$$
$$\tau_0 = \sum_j E_0 T_j \pi_j = n - \tfrac{3}{2}. \tag{5.42}$$

In continuous time we find

$$\begin{aligned}
P_1(X_t = 1) &= \frac{1}{2(n-1)}(1 + e^{-2t}) + \frac{n-2}{n-1}e^{-t} \\
P_1(X_t = i) &= \frac{1}{2(n-1)}(1 + e^{-2t}) - \frac{1}{n-1}e^{-t},\ i > 1 \\
P_1(X_t = 0) &= \tfrac{1}{2}(1 - e^{-2t}) \\
P_0(X_t = 0) &= \tfrac{1}{2}(1 + e^{-2t}) \\
P_0(X_t = 1) &= \frac{1}{2(n-1)}(1 - e^{-2t})
\end{aligned}$$

This leads to

$$\bar{d}(t) = e^{-t}, \quad d(t) = \frac{1}{2(n-1)}e^{-2t} + \frac{n-2}{n-1}e^{-t},$$

from which

$$\tau_1 = 1.$$

Clearly (isolate a leaf)

$$\tau_c = 1 - \frac{1}{2(n-1)}.$$

We shall see in Section 5.3.2 that the n-star attains the exact minimum of our parameters over all n-vertex trees.

Example 5.11 *The barbell.*

Here is a graph on $n = 2m_1 + m_2$ vertices ($m_1 \geq 2, m_2 \geq 0$). Start with two complete graphs on m_1 vertices. Distinguish vertices $v_l \neq v_L$ in one graph ("the left bell") and vertices $v_R \neq v_r$ in the other graph ("the right bell"). Then connect the graphs via a path $v_L - w_1 - w_2 - \cdots - w_{m_2} - v_R$ containing m_2 new vertices.

xxx picture

A point of the construction is that the mean time to go from a typical point v_l in the left bell to a typical point v_r in the right bell is roughly $m_1^2 m_2$. To argue this informally, it takes mean time about m_1 to hit v_L; then there is chance $1/m_1$ to hit w_1, so it takes mean time about m_1^2 to hit w_1; and from w_1 there is chance about $1/m_2$ to hit the right bell before returning into the left bell, so it takes mean time about $m_1^2 m_2$ to enter the right bell.

The exact result, argued below, is

$$\max_{ij} E_i T_j = E_{v_l} T_{v_r} = m_1(m_1-1)(m_2+1) + (m_2+1)^2 + 4(m_1-1) + 4\frac{m_2+1}{m_1}. \tag{5.43}$$

It is cleaner to consider asymptotics as

$$n \to \infty, \ m_1/n \to \alpha, \ m_2/n \to 1 - 2\alpha$$

with $0 < \alpha < 1/2$. Then

$$\begin{aligned} \max_{ij} E_i T_j &\sim \alpha^2(1-2\alpha)n^3 \\ &\sim \frac{n^3}{27} \text{ for } \alpha = 1/3 \end{aligned}$$

where $\alpha = 1/3$ is the asymptotic maximizer here and for the other parameters below. Similarly

$$\begin{aligned} \tau^* &\sim 2\alpha^2(1-2\alpha)n^3 \\ &\sim \frac{2n^3}{27} \text{ for } \alpha = 1/3. \end{aligned}$$

The stationary distribution π puts mass $\to 1/2$ on each bell. Also, by (5.45)–(5.47) below, $E_{v_l}T_v = o(E_{v_l}T_{v_r})$ uniformly for vertices v in the left bell and $E_{v_l}T_v \sim E_{v_l}T_{v_r} \sim \alpha^2(1-2\alpha)n^3$ uniformly for vertices v in the right bell. Hence

$$\tau_0 \equiv \sum_v \pi_v E_{v_l}T_v \sim \frac{1}{2}E_{v_l}T_{v_r} \sim \frac{1}{2}\alpha^2(1-2\alpha)n^3$$

and so we have proved the "τ_0" part of

$$\begin{aligned} \text{each of } \{\tau_0, \tau_1, \tau_2\} &\sim \tfrac{1}{2}\alpha^2(1-2\alpha)n^3 \\ &\sim \frac{n^3}{54} \text{ for } \alpha = 1/3. \end{aligned} \tag{5.44}$$

Consider the relaxation time τ_2. For the function g defined to be $+1$ on the left bell, -1 on the right bell and linear on the bar, the Dirichlet form gives

$$\mathcal{E}(g,g) = \frac{2}{(m_2+1)(m_1(m_1-1)+m_2+1)} \sim \frac{2}{\alpha^2(1-2\alpha)n^3}.$$

Since the variance of g tends to 1, the extremal characterization of τ_2 shows that $\frac{1}{2}\alpha^2(1-2\alpha)n^3$ is an asymptotic lower bound for τ_2. But in general $\tau_2 \le \tau_0$, so having already proved (5.44) for τ_0 we must have the same asymptotics for τ_2. Finally, without going into details, it is not hard to show that for fixed $t > 0$,

$$\bar{d}_n\left(\frac{1}{2}\alpha^2(1-2\alpha)n^3 t\right) \to e^{-t},\ d_n\left(\frac{1}{2}\alpha^2(1-2\alpha)n^3 t\right) \to \frac{1}{2}e^{-t}$$

from which the "τ_1" assertion of (5.44) follows.

jjj Proof? (It's not so terrifically easy, either! How much do we want to include?) I've (prior to writing this) carefully written out an argument similar to the present one, also involving approximate exponentiality of a hitting time distribution, for the balanced r-tree below. Here is a rough sketch for the argument for $\bar{d}$ here; note that it uses results about the next example (the lollipop). (The argument for d is similar.) The pair (v_l, v_r) of initial states achieves $\bar{d}(t)$ for every t (although the following can be

made to work without knowing this "obvious fact" a priori). Couple chains starting in these states by having them move symmetrically in the obvious fashion. Certainly these copies will couple by the time T the copy started at v_l has reached the center vertex $w_{m_2/2}$ of the bar. We claim that the distribution of T is approximately exponential, and its expected value is $\sim \frac{1}{2}m_1^2 m_2 \sim \frac{1}{2}\alpha^2(1-2\alpha)n^3$ by the first displayed result for the lollipop example, with m_2 changed to $m_2/2$ there. (In keeping with this observation, I'll refer to the "half-stick" lollipop in the next paragraph.)

jjj (cont.) To get approximate exponentiality for the distribution of T, first argue easily that it's approximately the same as that of $T_{w_{m/2}}$ for the half-stick lollipop started in stationarity. But that distribution is, in turn, approximately exponential by Chapter 3 Proposition yyy, since $\tau_2 = \Theta(n^2) = o(n^3)$ for the half-stick lollipop. ■

Proof of (5.43). The mean time in question is the sum of the following mean times:

$$E_{v_l}T_{v_L} = m_1 - 1 \tag{5.45}$$

$$E_{v_L}T_{v_R} = m_1(m_1-1)(m_2+1) + (m_2+1)^2 \tag{5.46}$$

$$E_{v_R}T_{v_r} = 3(m_1-1) + 4\frac{m_2+1}{m_1}. \tag{5.47}$$

Here (5.45) is just the result (5.34) for the complete graph. And (5.46) is obtained by summing over the edges of the "bar" the expression

$$E_{w_i}T_{w_{i+1}} = m_1(m_1-1) + 2i + 1, \quad i = 0, \ldots, m_2 \tag{5.48}$$

obtained from the general formula for mean hitting time across an essential edge of a graph (Lemma 5.1), where $w_0 = v_L$ and $w_{m_2+1} = v_R$. To argue (5.47), we start with the 1-step recurrence

$$E_{v_R}T_{v_r} = 1 + \frac{1}{m_1}E_{w_{m_2}}T_{v_r} + \frac{m_1-2}{m_1}E_xT_{v_r}$$

where x denotes a vertex of the right bell distinct from v_R and v_r. Now

$$E_{w_{m_2}}T_{v_r} = m_1(m_1-1) + 2m_2 + 1 + E_{v_R}T_{v_r}$$

using the formula (5.48) for the mean passage time from w_{m_2} to v_R. Starting from x, the time until a hit on either v_R or v_r has Geometric$(2/(m_1-1))$ distribution, and the two vertices are equally likely to be hit first. So

$$E_xT_{v_r} = (m_1-1)/2 + \tfrac{1}{2}E_{v_R}T_{v_r}.$$

The last three expressions give an equation for $E_{v_R}T_{v_r}$ whose solution is (5.47). And it is straightforward to check that $E_{v_l}T_{v_r}$ does achieve the maximum, using (5.45)–(5.47) to bound the general E_iT_j. ∎

It is straightforward to check

$$\tau_c \sim \frac{\alpha^2 n^2}{2}.$$

Example 5.12 *The lollipop.*

xxx picture

This is just the barbell without the right bell. That is, we start with a complete graph on m_1 vertices and add m_2 new vertices in a path. So there are $n = m_1 + m_2$ vertices, and w_{m_2} is now a leaf. In this example, by (5.45) and (5.46), with m_2 in place of $m_2 + 1$, we have

$$\max_{ij} E_iT_j = E_{v_l}T_{w_{m_2}} = m_1(m_1 - 1)m_2 + (m_1 - 1) + m_2^2.$$

In the asymptotic setting with

$$n \to \infty, m_1/n \to \alpha, m_2/n \to 1 - \alpha$$

where $0 < \alpha < 1$, we have

$$\begin{aligned} \max_{ij} E_iT_j &\sim \alpha^2(1-\alpha)n^3 \qquad (5.49)\\ &\sim \frac{4n^3}{27} \text{ for } \alpha = 2/3, \end{aligned}$$

where $\alpha = 2/3$ gives the asymptotic maximum.

Let us discuss the other parameters only briefly, in the asymptotic setting. Clearly $E_{w_{m_2}}T_{v_L} = m_2^2 \sim (1-\alpha)^2 n^2$ and it is not hard to check

$$E_{w_{m_2}}T_{v_l} = \max_v E_vT_{v_l} \sim (1-\alpha)^2 n^2, \qquad (5.50)$$

whence

$$\tau^* = \max_{ij}(E_iT_j + E_jT_i) = E_{v_l}T_{w_{m_2}} + E_{w_{m_2}}T_{v_l} \sim \alpha^2(1-\alpha)n^3.$$

Because the stationary distribution puts mass $\Theta(1/n)$ on the "bar", (5.50) is also enough to show that $\tau_0 = O(n^2)$. So by the general inequalities between our parameters, to show

$$\text{each of } \{\tau_0, \tau_1, \tau_2\} = \Theta(n^2) \qquad (5.51)$$

it is enough to show that $\tau_2 = \Omega(n^2)$. But for the function g defined to be 0 on the "bell", 1 at the end w_{m_2} of the "bar," and linear along the bar, a brief calculation gives

$$\mathcal{E}(g,g) = \Theta(n^{-3}), \quad \text{var}_{\pi} g = \Theta(n^{-1})$$

so that $\tau_2 \geq (\text{var}_{\pi} g)/\mathcal{E}(g,g) = \Omega(n^2)$, as required.

Finally, in the asymptotic setting it is straightforward to check that τ_c is achieved by $A = \{w_1, \ldots, w_{m_2}\}$, giving

$$\tau_c \sim 2(1-\alpha)n.$$

Remark. The barbell and lollipop are the natural candidates for the n-vertex graphs which maximize our parameters. The precise conjectures and known results will be discussed in Chapter 6.

jjj We need to put somewhere—Chapter 4 on τ_c? Chapter 6 on **max** parameters over n-vertex graphs? in the barbell example?—the fact that $\mathbf{max}\, \tau_c$ is attained, when n is even, by the barbell with $m_2 = 0$, the **max** value being $(n^2-2n+2)/8 \sim n^2/8$. Similarly, when n is odd, the **max**imizing graph is formed by joining complete graphs on $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ vertices respectively by a single edge, and the **max** value is easy to write down (I've kept a record) but not so pretty; however, this value too is $\sim n^2/8$, which is probably all we want to say anyway. Here is the first draft of a proof:

For random walk on an unweighted graph, τ_c is the maximum over nonempty proper subsets A of the ratio

$$\frac{(\deg A)(\deg A^c)}{2|\mathcal{E}|(A, A^c)}, \tag{5.52}$$

where $\deg A$ is defined to be the sum of the degrees of vertices in A and (A, A^c) is the number of directed cut edges from A to A^c.

jjj Perhaps it would be better for exposition to stick with *un*directed edges and introduce factor 1/2?

Maximizing now over choice of graphs, the **max** in question is no larger than the maximum M, over all choices of $n_1 > 0$, $n_2 > 0$, e_1, e_2, and e' satisfying $n_1 + n_2 = n$ and $0 \leq e_i \leq \binom{n_i}{2}$ for $i = 1, 2$ and $1 \leq e' \leq n_1 n_2$, of the ratio

$$\frac{(2e_1 + e')(2e_2 + e')}{2(e_1 + e_2 + e')e'}. \tag{5.53}$$

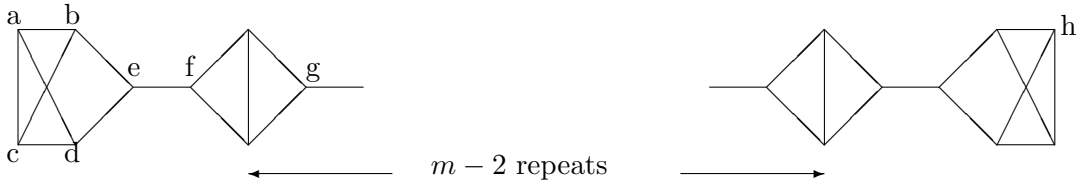
(We don't claim equality because we don't check that each n_i-graph is connected. But we'll show that M is in fact achieved by the connected graph claimed above.)

Simple calculus shows that the ratio (5.53) is (as one would expect) increasing in e_1 and e_2 and decreasing in e'. Thus, for given n_1, (5.53) is maximized by considering complete graphs of size n_1 and $n_2 = n - n_1$ joined by a single edge. Call the maximum value $M(n_1)$. If n is even, it is then easy to see that M_{n_1} is maximized by $n_1 = n/2$, giving $M = (n^2 - 2n + 2)/8$, as desired.

For the record, here are the slightly tricky details if n is odd. Write $\nu = n/2$ and $n_1 = \nu - y$ and put $x = y^2$. A short calculation gives $M(n_1) = 1 + g(x)$, where $g(x) \equiv [(a-x)(b-x)-1]/(2x+c)$ with $a = \nu^2$, $b = (\nu-1)^2$, and $c = 2\nu(\nu-1)+2$. Easy calculus shows that g is U-shaped over $[0, \nu]$ and then that $g(1/4) \geq g(\nu^2)$. Thus $M(n_1)$ is maximized when $n_1 = \nu - \frac{1}{2} = \lfloor n/2 \rfloor$. ■

Example 5.13 *The necklace.*

This graph, pictured below, is 3-regular with $n = 4m+2$ vertices, consisting of m subgraphs linked in a line, the two end subgraphs being different from the intervening ones. This is an artificial graph designed to mimic the n-path while being regular, and this construction (or some similar one) is the natural candidate for the n-vertex graph which asymptotically maximizes our parameters over regular graphs.



This example affords a nice illustration of use of the commute interpretation of resistance. Applying voltage 1 at vertex a and voltage 0 at e, a brief calculation gives the potentials at intervening vertices as

$$g(b) = 4/7,\ g(c) = 5/7,\ g(d) = 4/7$$

and gives the effective resistance $r_{ae} = 7/8$. Since the effective resistance between f and g equals 1, we see the maximal effective resistance is

$$r_{ah} = \tfrac{7}{8} + (2m-3) + \tfrac{7}{8} = 2m - \tfrac{5}{4}.$$

So

$$\tau^* = E_a T_h + E_h T_a = 3 \times (4m+2) \times \left(2m - \frac{5}{4}\right) \sim \frac{3n^2}{2}.$$

One could do elementary exact calculations of other parameters, but it is simpler to get asymptotics from the Brownian motion limit, which implies that the asymptotic ratio of each parameter (excluding τ_c) in this example and the n-path is the same for each parameter. So

$$\tau_0 \sim \frac{n^2}{4},\ \tau_2 \sim \frac{3n^2}{2\pi^2}.$$

jjj I haven't checked this carefully, and I also have abstained from writing anything further about τ_1.

Finally, it is clear that $\tau_c \sim 3n/4$, achieved by breaking a "link" between "beads" in the middle of the necklace.

Example 5.14 *The balanced r-tree.*

Take $r \geq 2$ and $h \geq 1$. The *balanced r-tree* is the rooted tree where all leaves are at distance h from the root, where the root has degree r, and where the other internal vertices have degree $r+1$. Call h the *height* of the tree. For $h = 1$ we have the $(r+1)$-star, and for $r = 2$ we have the balanced *binary* tree. The number of vertices is

$$n = 1 + r + r^2 + \cdots + r^h = (r^{h+1} - 1)/(r-1).$$

The chain $\hat{X}$ induced (in the sense of Chapter 4 Section yyy) by the function

$$f(i) = h - (\text{distance from } i \text{ to the root})$$

is random walk on $\{0, \ldots, h\}$, biased towards 0, with reflecting barriers, as in Example 5.5 with

$$\rho = 1/r.$$

In fact, the transition probabilities for X can be expressed in terms of $\hat{X}$ as follows. Given vertices v_1 and v_2 with $f(v_1) = f_1$ and $f(v_2) = f_2$, the paths $[\text{root}, v_1]$ and $[\text{root}, v_2]$ intersect in the path $[\text{root}, v_3]$, say, with $f(v_3) = f_3 \geq f_1 \vee f_2$. Then

$$P_{v_1}(X_t = v_2) = \sum_{m=f_3}^{h} P_{f_1}\left(\max_{0\leq s\leq t} \hat{X}_s = m, \hat{X}_t = f_2\right) r^{-(m-f_2)}.$$

As a special case, suppose that v_1 is on the path from the root to v_2; in this case $v_3 = v_1$. Using the essential edge lemma (or Theorem 5.20 below) we can calculate

$$E_{v_2} T_{v_1} = 2(r-1)^{-2}(r^{f_1+1} - r^{f_2+1}) - 2(r-1)^{-1}(f_1 - f_2) - (f_1 - f_2),$$

$$E_{v_1}T_{v_2} = 2(n-1)(f_1 - f_2) - E_{v_2}T_{v_1}. \tag{5.54}$$

Using this special case we can deduce the general formula for mean hitting times. Indeed, $E_{v_1}T_{v_2} = E_{v_1}T_{v_3} + E_{v_3}T_{v_2}$, which leads to

$$\begin{aligned} E_{v_1}T_{v_2} &= 2(n-1)(f_3 - f_2) + 2(r-1)^{-2}(r^{f_2+1} - r^{f_1+1}) \\ &\quad -2(r-1)^{-1}(f_2 - f_1) - (f_2 - f_1). \end{aligned} \tag{5.55}$$

The maximum value $2h(n-1)$ is attained when v_1 and v_2 are leaves and v_3 is the root. So

$$\tfrac{1}{2}\tau^* = \max_{v,x} E_v T_x = 2(n-1)h. \tag{5.56}$$

(The τ^* part is simpler via (5.88) below.) Another special case is that, for a leaf v,

$$E_v T_{\text{root}} = 2(r-1)^{-2}(r^{h+1} - r) - 2h(r-1)^{-1} - h \sim 2n/(r-1), \tag{5.57}$$

$$E_{\text{root}}T_v = 2(n-1)h - E_v T_{\text{root}} \sim 2nh \tag{5.58}$$

where asymptotics are as $h \to \infty$ for fixed r. Since $E_{\text{root}}T_w$ is decreasing in $f(w)$, it follows that

$$\tau_0 = \sum_w \pi_w E_{\text{root}} T_w \le (1 + o(1))2nh.$$

On the other hand, we claim $\tau_0 \ge (1 + o(1))2nh$, so that

$$\tau_0 \sim 2nh.$$

To sketch the proof, given a vertex w, let v be a leaf such that w lies on the path from v to the root. Then

$$E_{\text{root}}T_w = E_{\text{root}}T_v - E_w T_v,$$

and $E_w T_v \le 2(n-1)f(w)$ by (5.54). But the stationary distribution puts nearly all its mass on vertices w with $f(w)$ of constant order, and $n = o(nh)$.

We claim next that

$$\tau_1 \sim \tau_2 \sim 2n/(r-1).$$

Since $\tau_2 \le \tau_1$, it is enough to show

$$\tau_1 \le (1 + o(1))\frac{2n}{r-1} \tag{5.59}$$

and

$$\tau_2 \geq (1 + o(1))\frac{2n}{r-1}. \tag{5.60}$$

Proof of (5.59). Put

$$t_n \equiv \frac{2n}{r-1}$$

for brevity. We begin the proof by recalling the results (5.22) and (5.19) for the induced walk $\hat{X}$:

$$\hat{\tau}_2 \to \frac{(r+1)}{(r^{1/2}-1)^2},$$

$$E_{\hat{\pi}}\hat{T}_h \sim \frac{2r^{h+1}}{(r-1)^2} \sim t_n. \tag{5.61}$$

By Proposition yyy of Chapter 3,

$$\sup_t \left| P_{\hat{\pi}}(\hat{T}_h > t) - \exp\left(-\frac{t}{E_{\hat{\pi}}\hat{T}_h}\right)\right| \leq \frac{\hat{\tau}_2}{E_{\hat{\pi}}\hat{T}_h} = \Theta(n^{-1}) = o(1). \tag{5.62}$$

For $\hat{X}$ started at 0, let $\hat{S}$ be a stopping time at which the chain has exactly the stationary distribution. Then, for $0 \leq s \leq t$,

$$P_0(\hat{T}_h > t) \leq P_0(\hat{S} > s) + P_{\hat{\pi}}(\hat{T}_h > t - s).$$

Since $\hat{\tau}_1^{(2)} = O(h) = O(\log n)$ by (5.23), we can arrange to have $E_0\hat{S} = O(\log n)$. Fixing $\epsilon > 0$ and choosing $t = (1+\epsilon)t_n$ and (say) $s = (\log n)^2$, (5.62) and (5.61) in conjunction with Markov's inequality yield

$$\begin{aligned} P_0(\hat{T}_h > (1+\epsilon)t_n) &= \exp\left[-\frac{(1+\epsilon)t_n - (\log n)^2}{E_{\hat{\pi}}\hat{T}_h}\right] \\ &\quad + O((\log n)^{-1}) + O(n^{-1}) \\ &\to e^{-(1+\epsilon)}. \end{aligned}$$

Returning to the continuous-time walk on the tree, for n sufficiently large we have

$$P_v(T_{\text{root}} > (1+\epsilon)t_n) \leq P_0(\hat{T}_h > (1+\epsilon)t_n) \leq e^{-1}$$

for every vertex v. Now a simple coupling argument (**jjj** spell out details?: Couple the induced walks and the tree-walks will agree when the induced walk starting farther from the origin has reached the origin) shows that

$$\bar{d}_n((1+\epsilon)t_n) \leq e^{-1}$$

for all large n. Hence $\tau_1 \le (1+\epsilon)t_n$ for all large n, and (5.59) follows. ∎

Proof of (5.60).

jjj [This requires further exposition in both Chapters 3 and 4-1. In Chapter 3, it needs to be made clear that one of the inequalities having to do with CM hitting time distributions says precisely that $E_{\alpha_A}T_A \ge E_\pi T_A/\pi(A^c) \ge E_\pi T_A$. In Chapter 4-1 (2/96 version), it needs to be noted that Lemma 2(a) (concerning τ_2 for the joining of two copies of a graph) extends to the joining of any finite number of copies.]

Let G denote a balanced r-tree of height h. Let G'' denote a balanced r-tree of height $h-1$ with root y and construct a tree G' from G'' by adding an edge from y to an additional vertex z. We can construct G by joining r copies of G' at the vertex z, which becomes the root of G. Let π' and π'' denote the respective stationary distributions for the random walks on G' and G'', and use the notation T' and T'', respectively, for hitting times on these graphs. By Chapter 4 jjj,

$$\tau_2 = E_{\alpha'}T'_z \tag{5.63}$$

where α' is the quasistationary distribution on G' associated with the hitting time T'_z. By Chapter 3 jjj, the expectation (5.63) is no smaller than $E_{\pi'}T'_z$, which by the collapsing principle equals

$$\pi'(G'')\left(E_{\pi''}T''_y + E_y T'_z\right) = \pi'(G'')\left(E_{\pi''}T''_y + E_y T_z\right).$$

But it is easy to see that this last quantity equals $(1+o(1))E_\pi T_z$, which is asymptotically equivalent to $2n/(r-1)$ by (5.61). ∎

¿From the discussion at the beginning of Section 5.3.1, it follows that τ_c is achieved at any of the r subtrees of the root. This gives

$$\tau_c = \frac{(2r^h - r - 1)(2r^h - 1)}{2r(r^h - 1)} \sim \frac{2n}{r}.$$

An extension of the balanced r-tree example is treated in Section 5.2.1 below.

Example 5.15 *The d-cube.*

This is a graph with vertex-set $\mathbf{I} = \{0,1\}^d$ and hence with $n = 2^d$ vertices. Write $\mathbf{i} = (i_1, \ldots, i_d)$ for a vertex, and write $|\mathbf{i}-\mathbf{j}| = \sum_u |i_u - j_u|$. Then $(\mathbf{i},\mathbf{j})$ is an edge if and only if $|\mathbf{i}-\mathbf{j}| = 1$, and in general $|\mathbf{i}-\mathbf{j}|$ is the graph-distance between $\mathbf{i}$ and $\mathbf{j}$. Thus discrete-time random walk proceeds at each step by choosing a coordinate at random and changing its parity.

It is easier to use the continuized walk $\mathbf{X}(t) = (X_1(t), \ldots, X_d(t))$, because the component processes $(X_u(t))$ are independent as u varies, and each is in fact just the continuous-time random walk on the 2-path with transition rate $1/d$. This follows from an elementary fact about the superposition of (marked) Poisson processes.

Thus, in continuous time,

$$\begin{aligned} P_{\mathbf{i}}(\mathbf{X}(t) = \mathbf{j}) &= \prod_{u=1}^{d} \left[\frac{1}{2} \left(1 + (-1)^{|i_u - j_u|} e^{-2t/d} \right) \right] \\ &= 2^{-d} \left(1 - e^{-2t/d} \right)^{|\mathbf{i}-\mathbf{j}|} \left(1 + e^{-2t/d} \right)^{d-|\mathbf{i}-\mathbf{j}|}. \end{aligned} \quad (5.64)$$

By expanding the right side, we see that the continuous-time eigenvalues are

$$\lambda_k = 2k/d \text{ with multiplicity } \binom{d}{k}, \quad k = 0, 1, \ldots, d. \quad (5.65)$$

(Of course, this is just the general fact that the eigenvalues of a d-fold product of continuous-time chains are

$$(\lambda_{i_1} + \cdots + \lambda_{i_d}; 1 \le i_1, \ldots, i_d \le n) \quad (5.66)$$

where $(\lambda_i; 1 \le i \le n)$ are the eigenvalues of the marginal chain.)

In particular,

$$\tau_2 = d/2. \quad (5.67)$$

By the eigentime identity (Chapter 3 yyy)

$$\begin{aligned} \tau_0 = \sum_{m \ge 2} \frac{1}{\lambda_m} &= \frac{d}{2} \sum_{k=1}^{d} k^{-1} \binom{d}{k} \\ &= 2^d (1 + d^{-1} + O(d^{-2})), \end{aligned} \quad (5.68)$$

the asymptotics being easy analysis.

From (5.64) it is also straightforward to derive the discrete-time t-step transition probabilities:

$$P_{\mathbf{i}}(\mathbf{X}_t = j) = 2^{-d} \sum_{m=0}^{d} \left(1 - \frac{2m}{d} \right)^t \sum_r (-1)^r \binom{|\mathbf{i}-\mathbf{j}|}{r} \binom{d - |\mathbf{i}-\mathbf{j}|}{m-r}.$$

Starting the walk at $\mathbf{0}$, let $Y_t = |\mathbf{X}(t)|$. Then Y is the birth-and-death chain on states $\{0, 1, \ldots, d\}$ with transition rates (transition probabilities, in discrete time)

$$q_{i,i+1} = \frac{d-i}{d}, \; q_{i,i-1} = \frac{i}{d}, \quad 0 \le i \le d.$$

xxx box picture

This is the *Ehrenfest urn model* mentioned in many textbooks. In our terminology we may regard Y as random walk on the weighted linear graph (Section 5.1.2) with weights

$$w_i = \binom{d-1}{i-1}, \quad w = 2^d.$$

In particular, writing T^Y for hitting times for Y, symmetry and (5.13) give

$$\frac{1}{2}\tau^{*Y} = \frac{1}{2}(E_0 T_d^Y + E_d T_0^Y) = E_0 T_d^Y = 2^{d-1} \sum_{i=1}^{d} \frac{1}{\binom{d-1}{i-1}}.$$

On the d-cube, it is "obvious" that $E_{\mathbf{0}} T_{\mathbf{j}}$ is maximized by $\mathbf{j} = \mathbf{1}$, and this can be verified by observing in (5.64) that $P_{\mathbf{0}}(\mathbf{X}(t) = \mathbf{j})$ is minimized by $\mathbf{j} = \mathbf{1}$, and hence $Z_{\mathbf{0j}}$ is minimized by $\mathbf{j} = \mathbf{1}$, so we can apply Chapter 2 yyy. Thus

$$\frac{1}{2}\tau^* = \max_{\mathbf{ij}} E_{\mathbf{i}} T_{\mathbf{j}} = E_{\mathbf{0}} T_{\mathbf{1}} = 2^{d-1} \sum_{i=1}^{d} \frac{1}{\binom{d-1}{i-1}} \sim 2^d(1 + 1/d + O(1/d^2)). \quad (5.69)$$

The asymptotics are the same as in (5.68). In fact it is easy to use (5.64) to show

$$Z_{\mathbf{ii}} = 2^{-d}\tau_0 = 1 + d^{-1} + O(d^{-2})$$

$$Z_{\mathbf{ij}} = O(d^{-2}) \text{ uniformly over } |\mathbf{i} - \mathbf{j}| \geq 2$$

and then by Chapter 2 yyy

$$E_{\mathbf{i}} T_{\mathbf{j}} = 2^d(1 + d^{-1} + O(d^{-2})) \text{ uniformly over } |\mathbf{i} - \mathbf{j}| \geq 2.$$

Since

$$1 + E_1 T_0^Y = E_0 T_1^Y + E_1 T_0^Y = w/w_1 = 2^d,$$

it follows that

$$E_{\mathbf{i}} T_{\mathbf{j}} = 2^d - 1 \text{ if } |\mathbf{i} - \mathbf{j}| = 1.$$

xxx refrain from write out exact $E_{\mathbf{i}} T_{\mathbf{j}}$—refs

To discuss total variation convergence, we have by symmetry (and writing $\mathbf{d}$ to distinguish from dimension d)

$$\bar{\mathbf{d}}(t) = \|P_{\mathbf{0}}(\mathbf{X}(t) \in \cdot) - P_{\mathbf{1}}(\mathbf{X}(t) \in \cdot)\|$$

$$\mathbf{d}(t) = \|P_{\mathbf{0}}(\mathbf{X}(t) \in \cdot) - \pi(\cdot)\|.$$

Following Diaconis et al [116] we shall sketch an argument leading to

$$\mathbf{d}\left(\tfrac{1}{4}d\log d + sd\right) \to L(s) \equiv P\left(|Z| \le \tfrac{1}{2}e^{-2s}\right) \ , -\infty < s < \infty \qquad (5.70)$$

where Z has the standard Normal distribution. This implies

$$\tau_1 \sim \tfrac{1}{4}d\log d. \qquad (5.71)$$

For the discrete-time walk made aperiodic by incorporating chance $1/(d+1)$ of holding, (5.70) and (5.71) remain true, though rigorous proof seems complicated: see [116].

Fix u, and consider $\mathbf{j} = \mathbf{j}(u)$ such that $|\mathbf{j}| - d/2 \sim ud^{1/2}/2$. Using $1 - \exp(-\delta) \approx \delta - \frac{1}{2}\delta^2$ as $\delta \to 0$ in (5.64), we can calculate for $t = t(d) = \frac{1}{4}d\log d + sd$ with s fixed that

$$2^d P_{\mathbf{0}}(\mathbf{X}(t) = \mathbf{j}) \to \exp\left(-\frac{e^{-4s}}{2} - ue^{-2s}\right).$$

Note the limit is > 1 when $u < u_0(s) \equiv -e^{-2s}/2$. Now

$$\mathbf{d}(t) = \frac{1}{2}\sum_{\mathbf{j}} |P_{\mathbf{0}}(\mathbf{X}(t) = \mathbf{j}) - 2^{-d}| \sim \sum (P_{\mathbf{0}}(\mathbf{X}(t) = \mathbf{j}) - 2^{-d})$$

where the second sum is over $\mathbf{j}(u)$ with $u < u_0(s)$. But from (5.64) we can write this sum as

$$P\left(B\left(\tfrac{1}{2}(1 - d^{-1/2}e^{-2s})\right) \le |\mathbf{j}(u_0(s))|\right) - P\left(B\left(\tfrac{1}{2}\right) \le |\mathbf{j}(u_0(s))|\right)$$

where $B(p)$ denotes a Binomial(d,p) random variable. By the Normal approximation to Binomial, this converges to

$$P(Z \le -u_0(s)) - P(Z \le u_0(s))$$

as stated.

As an aside, symmetry and Chapter 4 yyy give

$$\tau_0 \le E_{\mathbf{0}}T_{\mathbf{1}} \le \tau_1^{(2)} + \tau_0$$

and so the difference $E_{\mathbf{0}}T_{\mathbf{1}} - \tau_0$ is $O(d\log d)$, which is much smaller than what the series expansions (5.68) and (5.69) imply.

The fact that the "half-cube" $A = \{\mathbf{i} \in \mathbf{I} : \ i_d = 0\}$, yielding

$$\tau_c = d/2,$$

achieves the *sup* in the definition of τ_c can be proved using a slightly tricky induction argument. However, the result follows immediately from (5.67) together with the general inequality $\tau_2 \ge \tau_c$.

Example 5.16 *Dense regular graphs.*

Consider an r-regular n-vertex graph with $r > n/2$. Of course here we are considering a class of graphs rather than a specific example. The calculations below show that these graphs necessarily mimic the complete graph (as far as smallness of the random walk parameters is concerned) in the asymptotic setting $r/n \to c > 1/2$.

The basic fact is that, for any pair i, j of vertices, there must be at least $2r - n$ other vertices k such that $i - k - j$ is a path. To prove this, let a_1 (resp., a_2) be the number of vertices $k \neq i, j$ such that exactly 1 (resp., 2) of the edges $(k, i), (k, j)$ exist. Then $a_1 + a_2 \leq n - 2$ by counting vertices, and $a_1 + 2a_2 \geq 2(r - 1)$ by counting edges, and these inequalities imply $a_2 \geq 2r - n$.

Thus, by Thompson's principle (Chapter 3, yyy) the effective resistance $r_{ij} \leq \frac{2}{2r-n}$ and so the commute interpretation of resistance implies

$$\tau^* \leq \frac{2rn}{2r - n} \sim \frac{2cn}{2c - 1}. \tag{5.72}$$

A simple "greedy coupling" argument (Chapter 14, Example yyy) shows

$$\tau_1 \leq \frac{r}{2r - n} \sim \frac{c}{2c - 1}. \tag{5.73}$$

This is also a bound on τ_2 and on τ_c, because $\tau_c \leq \tau_2 \leq \tau_1$ always, and special case 2 below shows that this bound on τ_c cannot be improved asymptotically (nor hence can the bound on τ_1 or τ_2). Because $E_\pi T_j \leq n\tau_2$ for regular graphs (Chapter 3 yyy), we get

$$E_\pi T_j \leq \frac{nr}{2r - n}.$$

This implies

$$\tau_0 \leq \frac{nr}{2r - n} \sim \frac{cn}{2c - 1}$$

which also follows from (5.72) and $\tau_0 \leq \tau^*/2$. We can also argue, in the notation of Chapter 4 yyy, that

$$\max_{ij} E_i T_j \leq \tau_1^{(2)} + \max_j E_\pi T_j \leq \frac{4e}{e - 1}\tau_1 + n\tau_1 \leq (1 + o(1))\frac{nr}{2r - n} \sim \frac{cn}{2c - 1}.$$

Special case 1. The orders of magnitude may change for $c = 1/2$. Take two complete $(n/2)$-graphs, break one edge in each (say edges (v_1, v_2) and

(w_1, w_2)) and add edges (v_1, w_1) and (v_2, w_2). This gives an n-vertex $((n/2)-1)$-regular graph for which all our parameters are $\Theta(n^2)$.

jjj I haven't checked this.

Special case 2. Can the bound $\tau_c \le r/(2r-n) \sim c/(2c-1)$ be asymptotically improved? Eric Ordentlicht has provided the following natural counterexample. Again start with two $(n/2)$-complete graphs on vertices (v_i) and (w_i). Now add the edges (v_i, w_j) for which $0 \le (j-i) \bmod (n/2) \le r-(n/2)$. This gives an n-vertex r-regular graph. By considering the set A consisting of the vertices v_i, a brief calculation gives

$$\tau_c \ge \frac{r}{2r-n+2} \sim \frac{c}{2c-1}.$$

Example 5.17 *The d-dimensional torus Z_m^d.*

The torus is the set of d-dimensional integers $\mathbf{i} = (i_1, \ldots, i_d)$ modulo m, considered in the natural way as a $2d$-regular graph on $n = m^d$ vertices. It is much simpler to work with the random walk in *continuous* time, $\mathbf{X}(t) = (X_1(t), \ldots, X_d(t))$, because its component processes $(X_u(t))$ are independent as u varies; and each is just continuous-time random walk on the m-cycle, slowed down by a factor $1/d$. Thus we can immediately write the time-t transition probabilities for $\mathbf{X}$ in terms of the corresponding probabilities $p_{0,j}(t)$ for continuous-time random walk on the m-cycle (see Example 5.7 above) as

$$p_{\mathbf{0},\mathbf{j}}(t) = \prod_{u=1}^{d} p_{0,j_u}(t/d).$$

Since the eigenvalues on the m-cycle are $(1 - \cos(2\pi k/m), 0 \le k \le m-1)$, by (5.66) the eigenvalues of $\mathbf{X}$ are

$$\lambda_{(k_1 \ldots k_d)} = \frac{1}{d} \sum_{u=1}^{d} (1 - \cos(2\pi k_u/m)), \ 0 \le k_u \le m-1.$$

In particular, we see that the relaxation time satisfies

$$\tau_2 \sim \frac{dm^2}{2\pi^2} = \frac{dn^{2/d}}{2\pi^2}$$

where here and below asymptotics are as $m \to \infty$ for fixed d. This relaxation time could more simply be derived from the N-cycle result via the general "product chain" result of Chapter 4 yyy. But writing out all the eigenvalues enables us to use the eigentime identity.

$$\tau_0 = \sum_{k_1} \cdots \sum_{k_d} 1/\lambda_{(k_1, \ldots, k_d)}$$

(the sum excluding $(0, \ldots, 0)$), and hence

$$\tau_0 \sim m^d R_d \tag{5.74}$$

where

$$R_d \equiv \int_0^1 \cdots \int_0^1 \frac{1}{\frac{1}{d}\sum_{u=1}^d (1 - \cos(2\pi x_u))}\, dx_1 \cdots dx_d \tag{5.75}$$

provided the integral converges. The reader who is a calculus whiz will see that in fact $R_d < \infty$ for $d \geq 3$ only, but this is seen more easily in the alternative approach of Chapter 15, Section yyy.

xxx more stuff: connection to transience, recurrent potential, etc

xxx new copy from lectures

xxx τ_1, τ_c

jjj David: I will let you develop the rest of this example. Note that τ_1 is considered very briefly in Chapter 15, eq. (17) in 3/6/96 version. Here are a few comments for τ_c. First suppose that $m > 2$ is even and $d \geq 2$. Presumably, τ_c is achieved by the following half-torus:

$$A := \{\mathbf{i} = (i_1, \ldots, i_d) \in Z_m^d : 0 \leq i_d < m/2\}.$$

In the notation of (5.52) observe

$$|\mathcal{E}| = dn, \quad \deg A = dn, \quad \deg A^c = dn, \quad (A, A^c) = 2m^{d-1} = 2n/m,$$

whence

$$\tau(A) = \tfrac{d}{4} n^{1/d}.$$

[By Example 5.15 (the d-cube) this last result is also true for $m = 2$, and (for even $m \geq 2$) it is by Example 5.7 (the n-cycle) also true for $d = 1$.] If we have correctly conjectured the maximizing A, then

$$\tau_c = \tfrac{d}{4} n^{1/d} \text{ if } m \text{ is even},$$

and presumably(??)

$$\tau_c \sim \tfrac{d}{4} n^{1/d}$$

in any case.

Example 5.18 *Chess moves.*

Here is a classic homework problem for an undergraduate Markov chains course.

> Start a knight at a corner square of an otherwise-empty chessboard. Move the knight at random, by choosing uniformly from the legal knight-moves at each step. What is the mean number of moves until the knight returns to the starting square?

It's a good question, because if you don't know Markov chain theory it looks too messy to do by hand, whereas using Markov chain theory it becomes very simple. The knight is performing random walk on a graph (the 64 squares are the vertices, and the possible knight-moves are the edges). It is not hard to check that the graph is connected, so by the elementary Chapter 3 yyy for a corner square v the mean return time is

$$E_v T_v^+ = \frac{1}{\pi_v} = \frac{2|\mathcal{E}|}{d_v} = |\mathcal{E}|,$$

and by drawing a sketch in the margin the reader can count the number of edges $|\mathcal{E}|$ to be 168.

Other chess pieces—queen, king, rook—define different graphs (the bishop's is of course not connected, and the pawn's not undirected). One might expect that the conventional ordering of the "strength" of the pieces as (queen, rook, knight, king) is reflected in parameters τ_0 and τ_2 (**jjj** how about the other taus?) being increasing in this ordering. The reader is invited to perform the computations. (**jjj**: an undergraduate project?) We have done so only for the rook's move, treated in the next example.

The computations for the queen, knight, and king are simplified if the walks are made on a toroidal chessboard. (There is no difference for the rook.)

jjj Chess on a bagel, anyone? Continue same paragraph:

Then Fourier analysis (see Diaconis [112]) on the abelian group Z_m^2 (with $m = 8$) can be brought to bear, and the eigenvalues are easy to compute. We omit the details, but the results for (queen, rook, knight, king) are asymptotically

$$\begin{aligned} \tau_0 &= (m^2 + \tfrac{7}{12}m + O(1),\ m^2 + m + O(1), \\ &\qquad \textbf{jjj}?(1 + o(1))c_{\text{knight}} m^2 \log m,\ \textbf{jjj}?(1 + o(1))c_{\text{king}} m^2 \log m) \\ \tau_2 &\sim \left(\tfrac{4}{3},\ 2,\ \tfrac{1}{5\pi^2}m^2,\ \tfrac{2}{3\pi^2}m^2\right) \end{aligned}$$

as $m \to \infty$, in conformance with our expectations, and numerically

$$\begin{aligned} \tau_0 &= (65.04,\ 67.38,\ 69.74,\ 79.36) \\ \tau_2 &= (1.29,\ 1.75,\ 1.55,\ 4.55) \end{aligned}$$

for $m = 8$. The only surprise is the inverted τ_2 ordering for (rook, knight).

Example 5.19 *Rook's random walk on an m-by-m chessboard.*

jjj Do we want to do this also on a d-dimensional grid? We need to mention how this is a serious example, used with Metropolis for sampling from log concave distributions; reference is [32]? [33]?

Number the rows and columns of the chessboard each 0 through $m-1$ in arbitrary fashion, and denote the square of the chessboard at row i_1 and column i_2 by $\mathbf{i} = (i_1, i_2)$. In continuous time, the rook's random walk $(\mathbf{X}(t))$ is the product of two continuous-time random walks on the complete graph K_m on m vertices, each run at rate 1/2. Thus (cf. Example 5.9)

$$P_{\mathbf{i}}(\mathbf{X}(t) = \mathbf{j}) = \prod_{u=1}^{2} \left[\frac{1}{m} + \left(\delta_{i_u, j_u} - \frac{1}{m} \right) \exp\left(-\frac{mt}{2(m-1)} \right) \right], \qquad (5.76)$$

which can be expanded to get the discrete-time multistep transition probabilities, if desired. We recall that the eigenvalues for discrete-time random walk on K_m are 1 with multiplicity 1 and $-1/(m-1)$ with multiplicity $m-1$. It follows [recall (5.66)] that the eigenvalues for the continuous-time rook's walk are

$$0, \ \frac{m}{2(m-1)}, \ \frac{m}{m-1} \quad \text{with resp. multiplicities} \ \ 1, \ 2(m-1), \ (m-1)^2.$$

In particular,

$$\tau_2 = \frac{2(m-1)}{m}, \qquad (5.77)$$

which equals 1.75 for $m = 8$ and converges to 2 as m grows. Applying the eigentime identity, a brief calculation gives

$$\tau_0 = \frac{(m-1)^2(m+3)}{m}, \qquad (5.78)$$

which equals 67.375 for $m = 8$ and $m^2 + m + O(1)$ for m large.

Starting the walk $\mathbf{X}$ at $\mathbf{0} = (0, 0)$, let $Y(t)$ denote the Hamming distance $H(\mathbf{X}(t), \mathbf{0})$ of $\mathbf{X}(t)$ from $\mathbf{0}$, i.e., the number of coordinates (0, 1, or 2) in which $\mathbf{X}(t)$ differs from $\mathbf{0}$. Then Y is a birth-and-death chain with transition rates

$$q_{01} = 1, \quad q_{10} = \frac{1}{2(m-1)}, \quad q_{12} = \frac{1}{2}, \quad q_{21} = \frac{1}{m-1}.$$

This is useful for computing mean hitting times. Of course

$$E_{\mathbf{i}} T_{\mathbf{j}} = 0 \text{ if } H(\mathbf{i}, \mathbf{j}) = 0.$$

Since

$$1 + E_1 T_0^Y = E_0 T_1^Y + E_1 T_0^Y = m^2,$$

it follows that

$$E_{\mathbf{i}} T_{\mathbf{j}} = m^2 - 1 \text{ if } H(\mathbf{i}, \mathbf{j}) = 1.$$

Finally, it is clear that $E_2 T_1^Y = m - 1$, so that

$$E_2 T_0^Y = E_2 T_1^Y + E_1 T_0^Y = m^2 + m - 2,$$

whence

$$E_{\mathbf{i}} T_{\mathbf{j}} = m^2 + m - 2 \text{ if } H(\mathbf{i}, \mathbf{j}) = 2.$$

These calculations show

$$\tfrac{1}{2}\tau^* = \max_{\mathbf{ij}} E_{\mathbf{i}} T_{\mathbf{j}} = m^2 + m - 2,$$

which equals 70 for $m = 8$, and they provide another proof of (5.78).

From (5.76) it is easy to derive

$$\bar{d}_m(t) = \left(2 - \frac{2}{m}\right) \exp\left(-\frac{mt}{2(m-1)}\right) - \left(1 - \frac{2}{m}\right) \exp\left(-\frac{mt}{m-1}\right)$$

and thence

$$\tau_1 = -2\frac{m-1}{m}\left[\ln\left(1 - \left(1 - e^{-1}\frac{m(m-2)}{(m-1)^2}\right)^{1/2}\right) + \ln\left(\frac{m-1}{m-2}\right)\right],$$

which rounds to 2.54 for $m = 8$ and converges to $-2\ln(1 - (1 - e^{-1})^{1/2}) \doteq 3.17$ as m becomes large.

Any set A of the form $\{(i_1, i_2) : i_u \in J\}$ with either $u = 1$ or $u = 2$ and J a nonempty proper subset of $\{0, \ldots, m-1\}$ achieves the value

$$\tau_c = 2\frac{m-1}{m}.$$

A direct proof is messy, but this follows immediately from the general inequality $\tau_c \le \tau_2$, (5.77), and a brief calculation that the indicated A indeed gives the indicated value.

xxx other examples left to reader? complete bipartite; ladders

jjj Note: I've worked these out and have handwritten notes. How much do we want to include, if at all? (I could at least put the results in the table.)

5.2.1 Biased walk on a balanced tree

Consider again the balanced r-tree setup of Example 5.14. Fix a parameter $0 < \lambda < \infty$. We now consider biased random walk (X_t) on the tree, where from each non-leaf vertex other than the root the transition goes to the parent with probability $\lambda/(\lambda+r)$ and to each child with probability $1/(\lambda+r)$. As in Example 5.14 (the case $\lambda = 1$), the chain $\hat{X}$ induced by the function

$$f(i) = h - (\text{distance from } i \text{ to the root})$$

is (biased) reflecting random walk on $\{0, \ldots, h\}$ with respective probabilities $\lambda/(\lambda + r)$ and $r/(\lambda + r)$ of moving to the right and left from any $i \neq 0, h$; the ratio of these two transition probabilities is

$$\rho = \lambda/r.$$

The stationary distribution $\hat{\pi}$ for $\hat{X}$ is a modified geometric:

$$\hat{\pi}_m = \frac{1}{\hat{w}} \times \begin{cases} 1 & \text{if } m = 0 \\ (1+\rho)\rho^{m-1} & \text{if } 1 \le m \le h-1 \\ \rho^{h-1} & \text{if } m = h \end{cases}$$

where

$$\hat{w} = 2\sum_{m=0}^{h-1} \rho^m = \begin{cases} 2(1-\rho^h)/(1-\rho) & \text{if } \rho \neq 1 \\ 2h & \text{if } \rho = 1. \end{cases}$$

Since the stationary distribution π for X is assigns the same probability to each of the $r^{h-f(v)}$ vertices v with a given value of $f(v)$, a brief calculation shows that $\pi_v p_{vx} = \lambda^{f(v)}/\hat{w}r^h$ for any edge ($v =$ child, $x =$ parent) in the tree. In the same notation, it follows that X is random walk on the balanced r-tree with edge weights $w_{vx} = \lambda^{f(v)}$ and total weight $w = \sum_{v,x} w_{vx} = \hat{w}r^h$.

The distribution $\hat{\pi}$ concentrates near the root-level if $\rho < 1$ and near the leaves-level if $\rho > 1$; it is nearly uniform on the h levels if $\rho = 1$. On the other hand, the weight assigned by the distribution π to an individual vertex v is a decreasing function of $f(v)$ (thus favoring vertices near the leaves) if $\lambda < 1$ (i.e., $\rho < 1/r$) and is an increasing function (thus favoring vertices near the root) if $\lambda > 1$; it is uniform on the vertices in the unbiased case $\lambda = 1$.

The mean hitting time calculations of Example 5.14 can all be extended to the biased case. For example, for $\lambda \neq 1$ the general formula (5.55) becomes [using the same notation as at (5.55)]

$$\begin{aligned} E_{v_1}T_{v_2} &= \hat{w}r^h \frac{\lambda^{-f_3} - \lambda^{-f_2}}{\lambda^{-1} - 1} + 2(\rho^{-1}-1)^{-2}\left(\rho^{-(f_2+1)} - \rho^{-(f_1+1)}\right) \\ &\quad -2(\rho^{-1}-1)^{-1}(f_2 - f_1) - (f_2 - f_1) \end{aligned} \tag{5.79}$$

if $\rho \neq 1$ and

$$E_{v_1} T_{v_2} = \hat{w} r^h \frac{\lambda^{-f_3} - \lambda^{-f_2}}{\lambda^{-1} - 1} + f_2^2 - f_1^2$$

if $\rho = 1$. The maximum value is attained when v_1 and v_2 are leaves and v_3 is the root. So if $\lambda \neq 1$,

$$\tfrac{1}{2}\tau^* = \max_{v,x} E_v T_x = \hat{w} r^h \frac{\lambda^{-h} - 1}{\lambda^{-1} - 1}. \tag{5.80}$$

The orders of magnitude for all of the τ-parameters (with r and λ, and hence ρ, fixed as h, and hence n, becomes large) are summarized on a case-by-case basis in the next table. Following are some of the highlights in deriving these results; the details, and derivation of exact formulas and more detailed asymptotic results, are left to the reader.

Orders of magnitude of parameters [$\tau = \Theta$(entry)]
for λ-biased walk on a balanced r-tree of height h ($\rho = \lambda/r$).

Value of ρ	τ^*	τ_0	τ_1	τ_2	τ_c
$\rho < 1/r$	ρ^{-h}	ρ^{-h}	ρ^{-h}	ρ^{-h}	ρ^{-h}
$\rho = 1/r$ ($\equiv$ Example 5.14)	nh	nh	n	n	n
$1/r < \rho < 1$	n	n	ρ^{-h}	ρ^{-h}	ρ^{-h}
$\rho = 1$	nh	n	h	h	h
$\rho > 1$	n	n	h	1	1

For $\tau_0 = \sum_x \pi_x E_{\text{root}} T_x$ we have $\tau_0 \leq E_{\text{root}} T_{\text{leaf}}$. If $\rho < 1/r$, this bound is tight:

$$\tau_0 \sim E_{\text{root}} T_{\text{leaf}} \sim \frac{2\rho^{-h}}{(1-\rho)^2(1-\lambda)}(\lambda - \rho);$$

for $\rho > 1/r$ a more careful calculation is required.

If $\rho < 1$, then the same arguments as for the unbiased case ($\rho = 1/r$) show

$$\tau_1 \sim \tau_2 \sim 2\rho^{-(h-1)}/(1-\rho)^2.$$

In this case it is not hard to show that

$$\tau_c = \Theta(\rho^{-h})$$

as well. If $\rho = 1$, then it is not hard to show that

$$\tau_1 = \Theta(h), \quad \tau_c \sim 2(1 - \tfrac{1}{r})h$$

with τ_c achieved at a branch of the root (excluding the root), and so

$$\tau_2 = \Theta(h)$$

as well. If $\rho > 1$, then since $\hat{X}$ has positive drift equal to $(\rho - 1)/(\rho + 1)$, it follows that

$$\tau_1 \sim \frac{\rho + 1}{\rho - 1} h.$$

The value τ_c is achieved by isolating a leaf, giving

$$\tau_c \to 1,$$

and so, by the inequalities $\tau_c \le \tau_2 \le 8\tau_c^2$ of Chapter 4, Section yyy,

$$\tau_2 = \Theta(1)$$

as well.

jjj Limiting value of τ_2 when $\rho > 1$ is that of τ_2 for biased infinite tree? Namely?

5.3 Trees

For random walk on a finite tree, we can develop explicit formulas for means and variances of first passage times, and for distributions of first hitting places. We shall only treat the unweighted case, but the formulas can be extended to the weighted case without difficulty.

xxx notation below —change w to x ? Used i, j, v, w, x haphazardly for vertices.

In this section we'll write r_v for the degree of a vertex v, and $d(v, x)$ for the distance between v and x. On a tree we may unambiguously write $[v, x]$ for the path from v to x. Given vertices $j, v_1, v_2, \ldots$ in a tree, the intersection of the paths $[j, v_1]$, $[j, v_2], \ldots$ is a (maybe trivial) path; write $d(j, v_1 \wedge v_2 \wedge \cdots) \ge 0$ for the length of this intersection path.

On an n-vertex tree, the random walk's stationary distribution is

$$\pi_v = \frac{r_v}{2(n-1)}.$$

Recall from the beginning of this chapter that an edge (v, x) of a graph is *essential* if its removal would disconnect the graph into two components $A(v, x)$ and $A(x, v)$, say, containing v and x respectively. Obviously, in a tree every edge is essential, so we get a lot of mileage out of the essential edge lemma (Lemma 5.1).

Theorem 5.20 *Consider discrete-time random walk on an n-vertex tree. For each edge (i,j),*

$$E_iT_j = 2|A(i,j)| - 1 \tag{5.81}$$

$$E_iT_j + E_jT_i = 2(n-1). \tag{5.82}$$

For arbitrary i,j,

$$E_iT_j = -d(i,j) + 2\sum_v d(j, i\wedge v) = \sum_v r_v d(j, i\wedge v) \tag{5.83}$$

$$E_iT_j + E_jT_i = 2(n-1)d(i,j). \tag{5.84}$$

For each edge (i,j),

$$\mathrm{var}_iT_j = -E_iT_j + \sum_{v\in A(i,j)}\sum_{w\in A(i,j)} r_v r_w (2d(j, v\wedge w) - 1). \tag{5.85}$$

For arbitrary i,j,

$$\mathrm{var}_iT_j = -E_iT_j + \sum_v\sum_w r_v r_w d(j, i\wedge v\wedge w)\left[2d(j, v\wedge w) - d(j, i\wedge v\wedge w)\right]. \tag{5.86}$$

Remarks. 1. There are several equivalent expressions for the sums above: we chose the most symmetric-looking ones. We've written sums over vertices, but one could rephrase in terms of sums over edges.

2. In continuous time, the terms "$-E_iT_j$" disappear from the variance formulas—see xxx.

Proof of Theorem 5.20. Equations (5.81) and (5.82) are rephrasings of (5.3) and (5.4) from the essential edge lemma. Equation (5.84) and the first equality in (5.83) follow from (5.82) and (5.81) by summing over the edges in the path $[i,j]$. Note alternatively that (5.84) can be regarded as a consequence of the commute interpretation of resistance, since the effective resistance between i and j is $d(i,j)$. To get the second equality in (5.83),consider the following deterministic identity (whose proof is obvious), relating sums over vertices to sums over edges.

Lemma 5.21 *Let f be a function on the vertices of a tree, and let j be a distinguished vertex. Then*

$$\sum_v r_v f(v) = \sum_{v\neq j}(f(v) + f(v^*))$$

where v^* is the first vertex (other than v) in the path $[v,j]$.

To apply to (5.83), note

$$\begin{aligned} d(j, i \wedge v^*) &= d(j, i \wedge v) \text{ if } v \notin [i, j] \\ &= d(j, i \wedge v) - 1 \text{ if } v \in [i, j],\ v \neq j. \end{aligned}$$

The equality in Lemma 5.21 now becomes the equality in (5.83).

We prove (5.85) below. To derive (5.86) from it, sum over the edges in the path $[i, j] = (i = i_0, i_1, \ldots, i_m = j)$ to obtain

$$\mathrm{var}_i T_j = -E_i T_j + \sum_v \sum_w \sum_l (2d(i_{l+1}, v \wedge w) - 1) \tag{5.87}$$

where $\sum_l$ denotes the sum over all $0 \leq l \leq m-1$ for which $A(i_l, i_{l+1})$ contains both v and w. Given vertices v and w, there exist unique smallest values of p and q so that $v \in A(i_p, i_{p+1})$ and $w \in A(i_q, i_{q+1})$. If $p \neq q$, then the sum $\sum_l$ in (5.87) equals

$$\begin{aligned} \sum_{l=p\vee q}^{m-1} (2d(i_{l+1}, i_{p\vee q}) - 1) &= \sum_{l=p\vee q}^{m-1} (2((l+1) - (p \vee q)) - 1) \\ &= (m - (p \vee q))^2 = d^2(j, v \wedge w) \\ &= d(j, i \wedge v \wedge w)\left[2d(j, v \wedge w) - d(j, i \wedge v \wedge w)\right], \end{aligned}$$

as required by (5.86). If $p = q$, then the sum $\sum_l$ in (5.87) equals

$$\sum_{l=p}^{m-1} (2d(i_{l+1}, i_p) + 2d(i_p, v \wedge w) - 1)$$

which again equals $d(j, i \wedge v \wedge w)\left[2d(j, v \wedge w) - d(j, i \wedge v \wedge w)\right]$ by a similar calculation.

So it remains to prove (5.85), for which we may suppose, as in the proof of Lemma 5.1, that j is a leaf. By considering the first step from j to i we have

$$\mathrm{var}_j T_j^+ = \mathrm{var}_i T_j.$$

Now yyy of Chapter 2 gives a general expression for $\mathrm{var}_j T_j^+$ in terms of $E_\pi T_j$, and in the present setting this becomes

$$\mathrm{var}_j T_j^+ = 2(n-1) - (2(n-1))^2 + \sum_v 2r_v E_v T_j.$$

Using the second equality in (5.83), we may rewrite the sum as

$$\sum_{v \neq j} \sum_{w \neq j} r_v r_w 2d(j, v \wedge w).$$

Also,

$$\sum_{v \neq j} r_v = 2(n-1) - 1.$$

Combining these expressions gives

$$\mathrm{var}_i T_j = -(2n-3) + \sum_{v \neq j} \sum_{w \neq j} r_v r_w (2d(j, v \wedge w) - 1).$$

But by (5.81), $E_i T_j = 2n - 3$. ∎

5.3.1 Parameters for trees

Here we discuss the five parameters of Chapter 4. Obviously by (5.84)

$$\tau^* = 2(n-1)\Delta \tag{5.88}$$

where Δ is the diameter of the tree. As for τ_c, it is clear that the *sup* in its definition is attained by $A(v,w)$ for some edge (v,w). Note that

$$\pi(A(v,w)) = \frac{2|A(v,w)| - 1}{2(n-1)}. \tag{5.89}$$

This leads to

$$\begin{aligned} \tau_c &= \max_{(v,w)} \frac{\frac{2|A(v,w)|-1}{2(n-1)} \frac{2|A(w,v)|-1}{2(n-1)}}{\frac{1}{2(n-1)}} \\ &= \max_{(v,w)} \frac{4|A(v,w)||A(w,v)| - 2n + 1}{2(n-1)}. \end{aligned} \tag{5.90}$$

Obviously the *max* is attained by an edge for which $|A(v,w)|$ is as close as possible to $n/2$. This is one of several notions of "centrality" of vertices and edges which arise in our discussion—see Buckley and Harary [81] for a treatment of centrality in the general graph context, and for the standard graph-theoretic terminology.

Proposition 5.22 *On an n-vertex tree,*

$$\tau_0 = \frac{1}{2} + \frac{2}{n} \sum_{(v,w)} \left[|A(v,w)||A(w,v)| - \frac{1}{2(n-1)} \left(|A(v,w)|^2 + |A(w,v)|^2 \right) \right]$$

where $\sum_{(v,w)}$ *denotes the sum over all undirected edges* (v,w).

Proof. Using the formula for the stationary distribution, for each i

$$\tau_0 = \frac{1}{2(n-1)} \sum_j r_j E_i T_j.$$

Appealing to Lemma 5.21 (with i as the distinguished vertex)

$$\tau_0 = \frac{1}{2(n-1)} \sum_j (2E_i T_j - a(i,j))$$

where $a(i,i) = 0$ and $a(i,j) = E_x T_j$, where (j,x) is the first edge of the path $[j,i]$. Taking the (unweighted) average over i,

$$\tau_0 = \frac{1}{2n(n-1)} \sum_i \sum_j (2E_i T_j - a(i,j)).$$

Each term $E_i T_j$ is the sum of terms $E_v T_w$ along the edges (v,w) of the path $[i,j]$. Counting how many times a directed edge (v,w) appears,

$$\tau_0 = \frac{1}{2n(n-1)} \sum (2|A(v,w)||A(w,v)| - |A(v,w)\,|) E_v T_w,$$

where we sum over *directed* edges (v,w). Changing to a sum over undirected edges, using $E_v T_w + E_w T_v = 2(n-1)$ and $E_v T_w = 2|A(v,w)| - 1$, gives

$$\begin{aligned} 2n(n-1)\tau_0 \;=\; & \sum_{(v,w)} [2|A(v,w)||A(w,v)|2(n-1) \\ & -|A(v,w)|(2|A(v,w)|-1) \\ & -|A(w,v)|(2|A(w,v)|-1)]\,. \end{aligned}$$

This simplifies to the assertion of the Proposition. ∎

For τ_1 we content ourselves with a result "up to equivalence".

Proposition 5.23 *There exist constants $K_1, K_2 < \infty$ such that*

$$\frac{1}{K_1} \min_i \max_j E_j T_i \le \tau_1 \le K_2 \min_i \max_j E_j T_i.$$

Of course the expectations can be computed by (5.83).

Proof. We work with the parameter

$$\tau_1^{(3)} \equiv \max_{i,j} \sum_k \pi_k |E_j T_k - E_i T_k|$$

which we know is equivalent to τ_1. Write

$$\sigma = \min_i \max_j E_j T_i.$$

Fix an i attaining the minimum. For arbitrary j we have (the first equality uses the random target lemma, cf. the proof of Chapter 4 Lemma yyy)

$$\begin{aligned}
\sum_k \pi_k |E_j T_k - E_i T_k| &= 2 \sum_k \pi_k (E_j T_k - E_i T_k)^+ \\
&\le 2 \sum_k \pi_k E_j T_i \text{ because } E_j T_k \le E_j T_i + E_i T_k \\
&\le 2\sigma
\end{aligned}$$

and so $\tau_1^{(3)} \le 4\sigma$.

For the converse, it is elementary that we can find a vertex i such that the size (n^*, say) of the largest branch from i satisfies $n^* \le n/2$. (This is another notion of "centrality". To be precise, we are excluding i itself from the branch.) Fix this i, and consider the j which maximizes $E_j T_i$, so that $E_j T_i \ge \sigma$ by definition. Let B denote the set of vertices in the branch from i which contains j. Then

$$E_j T_k = E_j T_i + E_i T_k, \ k \in B^c$$

and so

$$\tau_1^{(3)} \ge \sum_k \pi_k |E_j T_k - E_i T_k| \ge \pi(B^c) E_j T_i \ge \pi(B^c)\sigma.$$

But by (5.89) $\pi(B) = \frac{2n^*-1}{2(n-1)} \le \frac{1}{2}$, so we have shown $\tau_1^{(3)} \ge \sigma/2$. ∎

We do not know whether τ_2 has a simple expression "up to equivalence" analogous to Proposition 5.23. It is natural to apply the "distinguished paths" bound (Chapter 4 yyy). This gives the inequality

$$\begin{aligned}
\tau_2 &\le 2(n-1) \max_{(v,w)} \sum_{x \in A(v,w)} \sum_{y \in A(w,v)} \pi_x \pi_y d(x,y) \\
&= 2(n-1) \max_{(v,w)} \Big(\pi(A(v,w)) E \left[d(v,V) 1_{(V \in A(w,v))} \right] \\
&\qquad + \pi(A(w,v)) E \left[d(v,V) 1_{(V \in A(v,w))} \right] \Big)
\end{aligned}$$

where V has the stationary distribution π and where we got the equality by writing $d(x,y) = d(v,y) + d(v,x)$. The edge attaining the *max* gives yet another notion of "centrality."

xxx further remarks on τ_2.

5.3.2 Extremal trees

It is natural to think of the n-path (Example 5.8) and the n-star (Example 5.10) as being "extremal" amongst all n-vertex trees. The proposition below confirms that the values of τ^*, $\max_{i,j} E_i T_j$, τ_0, τ_2, and τ_c in those examples are the exact extremal values (minimal for the star, maximal for the path).

Proposition 5.24 *For any n-vertex tree with $n \geq 3$,*
(a) $4(n-1) \leq \tau^* \leq 2(n-1)^2$
(b) $2(n-1) \leq \max_{i,j} E_i T_j \leq (n-1)^2$
(c) $n - \frac{3}{2} \leq \tau_0 \leq (2n^2 - 4n + 3)/6$.
(d) $1 \leq \tau_2 \leq (1 - \cos(\pi/(n-1)))^{-1}$.
(e) $1 - \frac{1}{2(n-1)} \leq \tau_c \leq \frac{4\lfloor n^2/4 \rfloor - 2n + 1}{2(n-1)}$.

Proof. (a) is obvious from (5.88), because Δ varies between 2 for the n-star and $(n-1)$ for the n-path. The lower bound in (b) follows from the lower bound in (a). For the upper bound in (b), consider some path $i = v_0, v_1, \ldots, v_d = j$ in the tree, where plainly $d \leq (n-1)$. Now $|A(v_{d-1}, v_d)| \leq n - 1$ and so

$$|A(v_{d-i}, v_{d-i+1})| \leq n - i \text{ for all } i$$

because the left side decreases by at least 1 as i increases. So

$$\begin{aligned} E_i T_j &= \sum_{m=0}^{d-1} E_{v_m} T_{v_{m+1}} \\ &= \sum_{m=0}^{d-1} (2|A(v_m, v_{m+1})| - 1) \text{ by (5.81)} \\ &\leq \sum_{m=0}^{d-1} (2(m + n - d) - 1) \\ &\leq \sum_{l=1}^{n-1} (2l - 1) \\ &= (n-1)^2. \end{aligned}$$

To prove (c), it is enough to show that the sum in Proposition 5.22 is minimized by the n-star and maximized by the n-path. For each undirected edge (v, w), let

$$b(v, w) = \min(|A(v, w)|, |A(w, v)|) \leq n/2.$$

Let $\mathbf{b} = (b_1, b_2, \ldots, b_{n-1})$ be the non-decreasing rearrangement of these values $b(v, w)$. The summands in Proposition 5.22 are of the form

$$a(n-a) - \frac{1}{2(n-1)}(a^2 + (n-a)^2)$$

with a ranging over the b_i.

One can check that this quantity is an increasing function of $a \leq n/2$. Thus it is enough to show that the vector $\mathbf{b}$ on an arbitrary n-tree dominates coordinatewise the vector $\mathbf{b}$ for the n-star and is dominated by the vector $\mathbf{b}$ for the n-path. The former is obvious, since on the n-star $\mathbf{b} = (1, 1, \ldots, 1)$. The latter needs a little work. On the n-path $\mathbf{b} = (1, 1, 2, 2, 3, 3, \ldots)$. So we must prove that in any n-tree

$$b_i \leq \left\lfloor \frac{i+1}{2} \right\rfloor \text{ for all } i. \tag{5.91}$$

Consider a *rooted* tree on m vertices. Breaking an edge e gives two components; let $a(e)$ be the size of the component not containing the root. Let $(a_1, a_2, \ldots)$ be the non-decreasing rearrangement of $(a(e))$. For an m-path rooted at one leaf, $(a_1, a_2, \ldots) = (1, 2, 3, \ldots)$. We assert this is extremal, in that for any rooted tree

$$a_i \leq i \text{ for all } i. \tag{5.92}$$

This fact can be proved by an obvious induction on m, growing trees by adding leaves.

Now consider an unrooted tree, and let $\mathbf{b}$ be as above. There exists some vertex v, of degree $r \geq 2$, such that each of the r branches from v has size (excluding v) at most $n/2$. Consider these branches as trees rooted at v, apply (5.92), and it is easy to deduce (5.91).

For (d), the lower bound is easy. Fix a leaf v and let w be its neighbor. We want to apply the extremal characterization (Chapter 3 yyy) of τ_2 to the function

$$g(v) = 1 - \pi_v - \pi_w, g(w) = 0, g(\cdot) = -\pi_v \text{ elsewhere.}$$

For this function, $\sum \pi_x g(x) = 0$,

$$[g, g] = \pi_v(1 - \pi_v - \pi_w)^2 + (1 - \pi_v - \pi_w)\pi_v^2,$$

and by considering transitions out of w

$$\mathcal{E}(g, g) = \pi_v(1 - \pi_v - \pi_w)^2 + (\pi_w - \pi_v)\pi_v^2.$$

Since $\pi_w \le 1/2$ we have $[g,g] \ge \mathcal{E}(g,g)$ and hence $\tau_2 \ge [g,g]/\mathcal{E}(g,g) \ge 1$.

qqq Anyone have a short proof of upper bound in (d)?

Finally, (e) is clear from (5.90). ∎

Other extremal questions. Several other extremal questions have been studied. Results on cover time are given in Chapter 6. Yaron [340] shows that for leaves l the mean hitting time $E_\pi T_l$ is maximal on the n-path and minimal on the n-star. (He actually studies the variance of return times, but Chapter 2 yyy permits the rephrasing.) Finally, if we are interested in the mean hitting time $E_x T_A$ or the hitting place distribution, we can reduce to the case where A is the set L of leaves, and then set up recursively-solvable equations for $h(i) \equiv E_i T_L$ or for $f(i) = P_i(T_A = T_l)$ for fixed $l \in L$. An elementary treatment of such ideas is in Pearce [277], who essentially proved that (on n-vertex trees) $\max_x E_x T_L$ is minimized by the n-star and maximized by the n-path.

5.4 Notes on Chapter 5

Most of the material seems pretty straightforward, so we will give references sparingly.

Introduction. The essential edge lemma is one of those oft-rediscovered results which defies attribution.

Section 5.1.2. One can of course use the essential edge lemma to derive the formula for mean hitting times in the general birth-and-death process. This approach seems more elegant than the usual textbook derivation. Although we are fans of martingale methods, we didn't use them in Proposition 5.3(b), because to define the right martingale requires one to know the answer beforehand!

For a birth-and-death chain the spectral representation involves orthogonal polynomials. This theory was developed by Karlin and McGregor in the 1950s, and is summarized in Chapter 8 of Anderson [31]. It enables one to write down explicit formulas for $P_i(X_t = j)$ in special cases. But it is less clear how to gain qualitative insight, or inequalities valid over all birth-and-death chains, from this approach.

An alternative approach which is more useful for our purposes is based on *Siegmund duality* (see e.g. [31] Section 7.4). Associated with a birth-and-death process (X_t) is another birth-and-death process (Y_t) which is "dual" in the sense that

$$P_i(X_t \le j) = P_j(Y_t \ge i) \text{ for all } i, j, t$$

and whose transition rates have a simple specification in terms of those of (X_t). It is easy to see that τ_1 for (X_t) is equivalent to $\max_j E_j T_{0,n}$ for (Y_t), for which there is an explicit formula. This gives an alternative to (5.16).

Section 5.2.

That the barbell is a good candidate for an "extremal" graph with respect to random walk properties was realized by Landau and Odlyzko [219], who computed the asymptotics of τ_2, and by Mazo [261], who computed the asymptotics of the unweighted average of $(E_i T_j; i, j \in I)$, which in this example is asymptotically our τ_0. Note we were able to give a one-line argument for the asymptotics of τ_2 by relying on the general fact $\tau_2 \le \tau_0$.

Formulas for quantities associated with random walk on the d-cube and with the Ehrenfest urn model have been repeatedly rediscovered, and we certainly haven't given all the known results. Bingham [51] has an extensive bibliography. Palacios [275] uses the simple "resistance" argument used in the text, and notes that the same argument can be used on the Platonic graphs. Different methods of computing $E_{\mathbf{0}} T_{\mathbf{1}}$ lead to formulas looking different from our (5.69), for instance

$$\begin{aligned} E_{\mathbf{0}} T_{\mathbf{1}} &= d \sum_{i=1}^{d} 2^{i-1}/i && \text{[216], eq. (4.27)} \\ &= d \sum_{1 \le j \le d,\ j \text{ odd}} j^{-1} \binom{d}{j} && \text{[51].} \end{aligned}$$

Similarly, one can get different-looking expressions for τ_0. Wilf [337] lists 54 identities involving binomial coefficients—it would be amusing to see how many could be derived by calculating a random walk on the d-cube quantity in two different ways!

Comparing our treatment of dense regular graphs (Example 5.16) with that in [272] should convince the reader of the value of general theory.

Section 5.3. An early reference to formulas for the mean and variance of hitting times on a tree (Theorem 5.20) is Moon [264], who used less intuitive generating function arguments. The formulas for the mean have been repeatedly rediscovered.

Of course there are many other questions we can ask about random walk on trees. Some issues treated later are

xxx list.

xxx more sophisticated ideas in Lyons [245].

Chapter 6

Cover Times (October 31, 1994)

The maximal mean hitting time $\max_{i,j} E_i T_j$ arises in many contexts. In Chapter 5 we saw how to compute this in various simple examples, and the discussion of τ^* in Chapter 4 indicated general methods (in particular, the electrical resistance story) for upper bounding this quantity. But what we've done so far doesn't answer questions like "how large can τ^* be, for random walk on a n-vertex graph". Such questions are dealt with in this Chapter, in parallel with a slightly different topic. The *cover time* for a n-state Markov chain is the random time C taken for the entire state-space I to be visited. Formally,

$$C \equiv \max_j T_j.$$

It is sometimes mathematically nicer to work with the "cover-and-return" time

$$C^+ \equiv \min\{t \geq C : X_t = X_0\}.$$

There are several reasons why cover times are interesting.

- Several applications involve cover times directly: graph connectivity algorithms (section 6.8.2), universal traversal sequences (section 6.8.1), the "white screen problem" (Chapter 1 yyy)
- There remains an interesting "computability" open question (section 6.8.3)
- In certain "critical" graphs, the uncovered subset at the time when the graph is almost covered is believed to be "fractal" (see the Notes on Chapter 7).

We are ultimately interested in random walks on unweighted graphs, but some of the arguments have as their natural setting either reversible Markov chains or general Markov chains, so we sometimes switch to those settings. Results are almost all stated for discrete-time walks, but we occasionally work with continuized chains in the proofs, or to avoid distracting complications in statements of results. Results often can be simplified or sharpened under extra symmetry conditions, but such results and examples are deferred until Chapter 7.

xxx contents of chapter

6.1 The spanning tree argument

Except for Theorem 6.1, we consider in this section random walk on an n-vertex unweighted graph. Results can be stated in terms of the number of edges $|\mathcal{E}|$ of the graph, but to aid comparison with results involving minimal or maximal degree it is helpful to state results in terms of average degree $\bar{d}$:

$$\bar{d} = \frac{2|\mathcal{E}|}{n}; \qquad |\mathcal{E}| = n\bar{d}/2.$$

The argument for Theorem 6.1 goes back to Aleliunas et al [25]. Though elementary, it can be considered the first (both historically and logically) result which combines Markov chain theory with graph theory in a non-trivial way.

Consider random walk on a weighted graph. Recall from Chapter 3 yyy the *edge-commute inequality*: for an edge (v, x)

$$\begin{aligned} E_v T_x + E_x T_v &\leq w/w_{vx} \qquad \text{(weighted)} && (6.1)\\ &\leq \bar{d}n \qquad \text{(unweighted).} && (6.2) \end{aligned}$$

One can alternatively derive these inequalities from the commute interpretation of resistance (Chapter 3 yyy), since the resistance between x and v is at most $1/w_{vx}$.

Theorem 6.1 *For random walk on a weighted graph,*

$$\max_v E_v C^+ \leq w \min_{\mathcal{T}} \sum_{e \in \mathcal{T}} 1/w_e$$

where the min *is over spanning trees* $\mathcal{T}$*. In the unweighted case*

$$\max_v E_v C^+ \leq \bar{d}n(n-1).$$

Proof. Given a spanning tree $\mathcal{T}$ and a vertex v, there is a path $v = v_0, v_1, \ldots, v_{2n-2} = v$ which traverses each edge of the tree once in each direction, and in particular visits every vertex. So

$$\begin{aligned} E_v C^+ &\leq \sum_{j=0}^{2n-3} E_{v_j} T_{v_{j+1}} \\ &= \sum_{e=(v,x)\in\mathcal{T}} (E_v T_x + E_x T_v) \\ &\leq \sum_{e\in\mathcal{T}} w/w_e \text{ by (6.1)} \end{aligned}$$

This gives the weighted case, and in the unweighted case $w = \bar{d}n$ and each spanning tree has $\sum_{e\in\mathcal{T}} 1/w_e = n - 1$. □

Note that in the unweighted case, the bound is at most $n(n-1)^2$. On the barbell (Chapter 5 Example yyy) it is easy to see that $\min_i E_i C = \Omega(n^3)$, so the maximal values of any formalization of "mean cover time", over n-vertex graphs, is $\Theta(n^3)$. Results and conjectures on the optimal numerical constants in the $\Theta(n^3)$ upper bounds are given in section 6.3.

Corollary 6.2 *On an unweighted n-vertex tree, $E_v C_v^+ \leq 2(n-1)^2$, with equality iff the tree is the n-path and v is a leaf.*

Proof. The inequality follows from Theorem 6.1. On the n-path with leaves v, z we have $E_v C_v^+ = E_v T_z + E_z T_v = 2(n-1)^2$. □

It is worth dissecting the proof of Theorem 6.1. Two different inequalities are used in the proof. Inequality (6.2) is an equality iff the edge is essential, so the second inequality in the proof is an equality iff the graph is a tree. But the first inequality in the proof bounds C^+ by the time to traverse a spanning tree *in a particular order*, and is certainly not sharp on a general tree, but only on a path. This explains Corollary 6.2. More importantly, these remarks suggest that the bound $\bar{d}n(n-1)$ in Theorem 6.1 will be good iff there is some fixed "essential path" in the graph, and the dominant contribution to C is from the time taken to traverse that path (as happens on the barbell).

There are a number of variations on the theme of Theorem 6.1, and we will give two. The first (due to Zuckerman [342], whose proof we follow) provides a nice illustration of probabilistic technique.

Proposition 6.3 *Write C_e for the time to cover all edges of an unweighted graph, i.e. until each edge (v, w) has been traversed in each direction. Then*

$$\max_v E_v C_e \leq 11\bar{d}n^2.$$

Proof. Fix a vertex v and a time t_0. Define "excursions", starting and ending at v, as follows. In each excursion, wait until all vertices have been visited, then wait t_0 longer, then end the excursion at the next visit to v. Writing S_i for the time at which the i'th excursion ends, and N for the (random) number of excursions required to cover each edge in each direction, we have

$$S_N = \min\{S_i : S_i \geq C_e\}$$

and so by Wald's identity (yyy refs)

$$E_v C_e \leq E_v S_N = E_v N \times E_v S_1. \tag{6.3}$$

Clearly

$$E_v S_1 \leq E_v C + t_0 + \max_w E_w T_v \leq t_0 + 2 \max_i E_i C.$$

To estimate the other factor, we shall first show

$$P_v(N > 2) \leq m^3/t_0^2 \tag{6.4}$$

where $m \equiv \bar{d}n$ is the number of directed edges. Fix a directed edge (w, x), say. By Chapter 3 Lemma yyy the mean time, starting at x, until (w, x) is traversed equals m. So the chance, starting at x, that (w, x) is not traversed before time t_0 is at most m/t_0. So using the definition of excursion, the chance that (v, w) is not traversed during the first excursion is at most m/t_0, so the chance it is not traversed during the first *two* excursions is at most $(m/t_0)^2$. Since there are m directed edges, (6.4) follows.

Repeating the argument for (6.4) gives

$$P_v(N > 2j) \leq \left(\frac{m^3}{t_0^2}\right)^j ; \ j \geq 0$$

and hence, assuming $m^3 < t_0^2$,

$$E_v N \leq \frac{2}{1 - m^3/t_0^2}.$$

Putting $t_0 = \lceil 2m^{3/2} \rceil$ gives $E_v N \leq 8/3$. Substituting into (6.3),

$$\max_v E_v C_e \leq \frac{8}{3}(\lceil 2m^{3/2} \rceil + 2 \max_v E_v C).$$

Now Theorem 6.1 says $\max_v E_v C \le m(n-1) \le mn - 1$, so

$$\begin{aligned} \max_v E_v C_e &\le \frac{8}{3}(2m^{3/2} + 2mn) \\ &= \frac{16}{3} m(m^{1/2} + n) \\ &\le \frac{32}{3} mn \end{aligned}$$

establishing the Proposition. □

Another variant of Theorem 6.1, due to Kahn et al [205] (whose proof we follow), uses a graph-theoretical lemma to produce a "good" spanning tree in graphs of high degree.

Theorem 6.4 *Writing* $d_* = \min_v d_v$,

$$\max_v E_v C^+ \le \frac{6\bar{d}n^2}{d_*} \tag{6.5}$$

and so on a regular graph

$$\max_v E_v C^+ \le 6n^2. \tag{6.6}$$

To appreciate (6.6), consider

Example 6.5 Take an even number $j \ge 2$ cliques of size $d \ge 3$, distinguish two vertices v_i, v_i' in the i'th clique (for each $0 \le i < j$), remove the edges (v_i, v_i') and add the edges $(v_i', v_{(i+1) \bmod j})$. This creates a $(d-1)$-regular graph with $n = jd$ vertices.

Arguing as in the barbell example (Chapter 5 yyy), as $d \to \infty$ with j varying arbitrarily,

$$\max_{v,w} E_v T_w \sim \frac{d}{2} \times d \times \frac{j^2}{4} \sim \frac{n^2}{8}.$$

Thus the $O(n^2)$ bound in (6.6) can't be improved, even as a bound for the smaller quantity $\max_{v,w} E_v T_w$. (Note that in the example, $d/n \le 1/2$. From the results in Chapter 5 Example yyy and Matthews' method one gets $EC = O(n \log n)$ for regular graphs with d/n bounded above $1/2$.)

Here is the graph-theory lemma needed for the proof of Theorem 6.4.

Lemma 6.6 *Let G be an n-vertex graph with minimal degree d_*. There exists a family of $\lceil d_*/2 \rceil$ spanning forests F_i such that*
(i) Each edge of G appears in at most 2 forests
(ii) Each component of each forest has size at least $\lceil d_*/2 \rceil$.

Proof. Replace each edge (i, j) of the graph by two directed edges $(i \to j)$, $(j \to i)$. Pick an arbitrary v_1 and construct a path $v_1 \to v_2 \to \ldots v_q$ on distinct vertices, stopping when the path cannot be extended. That is the first stage of the construction of F_1. For the second stage, pick a vertex v_{q+1} not used in the first stage and construct a path $v_{q+1} \to v_{q+2} \to \ldots v_r$ in which no second-stage vertex is revisited, stopping when a first-stage vertex is hit or when the path cannot be extended. Continue stages until all vertices have been touched. This creates a directed spanning forest F_1. Note that all the neighbors of v_q must be amongst $\{v_1, \ldots, v_{q-1}\}$, and so the size of the component of F_1 containing v_1 is at least $d_* + 1$, and similarly for the other components of F_1.

Now delete from the graph all the directed edges used in F_1. Inductively construct forests $F_2, F_3, \ldots, F_{\lceil d_*/2 \rceil}$ in the same way. The same argument shows that each component of F_i has size at least $d_* + 2 - i$, because at a "stopping" vertex v at most $i - 1$ of the directed edges out of v were used in previous forests.

Proof of Theorem 6.4. Write m for the number of (undirected) edges. For an edge $e = (v, x)$ write $b_e = E_v T_x + E_x T_v$. Chapter 3 Lemma yyy says $\sum_e b_e = 2m(n-1)$. Now consider the $\lceil d_*/2 \rceil$ forests F_i given by Lemma 6.6. Since each edge appears in at most two forests,

$$\sum_i \sum_{e \in F_i} b_e \le 2 \sum_e b_e \le 4mn,$$

and so there exists a forest F with $\sum_{e \in F} b_e \le 4mn/\lceil d_*/2 \rceil \le 8mn/d_*$. But each component of F has size at least $\lceil d_*/2 \rceil$, so F has at most $2n/d_*$ components. So to extend F to a tree $\mathcal{T}$ requires adding at most $2n/d_* - 1$ edges (e_j), and for each edge e we have $b_e \le 2m$ by (6.2). This creates a spanning tree $\mathcal{T}$ with $\sum_{e \in \mathcal{T}} b_e \le 12mn/d_*$. As in the proof of Theorem 6.1, this is an upper bound for $E_v C^+$.

6.2 Simple examples of cover times

There are a few (and only a few) examples where one can study EC by bare-hands exact calculations. Write h_n for the *harmonic sum*

$$h_n = \sum_{i=1}^{n} i^{-1} \sim \log n. \tag{6.7}$$

(a) *The coupon collector's problem.* Many textbooks discuss this classical problem, which involves C for the chain $(X_t; t \ge 0)$ whose values are

independent and uniform on an n-element set, i.e. random walk on the complete graph with self-loops. Write (cf. the proof of Matthews' method, Chapter 2 yyy) C^m for the first time at which m distinct vertices have been visited. Then each step following time C^m has chance $(n-m)/n$ to hit a new vertex, so $E(C^{m+1} - C^m) = n/(n-m)$, and so

$$EC = \sum_{m=1}^{n-1} E(C^{m+1} - C^m) = nh_{n-1}. \tag{6.8}$$

(By symmetry, $E_v C$ is the same for each initial vertex, so we just write EC) It is also a textbook exercise (e.g. [133] p. 124) to obtain the limit distribution

$$n^{-1}(C - n\log n) \stackrel{d}{\to} \xi \tag{6.9}$$

where ξ has the extreme value distribution

$$P(\xi \le x) = \exp(-e^{-x}), \quad -\infty < x < \infty. \tag{6.10}$$

We won't go into the elementary derivations of results like (6.9) here, because in Chapter 7 yyy we give more general results.

(b) *The complete graph.* The analysis of C for random walk on the complete graph (i.e. without self-loops) is just a trivial variation of the analysis above. Each step following time C^m has chance $(n-m)/(n-1)$ to hit a new vertex, so

$$EC = (n-1)h_n \sim n\log n. \tag{6.11}$$

And the distribution limit (6.9) still holds. Because $E_v T_w = n-1$ for $w \ne v$, we also have

$$EC^+ = EC + (n-1) = (n-1)(1 + h_{n-1}) \sim n\log n. \tag{6.12}$$

(c) *The n-star (Chapter 5 Example yyy).* Here the visits to the leaves (every second step) are exactly i.i.d., so we can directly apply the coupon collector's problem. For instance, writing v for the central vertex and l for a leaf,

$$\begin{aligned} E_l C &= 2(n-1)h_{n-2} \sim 2n\log n \\ E_v C^+ &= 1 + E_l C + 1 = 2(n-1)h_{n-1} \sim 2n\log n \end{aligned}$$

and $C/2$ satisfies (6.9). Though we won't give the details, it turns out that a clever inductive argument shows these are the minima over all trees.

Proposition 6.7 (Brightwell - Winkler [61]) *On an n-vertex tree,*

$$\min_v E_v C \geq 2(n-1)h_{n-2}$$

$$\min_v E_v C^+ \geq 2(n-1)h_{n-1}.$$

(d) *The n-cycle.* Random walk on the n-cycle is also easy to study. At time C^m the walk has visited m distinct vertices, and the set of visited vertices must form an interval $[j, j+m-1]$, say, where we add modulo n. At time C^m the walk is at one of the endpoints of that interval, and $C^{m+1} - C^m$ is the time until the first of $\{j-1, j+m\}$ is visited, which by Chapter 5 yyy has expectation $1 \times m$. So

$$EC = \sum_{m=1}^{n-1} E(C^{m+1} - C^m) = \sum_{i=1}^{n-1} i = \frac{1}{2}n(n-1).$$

There is also an expression for the limit distribution (see Notes).

The n-cycle also has an unexpected property. Let V denote the last vertex to be hit. Then

$$\begin{aligned} P_0(V = v) &= P_0(T_{v-1} < T_{v+1}) P_{v-1}(T_{v+1} < T_v) \\ &\quad + P_0(T_{v+1} < T_{v-1}) P_{v+1}(T_{v-1} < T_v) \\ &= \frac{n-(v+1)}{n-2}\frac{1}{n-1} + \frac{v-1}{n-2}\frac{1}{n-1} \\ &= \frac{1}{n-1}. \end{aligned}$$

In other words, the n-cycle has the property

> For any initial vertex v_0, the last-visited vertex V is uniform on the states excluding v_0.

Obviously the complete graph has the same property, by symmetry. Lovasz and Winkler [240] gave a short but ingenious proof that these are the *only* graphs with that property, a result rediscovered in [179].

6.3 More upper bounds

We remain in the setting of random walk on an unweighted graph. Theorems 6.1 and 6.4 show that the mean cover times, and hence mean hitting times, are $O(n^3)$ on irregular graphs and $O(n^2)$ on regular graphs, and examples

such as the barbell and the n-cycle show these bounds are the right order of magnitude. Quite a lot of attention has been paid to sharpening the constants in such bounds. We will not go into details, but will merely record a very simple argument in section 6.3.1 and the best known results in section 6.3.2.

6.3.1 Simple upper bounds for mean hitting times

Obviously $\max_j(E_iT_j + E_jT_i) \le E_iC^+$, so maximizing over i gives

$$\tau^* \le \max_i E_iC^+ \tag{6.13}$$

and the results of section 6.1 imply upper bounds on τ^*. But implicit in earlier results is a direct bound on τ^*. The edge-commute inequality implies that, for arbitrary v, x at distance $\Delta(v, x)$,

$$E_vT_x + E_xT_v \le \bar{d}n\Delta(v, x) \tag{6.14}$$

and hence

Corollary 6.8 *$\tau^* \le \bar{d}n\Delta$, where Δ is the diameter of the graph.*

It is interesting to compare the implications of Corollary 6.8 with what can be deduced from (6.13) and the results of section 6.1. To bound Δ in terms of n alone, we have $\Delta \le n-1$, and then Corollary 6.8 gives the same bound $\tau^* \le \bar{d}n(n-1)$ as follows from Theorem 6.1. On the other hand, the very simple graph-theoretic Lemma 6.10 gives (with Corollary 6.8) the following bound, which removes a factor of 2 from the bound implied by Theorem 6.4.

Corollary 6.9 *$\tau^* \le \frac{3\bar{d}n^2}{d_*}$ and so on a regular graph $\tau^* \le 3n^2$.*

Lemma 6.10 *$\Delta \le 3n/d_*$.*

Proof. Consider a path $v_0, v_1, \ldots, v_\Delta$, where vertices v_0 and v_Δ are distance Δ apart. Write A_i for the set of neighbors of v_i. Then A_i and A_j must be disjoint when $|j-i| \ge 3$. So a given vertex can be in at most 3 of the A's, giving the final inequality of

$$(\Delta+1)d_* \le \sum_{i=0}^{\Delta} d_{v_i} = \sum_{i=0}^{\Delta} |A_i| \le 3n. \qquad \square$$

6.3.2 Known and conjectured upper bounds

Here we record results without giving proofs. Write **max** for the maximum over n-vertex graphs. The next result is the only case where the exact extremal graph is known.

Theorem 6.11 (Brightwell-Winkler [62]) **max** $\max_{v,x} E_v T_x$ *is attained by the lollipop (Chapter 5 Example yyy) with* $m_1 = \lfloor (2n+1)/3 \rfloor$*, taking* x *to be the leaf.*

Note that the implied asymptotic behavior is

$$\mathbf{max} \max_{v,w} E_v T_w \sim \frac{4}{27} n^3. \tag{6.15}$$

Further asymptotic results are given by

Theorem 6.12 (Feige [143, 144])

$$\mathbf{max} \max_v E_v C^+ \sim \frac{4}{27} n^3 \tag{6.16}$$

$$\mathbf{max} \min_v E_v C^+ \sim \frac{3}{27} n^3 \tag{6.17}$$

$$\mathbf{max} \min_v E_v C \sim \frac{2}{27} n^3 \tag{6.18}$$

The value in (6.16) is asymptotically attained on the lollipop, as in Theorem 6.11. Note that (6.15) and (6.16) imply the same $4n^3/27$ behavior for intermediate quantities such as τ^* and $\max_v E_v C$. The values in (6.17) and (6.18) are asymptotically attained by the graph consisting of a $n/3$-path with a $2n/3$-clique attached at the middle of the path.

The corresponding results for τ_0 and τ_2 are not known. We have $\tau_2 \le \tau_0 \le \min_v E_v C$, the latter inequality from the random target lemma, and so (6.18) implies

$$\mathbf{max}\tau_0 \text{ and } \mathbf{max}\tau_2 \ \le (\frac{2}{27} + o(1))n^3. \tag{6.19}$$

But a natural guess is that the asymptotic behavior is that of the barbell, giving the values below.

Open Problem 6.13 Prove the conjectures

$$\mathbf{max}\ \tau_0 \sim \frac{1}{54} n^3, \quad \mathbf{max}\ \tau_2 \sim \frac{1}{54} n^3.$$

For regular graphs, none of the asymptotic values are known exactly. A natural candidate for extremality is the necklace graph (Chapter 5 yyy), where the time parameters are asymptotically 3/4 times the parameters for the n-path. So the next conjecture uses the numerical values from the necklace graph.

Open Problem 6.14 Prove the conjectures that, over the class of regular n-vertex graphs

$$\mathbf{max} \max_{i,j} E_i T_j \sim \frac{3}{4} n^2$$

$$\mathbf{max}\ \tau^* \sim \frac{3}{2} n^2$$

$$\mathbf{max} \max_{v} E_v C^+ \sim \frac{3}{2} n^2$$

$$\mathbf{max} \max_{v} E_v C \sim \frac{15}{16} n^2$$

$$\mathbf{max} \min_{v} E_v C \sim \frac{3}{4} n^2$$

$$\mathbf{max}\ \tau_0 \sim \frac{1}{4} n^2$$

$$\mathbf{max}\ \tau_2 \sim \frac{3}{2\pi^2} n^2$$

The best bounds known are those implied by the following result.

Theorem 6.15 (Feige [144]) *On a d-regular graph,*

$$\begin{aligned} \max_v E_v C &\leq && 2n^2 \\ \max_v E_v C_v^+ &\leq && 2n^2 \left(1 + \tfrac{d-2}{(d+1)^2}\right) \quad \leq 13n^2/6. \end{aligned}$$

6.4 Short-time bounds

It turns out that the bound "$\tau^* \leq 3n^2$ on a regular graph" given by Corollary 6.9 can be used to obtain bounds concerning the short-time behavior of random walks. Such bounds, and their applications, are the focus of this section. We haven't attempted to optimize numerical constants (e.g. Theorem 6.15 implies that $\tau^* \leq 13n^2/6$ on regular graphs). More elaborate arguments (see Notes) can be used to improve constants and to deal with the irregular case, but we'll restrict attention to the regular case for simplicity.

Write $N_i(t)$ for the number of visits to i before time t, i.e. during $[0, t-1]$.

Proposition 6.16 *Consider random walk on an n-vertex regular graph G. Let A be a proper subset of vertices and let $i \in A$.*
(i) $E_i T_{A^c} \le 4|A|^2$.
(ii) $E_i N_i(T_{A^c}) \le 5|A|$.
(iii) $E_i N_i(t) \le 5t^{1/2},\ 0 \le t < 5n^2$.
(iv) $P_\pi(T_i < t) \ge \frac{1}{5n} \min(t^{1/2}, n)$.

Remarks. For part (i) we give a slightly fussy argument repeating ingredients of the proof of Corollary 6.9, since these are needed for (ii). The point of (iv) is to get a bound for $t \ll E_\pi T_i$. On the n-cycle, it can be shown that the probability in question really is $\Theta(\min(t^{1/2}/n, 1))$, uniformly in n and t.

Proof of Proposition 6.16. Choose a vertex $b \in A^c$ at minimum distance from i, and let $i = i_0, i_1, \ldots, i_j, i_{j+1} = b$ be a minimum-length path. Let G^* be the subgraph on vertex-set A, and let G^{**} be the subgraph on vertex-set A together with all the neighbors of i_j. Write superscripts * and ** for the random walks on G^* and G^{**}. Then

$$E_i T_{A^c} \le E_i T^{**}_{A^c} = E_i T^*_{i_j} + E_{i_j} T^{**}_{A^c}$$

The inequality holds because we can specify the walk on G in terms of the walk on G^{**} with possibly extra chances of jumping to A^c at each step (this is a routine *stochastic comparison* argument, written out as an example in Chapter 14 yyy). The equality holds because the only routes in G^{**} from i to A^c are via i_j, by the minimum-length assumption. Now write $\mathcal{E}, \mathcal{E}^*, \mathcal{E}^{**}$ for the edge-sets. Using the commute interpretation of resistance,

$$E_i T^*_{i_j} \le 2|\mathcal{E}^*| j. \tag{6.20}$$

Writing $q \ge 1$ for the number of neighbors of i_j in A^c, the effective resistance in G^{**} between i_j and A^c is $1/q$, so the commute interpretation of resistance give the first equality in

$$E_{i_j} T^{**}_{A^c} = 2|\mathcal{E}^{**}| \frac{1}{q} - 1 = 2\frac{|\mathcal{E}^*|}{q} + 1 \le 2|\mathcal{E}^*| + 1 \le |A|^2.$$

The neighbors of $i_0, i_1, \ldots, i_{j-1}$ are all in A, so the proof of Lemma 6.10 implies

$$j \le 3|A|/d \tag{6.21}$$

where d is the degree of G. Since $2|\mathcal{E}^*| \le d|A|$, the bound in (6.20) is at most $3|A|^2$, and part (i) follows.

For part (ii), by the electrical network analogy (Chapter 3 yyy) the quantity in question equals

$$\frac{1}{P_i(T_{A^c} < T_i^+)} = w_i r(i, A^c) = d r(i, A^c) \tag{6.22}$$

where $r(i, A^c)$ is the effective resistance in G between i and A^c. Clearly this effective resistance is at most the distance ($j + 1$, in the argument above) from i to A^c, which by (6.21) is at most $3|A|/d + 1$. Thus the quantity (6.22) is at most $3|A| + d$, establishing the desired result in the case $d \le 2|A|$. If $d > 2|A|$ then there are at least $d - |A|$ edges from i to A^c, so $r(i, A^c) \le \frac{1}{d-|A|}$ and the quantity (6.22) is at most $\frac{d}{d-|A|} \le 2 \le 5|A|$.

For part (iii), fix a state i and an integer time t. Write $N_i(t)$ for the number of visits to i before time t, i.e. during times $\{0, 1, \ldots, t-1\}$. Then

$$\frac{t}{n} = E_\pi N_i(t) \le P_\pi(T_i < t) E_i N_i(t) \tag{6.23}$$

the inequality by conditioning on T_i. Now choose real s such that $ns \ge t$. Since $\sum_j E_i N_j(t) = t$, the set

$$A \equiv \{j : E_i N_j(t) > s\}$$

has $|A| < t/s \le n$, so part (ii) implies

$$E_i N_i(T_{A^c}) \le 5t/s. \tag{6.24}$$

Now by regularity we can rewrite A as $\{j : E_j N_i(t) > s\}$, and so by conditioning on T_{A^c}

$$E_i N_i(t) \le E_i N_i(T_{A^c}) + s.$$

Setting $s = \sqrt{5t}$ and combining with (6.24) gives (iii). The bound in (iv) now follows from (iii) and (6.23).

6.4.1 Covering by multiple walks

The first application is a variant of work of Broder et al [67] discussed further in section 6.8.2.

Proposition 6.17 *On a regular n-vertex graph, consider K independent random walks, each started at a uniform random vertex. Let $C^{[K]}$ be the time until every vertex has been hit by some walk. Then*

$$EC^{[K]} \le \frac{(25 + o(1))n^2 \log^2 n}{K^2} \text{ as } n \to \infty \text{ with } K \ge 6 \log n.$$

Remarks. The point is the $\frac{1}{K^2}$ dependence on K. On the n-cycle, for $K \sim \varepsilon n$ it can be shown that initially the largest gap between adjacent walkers is $\Theta(\log n)$ and that $EC^{[K]} = \Theta(\log^2 n)$, so in this respect the bound is sharp. Of course, for $K \leq \log n$ the bound would be no improvement over Theorem 6.4.

Proof. As usual write T_i for the hitting time on i for a single walk, and write $T_i^{[K]}$ for the first time i is visited by some walk. Then

$$\begin{aligned} P_\pi(T_i^{[K]} \geq t) &= (P_\pi(T_i \geq t))^K \\ &= (1 - P_\pi(T_i < t))^K \\ &\leq \exp(-K P_\pi(T_i < t)) \\ &\leq \exp\left(-\frac{K t^{1/2}}{5n}\right) \end{aligned}$$

by Proposition 6.16 (iii), provided $t \leq n^2$. So

$$P(C^{[K]} \geq t) \leq \sum_i P(T_i^{[K]} \geq t) \leq n \exp\left(-\frac{K t^{1/2}}{5n}\right), \ t \leq n^2. \qquad (6.25)$$

The bound becomes 1 for $t_0 = \frac{25n^2}{K^2} \log^2 n$. So

$$\begin{aligned} EC^{[K]} &= \sum_{t=1}^{\infty} P(C^{[K]} \geq t) \\ &\leq \lceil t_0 \rceil + \sum_{t=\lceil t_0 \rceil + 1}^{n^2 - 1} n \exp\left(-\frac{K t^{1/2}}{5n}\right) + \sum_{t=n^2}^{\infty} P(C^{[K]} \geq t) \\ &= \lceil t_0 \rceil + S_1 + S_2, \text{ say,} \end{aligned}$$

and the issue is to show that S_1 and S_2 are $o(t_0)$. To handle S_2, split the set of K walks into subsets of sizes $K-1$ and 1. By independence, for $t \geq n^2$ we have $P(C^{[K]} \geq t) \leq P(C^{[K-1]} \geq n^2) P(C^{[1]} \geq t)$. Then

$$\begin{aligned} S_2 &\leq P(C^{[K-1]} \geq n^2) EC^{[1]} \text{ by summing over } t \\ &\leq n \exp(-(K-1)/5) \cdot 6n^2 \text{ by (6.25) and Theorem 6.4} \\ &= o(t_0) \text{ using the hypothesis } K \geq 6 \log n. \end{aligned}$$

To bound S_1 we start with a calculus exercise: for $u > 1$

$$\int_{u^2}^{\infty} \exp(-x^{1/2}) \, dx = \int_u^{\infty} 2y \, \exp(-y) \, dy \text{ by putting } x = y^2$$

$$\begin{aligned} &\leq 2e^{-1}u\int_u^\infty \exp(-\left(\frac{u-1}{u}\right)y)\ dy \text{ , using } \frac{y}{u} \leq \exp(\frac{y}{u}-1) \\ &= \frac{2u^2\exp(-u)}{u-1}. \end{aligned}$$

The sum S_1 is bounded by the corresponding integral over $[t_0, \infty)$ and the obvious calculation, whose details we omit, bounds this integral by $2t_0/(\log n - 1)$.

6.4.2 Bounding point probabilities

Our second application is to universal bounds on point probabilities. A quite different universal bound will be given in Chapter yyy.

Proposition 6.18 *For continuous-time random walk on a regular n-vertex graph,*

$$\begin{aligned} P_i(X_t = j) &\leq 5t^{-1/2},\ t \leq n^2 \\ &\leq \frac{1}{n} + \frac{K_1}{n}\exp\left(\frac{-t}{K_2 n^2}\right),\ t \geq n^2 \end{aligned}$$

where K_1 and K_2 are absolute constants.

In discrete time one can get essentially the same result, but with the bounds multiplied by 2, though we shall not give details (see Notes).

Proof. $P_i(X_t = i)$ is decreasing in t, so

$$P_i(X_t = i) \leq t^{-1}\int_0^t P_i(X_s = i)ds = t^{-1}E_i N_i(t) \leq 5t^{-1/2}$$

where the last inequality is Proposition 6.16 (iii), whose proof is unchanged in continuous time, and which holds for $t \leq n^2$. This gives the first inequality when $i = j$, and the general case follows from Chapter 3 yyy.

For the second inequality, recall the definition of *separation* $s(t)$ from Chapter 4 yyy. Given a vertex i and a time t, there exists a probability distribution θ such that

$$P_i(X_t \in \cdot) = (1 - s(t))\pi + s(t)\theta.$$

Then for $u \geq 0$,

$$P_i(X_{t+u} = j) - \frac{1}{n} = s(t)\left(P_\theta(X_u = j) - \frac{1}{n}\right).$$

Thus, defining $q(t) = \max_{i,j} \left(P_i(X_t = j) - \frac{1}{n} \right)$, we have proved

$$q(t+u) \le s(t)q(u);\ t, u \ge 0. \tag{6.26}$$

Now $q(n^2) \le 4/n$ by the first inequality of the Proposition, and $s(\tau_1^{(1)}) = e^{-1}$ by definition of $\tau_1^{(1)}$ in Chapter 4 yyy, so by iterating (6.26) we have

$$q(n^2 + m\tau_1^{(1)}) \le \frac{4}{n}\, e^{-m},\ m \ge 1. \tag{6.27}$$

But by Chapter 4 yyy we have $\tau_1^{(1)} \le K\tau^*$ for an absolute constant K, and then by Corollary 6.9 we have $\tau_1^{(1)} \le 3Kn^2$. The desired inequality now follows from (6.27).

6.4.3 A cat and mouse game

Here we reconsider the cat and mouse game discussed in Chapter 4 section yyy. Recall that the cat performs continuous-time random walk on a n-vertex graph, and the mouse moves according to some arbitrary deterministic strategy. Let M be the first meeting time, and let m^* be the maximum of EM over all pairs of initial vertices and all strategies for the mouse.

Proposition 6.19 *On a regular graph, $m^* \le KN^2$ for some absolute constant K.*

Proof. The proof relies on Proposition 6.18, whose conclusion implies there exists a constant K such that

$$p^*(t) \equiv \max_{x,v} p_{vx}(t) \le \frac{1}{n} + Kt^{-1/2};\ 0 \le t < \infty.$$

Consider running the process forever. The point is that, regardless of the initial positions, the chance that the cat and mouse are "together" (i.e. at the same vertex) at time u is at most $p^*(u)$. So in the case where the cat starts with the (uniform) stationary distribution,

$$\begin{aligned} P(\text{ together at time } s\ ) &= \int_0^s f(u) P(\text{ together at time } s | M = u)\ du \\ &\quad \text{where } f \text{ is the density function of } M \\ &\le \int_0^s f(u) p^*(s-u) du \\ &\le \frac{1}{n} P(M \le s)\ + K \int_0^s f(u)(s-u)^{-1/2} du. \end{aligned}$$

So

$$\begin{aligned}
\frac{t}{n} &= \int_0^t P(\text{ together at time } s\ )\, ds \text{ by stationarity} \\
&\le \frac{1}{n}\int_0^t P(M \le s) ds \ + K \int_0^t f(u) du \int_u^t (s-u)^{-1/2} ds \\
&= \frac{t}{n} - \frac{1}{n}\int_0^t P(M > s) ds + 2K \int_0^t f(u)(t-u)^{1/2} du \\
&\le \frac{t}{n} - \frac{1}{n} E \min(M, t) + 2Kt^{1/2}.
\end{aligned}$$

Rearranging, $E\min(M,t) \le 2Knt^{1/2}$. Writing $t_0 = (4Kn)^2$, Markov's inequality gives $P(M \le t_0) \ge 1/2$. This inequality assumes the cat starts with the stationary distribution. When it starts at some arbitrary vertex, we may use the definition of separation $s(u)$ (recall Chapter 4 yyy) to see $P(M \le u + t_0) \ge (1 - s(u))/2$. Then by iteration, $EM \le \frac{2(u+t_0)}{1-s(u)}$. So appealing to the definition of $\tau_1^{(1)}$,

$$m^* \le \frac{2}{1-e^{-1}}(t_0 + \tau_1^{(1)}).$$

But results from Chapter 4 and this chapter show $\tau_1^{(1)} = O(\tau^*) = O(n^2)$, establishing the Proposition.

6.5 Hitting time bounds and connectivity

The results so far in this chapter may be misleading in that upper bounds accommodating extremal graphs are rather uninformative for "typical" graphs. For a family of n-vertex graphs with $n \to \infty$, consider the property

$$\tau^* = O(n). \tag{6.28}$$

(in this order-of-magnitude discussion, τ^* is equivalent to $\max_{v,x} E_v T_x$). Recalling from Chapter 3 yyy that $\tau^* \ge 2(n-1)$, we see that (6.28) is equivalent to $\tau^* = \Theta(n)$. By Matthews' method (repeated as Theorem 2.6 below), (6.28) implies $EC = O(n \log n)$, and then by Theorem 6.31 we have $EC = \Theta(n \log n)$. Thus understanding when (6.28) holds is fundamental to understanding order-of-magnitude questions about cover times. But surprisingly, this question has not been studied very carefully. An instructive example in the d-dimensional torus (Chapter 5 Example yyy), where (6.28) holds iff $d \ge 3$. This example, and other examples of vertex-transitive

graphs satisfying (6.28) discussed in Chapter 8, suggest that (6.28) is frequently true. More concretely, the torus example suggests that the following condition ("the isoperimetric property in $2+\varepsilon$ dimensions") may be sufficient.

Open Problem 6.20 *Show that for real $1/2 < \gamma < 1$ and $\delta > 0$, there exists a constant $K_{\gamma,\delta}$ with the following property. Let G be a regular n-vertex graph such that, for any subset A of vertices with $|A| \le n/2$, there exist at least $\delta|A|^\gamma$ edges between A and A^c. Then $\tau^* \le K_{\gamma,\delta}\ n$.*

The $\gamma = 1$ case is implicit in results from previous chapters. Chapter 3 yyy gave the bound $\max_{i,j} E_i T_j \le 2 \max_j E_\pi T_j$, and Chapter 3 yyy gave the bound $E_\pi T_j \le \tau_2/\pi_j$. This gives the first assertion below, and the second follows from Cheeger's inequality.

Corollary 6.21 *On a regular graph,*

$$\max_{v,x} E_v T_x \le 2n\tau_2 \le 16 n \tau_c^2.$$

Thus the "expander" property that $\tau_2 = O(1)$, or equivalently that $\tau_c = O(1)$, is sufficient for (6.28), and the latter is the $\gamma = 1$ case of Open Problem 6.20.

6.5.1 Edge-connectivity

At the other end of the spectrum from expanders, we can consider graphs satisfying only a little more than connectivity.

xxx more details in proofs – see Fill's comments.

Recall that a graph is *r-edge-connected* if for each proper subset A of vertices there are at least r edges linking A with A^c. By a variant of Menger's theorem (e.g. [86] Theorem 5.11), for each pair (a, b) of vertices in such a graph, there exist r paths $(a = v_0^i, v_1^i, v_2^i, \ldots, v_{m_i}^i = b)$, $i = 1, \ldots, r$ for which the edges (v_j^i, v_{j+1}^i) are all distinct.

Proposition 6.22 *For a r-edge-connected graph,*

$$\tau^* \le \frac{\bar{d} n^2 \psi(r)}{r^2}$$

where ψ is defined by

$$\psi\left(\frac{i(i+1)}{2}\right) = i$$

$$\psi(\cdot) \text{ is linear on } \left[\frac{i(i+1)}{2}, \frac{(i+1)(i+2)}{2}\right].$$

Note $\psi(r) \sim \sqrt{2r}$. So for a d-regular, d-edge-connected graph, the bound becomes $\sim 2^{1/2} d^{-1/2} n^2$ for large d, improving on the bound from Corollary 6.9. Also, the Proposition improves on the bound implied by Chapter 4 yyy in this setting.

Proof. Given vertices a, b, construct a unit flow from a to b by putting flow $1/r$ along each of the r paths $(a = v_0^i, v_1^i, v_2^i, \ldots, v_{m_i}^i = b)$. By Chapter 3 Theorem yyy

$$E_a T_b + E_b T_a \le \bar{d} n (1/r)^2 M$$

where $M = \sum_i m_i$ is the total number of edges in the r paths. So the issue is bounding M. Consider the digraph of all edges (v_j^i, v_{j+1}^i). If this digraph contained a directed cycle, we could eliminate the edges on that cycle, and still create r paths from a to b using the remaining edges. So we may assume the digraph is acyclic, which implies we can label the vertices as $a = 1, 2, 3, \ldots, n = b$ in such a way that each edge (j, k) has $k > j$. So the desired result follows from

Lemma 6.23 *In a digraph on vertices $\{1, 2, \ldots, n\}$ consisting of r paths $1 = v_0^i < v_1^i < v_2^i < \ldots v_{m_i}^i = n$ and where all edges are distinct, the total number of edges is at most $n\psi(r)$.*

Proof.

xxx give proof and picture.

Example 6.24 Take vertices $\{0, 1, \ldots, n-1\}$ and edges $(i, i+u \bmod n)$ for all i and all $1 \le u \le \kappa$.

This example highlights the "slack" in Proposition 6.22. Regard κ as large and fixed, and $n \to \infty$. Random walk on this graph is classical random walk (i.e. sums of independent steps) on the n-cycle, where the steps have variance $\sigma^2 = \frac{1}{\kappa} \sum_{u=1}^{\kappa} u^2$, and it is easy to see

$$\tau^* = 2 E_0 T_{\lfloor n/2 \rfloor} \sim \frac{(n/2)^2}{\sigma^2} = \Theta(n^2/\kappa^2).$$

This is the bound Proposition 6.22 would give if the graph were $\Theta(\kappa^2)$-edge-connected. And for a "typical" subset A such as an interval of length greater than κ there are indeed $\Omega(\kappa^2)$ edges crossing the boundary of A. But by considering a singleton A we see that the graph is really only 2κ-edge-connected, and Proposition 6.22 gives only the weaker $O(n^2/\kappa^{1/2})$ bound.

xxx tie up with similar discussion of τ_2 and connectivity being affected by small sets; better than bound using τ_c only.

6.5.2 Equivalence of mean cover time parameters

Returning to the order-of-magnitude discussion at the start of section 6.5, let us record the simple equivalence result. Recall (cf. Chapter 4 yyy) we call parameters *equivalent* if their ratios are bounded by absolute constants.

Lemma 6.25 *The parameters* $\max_i E_i C^+$, $E_\pi C^+$, $\min_i E_i C^+$, $\max_i E_i C$ *and* $E_\pi C$ *are equivalent for reversible chains, but* $\min_i E_i C$ *is not equivalent to these.*

Proof. Of the five parameters asserted to be equivalent, it is clear that $\max_i E_i C$ is the largest, and that either $\min_i E_i C^+$ or $E_\pi C$ is the smallest, so it suffices to prove

$$\max_i E_i C^+ \le 4 E_\pi C \tag{6.29}$$

$$\max_j E_j C^+ \le 3 \min_i E_i C^+. \tag{6.30}$$

Inequality (6.30) holds by concatenating three "cover-and-return" cycles starting at i and considering the first hitting time on j in the first and third cycles. In more detail, write

$$\Gamma(s) = \min\{u > s : (X_t : s \le t \le u) \text{ covers all states}\}.$$

For the chain started at i write $C^{++} = \Gamma(C^+)$ and $C^{+++} = \Gamma(C^{++})$. Since $T_j < C^+$ we have $\Gamma(T_j) \le C^{++}$. So the chain started at time T_j has covered all states and returned to j by time C^{+++}, implying $E_j C^+ \le E C^{+++} = 3 E_i C^+$. For inequality (6.29), recall the *random target lemma*: the mean time to hit a π-random state V equals τ_0, regardless of the initial distribution. The inequality

$$E_i C^+ \le \tau_0 + E_\pi C + \tau_0 + E_\pi T_i$$

follows from the four-step construction:
(i) Start the chain at i and run until hitting a π-random vertex V at time T_V;
(ii) continue until time $\Gamma(T_V)$;
(iii) continue until hitting an independent π-random vertex V';
(iv) continue until hitting i.
But $E_\pi T_i \le E_\pi C$, and then by the random target lemma $\tau_0 \le E_\pi C$, so (6.29) follows.

For the final assertion, on the lollipop graph (Chapter 5 Example yyy) one has $\min_i E_i C = \Theta(n^2)$ while the other quantities are $\Theta(n^3)$. One can also give examples on regular graphs (see Notes).

6.6 Lower bounds

6.6.1 Matthews' method

We restate Matthews' method (Chapter 2 yyy) as follows. The upper bound is widely useful: we have already used it several times in this chapter, and will use it several more times in the sequel.

Theorem 6.26 *For a general Markov chain,*

$$\max_v E_v C \le h_{n-1} \max_{i,j} E_i T_j.$$

And for any subset A of states,

$$\min_v E_v C \ge h_{|A|-1} \min_{i \ne j:\ i,j \in A} E_i T_j.$$

In Chapter 2 we proved the lower bound in the case where A was the entire state space, but the result for general A follows by the same proof, taking the J's to be a uniform random ordering of the states in A. One obvious motivation for the more general formulation comes from the case of trees, where for a leaf l we have $\min_j E_l T_j = 1$, so the lower bound with A being the entire state space would be just h_{n-1}. We now illustrate use of the more general formulation.

6.6.2 Balanced trees

We are accustomed to finding that problems on trees are simpler than problems on general graphs, so it is a little surprising to discover that one of the graphs where studying the mean cover time is difficult is the balanced r-tree of height H (Chapter 5 Example yyy). Recall this tree has $n = (r^{H+1} - 1)/(r - 1)$ vertices, and that (by the commute interpretation of resistance)

$$E_i T_j = 2m(n-1) \text{ for leaves } (i,j) \text{ distance } 2m \text{ apart.}$$

Now clearly $E_i T_j$ is maximized by some pair of leaves, so $\max_{i,j} E_i T_j = 2H(n-1)$. Theorem 6.26 gives

$$\max_v E_v C \le 2H(n-1)h_{n-1} \sim 2Hn \log n.$$

To get a lower bound, consider the set S_m of r^{H+1-m} vertices at depth $H + 1 - m$, and let A_m be a set of leaves consisting of one descendant of

each element of S_m. The elements of A_m are at least $2m$ apart, so applying the lower bound in Theorem 6.26

$$\begin{aligned} \min_v E_v C &\geq \max_m 2m(n-1)\ h_{r^{H+1-m}} \\ &\sim 2n \log r\ \max_m m(H-m) \\ &\sim \frac{1}{2} H^2 n \log r. \end{aligned}$$

It turns out that this lower bound is asymptotically off by a factor of 4, while the upper bound is asymptotically correct.

Theorem 6.27 ([16]) *On the balanced r-tree, as $H \to \infty$ for arbitrary starting vertex,*

$$EC \sim 2Hn \log n \sim \frac{2H^2 r^{H+1} \log r}{r-1}$$

Improving the lower bound to obtain this result is not easy. The natural approach (used in [16]) is to seek a recursion for the cover time distribution $C^{(H+1)}$ in terms of $C^{(H)}$. But the appropriate recursion is rather subtle (we invite the reader to try to find it!) so we won't give the statement or analysis of the recursion here.

6.6.3 A resistance lower bound

Our use of the commute interpretation of resistance has so far been only to obtain *upper* bounds on commute times. One can also use "shorting" ideas to obtain lower bounds, and here is a very simple implementation of that idea.

Lemma 6.28 *The effective resistance between $r(v,x)$ between vertices v and x in a weighted graph satisfies*

$$\frac{1}{r(v,x)} \leq w_{v,x} + \frac{1}{\frac{1}{w_v - w_{v,x}} + \frac{1}{w_x - w_{v,x}}}.$$

In particular, on an unweighted graph

$$\begin{aligned} r(v,x) &\geq \frac{d_v + d_x - 2}{d_v d_x\ - 1} \text{ if } (v,x) \text{ is an edge} \\ &\geq \frac{1}{d_v} + \frac{1}{d_w} \text{ if not} \end{aligned}$$

and on an unweighted d-regular graph

$$\begin{aligned} r(v,x) &\geq \frac{2}{d+1} \text{ if } (v,x) \text{ is an edge} \\ &\geq \frac{2}{d} \text{ if not .} \end{aligned}$$

So on an unweighted d-regular n-vertex graph,

$$\begin{aligned} E_v T_x + E_x T_v &\geq \frac{2dn}{d+1} \text{ if } (v,x) \text{ is an edge} \\ &\geq 2n \text{ if not .} \end{aligned}$$

Proof. We need only prove the first assertion, since the others follow by specialization and by the commute interpretation of resistance. Let A be the set of vertices which are neighbors of either v or x, but exclude v and x themselves from A. Short the vertices of A together, to form a single vertex a. In the shorted graph, the only way current can flow from v to x is directly $v \to x$ or indirectly as $v \to a \to x$. So, using $'$ to denote the shorted graph, the effective resistance $r'(v,x)$ in the shorted graph satisfies

$$\frac{1}{r'(v,x)} = w'_{v,x} + \frac{1}{\frac{1}{w'_{v,a}} + \frac{1}{w'_{x,a}}}.$$

Now $w'_{x,v} = w_{x,v}$, $w'_{v,a} = w_v - w_{v,x}$ and $w'_{x,a} = w_x - w_{v,x}$. Since shorting decreases resistance, $r'(v,x) \leq r(v,x)$, establishing the first inequality.

6.6.4 General lower bounds

Chapter 3 yyy shows that, over the class of random walks on n-vertex graphs or the larger class of reversible chains on n states, various mean hitting time parameters are minimized on the complete graph. So it is natural to anticipate a similar result for cover time parameters. But the next example shows that some care is required in formulating conjectures.

Example 6.29 Take the complete graph on n vertices, and add an edge (v,l) to a new leaf l.

Since random walk on the complete graph has mean cover time $(n-1)h_{n-1}$, random walk on the enlarged graph has

$$E_l C = 1 + (n-1)h_{n-1} + 2\mu$$

where μ is the mean number of returns to l before covering. Now after each visit to v, the walk has chance $1/n$ to visit l on the next step, and so the mean number of visits to l before visiting some other vertex of the complete graph equals $1/(n-1)$. We may therefore write μ in terms of expectations for random walk on the complete graph as

$$\begin{aligned}
\mu &= \frac{1}{n-1} E_v(\text{ number of visits to } v \text{ before } C) \\
&= \frac{1}{n-1} E_v(\text{ number of visits to } v \text{ before } C^+) \\
&= \frac{1}{n-1}\ \frac{1}{n} E_v C^+ \text{ by Chapter 2 Proposition yyy} \\
&= \frac{1+h_{n-1}}{n} \text{ by (6.12).}
\end{aligned}$$

This establishes an expression for $E_l C$, which (after a brief calculation) can be rewritten as

$$E_l C = n h_n - \left(1 - \frac{2}{n}\right)\left(h_n - \frac{1}{n}\right).$$

Now random walk on the complete $(n+1)$-graph has mean cover time nh_n, so $E_l C$ is smaller in our example than in the complete graph.

The example motivates the following as the natural "exact extremal conjecture".

Open Problem 6.30 *Prove that, for any reversible chain on n states,*

$$E_\pi C \geq (n-1) h_{n-1}$$

(the value for random walk on the complete graph).

The related asymptotic question was open for many years, and was finally proved by Feige [142].

Theorem 6.31 *For random walk on an unweighted n-vertex graph,*

$$\min_v E_v C \geq c_n,$$

where $c_n \sim n \log n$ as $n \to \infty$.

The proof is an intricate mixture of many of the techniques we have already described.

6.7 Distributional aspects

In many examples one can apply the following result to show that hitting time distributions become exponential as the size of state space increases.

Corollary 6.32 *Let i, j be arbitrary states in a sequence of reversible Markov chains.*
(i) If $E_\pi T_j/\tau_2 \to \infty$ then

$$P_\pi\left(\frac{T_j}{E_\pi T_j} > x\right) \to e^{-x},\ 0 < x < \infty.$$

(ii) If $E_i T_j/\tau_1 \to \infty$ and $E_i T_j \geq (1 - o(1))E_\pi T_j$ then $E_i T_j/E_\pi T_j \to 1$ and

$$P_i\left(\frac{T_j}{E_i T_j} > x\right) \to e^{-x},\ 0 < x < \infty.$$

Proof. In continuous time, assertion (i) is immediate from Chapter 3 Proposition yyy. The result in discrete time now holds by continuization: if T_j is the hitting time in discrete time and T'_j in continuous time, then $E_\pi T'_j = E_\pi T_j$ and $T'_j - T_j$ is order $\sqrt{E_\pi T_j}$. For (ii) we have (cf. Chapter 4 section yyy) $T_j \leq U_i + T^*_j$ where T_j is the hitting time started at i, T^*_j is the hitting time started from stationarity, and $E_i U_i \leq \tau_1^{(2)}$. So $ET_j \leq ET^* + O(\tau_1)$, and the hypotheses of (ii) force $ET_j/ET^*_j \to 1$ and force the limit distribution of T_j/ET_j to be the same as the limit distribution of T^*_j/ET^*_j, which is the exponential distribution by (i) and the relation $\tau_2 \leq \tau_1$. □

In the complete graph example, C has mean $\sim n \log n$ and s.d. $\Theta(n)$, so that $C/EC \to 1$ in distribution, although the convergence is slow. The next result shows this "concentration" result holds whenever the mean cover time is essentially larger than the maximal mean hitting time.

Theorem 6.33 ([17]) *For states i in a sequence of (not necessarily reversible) Markov chains,*

$$\text{if } E_i C/\tau^* \to \infty \text{ then } P_i\left(\left|\frac{C}{E_i C} - 1\right| > \varepsilon\right) \to 0,\ \varepsilon > 0.$$

The proof is too long to reproduce.

6.8 Algorithmic aspects

Many of the mathematical results in this chapter arose originally from algorithmic questions, so let me briefly describe the questions and their relation to the mathematical results.

6.8.1 Universal traversal sequences

This was one motivation for the seminal paper [25]. Consider an n-vertex d-regular graph G, with a distinguished vertex v_0, and where for each vertex v the edges at v are labeled as $1, 2, \ldots, d$ in some way – it is not required that the labels be the same at both ends of an edge. Now consider a sequence $\mathbf{i} = (i_1, i_2, \ldots, i_L) \in \{1, \ldots, d\}^L$. The sequence defines a deterministic walk (x_i) on the vertices of G via

$x_0 = v_0$

(x_{j-1}, x_j) is the edge at x_{j-1} labeled i_j.

Say $\mathbf{i}$ is a *traversal sequence* for G if the walk $(x_i : 0 \le i \le L)$ visits every vertex of G. Say $\mathbf{i}$ is a *universal traversal sequence* if it is a traversal sequence for every graph G in the set $\mathcal{G}_{n,d}$ of edge-labeled graphs with distinguished vertices.

Proposition 6.34 (Aleliunas et al [25]) *There exists a universal traversal sequence of length* $(6e + o(1))dn^3 \log(nd)$ *as* $n \to \infty$ *with* d *varying arbitrarily.*

Proof. It is enough to show that a uniform random sequence of that length has non-zero chance to be a universal traversal sequence. But for such a random sequence, the induced walk on a fixed G is just simple random walk on the vertices of G. Writing $t_0 = \lceil 6en^2 \rceil$, Theorem 6.4 implies

$$P_v(C > t_0) \le \frac{E_v C}{t_0} \le \frac{6n^2}{t_0} \le e^{-1} \text{ for all initial } v$$

and so inductively (cf. Chapter 2 section yyy)

$$P_{v_0}(C > Kt_0) \le e^{-K}, \quad K \ge 1 \text{ integer} .$$

Thus by taking K sufficiently large that

$$e^{-K}|\mathcal{G}_{n,d}| < 1$$

there is non-zero chance that the induced walk on *every* G covers before time Kt_0. The crude bound $|\mathcal{G}_{n,d}| \le (nd)^{nd}$ means we may take $K = \lceil nd \log(nd) \rceil$.

6.8.2 Graph connectivity algorithms

Another motivation for the seminal paper [25] was the time-space tradeoff in algorithms for determining connectivity in graphs. Here is a highly informal presentation, illustrated by the two *Mathematician graphs*. The vertices are all mathematicians (living or dead). In the first graph, there is an edge between two mathematicians if they have written a joint paper; in the second, there is an edge if they have written two or more joint papers. A well known Folk Theorem asserts that the first graph has a giant component containing most famous mathematicians; a lesser known and more cynical Folk Theorem asserts that the second graph doesn't. Suppose we actually want to answer a question of that type – specifically, take two mathematicians (say, the reader and Paul Erdos) and ask if they are in the same component of the first graph. Suppose we have a database which, given a mathematician's name, will tell us information about their papers and in particular will list all their co-authors.

xxx continue story

Broder et al [67]

6.8.3 A computational question

Consider the question of getting a numerical value for E_iC (up to error factor $1 \pm \varepsilon$, for fixed ε) for random walk on a n-vertex graph. Using Theorem 6.1 it's clear we can do this by Monte Carlo simulation in $O(n^3)$ steps.

xxx technically, using s.d./mean bounded by submultiplicitivity.

Open Problem 6.35 Can E_iC be deterministically calculated in a polynomial (in n) number of steps?

It's clear one can compute mean *hitting* times on a n-step chain in polynomial time, but to set up the computation of E_iC as a hitting-time problem one has to incorporate the subset of already-visited states into the "current state", and thus work with hitting times for a $n \times 2^{n-1}$-state chain.

6.9 Notes on Chapter 6

Attributions for what I regard as the main ideas were given in the text. The literature contains a number of corollaries or variations of these ideas, some of which I've used without attribution, and many of which I haven't mentioned at all. A number of these ideas can be found in Zuckerman

[341, 343], Palacios [276, 273] and the Ph.D. thesis of Sbihi [306], as well as papers cited elsewhere.

Section 6.1. The conference proceedings paper [25] proving Theorem 6.1 was not widely known, or at least its implications not realized, for some years. Several papers subsequently appeared proving results which are consequences (either obvious, or via the general relations of Chapter 4) of Theorem 6.1. I will spare their authors embarrassment by not listing them all here!

The spanning tree argument shows, writing b_e for the mean commute time across an edge e, that

$$\max_v E_v C^+ \le \min_{\mathcal{T}} \sum_{e \in \mathcal{T}} b_e.$$

Coppersmith et al [100] give a deeper study and show that the right side is bounded between γ and $10\gamma/3$, where

$$\gamma = \left(\sum_v d_v \right) \left(\sum_v \frac{1}{d_v + 1} \right).$$

The upper bound is obtained by considering a random spanning tree, cf. Chapter yyy.

Section 6.2. The calculations in these examples, and the uniformity property of V on the n-cycle, are essentially classical. For the cover time C_n on the n-cycle there is a non-degenerate limit distribution $n^{-2} C_n \stackrel{d}{\to} C$. From the viewpoint of weak convergence (Chapter yyy), C is just the cover time for Brownian motion on the circle of unit circumference, and its distribution is known as part of a large family of known distributions for maximal-like statistics of Brownian motion: Imhof [187] eq. (2.4) gives the density as

$$f_C(t) = 2^{3/2} \pi^{-1/2} t^{-3/2} \sum_{m=1}^{\infty} (-1)^{m-1} m^2 \exp(-\frac{m^2}{2t}).$$

Sbihi [306] gives a direct derivation of a different representation of f_C.

Section 6.4. Use of Lemma 6.10 in the random walk context goes back at least to Flatto et al [152].

Barnes and Feige [42] give a more extensive treatment of short-time bounds in the irregular setting, and their applications to covering with multiple walks (cf. Proposition 6.17 and section 6.8.2). They also give bounds on the mean time taken to cover μ different edges or ν different vertices – their bound for the latter becomes $O(\nu^2 \log \nu)$ on regular graphs.

Proposition 6.18 implies that on an infinite regular graph $P_i(X_t = j) \leq Kt^{-1/2}$. Carlen et al [84] Theorem 5.14 prove this as a corollary of results using more sophisticated machinery. Our argument shows the result is fairly elementary. In discrete time the analog of the first inequality can be proved using the "CM proxy" property than $P_i(X_{2t} = i) + P_i(X_{2t+1} = i)$ is decreasing, but the analog of the second inequality requires different arguments because we cannot exploit the $\tau_1^{(1)}$ inequalities.

Section 6.5. Variations on Corollary 6.21 are given in Broder and Karlin [66] and Chandra et al. [85].

Upper bounds on mean hitting times imply upper bounds on the relaxation time τ_2 via the general inequalities $\tau_2 \leq \tau_0 \leq \frac{1}{2}\tau^*$. In most concrete examples these bounds are too crude to be useful, but in "extremal" settings these bounds are essentially as good as results seeking to bound τ_2 directly. For instance, in the setting of a d-regular r-edge-connected graph, a direct bound (Chapter 4 Proposition yyy) gives

$$\tau_2 \leq \frac{d}{4r \sin^2 \frac{\pi}{2n}} \sim \frac{dn^2}{\pi^2 r}.$$

Up to the numerical constant, the same bound is obtained from Proposition 6.22 and the general inequality $\tau_2 \leq \tau^*/2$.

xxx contrast with potential and Cheeger-like arguments ?

To sketch an example of a regular graph where $\min_i E_i C$ has a different order than $\max_i E_i C$, make a regular $m_1 + m_2$-vertex graph from a m_1-vertex graph with mean cover time $\Theta(m_1 \log m_1)$ and a m_2-vertex graph (such as the necklace) with mean cover time $\Theta(m_2^2)$, for suitable values of the m's. Starting from a typical vertex of the former, the mean cover time is $\Theta(m_1 \log m_1 + m_1 m_2 + m_2^2)$ whereas starting from the unattached end of the necklace the mean cover time is $\Theta(m_1 \log m_1 + m_2^2)$. Taking $m_1 \log m_1 + m_2^2) = o(m_1 m_2)$ gives the desired example.

Section 6.6. The "subset" version of Matthews' lower bound (Theorem 2.6) and its application to trees were noted by Zuckerman [343], Sbihi [306] and others. As well as giving a lower bound for balanced trees, these authors give several lower bounds for more general trees satisfying various constraints (cf. the unconstrained result, Proposition 6.7). As an illustration, Devroye - Sbihi [111] show that on a tree

$$\min_v E_v C \geq \frac{(1 + o(1))n \log^2 n}{2 \log(d^* - 1)} \text{ if } d^* \equiv \max_v d_v = n^{o(1)}.$$

I believe that the recursion set-up in [16] can be used to prove Open Problem 6.35 on trees, but I haven't thought carefully about it.

The "shorting" lower bound, Lemma 6.28, was apparently first exploited by Coppersmith et al [100].

Section 6.7. Corollary 6.32 encompasses a number of exponential limit results proved in the literature by *ad hoc* calculations in particular examples.

Section 6.8.1. Proposition 6.34 is one of the neatest instances of "Erdos's Probabilistic Method in Combinatorics", though surprisingly it isn't in the recent book [29] on that subject. Constructing explicit universal traversal sequences is a hard open problem: see Borodin et al [56] for a survey.

Section 6.8.2. See [67] for a more careful discussion of the issues. The alert reader of our example will have noticed the subtle implication that the reader has written fewer papers than Paul Erdos, otherwise (why?) it would be preferable to do the random walk in the other direction.

Miscellaneous. Condon and Hernek [98] study cover times in the following setting. The edges of a graph are colored, a sequence (c_t) of colors is prespecified and the "random walk" at step t picks an edge uniformly at random from the color-c_t edges at the current vertex.

Chapter 7

Symmetric Graphs and Chains (January 31, 1994)

In this Chapter we show how general results in Chapters 3, 4 and 6 can sometimes be strengthened when symmetry is present. Many of the ideas are just simple observations. Since the topic has a "discrete math" flavor our default convention is to work in discrete time, though as always the continuous-time case is similar. Note that we use the word "symmetry" in the sense of *spatial* symmetry (which is the customary use in mathematics as a whole) and not as a synonym for time-reversibility. Note also our use of "random flight" for what is usually called "random walk" on a group.

Biggs [48] contains an introductory account of symmetry properties for graphs, but we use little more than the definitions. I have deliberately not been overly fussy about giving weakest possible hypotheses. For instance many results for symmetric reversible chains depend only of the symmetry of mean hitting times (7.7), but I haven't spelt this out. Otherwise one can end up with more definitions than serious results! Instead, we focus on three different strengths of symmetry condition. Starting with the weakest, section 7.1 deals with symmetric reversible chains, a minor generalization of what is usually called "symmetric random walk on a finite group". In the graph setting, this specializes to random walk on a Cayley or vertex-transitive graph. Section 7.2 deals with random walk on an arc-transitive graph, encompassing what is usually called "random walk on a finite group with steps uniform on a conjugacy class". Section 5.16 deals with random walk on a distance-regular graph, which roughly corresponds to nearest-neighbor isotropic random walk on a discrete Gelfand pair.

This book focuses on inequalities rather than exact calculations, and the

limitation of this approach is most apparent in this chapter. Group representation theory, though of course developed for non-probabilistic reasons, turns out to be very well adapted to the study of many questions concerning random walks on groups. I lack the space (and, more importantly, the knowledge) to give a worthwhile treatment here, and in any case an account which is both introductory and gets to interesting results is available in Diaconis [112]. In many concrete examples, eigenvalues are known by group representation theory, and so in particular our parameters τ_2 and τ_0 are known. See e.g. section 7.2.1. In studying a particular example, after investigating eigenvalues one can seek to study further properties of the chain by either
(i) continuing with calculations specific to the example; or
(ii) using general inequalities relating other aspects of the chain to τ_2 and τ_0.
The purpose of this Chapter is to develop option (ii). Of course, the more highly-structured the example, the more likely one can get stronger explicit results via (i). For this reason we devote more space to the weaker setting of section 7.1 than to the stronger settings of sections 7.2 and 5.16.

xxx scattering of more sophisticated math in Chapter 10.

7.1 Symmetric reversible chains

7.1.1 Definitions

Consider an irreducible transition matrix $\mathbf{P} = (p_{ij})$ on a finite state space I. A *symmetry* of $\mathbf{P}$ is a $1-1$ map $\gamma : I \to I$ such that

$$p_{\gamma(i)\gamma(j)} = p_{ij} \text{ for all } i, j.$$

The set Γ of symmetries forms a group under convolution, and in our (non-standard) terminology a *symmetric* Markov transition matrix is one for which Γ acts transitively, i.e.

$$\text{for all } i, j \in I \text{ there exists } \gamma \in \Gamma \text{ such that } \gamma(i) = j.$$

Such a chain need not be reversible; a *symmetric reversible chain* is just a chain which is both symmetric and reversible. A natural setting is where I is itself a group under an operation $(i, j) \to ij$ which we write multiplicitively. If μ is a probability distribution on I and $(Z_t; t \geq 1)$ are i.i.d. I-valued with distribution μ then

$$X_t = x_0 Z_1 Z_2 \ldots Z_t \tag{7.1}$$

is the symmetric Markov chain with transition probabilities

$$p_{ij} = \mu(i^{-1} * j)$$

started at x_0. This chain is reversible iff

$$\mu(i) = \mu(i^{-1}) \text{ for all } i. \tag{7.2}$$

We have rather painted ourselves into a corner over terminology. The usual terminology for the process (7.1) is "random walk on the group I" and if (7.2) holds then it is a

$$\text{"symmetric random walk on the group } I\text{"} \ . \tag{7.3}$$

Unfortunately in this phrase, both "symmetric" and "walk" conflict with our conventions, so we can't use the phrase. Instead we will use "random flight on the group I" for a process (7.1), and "reversible random flight on the group I" when (7.2) also holds. Note that we always assume chains are irreducible, which in the case of a random flight holds iff the support of μ generates the whole group I. Just keep in mind that the topic of this section, symmetric reversible chains, forms a minor generalization of the processes usually described by (7.3).

On an graph $(\mathcal{V}, \mathcal{E})$, a *graph automorphism* is a $1-1$ map $\gamma : \mathcal{V} \to \mathcal{V}$ such that

$$(\gamma(w), \gamma(v)) \in \mathcal{E} \text{ iff } (w, v) \in \mathcal{E}.$$

The graph is called *vertex-transitive* if the automorphism group acts transitively on vertices. Clearly, random walk on a (unweighted) graph is a symmetric reversible chain iff the graph is vertex-transitive. We specialize to this case in section 7.1.8. A further specialization is to random walk on a *Cayley graph*. If $\mathcal{G} = (g_i)$ is a set of generators of a group I, which we always assume to satisfy

$$g \in \mathcal{G} \text{ implies } g^{-1} \in \mathcal{G}$$

then the associated Cayley graph has vertex-set I and edge-set

$$\{(v, vg) : v \in I, g \in \mathcal{G}\}.$$

A Cayley graph is vertex-transitive.

Finally, recall from Chapter 3 yyy that we can identify a reversible chain with a random walk on a weighted graph. With this identification, a symmetric reversible chain is one where the weighted graph is vertex-transitive, in the natural sense.

7.1.2 This section goes into Chapter 3

Lemma 7.1 *For an irreducible reversible chain, the following are equivalent.*

(a) $P_i(X_t = i) = P_j(X_t = j), i, j \in I, t \geq 1$
(b) $P_i(T_j = t) = P_j(T_i = t), i, j \in I, t \geq 1.$

Proof. In either case the stationary distribution is uniform – under (a), by letting $t \to \infty$, and under (b) by taking $t = 1$, implying $p_{ij} \equiv p_{ji}$. So by reversibility $P_i(X_t = j) = P_j(X_t = i)$ for $i \neq j$ and $t \geq 1$. But recall from Chapter 2 Lemma yyy that the generating functions $G_{ij}(z) = \sum_t P_i(X_t = j)z^j$ and $F_{ij}(z) = \sum_t P_i(T_t = j)z^j$ satisfy

$$F_{ij} = G_{ij}/G_{jj}. \tag{7.4}$$

For $i \neq j$ we have seen that $G_{ij} = G_{ji}$, and hence by (7.4)

$$F_{ij} = F_{ji} \text{ iff } G_{jj} = G_{ii},$$

which is the assertion of Lemma 7.1.

7.1.3 Elementary properties

Our standing assumption is that we have an irreducible symmetric reversible n-state chain. The symmetry property implies that the stationary distribution π is uniform, and also implies

$$P_i(X_t = i) = P_j(X_t = j), i, j \in I, t \geq 1. \tag{7.5}$$

But by Chapter 3 Lemma yyy, under reversibility (7.5) is equivalent to

$$P_i(T_j = t) = P_j(T_i = t), i, j \in I, t \geq 1. \tag{7.6}$$

And clearly (7.6) implies

$$E_i T_j = E_j T_i \text{ for all } i, j. \tag{7.7}$$

We make frequent use of these properties. Incidently, (7.7) is in general strictly weaker than (7.6): van Slijpe [330] p. 288 gives an example with a 3-state reversible chain.

We also have, from the definition of symmetric, that $E_\pi T_i$ is constant in i, and hence

$$E_\pi T_i = \tau_0 \text{ for all } i. \tag{7.8}$$

So by Chapter 4 yyy

$$\tau^* \le 4\tau_0. \tag{7.9}$$

The formula for $E_\pi T_i$ in terms of the fundamental matrix (Chapter 2 yyy) can be written as

$$\tau_0/n = 1 + \sum_{t=1}^{\infty}(P_i(X_t = i) - 1/n). \tag{7.10}$$

Approximating τ_0 by the first few terms is what we call the *local transience heuristic*. See Chapter xxx for rigorous discussion.

Lemma 7.2 *(i)* $E_iT_j \ge \frac{n}{1+p(i,j)}$, $j \ne i$.
(ii) $\max_{i,j} E_iT_j \le 2\tau_0$

Proof. (i) This is a specialization of Chapter 6 xxx.
(ii) For any i, j, k,

$$E_iT_j \le E_iT_k + E_kT_j = E_iT_k + E_jT_k.$$

Averaging over k, the right side becomes $2\tau_0$.

Recall that a simple Cauchy-Schwartz argument (Chapter 3 yyy) shows that, for any reversible chain whose stationary distribution is uniform,

$$P_i(X_{2t} = j) \le \sqrt{P_i(X_{2t} = i)P_j(X_{2t} = j)}.$$

So by (7.5), for a symmetric reversible chain, the most likely place to be after $2t$ steps is where you started:

Corollary 7.3 $P_i(X_{2t} = j) \le P_i(X_{2t} = i)$, *for all* $i, j, \in I, t \ge 1$.

This type of result is nicer in continuous time, where the inequality holds for all times.

7.1.4 Hitting times

Here is our first non-trivial result, from Aldous [12].

Theorem 7.4 *Suppose a sequence of symmetric reversible chains satisfies* $\tau_2/\tau_0 \to 0$. *Then*

(a) For the stationary chain, and for arbitrary j, *we have* $T_j/\tau_0 \overset{d}{\to} \xi$ *and* var $(T_j/\tau_0) \to 1$, *where* ξ *has exponential(1) distribution.*
(b) $\max_{i,j} E_iT_j/\tau_0 \to 1$.
(c) If (i_n, j_n) *are such that* $E_{i_n}T_{j_n}/\tau_0 \to 1$ *then* $P_{i_n}(T_{j_n}/\tau_0 \in \cdot) \overset{d}{\to} \xi$.

Note that, because $\tau_2 \le \tau_1 + 1$ and $\tau_0 \ge (n-1)^2/n$, the hypothesis "$\tau_2/\tau_0 \to 0$" is weaker than either "$\tau_2/n \to 0$" or "$\tau_1/\tau_0 \to 0$".

Part (a) is a specialization of Chapter 3 Proposition yyy and its proof. Parts (b) and (c) use refinements of the same technique. Part (b) implies

$$\text{if } \tau_2/\tau_0 \to 0 \text{ then } \tau^* \sim 2\tau_0.$$

Because this applies in many settings in this Chapter, we shall rarely need to discuss τ^* further.

xxx give proof

In connection with (b), note that

$$E_v T_w \le \tau_1^{(2)} + \tau_0 \tag{7.11}$$

by definition of $\tau_1^{(2)}$ and vertex-transitivity. So (b) is obvious under the slightly stronger hypothesis $\tau_1/\tau_0 \to 0$.

Chapter 3 Proposition yyy actually gives information on hitting times T_A to more general subsets A of vertices. Because (Chapter 3 yyy) $E_\pi T_A \ge \frac{(1-\pi(A))^2}{\pi(A)}$, we get (in continuous time) a quantification of the fact that T_A has approximately exponential distribution when $|A| \ll n/\tau_2$ and when the chain starts with the uniform distribution:

$$\sup_t |P_\pi(T_A > t) - \exp(-t/E_\pi T_A)| \le \frac{\tau_2 n}{|A|} \left(1 - \frac{|A|}{n}\right)^{-2}.$$

7.1.5 Cover times

Recall the cover time C from Chapter 6. By symmetry, in our present setting $E_i C$ doesn't depend on the starting place i, so we can write EC. In this section we combine results on hitting times with various forms of Matthews method to obtain asymptotics for cover times in the setting of a sequence of symmetric reversible chains. Experience, and the informal argument above (7.15), suggest the principle

$$EC \sim \tau_0 \log n, \text{ except for chains resembling random walk on the } n\text{-cycle}. \tag{7.12}$$

The results in this chapter concerning cover times go some way towards formalizing this principle.

Corollary 7.5 *For a sequence of symmetric reversible chains*

(a) $\frac{1-o(1)}{1+p^*}\, n \log n \le EC \le (2+o(1))\tau_0 \log n$, *where* $p^* \equiv \max_{j \ne i} p_{i,j}$.

(b) If $\tau_2/\tau_0 \to 0$ *then* $EC \le (1+o(1))\tau_0 \log n$.

(c) If $\tau_2/\tau_0 = O(n^{-\beta})$ *for fixed* $0 < \beta < 1$ *then*

$$EC \geq (\beta - o(1))\tau_0 \log n.$$

Proof. Using the basic form of Matthews method (Chapter 2 yyy), (a) follows from Lemma 7.2 and (b) from Theorem 7.4. To prove (c), fix a state j and $\varepsilon > 0$. Using (7.11) and Markov's inequality,

$$\pi\{i : E_i T_j \leq (1-\varepsilon)\tau_0\} \leq \frac{\tau_1^{(2)}}{\varepsilon\tau_0} \equiv \alpha, \text{ say.}$$

So we can inductively choose $\lceil \alpha^{-1} \rceil$ vertices i_k such that

$$E_{i_k} T_{i_l} > (1-\varepsilon)\tau_0; \ 1 \leq k < l \leq \lceil \alpha^{-1} \rceil.$$

By the extended form of Matthews method (Chapter 6 Corollary yyy)

$$EC \geq (1-\varepsilon)\tau_0 h_{\lceil \alpha^{-1} \rceil - 1}.$$

From Chapter 4 yyy, $\tau_1 \leq \tau_2(1+\log n)$ and so the hypothesis implies $\tau_1/\tau_0 = O(n^{\varepsilon-\beta})$. So the asymptotic lower bound for EC becomes $(1-\varepsilon)\tau_0(\beta - \varepsilon)\log n$, and since ε is arbitrary the result follows.

Since the only natural examples with $\tau_1/\tau_0 \not\to 0$ are variations of random walk on the n-cycle, for which $EC = \Theta(\tau_0)$ without the "$\log n$" term, we expect a positive answer to

Open Problem 7.6 *In the setting of Corollary 7.5, is* $EC \leq (1+o(1))\tau_0 \log n$ *without further hypotheses?*

Here is an artificial example to illustrate the bound in (c).

Example 7.7 *Two time scales.*

Take $m_1 = m_1(n), m_2 = m_2(n)$ such that $m_1 \sim n^{1-\beta}, m_1 m_2 \sim n$. The underlying idea is to take two continuous-time random walks on the complete graphs K_{m_1} and K_{m_2}, but with the walks run on different time scales. To set this up directly in discrete time, take state space $\{(x,y) : 1 \leq x \leq m_1, 1 \leq y \leq m_2\}$ and transition probabilities

$$\begin{aligned} (x,y) &\to (x',y) \quad \text{chance } (m_1-1)^{-1}\left(1 - \frac{1}{am_1 \log m_1}\right), \ x' \neq x \\ &\to (x,y') \quad \text{chance } (m_2-1)^{-1}\frac{1}{am_1 \log m_1}, \ y' \neq y \end{aligned}$$

where $a = a(n) \uparrow \infty$ slowly. It is not hard to formalize the following analysis. Writing the chain as (X_t, Y_t), the Y-component stays constant for time $\Theta(am_1 \log m_1)$, during which time every x-value is hit, because the cover time for K_{m_1} is $\sim m_1 \log m_1$. And $m_2 \log m_2$ jumps of the Y-component are required to hit every y-value, so

$$EC \sim (m_2 \log m_2) \times (am_1 \log m_1) \sim an(\log m_1)(\log m_2). \tag{7.13}$$

Now $\tau_2 \sim am_1 \log m_1$, and because the mean number of returns to the starting point before the first Y-jump is $\sim a \log m_1$ we can use the local transience heuristic (7.10) to see $\tau_0 \sim (a \log m_1) \times n$. So $\tau_2/\tau_0 \sim m_1/n \sim n^{-\beta}$, and the lower bound from (c) is

$$(\beta - o(1))(a \log m_1) n \log n.$$

But this agrees with the exact limit (7.13), because $m_2 \sim n^\beta$.

We now turn to sharper distributional limits for C. An (easy) background fact is that, for independent random variables (Z_i) with exponential, mean τ, distribution,

$$\frac{\max(Z_1, \ldots, Z_n) - \tau \log n}{\tau} \stackrel{d}{\to} \eta$$

where η has the extreme value distribution

$$P(\eta \le x) = \exp(-e^{-x}), \ -\infty < x < \infty. \tag{7.14}$$

Now the cover time $C = \max_i T_i$ is the *max* of the hitting times, and with the uniform initial distribution the T_i's have mean τ_0. So if the T_i's have approximately exponential distribution and are roughly independent of each other then we anticipate the limit result

$$\frac{C - \tau_0 \log n}{\tau_0} \stackrel{d}{\to} \eta. \tag{7.15}$$

Theorem 7.4 has already given us a condition for limit exponential distributions, and we shall build on this result to give (Theorem 7.9) conditions for (7.15) to hold.

The extreme value distribution (7.14) has transform

$$E \exp(\theta\eta) = \Gamma(1 - \theta), \ -\infty < \theta < 1. \tag{7.16}$$

Classical probability theory (see Notes) says that to prove (7.15) it is enough to show that transforms converge, i.e. to show

$$E \exp(\theta C/\tau_0) \sim n^{-\theta}\Gamma(1 - \theta), \ -\infty < \theta < 1. \tag{7.17}$$

But Matthews method, which previously we have used on expectations, can just as well be applied to transforms. By essentially the same argument as in Chapter 2 Theorem yyy, Matthews [257] obtained

Proposition 7.8 *The cover time C in a not-necessarily-reversible Markov chain with arbitrary initial distribution satisfies*

$$\frac{\Gamma(n+1)\Gamma(1/f_*(\beta))}{\Gamma(n+1/f_*(\beta))} \le E\exp(\beta C) \le \frac{\Gamma(n+1)\Gamma(1/f^*(\beta))}{\Gamma(n+1/f^*(\beta))}$$

where

$$f^*(\beta) \equiv \max_{j\neq i} E_i \exp(\beta T_j)$$
$$f_*(\beta) \equiv \min_{j\neq i} E_i \exp(\beta T_j).$$

Substituting into (7.17), and using the fact

$$\frac{\Gamma(n+1)}{\Gamma(n+1-s_n)} \sim n^s \text{ as } n\to\infty,\ s_n \to s$$

we see that to establish (7.15) it suffices to prove that for arbitrary $j_n \neq i_n$ and for each fixed $-\infty < \theta < 1$,

$$E_{i_n} \exp(\theta T_{j_n}/\tau_0) \to \frac{1}{1-\theta}. \tag{7.18}$$

Theorem 7.9 *For a sequence of symmetric reversible chains, if*

(a) $\min_{j\neq i} E_i T_j = \tau_0(1-o(1))$

(b) $\tau_2/\tau_0 = o\left(\frac{1}{\log n}\right)$

then

$$\frac{C-\tau_0 \log n}{\tau_0} \ \stackrel{d}{\to}\ \eta.$$

Proof. By hypothesis (a) and Theorem 7.4 (b,c), for arbitrary $j_n \neq i_n$ we have $P_{i_n}(T_{j_n}/\tau_0 \in \cdot) \stackrel{d}{\to} \xi$. This implies (7.18) for $\theta \le 0$, and also by Fatou's lemma implies $\liminf_n E_{i_n}\exp(\theta T_{j_n}/\tau_0) \ge \frac{1}{1-\theta}$ for $0<\theta<1$. Thus it is sufficient to prove

$$\max_{j\neq i} E_i \exp(\theta T_j/\tau_0) \le \frac{1+o(1)}{1-\theta},\ 0<\theta<1. \tag{7.19}$$

The proof exploits some of our earlier general inequalities. Switch to continuous time. Fix $\beta > 0$. By conditioning on the position at some fixed time s,

$$E_i \exp(\beta(T_j - s)^+) \le \max_x (nP_i(X_s = x)) \times E_\pi \exp(\beta T_j).$$

By Corollary 7.3 the *max* is attained by $x = i$, and so

$$E_i \exp(\beta T_j) \le (nP_i(X_s = i)e^{\beta s}) \times E_\pi \exp(\beta T_j).$$

We now apply some general inequalities. Chapter 4 yyy says $nP_i(X_s = i) \le 1 + n\exp(-s/\tau_2)$. Writing α_j for the quasistationary distribution on $\{j\}^c$, Chapter 3 (yyy) implies $P_\pi(T_j > t) \le \exp(-t/E_{\alpha_j} T_j)$ and hence

$$E_\pi \exp(\beta T_j) \le \frac{1}{1 - \beta E_{\alpha_j} T_j}.$$

But Chapter 3 Theorem yyy implies $E_{\alpha_j} T_j \le \tau_0 + \tau_2$. So setting $\beta = \theta/\tau_0$, these inequalities combine to give

$$E_i \exp(\theta T_j/\tau_0) \le (1 + n\exp(-s/\tau_2)) \times \exp(\theta s/\tau_0) \times \frac{1}{1 - \theta(1 + \tau_2/\tau_0)}.$$

But by hypothesis (b) we can choose $s = o(\tau_0) = \Omega(\tau_2 \log n)$ so that each of the first two terms in the bound tends to 1, establishing (7.19). Finally, the effect of continuization is to change C by at most $O(\sqrt{EC})$, so the asymptotics remain true in discrete time.

Remark. Presumably (c.f. Open Problem 7.6) the Theorem remains true without hypothesis (b).

In view of Chapter 6 yyy it is surprising that there is no obvious example to disprove

Open Problem 7.10 *Let V denote the last state to be hit. In a sequence of vertex-transitive graphs with $n \to \infty$, is it always true that V converges (in variation distance, say) to the uniform distribution?*

7.1.6 Product chains

In our collection of examples in Chapter 5 of random walks on graphs, the examples with enough symmetry to fit into the present setting have in fact extra symmetry, enough to fit into the arc-transitive setting of section 7.2. So in a sense, working at the level of generality of symmetric reversible chains merely serves to illustrate what properties of chains depend only on this minimal level of symmetry. But let us point out a general construction. Suppose we have symmetric reversible chains $X^{(1)}, \ldots, X^{(d)}$ on state spaces $I^{(1)}, \ldots, I^{(d)}$. Fix constants $a_1, \ldots, a_d$ with each $a_i > 0$ and with $\sum_i a_i = 1$. Then (c.f. Chapter 4 section yyy) we can define a "product chain" with state-space $I^{(1)} \times \ldots \times I^{(d)}$ and transition probabilities

$$(x_1, \ldots, x_d) \to (x_1, \ldots, x_i', \ldots, x_d) \text{: probability } a_i P(X_1^{(i)} = x_i' | X_0^{(i)} = x_i).$$

This product chain is also symmetric reversible. But if the underlying chains have extra symmetry properties, these extra properties are typically lost when one passes to the product chain. Thus we have a general method of constructing symmetric reversible chains which lack extra structure. Example 7.14 below gives a case with distinct underlying components, and Example 7.11 gives a case with a non-uniform product. In general, writing $(\lambda_u^{(i)} : 1 \le u \le |I^{(i)}|)$ for the continuous-time eigenvalues of $X^{(i)}$, we have (Chapter 4 yyy) that the continuous-time eigenvalues of the product chain are

$$\lambda_{\mathbf{u}} = a_1\lambda_{u_1}^{(1)} + \ldots + a_d\lambda_{u_d}^{(d)}$$

indexed by $\mathbf{u} = (u_1, \ldots, u_d) \in \{1, \ldots, |I^{(1)}|\} \times \ldots \times \{1, \ldots, |I^{(d)}|\}$. So in particular

$$\tau_2 = \max_i \frac{\tau_2^{(i)}}{a_i}$$

$$\tau_0 = \sum_{\mathbf{u} \neq (1,\ldots,1)} \frac{1}{a_1\lambda_{u_1}^{(1)} + \ldots + a_d\lambda_{u_d}^{(d)}}$$

and of course these parameters take the same values in discrete time.

Example 7.11 *Coordinate-biased random walk on the d-cube.*

Take $I = \{0,1\}^d$ and fix $0 < a_1 \le a_2 \le \ldots \le a_d$ with $\sum_i a_i = 1$. Then the chain with transitions

$$(b_1, \ldots, b_d) \to (b_1, \ldots, 1 - b_i, \ldots, b_d) : \text{probability } a_i$$

is the weighted product of two-state chains. Most of the calculations for simple symmetric random walk on the d-cube done in Chapter 5 Example yyy extend to this example, with some increase of complexity. In particular,

$$\tau_2 = \frac{1}{2a_1}$$

$$\tau_0 = \frac{1}{2} \sum_{\mathbf{u} \in I, \mathbf{u} \neq \mathbf{0}} \frac{1}{\sum_{i=1}^d u_i a_i}.$$

In continuout time we still get the product form for the distribution at time t:

$$P_b(X_t = b') = 2^{-d} \prod_i (1 + \eta_i \exp(-2a_i t)) \ ; \ \eta_i = 1 \text{ if } b_i' = b_i, \ = 0 \text{ if not.}$$

So in a sequence of continuous time chains with $d \to \infty$, the "separation" parameter $\tau_1^{(1)}$ of Chapter 3 section yyy is asymptotic to the solution t of

$$\sum_i \exp(-2a_i t) = -\log(1 - e^{-1}).$$

More elaborate calculations can be done to study τ_1 and the discrete-time version.

7.1.7 The cutoff phenomenon and the upper bound lemma

Chapter 2 yyy and Chapter 4 yyy discussed quantifications of notions of "time to approach stationarity" using variation distance. The emphasis in Chapter 4 yyy was on inequalities which hold up to universal constants. In the present context of symmetric reversible chains, one can seek to do sharper calculations. Thus for random walk on the d-cube (Chapter 5 Example yyy), with chances $1/(d+1)$ of making each possible step or staying still, writing $n = 2^d$ and $c_n = \frac{1}{4} d \log d$, we have (as $n \to \infty$) not only the fact $\tau_1 \sim c_n$ but also the stronger result

$$d((1+\varepsilon)c_n) \to 0 \text{ and } d((1-\varepsilon)c_n) \to 1, \text{ for all } \varepsilon > 0. \tag{7.20}$$

We call this the *cutoff phenomenon*, and when a sequence of chains satisfies (7.20) we say the sequence has "variation cutoff at c_n". As mentioned at xxx, the general theory of Chapter 4 works smoothly using $\bar{d}(t)$, but in examples it is more natural to use $d(t)$, which we shall do in this chapter. Clearly, (7.20) implies the same result for $\bar{d}$ and implies $\tau_1 \sim c_n$. Also, our convention in this chapter is to work in discrete time, whereas the Chapter 4 general theory worked more smoothly in continuous time. (Clearly (7.20) in discrete time implies the same result for the continuized chains, provided $c_n \to \infty$). Note that, in the context of symmetric reversible chains,

$$d(t) = d_i(t) = ||P_i(X_t \in \cdot) - \pi(\cdot)|| \text{ for each } i.$$

We also can discuss *separation distance* (Chapter 4 yyy) which in this context is

$$s(t) = 1 - n \min_j P_i(X_t = j) \text{ for each } i,$$

and introduce the analogous notion of *separation threshold*.

It turns out that these cut-offs automatically appear in sequences of chains defined by repeated products. An argument similar to the analysis of the d-cube (see [22] for a slightly different version) shows

Lemma 7.12 *Fix an aperiodic symmetric reversible chain with m states and with relaxation time $\tau_2 = 1/(1-\lambda_2)$. Consider the d-fold product chain with $n = m^d$ states and transition probabilities*

$$(x_1, \ldots, x_d) \to (x_1, \ldots, y_i, \ldots, y_d) \ : \ \textit{probability} \ \frac{1}{d} \ p_{x_i, y_i}.$$

As $d \to \infty$, this sequence of chains has variation cutoff $\frac{1}{2}\tau_2 d \log d$ and separation cut-off $\tau_2 d \log d$.

xxx discuss upper bound lemma
xxx heuristics
xxx mention later examples

7.1.8 Vertex-transitive graphs and Cayley graphs

So far we have worked in the setting of symmetric reversible chains, and haven't used any graph theory. We now specialize to the case of random walk on a vertex-transitive or Cayley graph $(\mathcal{V}, \mathcal{E})$. As usual, we won't write out all specializations of the previous results, but instead emphasize what extra we get from graph-theoretic arguments. Let d be the degree of the graph.

Lemma 7.13 *For random walk on a vertex-transitive graph,*
(i) $E_v T_x \geq n$ if $(v,x) \notin \mathcal{E}$
(ii) $\frac{2dn}{d+1} - d \geq E_v T_x \geq \frac{dn}{d+1}$ if $(v,x) \in \mathcal{E}$

Proof. The lower bounds are specializations of Lemma 7.2(i), i.e. of Chapter 6 xxx. For the upper bound in (ii),

$$\begin{aligned} n - 1 &= \frac{1}{d} \sum_{y \sim x} E_y T_x \qquad (7.21) \\ &\geq \frac{1}{d} \left(E_v T_x + (d-1) \frac{dn}{d+1} \right) \text{ by the lower bound in (ii).} \end{aligned}$$

Rearrange.

xxx mention general lower bound $\tau_0 \geq (1 - o(1)) nd/(d-2)$ via tree-cover.

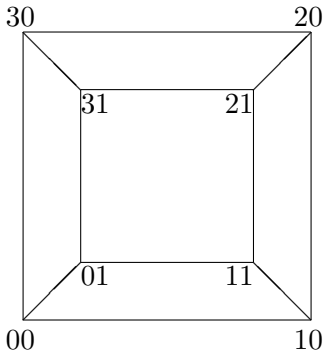
It is known (xxx ref) that a Cayley graph of degree d is d-edge-connected, and so Chapter 6 Proposition yyy gives

$$\tau^* \leq n^2 \psi(d)/d$$

where $\psi(d)/d \approx \sqrt{2/d}$.

Example 7.14 *A Cayley graph where E_vT_w is not the same for all edges (v, w).*

Consider $Z_m \times Z_2$ with generators $(1,0), (-1,0), (0,1)$. The figure illustrates the case $m = 4$.



Let's calculate $E_{00}T_{10}$ using the resistance interpretation. Put unit voltage at 10 and zero voltage at 00, and let a_i be the voltage at $i0$. By symmetry the voltage at $i1$ is $1 - a_i$, so we get the equations

$$a_i = \frac{1}{3}(a_{i-1} + a_{i+1} + (1 - a_i)),\ 1 \le i \le m-1$$

with $a_0 = a_m = 0$. But this is just a linear difference equation, and a brief calculation gives the solution

$$a_i = \frac{1}{2} - \frac{1}{2}\frac{\theta^{m/2-i} + \theta^{i-m/2}}{\theta^{m/2} + \theta^{-m/2}}$$

where $\theta = 2 - \sqrt{3}$. The current flow is $1 + 2a_1$, so the effective resistance is $r = (1 + 2a_1)^{-1}$. The commute interpretation of resistance gives $2E_{00}T_{01} = 3nr$, and so

$$E_{00}T_{01} = \frac{3n}{2(1 + 2a_1)}$$

where $n = 2m$ is the number of vertices. In particular,

$$n^{-1}\ E_{00}T_{01} \to \gamma \equiv \frac{3}{1 + \sqrt{3}} \text{ as } n \to \infty.$$

Using the averaging property (7.21)

$$n^{-1}\ E_{00}T_{10} \to \gamma' \equiv \frac{3\sqrt{3}}{2(1 + \sqrt{3})} \text{ as } n \to \infty.$$

Turning from hitting times to mixing times, recall the Cheeger constant

$$\tau_c \equiv \sup_A c(A)$$

where A is a proper subset of vertices and

$$c(A) \equiv \frac{\pi(A^c)}{P_\pi(X_1 \in A^c | X_0 \in A)}.$$

For random walk on a Cayley graph one can use simple "averaging" ideas to bound $c(A)$. This is Proposition 7.15 below. The result in fact extends to vertex-transitive graphs by a covering graph argument - see xxx.

Consider a n-vertex Cayley graph with degree d and generators $\mathcal{G} = \{g_1, \ldots, g_d\}$, where $g \in \mathcal{G}$ implies $g^{-1} \in \mathcal{G}$. Then

$$P_\pi(X_1 \in A^c | X_0 \in A) = \frac{1}{d} \sum_{g \in \mathcal{G}} \frac{|Ag \setminus A|}{|A|}$$

where $Ag = \{ag : a \in A\}$. Lower bounding the sum by its maximal term, we get

$$c(A) \le \frac{d}{n} \frac{|A|\ |A^c|}{\max_{g \in \mathcal{G}} |Ag \setminus A|}. \tag{7.22}$$

Proposition 7.15 *On a Cayley graph of degree d*
(i) $\tau_c \le d\Delta$, where Δ is the diameter of the graph.
(ii) $c(A) \le 2d\rho(A)$ for all A with $\rho(A) \ge 1$, where

$$\rho(A) \equiv \min_{v \in \mathcal{V}} \max_{w \in A} d(v, w)$$

is the radius of A.

Note that $\sup_A \rho(A)$ is bounded by Δ but not in general by $\Delta/2$ (consider the cycle), so that (ii) implies (i) with an extra factor of 2. Part (i) is from Aldous [10] and (ii) is from Babai [36].

Proof. (i) Fix A. Because

$$\frac{1}{n} \sum_{v \in \mathcal{V}} |A \cap Av| = |A|^2/n$$

there exists some $v \in \mathcal{V}$ such that $|A \cap Av| \le |A|^2/n$, implying

$$|Av \setminus A| \ge |A||A^c|/n. \tag{7.23}$$

We can write $v = g_1 g_2 \dots g_\delta$ for some sequence of generators (g_i) and some $\delta \le \Delta$, and

$$|Av \setminus A| \le \sum_{i=1}^{\delta} |Ag_1 \dots g_i \setminus Ag_1 \dots g_{i-1}| = \sum_{i=1}^{\delta} |Ag_i \setminus A|.$$

So there exists $g \in \mathcal{G}$ with $|Ag \setminus A| \ge \frac{1}{\Delta} \times |A||A^c|/n$, and so (i) follows from (7.22). For part (ii), fix A with $|A| \le n/2$, write $\rho = \rho(A)$ and suppose

$$\max_{g \in \mathcal{G}} |Ag \setminus A| < \frac{1}{4\rho}|A|. \tag{7.24}$$

Fix v with $\max_{w \in A} d(w, v) = \rho$. Since $|Ag \setminus A| < \frac{1}{4\rho}|A|$ and

$$A \setminus Axg \subseteq (A \setminus Ag) \cup (A \setminus Ax)g$$

we have by induction

$$|A \setminus Ax| < \frac{1}{4\rho}|A| d(x, v). \tag{7.25}$$

Write $B^r \equiv \{vg_1 \dots g_i; i \le r, g_i \in \mathcal{G}\}$ for the ball of radius r about v. Since $(2\rho+1)/(4\rho) < 1$, inequality (7.25) shows that $A \cap Ax$ is non-empty for each $x \in B^{2\rho+1}$, and so $B^{2\rho+1} \subseteq A^{-1}A$. But by definition of ρ we have $A \subseteq B^\rho$, implying $B^{2\rho+1} \subseteq B^{2\rho}$, which in turn implies $B^{2\rho}$ is the whole group. Now (7.25) implies that for every x

$$|Ax \setminus A| < \frac{1}{2}|A| \le \frac{|A||A^c|}{n}.$$

But this contradicts (7.23). So (7.24) is false, i.e.

$$\max_{g \in \mathcal{G}} |Ag \setminus A| \ge \frac{1}{4\rho}|A| \ge \frac{1}{2\rho} \ \frac{|A||A^c|}{n}.$$

By complementation the final inequality remains true when $|A| > n/2$, and the result follows from (7.22).

7.1.9 Comparison arguments for eigenvalues

The "distinguished paths" method of bounding relaxation times (Chapter 4 yyy) can also be used to compare relaxation times of two random flights on the same group, and hence to bound one "unknown" relaxation time

in terms of a second "known" relaxation time. This approach has been developed in great depth in

xxx ref Diaconis Saloff-Coste papers.

Here we give only the simplest of their results, from [117].

Consider generators $\mathcal{G}$ of a group I, and consider a reversible random flight with step-distribution μ supported on $\mathcal{G}$. Write $d(x, \mathrm{id})$ for the distance from x to the identity in the Cayley graph, i.e. the minimal length of a word

$$x = g_1 g_2 \dots g_d;\ g_i \in \mathcal{G}.$$

For each x choose some minimal-length word as above and define $N(g, x)$ to be the number of occurences of g in the word. Now consider a different reversible random flight on I with some step-distribution $\tilde{\mu}$, not necessarily supported on $\mathcal{G}$. If we know $\tilde{\tau_2}$, the next result allows us to bound τ_2.

Theorem 7.16

$$\frac{\tau_2}{\tilde{\tau_2}} \le K \equiv \max_{g \in \mathcal{G}} \frac{1}{\mu(g)} \sum_{x \in I} d(x, id) N(g, x) \tilde{\mu}(x).$$

xxx give proof – tie up with L^2 discussion

Perhaps surprisingly, Theorem 7.16 gives information even when the comparison walk is the "trivial" walk whose step-distribution $\tilde{\mu}$ is uniform on the group. In this case, both $d(x, \mathrm{id})$ and $N(g, x)$ are bounded by the diameter Δ, giving

Corollary 7.17 *For reversible flight with step-distribution μ on a group I,*

$$\tau_2 \le \frac{\Delta^2}{\min_{g \in \mathcal{G}} \mu(g)},$$

where $\mathcal{G}$ is the support of μ and Δ is the diameter of the Cayley graph associated with $\mathcal{G}$.

When μ is uniform on $\mathcal{G}$ and $|\mathcal{G}| = d$, the Corollary gives the bound $d\Delta^2$, which improves on the bound $8d^2\Delta^2$ which follows from Proposition 7.15 and Cheeger's inequality (Chapter 4 yyy). The examples of the torus Z_N^d show that Δ^2 enters naturally, but one could hope for the following variation.

Open Problem 7.18 *Write $\tau_* = \tau_*(I, \mathcal{G})$ for the minimum of τ_2 over all symmetric random flights on I with step-distribution supported on $\mathcal{G}$. Is it true that $\tau_* = O(\Delta^2)$?*

7.2 Arc-transitivity

Example 7.14 shows that random walk on a Cayley graph does not necessarily have the property that $E_v T_w$ is the same for all edges (v, w). It is natural to consider some stronger symmetry condition which does imply this property. Call a graph *arc-transitive* if for each 4-tuple of vertices (v_1, w_1, v_2, w_2) such that (v_1, w_1) and (v_2, w_2) are edges, there exists an automorphism γ such that $\gamma(v_1) = w_1, \gamma(v_2) = w_2$. Arc-transitivity is stronger than vertex-transitivity, and immediately implies that $E_v T_w$ is constant over edges (v, w).

Lemma 7.19 *On a n-vertex arc-transitive graph,*
(i) $E_v T_w = n - 1$ *for each edge* (v, w).
(ii) $E_v T_w \geq n - 2 + d(v, w)$ *for all* $w \neq v$.

Proof. (i) follows from $E_v T_v^+ = n$. For (ii), write $N(w)$ for the set of neighbors of w. Then

$$E_v T_w = E_v T_{N(w)} + (n - 1)$$

and $T_{N(w)} \geq d(v, N(w)) = d(v, w) - 1$.

In particular, $\min_{w \neq v} E_v T_w = n - 1$, which gives the following bounds on mean cover time EC. The first assertion uses Matthews method for expectations (Chapter 2 yyy) and the second follows from Theorem 7.9.

Corollary 7.20 *On a n-vertex arc-transitive graph,* $EC \geq (n - 1)h_{n-1}$. *And if* $\tau_0/n \to 1$ *and* $\tau_2 = o(n/\log n)$ *then*

$$\frac{C - \tau_0 \log n}{\tau_0} \stackrel{d}{\to} \eta \tag{7.26}$$

Note that the lower bound $(n - 1)h_{n-1}$ is attained on the complete graph. It is not known whether this exact lower bound remains true for vertex-transitive graphs, but this would be a consequence of Chapter 6 Open Problem yyy. Note also that by xxx the hypothesis $\tau_0/n \to 1$ can only hold if the degrees tend to infinity.

Corollary 7.20 provides easily-checkable conditions for the distributional limit for cover times, in examples with ample symmetry, such as the card-shuffling examples in the next section. Note that

$$(7.26) \text{ and } \tau_0 = n\left(1 + \frac{b + o(1)}{\log n}\right) \text{ imply } \frac{C - n \log n - bn}{n} \stackrel{d}{\to} \eta.$$

Thus on the d-cube (Chapter 5 yyy) $\tau_0 = n\left(1 + \frac{1+o(1)}{d}\right) = n\left(1 + \frac{\log 2 + o(1)}{\log n}\right)$ and so

$$\frac{C - n\log n - n\log 2}{n} \xrightarrow{d} \eta.$$

7.2.1 Card-shuffling examples

These examples are formally random flights on the permutation group, though we shall describe them informally as models for random shuffles of a m-card deck. Write X_t for the configuration of the deck after t shuffles, and write $Y_t = f_1(X_t)$ for the position of card 1 after t shuffles. In most examples (and all those we discuss) Y_t is itself a Markov chain on $\{1, 2, \ldots, m\}$. Example 7.21, mentioned in Chapter 1 xxx, has become the prototype for use of group representation methods.

Example 7.21 *Card-shuffling via random transpositions.*

The model is

> Make two independent uniform choices of cards, and interchange the positions of the two cards.

With chance $1/m$ the same card is chosen twice, so the "interchange" has no effect. This model was studied by Diaconis and Shahshahani [122], and more concisely in the book Diaconis [112] Chapter 3D. The chain Y_t has transition probabilities

$$\begin{array}{lll} i \to j & \text{probability} & 2/m^2,\ j \neq i \\ i \to i & \text{probability} & 1 - \dfrac{2(m-1)}{m^2} \end{array}$$

This is essentially random walk on the complete m-graph (precisely: the continuized chains are deterministic time-changes of each other) and it is easy to deduce that (Y_t) has relaxation time $m/2$. So by the contraction principle xxx the card-shuffling process has $\tau_2 \geq m/2$, and group representation methods show

$$\tau_2 = m/2. \tag{7.27}$$

Since the chance of being in the initial state after 1 step is $1/m$ and after 2 steps in $O(1/m^2)$, the local transience heuristic (7.10) suggests

$$\tau_0 = m!(1 + 1/m + O(1/m^2)) \tag{7.28}$$

which can be verified by group representation methods (see Flatto et al [152]). The general bound on τ_1 in terms of τ_2 gives only $\tau_1 = O(\tau_2 \log m!) = O(m^2 \log m)$. In fact group group representation methods ([112]) show

$$\text{there is a variation cutoff at } \frac{1}{2} m \log m. \tag{7.29}$$

Example 7.22 *Card-shuffling via random adjacent transpositions.*

The model is

> With probability $1/(m+1)$ do nothing. Otherwise, choose one pair of adjacent cards (counting the top and bottom cards as adjacent), with probability $1/(m+1)$ for each pair, and interchange them.

The chain Y_t has transition probabilities

$$\begin{array}{lll} i \to & i+1 & \text{probability } 1/(m+1) \\ i \to & i-1 & \text{probability } 1/(m+1) \\ i \to & i & \text{probability } (m-1)/(m+1) \end{array}$$

with $i \pm 1$ counted modulo m. This chain is (in continuous time) just a time-change of random walk on the m-cycle, so has relaxation time

$$a(m) \equiv \frac{m+1}{2} \frac{1}{1-\cos(2\pi/m)} \sim \frac{m^3}{4\pi^2}.$$

So by the contraction principle xxx the card-shuffling process has $\tau_2 \geq a(m)$, and (xxx unpublished Diaconis work) in fact

$$\tau_2 = a(m) \sim m^3/4\pi^2.$$

A coupling argument which we shall present in Chapter xxx gives an upper bound $\tau_1 = O(m^3 \log m)$ and (xxx unpublished Diaconis work) in fact

$$\tau_1 = \Theta(m^3 \log m).$$

The local transience heuristic (7.10) again suggests

$$\tau_0 = m!(1 + 1/m + O(1/m^2))$$

but this has not been studied rigorously.

Many variants of these examples have been studied, and we will mention a generalization of Examples 7.21 and 7.22 in Chapter xxx. Here is another example, from Diaconis and Saloff-Coste [117], which illustrates the use of comparison arguments.

Example 7.23 *A slow card-shuffling scheme.*

The model is: with probability 1/3 each, either
(i) interchange the top two cards
(ii) move the top card to the bottom
(iii) move the bottom card to the top.
This process is random walk on a certain Cayley graph, which (for $m \geq 3$) is not arc-transitive. Writing d for distances in the graph and writing

$$\beta = \max(d(\sigma, \mathrm{id}) : \sigma \text{ a transposition }),$$

it is easy to check that $\beta \leq 3m$. Comparing the present chain with the "random transpositions" chain (Example 7.21), denoted by $\tilde{}$, Theorem 7.16 implies

$$\frac{\tau_2}{\tilde{\tau_2}} \leq 3\beta^2.$$

Since $\tilde{\tau_2} = m/2$ we get

$$\tau_2 \leq \frac{27m^3}{2}.$$

7.2.2 Cover times for the d-dimensional torus Z_N^d.

This is Example yyy from Chapter 5, with $n = N^d$ vertices, and is clearly arc-transitive. Consider asymptotics as $N \to \infty$ for d fixed. We studied mean hitting times in this example in Chapter 5. Here $\tau_0/n \not\to 1$, so we cannot apply Corollary 7.20. For $d = 1$ the graph is just the d-cycle, treated in Chapter 6 yyy. For $d \geq 3$, Chapter 5 yyy gave

$$E_0 T_i \sim nR_d \text{ as } N \to \infty,\ |i| \to \infty$$

where $|i|$ is Euclidean distance on the torus, i.e.

$$|(i_1, \ldots, i_d)|^2 = \sum_{u=1}^{d} (\min(i_u, N - i_u))^2.$$

So EC has the asymptotic upper bound $R_d n \log n$. Now if we apply the subset form of Matthews method (Chapter 6 yyy) to the subset

$$A = \{(j_1 m, \ldots, j_d m) : 1 \leq j_i \leq \frac{N}{m}\} \tag{7.30}$$

then we get a lower bound for EC asymptotic to

$$\log |A| \times nR_d.$$

By taking $m = m(n) \uparrow \infty$ slowly, this agrees with the upper bound, so we find

Corollary 7.24 *On the d-dimensional torus with $d \geq 3$,*

$$EC \sim R_d n \log n.$$

Perhaps surprisingly, the case $d = 2$ turns out to be the hardest of all explicit graphs for the purposes of estimating cover times. (Recall this case is the *white screen problem* Chapter 1 xxx.) Loosely, the difficulty is caused by the fact that $\tau_2 = \Theta(n \log n)$ – recall from Chapter 6 yyy that another example with this property, the balanced tree, is also hard. Anyway, for the case $d = 2$ the calculations in Chapter 5 yyy gave

$$E_0 T_i \sim n \left(\frac{2}{\pi} \log |i| \ + O(1) \right).$$

This leads to the upper bound in Corollary 7.25 below. For the lower bound, we repeat the $d \geq 3$ argument using a subset of the form (7.30) with $m \to \infty$, and obtain a lower bound asymptotic to

$$\frac{2}{\pi} \log m \ \times \ \log(n^2/m^2).$$

The optimal choice is $m \sim n^{1/2}$, leading to the lower bound below.

Corollary 7.25 *On the 2-dimensional torus Z_N^2,*

$$\left(\frac{1}{4\pi} - o(1) \right) n \log^2 n \leq EC \leq \left(\frac{1}{\pi} + o(1) \right) n \log^2 n.$$

Lawler [221] has improved the constant in the lower bound to $\frac{1}{2\pi}$ – see Notes. It is widely believed that the upper bound is in fact the limit.

Open Problem 7.26 *Prove that, on the 2-dimensional torus Z_N^2,*

$$EC \sim \frac{1}{\pi} \ n \log^2 n.$$

The usual distributional limit

$$\frac{C - \tau_0 \log n}{\tau_0} \ \stackrel{d}{\to} \ \eta$$

certainly fails in $d = 1$ (see Chapter 6 yyy). It has not been studied in $d \geq 2$, but the natural conjecture it that it is true for $d \geq 3$ but false in $d = 2$. Note that (by Chapter 6 yyy) the weaker concentration result

$$C/EC \ \stackrel{d}{\to} \ 1$$

holds for all $d \geq 2$.

7.2.3 Bounds for the parameters

In Chapter 6 we discussed upper bounds on parameters τ for regular graphs. One can't essentially improve these bounds by imposing symmetry conditions, because the bounds are attained (up to constants) by the n-cycles. But what if we exclude the n-cycles? Example 7.14 shows that one can invent vertex-transitive graphs which mimic the n-cycle, but it is not clear whether such arc-transitive graphs exist. So perhaps the next-worst arc-transitive graph is Z_m^2.

Open Problem 7.27 *Is it true that, over arc-transitive graphs excluding the n-cycles, $\tau^* = O(n \log n)$, $\tau_2 = O(n)$ and $\frac{\tau^*}{2\tau_0} = 1 + o(1)$?*

7.2.4 Group-theory set-up

Recall that the Cayley graph associated with a set $\mathcal{G}$ of generators of a group I has edges

$$\{(v, vg); v \in I, g \in \mathcal{G}\}$$

where we assume $\mathcal{G}$ satisfies
(i) $g \in \mathcal{G}$ implies $g^{-1} \in \mathcal{G}$.
To ensure that the graph is arc-transitive, it is sufficient to add the condition

(ii) for each pair g_1, g_2 in $\mathcal{G}$, there exists a group automorphism γ such that $\gamma(\mathrm{id}) = \mathrm{id}$ and $\gamma(g_1) = g_2$.

In words, "the stabilizer acts transitively on $\mathcal{G}$". This is essentially the general case: see [71] Prop. A.3.1.

As a related concept, recall that elements x, y of a group I are *conjugate* if $x = g^{-1}yg$ for some group element g. This is an equivalence relation which therefore defines *conjugacy classes*. It is easy to check that a conjugacy class must satisfy condition (ii). Given a conjugacy class C one can consider the uniform distribution μ_C on C and then consider the random flight with step distribution μ_C. Such random flights fit into the framework of section 7.2, and Example 7.21 and the torus Z_N^d are of this form. On the other hand, Example 7.22 satisfies (i) and (ii) but are not random flights with steps uniform on a conjugacy class.

7.3 Distance-regular graphs

A graph is called *distance-transitive* if for each 4-tuple v_1, w_1, v_2, w_2 with $d(v_1, w_1) = d(v_2, w_2)$ there exists an automorphism γ such that $\gamma(v_1) =$

$w_1, \gamma(v_2) = w_2$. Associated with such a graph of diameter Δ are the *intersection numbers* $(a_i, b_i, c_i; 0 \le i \le \Delta)$ defined as follows. For each i choose (v, w) with $d(v, w) = i$, and define

$$\begin{aligned} c_i = &\text{ number of neighbors of } w \text{ at distance } i-1 \text{ from } v \\ a_i = &\text{ number of neighbors of } w \text{ at distance } i \text{ from } v \\ b_i = &\text{ number of neighbors of } w \text{ at distance } i+1 \text{ from } v. \end{aligned}$$

The distance-transitive property ensures that (a_i, b_i, c_i) does not depend on the choice of (v, w). A graph for which such intersection numbers exist is called *distance-regular*, and distance-regularity turns out to be strictly weaker than distance-transitivity. An encyclopedic treatment of such graphs has been given by Brouwer et al [71]. The bottom line is that there is almost a complete characterization (i.e. list of families and sporadic examples) of distance-regular graphs. Anticipating a future completion of the characterization, one could seek to prove inequalities for random walks on distance-regular graphs by simply doing explicit calculations with all the examples, but (to quote Biggs [49]) "this would certainly not find a place in *The Erdos Book* of ideal proofs". Instead, we shall just mention some properties of random walk which follow easily from the definitions.

Consider random walk (X_t) on a distance-regular graph started at v_0, and define $D_t = d(v_0, X_t)$. Then (D_t) is itself a Markov chain on states $\{0, 1, \ldots, \Delta\}$, and is in fact the birth-and-death chain with transition probabilities

$$p_{i,i-1} = c_i/r, \; p_{i,i} = a_i/r, \; p_{i,i+1} = b_i/r.$$

xxx b-and-d with holds

Finding exact t-step transition probabilities is tantamount to finding the orthogonal polynomials associated with the distance-regular graph – references to the latter topic can be found in [71], but we shall not pursue it.

7.3.1 Exact formulas

A large number of exact formulas can be derived by combining the standard results for birth-and-death chains in Chapter 5 section yyy with the standard renewal-theoretic identities of Chapter 2 section yyy. We present only the basic ones.

Fix a state $\mathbf{0}$ in a distance-regular graph. Let n_i be the number of states at distance i from $\mathbf{0}$. The number of edges with one end at distance i and

the other at distance $i+1$ is $n_i b_i = n_{i+1} c_{i+1}$, leading to the formula

$$n_i = \prod_{j=1}^{i} \frac{b_{j-1}}{c_j};\ 0 \le i \le \Delta.$$

The chain D_t has stationary distribution

$$\rho_i = n_i/n = n^{-1} \prod_{j=1}^{i} \frac{b_{j-1}}{c_j};\ 0 \le i \le \Delta.$$

Switching to the notation of Chapter 5 yyy, the chain D_t is random walk on a weighted linear graph, where the weight w_i on edge $(i-1, i)$ is

$$w_i = \frac{n_{i-1} b_{i-1}}{2n} = \frac{n_i c_i}{2n},\ 1 \le i \le \Delta$$

and total weight $w = 1$. This graph may have self-loops, but they don't affect the formulas. Clearly hitting times on the graph are related to hitting times of (D_t) by

$$E_v T_x = h(d(v,x)) \text{ , where } h(i) \equiv \tilde{E}_i T_0 \tag{7.31}$$

and where we write $\tilde{}$ to refer to expectations for D_t. Clearly $h(\cdot)$ is strictly increasing. Chapter 5 yyy gives the formula

$$h(i) = i + 2 \sum_{j=1}^{i} \sum_{i=j+1}^{\Delta} w_i w_j^{-1}. \tag{7.32}$$

And Chapter 5 yyy gives the last equality in

$$\tau_0 = E_\pi T_{\mathbf{0}} = \tilde{E}_\rho T_0 = \frac{1}{2} \sum_{i=1}^{\Delta} w_i^{-1} (\sum_{j \ge i} \rho_j)^2. \tag{7.33}$$

Finally, Chapter 5 yyy gives

$$\tilde{E}_0 T_\Delta + \tilde{E}_\Delta T_0 = \sum_{i=1}^{\Delta} 1/w_i. \tag{7.34}$$

Thus if the graph has the property that there exists a unique vertex $\mathbf{0}^*$ at distance Δ from $\mathbf{0}$, then we can pull back to the graph to get

$$\frac{\tau^*}{2} = \max_{x \ne v} E_x T_v = E_{\mathbf{0}} T_{\mathbf{0}^*} = \frac{1}{2} \sum_{i=1}^{\Delta} 1/w_i. \tag{7.35}$$

If the graph lacks that property, we can use (7.31) to calculate $h(\Delta)$.

The general identities of Chapter 3 yyy can now be used to give formulas for quantities such as $P_x(T_y < T_z)$ or E_x (number of visits to y before T_z).

7.3.2 Examples

Many treatments of random walk on sporadic examples such as regular polyhedra have been given, e.g. [227, 228, 275, 319, 320, 330, 331], so I shall not repeat them here. Of infinite families, the complete graph was discussed in Chapter 5 yyy, and the complete bipartite graph is very similar. The d-cube also was treated in Chapter 5. Closely related to the d-cube is a model arising in several contexts under different names,

Example 7.28 c-*subsets of a* d-*set.*

The model has parameters (c, d), where $1 \le c \le d-1$. Formally, we have random walk on the distance-transitive graph whose vertices are the $\frac{d!}{c!(d-c)!}$ c-element subsets $A \subset \{1, 2, \ldots, d\}$, and where (A, A') is an edge iff $|A \triangle A'| = 2$. More vividly, d balls $\{1, 2, \ldots, d\}$ are distributed between a left urn and a right urn, with c balls in the left urn, and at each stage one ball is picked at random from each urn, and the two picked balls are interchanged. The induced birth-and-death chain is often called the *Bernouilli-Laplace diffusion model*. The analysis is very similar to that of the d-cube. See [123, 127] and [112] Chapter 3F for details on convergence to equilibrium and [110] for hitting and cover times.

7.3.3 Monotonicity properties

The one result about random walk on distance-regular graphs we wish to highlight is the monotonicity property given in Proposition 7.29 below. Part (ii) can be viewed as a strengthening of the monotonicity property for mean hitting times (by integrating over time and using the formula relating mean hitting times to the fundamental matrix).

Proposition 7.29 *For random walk (X_t) on a distance-regular graph in continuous time, $P_v(X_t = w) = q(t, d(v, w))$, where the function $d \to q(t, d)$ satisfies*
(i) $d \to q(t, d)$ in non-increasing, for fixed t.
(ii) $q(t, d)/q(t, 0)$ in non-decreasing in t, for fixed d.

xxx proof – coupling – defer to coupling Chapter ??

Proposition 7.29 is a simple example of what I call a "geometric" result about a random walk. Corollary 7.3 gave a much weaker result in a more general setting. It's natural to ask for intermediate results, e.g.

Open Problem 7.30 *Does random walk on an arc-transitive graph have some monotonicity property stronger than that of Corollary 7.3?*

7.3.4 Extremal distance-regular graphs

Any brief look at examples suggests

Open Problem 7.31 *Prove that, over distance-regular graphs excluding the n-cycles,* $\tau_0 = O(n)$.

Of course this would imply $\tau^* = O(n)$ and $EC = O(n \log n)$. As mentioned earlier, one can try to tackle problems like this by using the list of known distance-regular graphs in [71]. Biggs [49] considered the essentially equivalent problem of the maximum value of $\max_{i,j} E_i T_j/(n-1)$, and found the value 195/101 taken on the cubic graph with 102 vertices, and outlined an argument that this may be the *max* over known distance-regular graphs.

xxx in same setting is $\tau_2 = O(\log n)$?

7.3.5 Gelfand pairs and isotropic flights

On a distance-regular graph, a natural generalization of our nearest-neighbor random walks is to *isotropic* random flight on the graph. Here one specifies a probability distribution $(s_0, s_1, \ldots, s_\Delta)$ for the step-length S, and each step moves to a random vertex at distance S from the previous vertex. Precisely, it is the chain with transition probabilities

$$p(v,w) = \frac{s_{d(v,w)}}{n_{d(v,w)}}. \tag{7.36}$$

The notion of isotropic random flight also makes sense in continuous space. For an isotropic random flight in R^d, the steps have some arbitrary specified random length S and a direction θ which is uniform and independent of S. A similar definition can be made on the d-dimensional sphere. The abstract notion which captures distance-regular graphs and their continuous analogs is a *Gelfand pair*. Isotropic random flights on Gelfand pairs can be studied in great detail by analytic methods. Brief accounts can be found in Letac [224, 225] and Diaconis [112] Chapter 3F, which contains an extensive annotated bibliography.

7.4 Notes on Chapter 7

Diaconis [112] Chapter 3 discusses random walks on groups, emphasizing use of the upper bound lemma to establish bounds on τ_1 and $d(t)$, and containing extensive references to previous work using group-theoretic methods. We have only mentioned reversible examples, but many natural non-reversible examples can also be handled by group representation methods.

Also, in Example 7.21 and related examples, group representation methods give stronger information about $d(t)$ then we have quoted.

Elementary properties of hitting and cover times on graphs with symmetry structure have been noted by many authors, a particularly comprehensive treatment being given in the Ph.D. thesis Sbihi [306]. Less extensive treatments and specific elementary results can be found in many of the papers cited later, plus [274, 330, 331]

Section 7.1. The phrase "random flight" is classically used for R^d. I have used it (as did Takacs [319, 320]) in place of "random walk' to emphasize it is not necessarily a nearest-neighbor random walk.

Section 7.1.3. Other elementary facts about symmetric reversible chains are

$$E_\pi \min(T_i, T_j) = \frac{n}{2}(Z_{ii} + Z_{ij}).$$

$$P_i(X_{2t} = i) + P_i(X_{2t} = j) \geq 2/n.$$

Chapter 6 yyy showed that on any regular graph, $\max_{i,j} E_i T_j \leq 3n^2$. On a vertex-transitive graph the constant "3" can be improved to "2", by an unpublished argument of the author, but this is still far from the natural conjecture of 1/4.

Section 7.1.4. Another curious result from [12]is that for a symmetric reversible chain the first passage time cannot be concentrated around its mean:

$$\frac{\text{var }_i T_j}{(E_i T_j)^2} \geq \frac{e-2}{e-1} - \frac{1}{E_i T_j}.$$

Section 7.1.5. Before Matthews method was available, a result like Corollary 7.5 (c) required a lot of work – see Aldous [8] for a result in the setting of non-reversible random flight on a group. The present version of Corollary 7.5 (c) is a slight polishing of ideas in Zuckerman [343] section 6.

The fact that (7.17) implies (7.15) is a slight variation of the usual textbook forms of the continuity theorem ([133] 2.3.4 and 2.3.11) for Fourier and Laplace transforms. By the same argument as therein, it is enough for the limit transform to be continuous at $\theta = 0$, which holds in our setting.

Matthews [255, 257] introduced Proposition 7.8 and used it to obtain the limiting cover time distribution for the d-cube and for card-shuffling examples. Devroye and Sbihi [110] applied it to generalized hypercubes and to Example 7.28. Our implementation in Theorem 7.9 and Corollary 7.20 reduces the need for ad hoc calculations in particular examples.

Section **??** Example 7.11 has been studied in the reliability literature (e.g. [212]) from the viewpoint of the exponential approximation for hitting times.

Section 7.1.7. The factor of 2 difference between the variation and separation cutoffs which appears in Lemma 7.12 is the largest possible – see Aldous and Diaconis [22].

Section 7.1.8. xxx walk-regular example – McKay paper.

Section 7.1.9. Diaconis and Saloff-Coste [117] give many other applications of Theorem 7.16. We mention some elsewhere; others include

xxx list.

Section 7.2. The name "arc-transitive" isn't standard: Biggs [48] writes "symmetric" and Brouwer et al [71] write "flag-transitive". Arc-transitivity is not necessary for the property "$E_v T_w$ is constant over edges". For instance, a graph which is vertex-transitive and edge-transitive (in the sense of undirected edges) has the property, but is not necessarily arc-transitive [182]. Gobel and Jagers [168] observed that the property

$$E_v T_w + E_w T_v = 2(n-1) \text{ for all edges } (v, w)$$

(equivalently: the effective resistance across each edge is constant) holds for arc-transitive graphs and for trees.

Section 7.2.2. Sbihi [306] and Zuckerman [343] noted that the subset version of Matthews method could be applied to the d-torus to give Corollaries 7.24 and 7.25.

The related topic of the time taken by random walk on the infinite lattice Z^d to cover a ball centered at the origin has been studied independently – see Revesz [288] Chapter 22 and Lawler [221], who observed that similar arguments could be applied to the d-torus, improving the lower bound in Corollary 7.25. It is easy to see an informal argument suggesting that, for random walk on the 2-torus, when n^α vertices are unvisited the set of unvisited vertices has some kind of fractal structure. No rigorous results are known, but heuristics are given in Brummelhuis and Hilhorst [75].

Section 7.3.1. Deriving these exact formulas is scarcely more than undergraduate mathematics, so I am amazed to see that research papers have continued to be published in the 1980s and 1990s claiming various special or general cases as new or noteworthy.

Section 7.3.5. In the setting of isotropic random flight (7.36) with steplength distribution q, it is natural to ask what conditions on q and q' imply that $\tau(q) \geq \tau(q')$ for our parameters τ. For certain distributions on the d-cube, detailed explicit calculations by Karlin et al [207] establish an ordering of the entire eigenvalue sequences, which in particular implies this inequality for τ_2 and τ_0. Establishing results of this type for general Gelfand pairs seems an interesting project.

Miscellaneous. On a finite *field*, such as Z_p for prime p, one can consider "random walks" with steps of the form $x \to \alpha x + \beta$, with a specified joint distribution for (α, β). Chung et al [94] treat one example in detail.

Chapter 8

Advanced L^2 Techniques for Bounding Mixing Times (May 19 1999)

xxx In next revision, we should change the definition [in Chapter 4, yyy:(14)] **of** $\hat{d}(t)$ **so that what is now** $\sqrt{\hat{d}(2t)}$ **becomes** $\hat{d}(t)$**.**

This chapter concerns advanced L^2-based techniques, developed mainly by Persi Diaconis and Laurent Saloff-Coste [117, 118, 119, 120] for bounding mixing times for (finite, irreducible) reversible Markov chains. For convenience, we will work in continuous time throughout this chapter, unless otherwise noted. Many of the results are conveniently expressed in terms of an "L^2 threshold time" $\hat{\tau}$ (xxx use different notation?) defined by

$$\hat{\tau} := \inf\{t > 0 : \max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le e^{-1}\}. \tag{8.1}$$

xxx For NOTES: Discussion of discrete time, esp. negative eigenvalues.

Several preliminary comments are in order here. First, the definition of the L^2 distance $\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2$ may be recalled from Chapter 2 section yyy:6.2, and Chapter 3 yyy:(55) and the spectral representation give useful reexpressions:

$$\begin{aligned} \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2^2 &= \sum_j \pi_j \left(\frac{p_{ij}(t)}{\pi_j} - 1 \right)^2 \\ &= \frac{p_{ii}(2t)}{\pi_i} - 1 \end{aligned} \tag{8.2}$$

$$= \pi_i^{-1} \sum_{m=2}^{n} \exp(-2\lambda_m t) u_{im}^2.$$

Second, from (8.2) and Chapter 4 yyy:(14) we may also write the maximum L^2 distance appearing in (8.1) using

$$\max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2^2 = \max_i \frac{p_{ii}(2t)}{\pi_i} - 1 = \max_{i,j} \frac{p_{ij}(2t)}{\pi_j} - 1 = \hat{d}(2t).$$

Third, by the application of the Cauchy–Schwarz lemma in Chapter 4 Lemma yyy:8, variation distance can be bounded by L^2 distance:

$$4d_i^2(t) := 4\|P_i(X_t \in \cdot) - \pi(\cdot)\|^2 \le \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2^2, \tag{8.3}$$

$$4d^2(t) := 4 \max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|^2 \le \hat{d}(2t); \tag{8.4}$$

these inequalities are the primary motivation for studying L^2 distance.

As argued in Chapter 4 yyy:just following (23),

$$\hat{d}(2t) \le \pi_*^{-1} e^{-2t/\tau_2}, \tag{8.5}$$

where $\tau_2 := \lambda_2^{-1}$ is the relaxation time and $\pi_* := \min_i \pi_i$. Thus if

$$t \ge \tau_2 \left(\frac{1}{2} \log \frac{1}{\pi_*} + c \right),$$

then

$$d(t) \le \tfrac{1}{2} \sqrt{\hat{d}(2t)} \le \tfrac{1}{2} e^{-c}, \tag{8.6}$$

which is small if c is large; in particular, (8.6) gives the upper bound in

$$\tau_2 \le \hat{\tau} \le \tau_2 \left(\frac{1}{2} \log \frac{1}{\pi_*} + 1 \right), \tag{8.7}$$

and the lower bound follows easily.

For many simple chains (see Chapter 5), τ_2 can be computed exactly. Typically, however, τ_2 can only be bounded. This can be done using the "distinguished paths" method of Chapter 4 Section yyy:3. In Section 1 we will see that that method may be regarded as a special case of a "comparison method" whereby a chain with "unknown" relaxation time is compared to a second chain with "known" relaxation time. The greater generality often leads to improved bounds on τ_2. As a bonus, the comparison method also gives bounds on the other "unknown" eigenvalues, and such bounds in turn

can sometimes further decrease the time t required to guarantee that $\hat{d}(2t)$, and hence also $d(t)$, is small.

A second set of advanced techniques, encompassing the notions of Nash inequalities, moderate growth, and local Poincaré inequaltities, is described in Section 3. The development there springs from the inequality

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le N(s) e^{-(t-s)/\tau_2}, \tag{8.8}$$

established for all $0 \le s \le t$ in Section 2, where

$$N(t) = \max_i \|P_i(X_t \in \cdot)\|_2 = \max_i \sqrt{\frac{p_{ii}(2t)}{\pi_i}}, \quad t \ge 0. \tag{8.9}$$

Choosing $s = 0$ in (8.8) gives

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le \pi_i^{-1/2} e^{-t/\tau_2},$$

and maximizing over i recaptures (8.5). The point of Section 3, however, is that one can sometimes reduce the bound by a better choice of s and suitable estimates of the decay rate of $N(\cdot)$. Such estimates can be provided by so-called Nash inequalities, which are implied by (1) moderate growth conditions and (2) local Poincaré inequalities. Roughly speaking, for chains satisfying these two conditions, judicious choice of s shows that variation mixing time and $\hat{\tau}$ are both of order Δ^2, where Δ is the diameter of the graph underlying the chain.

xxx Might not do (1) or (2), so need to modify the above.

To outline a third direction of improvement, we begin by noting that neither of the bounds in (8.7) can be much improved in general. Indeed, ignoring $\Theta(1)$ factors as usual, the lower bound is equality for the n-cycle (Chapter 5, Example yyy:7) and the upper bound is equality for the M/M/1/n queue (Chapter 5, Example yyy:6) with traffic intensity $\rho \in (0, 1)$.

In Section 4 we introduce the *log-Sobolev time* τ_l defined by

$$\tau_l := \sup\{\mathcal{L}(g)/\mathcal{E}(g, g) : g \not\equiv \text{constant}\} \tag{8.10}$$

where $\mathcal{L}(g)$ is the entropy-like quantity

$$\mathcal{L}(g) := \sum_i \pi_i g^2(i) \log(|g(i)|/\|g\|_2),$$

recalling $\|g\|_2^2 = \sum_i \pi_i g^2(i)$. Notice the similarity between (8.10) and the extremal characterization of τ_2 (Chapter 3 Theorem yyy:22):

$$\tau_2 = \sup\{\|g\|_2^2/\mathcal{E}(g, g) : \sum_i \pi_i g(i) = 0, \quad g \not\equiv 0\}.$$

We will see that

$$\tau_2 \le \tau_l \le \tau_2 \frac{\log\left(\frac{1}{\pi_*} - 1\right)}{2(1 - 2\pi_*)}$$

and that $\hat{\tau}$ is more closely related to τ_l than to τ_2, in the sense that

$$\tau_l \le \hat{\tau} \le \tau_l \left(\frac{1}{2} \log\log \frac{1}{\pi_*} + 2\right). \tag{8.11}$$

To illustrate the improvement over (8.7), from the knowledge for the d-cube (Chapter 5, Example yyy:15) that $\tau_2 = d/2$, one can deduce from (8.7) that

$$\tfrac{1}{2}d \le \hat{\tau} \le \tfrac{1}{4}(\log 2)d^2 + \tfrac{1}{2}d. \tag{8.12}$$

In Section 4.4 (Example 27) we will see that $\tau_l = d/2$; then from (8.11) we can deduce the substantial improvement

$$\frac{1}{2}d \le \hat{\tau} \le \frac{1}{4}d\log d + \left(1 - \frac{1}{4}\log\frac{1}{\log 2}\right)d \tag{8.13}$$

upon (8.12).

ZZZ!: Recall also the corrections in my notes on pages 8.2.11–12 (and 8.4.27). Continue same paragraph:

The upper bound here is remarkably tight: from Chapter 5 yyy:(65),

$$\hat{\tau} = \frac{1}{4}d\log d + \left(\frac{1}{4}\log\frac{1}{\log(1 + e^{-2})}\right)d + o(d) \quad \text{as } d \to \infty.$$

ZZZ!: In fact, the remainder term is $O(1)$. Continue same paragraph:

Thus log-Sobolev techniques provide another means of improving mixing time bounds, both in L^2 and, because of (8.3)–(8.4), in variation. As will be seen, these techniques can also be combined usefully with comparison methods and Nash inequalities.

8.1 The comparison method for eigenvalues

xxx Revise Chapter 7, Sections 1.9 and 4, in light of this section?

The comparison method, introduced by Diaconis and Saloff-Coste [117, 118], generalizes the distinguished path method of Chapter 4, Section yyy:3 for bounding the relaxation time of a reversible Markov chain. As before, we first (xxx: delete word?) work in the setting of random walks on weighted graphs. We will proceed for given state space (vertex set) I by comparing a collection (w_{ij}) of weights of interest to another collection $(\tilde{w}_{ij})$; the idea

will be to use known results for the random walk with weights $(\tilde{w}_{ij})$ to derive corresponding results for the walk of interest. We assume that the graph is connected under each set of weights. As in Chapter 4, Section yyy:4.3, we choose ("distinguish") paths γ_{xy} from x to y. Now, however, this need be done only for those (x, y) with $x \neq y$ and $\tilde{w}_{xy} > 0$, but we impose the additional constraint $w_e > 0$ for each edge e in the path. (Here and below, e denotes a directed edge in the graph of interest.) In other words, roughly put, we need to construct a (w_{ij})-path to effect each given $(\tilde{w}_{xy})$-edge. Recall from Chapter 3 yyy:(71) the definition of Dirichlet form:

$$\begin{aligned} \mathcal{E}(g,g) &= \frac{1}{2}\sum_i \sum_{j\neq i} \frac{w_{ij}}{w}(g(j)-g(i))^2, \qquad (8.14)\\ \tilde{\mathcal{E}}(g,g) &= \frac{1}{2}\sum_i \sum_{j\neq i} \frac{\tilde{w}_{ij}}{\tilde{w}}(g(j)-g(i))^2. \end{aligned}$$

Theorem 8.1 (comparison of Dirichlet forms) *For each ordered pair (x, y) of distinct vertices with $\tilde{w}_{xy} > 0$, let γ_{xy} be a path from x to y with $w_e > 0$ for every $e \in \gamma_{xy}$. Then the Dirichlet forms (8.14) satisfy*

$$\tilde{\mathcal{E}}(g,g) \le A\mathcal{E}(g,g) = \mathcal{E}(g,g)\frac{w}{\tilde{w}}\max_e \frac{1}{w_e}\sum_x \sum_{y\neq x} \tilde{w}_{xy}|\gamma_{xy}|1_{(e\in\gamma_{xy})}$$

for every g.

Proof. For an edge $e = (i, j)$ write $\Delta g(e) = g(j) - g(i)$. Then

$$\begin{aligned} 2\tilde{w}\tilde{\mathcal{E}}(g,g) &= \sum_x \sum_{y\neq x} \tilde{w}_{xy}(g(y)-g(x))^2\\ &= \sum_x \sum_{y\neq x} \tilde{w}_{xy}\left(\sum_{e\in\gamma_{xy}} \Delta g(e)\right)^2\\ &\le \sum_x \sum_{y\neq x} \tilde{w}_{xy}|\gamma_{xy}| \sum_{e\in\gamma_{xy}} (\Delta g(e))^2 \quad \text{by Cauchy–Schwarz}\\ &\le A\sum_e w_e(\Delta g(e))^2 = A\cdot 2w\mathcal{E}(g,g). \quad \blacksquare \end{aligned}$$

Remarks. (a) Suppose the comparison weights $(\tilde{w}_{ij})$ are given by

$$\tilde{w}_{ij} = w_i w_j / w \quad \text{for } i, j \in I.$$

The corresponding discrete-time random walk is then the "trivial" walk with $\tilde{w} = w$ and

$$\tilde{w}_i = w_i, \quad \tilde{p}_{ij} = \pi_j, \quad \tilde{\pi}_j = \pi_j$$

for all i, j, and

$$\begin{aligned}\tilde{\mathcal{E}}(g,g) &= \frac{1}{2}\sum_i \sum_{j\neq i} \pi_i \pi_j (g(j)-g(i))^2 = \mathrm{var}_\pi\, g \\ &= \|g\|_2^2 \quad \text{provided } \textstyle\sum_i \pi_i g(i) = 0.\end{aligned}$$

In this case the conclusion of Theorem 1 reduces to

$$\|g\|_2^2 \le \mathcal{E}(g,g) w \max_e \tfrac{1}{w_e} \sum_x \sum_{y\neq x} \pi_x \pi_y |\gamma_{xy}| 1_{(e\in\gamma_{xy})}.$$

This inequality was established in the proof of the distinguished path theorem (Chapter 4 Theorem yyy:32), and that theorem was an immediate consequence of the inequality. Hence the comparison Theorem 1 may be regarded as a generalization of the distinguished path theorem.

[xxx For NOTES: We've used simple Sinclair weighting. What about other weighting in use of Cauchy–Schwarz? Hasn't been considered, as far as I know.]

(b) When specialized to the setting of reversible random flights on Cayley graphs described in Chapter 7 Section yyy:1.9, Theorem 1 yields Theorem yyy:14 of Chapter 7. To see this, adopt the setup in Chapter 7 Section yyy:1.9, and observe that the word

$$x = g_1 g_2 \cdots g_d \quad (\text{with each } g_i \in \mathcal{G}) \tag{8.15}$$

corresponds uniquely to a path

$$\gamma_{\mathrm{id},x} = (\mathrm{id}, g_1, g_1 g_2, \ldots, g_1 g_2 \cdots g_d = x) \tag{8.16}$$

in the Cayley graph corresponding to the generating set $\mathcal{G}$ of interest. Having built paths $\gamma_{\mathrm{id},x}$ for each $x \in I$, we then can build paths γ_{yz} for $y, z \in I$ by exploiting vertex-transitivity, to wit, by setting

$$\gamma_{yz} = (y, yg_1, yg_1 g_2, \ldots, yg_1 g_2 \cdots g_d = z)$$

where $y^{-1}z = x$ and the path $\gamma_{\mathrm{id},x}$ is given by (8.16). In Theorem 1 we then have both stationary distributions π and $\tilde{\pi}$ uniform,

$$\tilde{w}_{xy} = \tilde{\mu}(x^{-1}y)/n, \quad \tilde{w} = 1, \quad |\gamma_{xy}| = |\gamma_{\mathrm{id},x^{-1}y}| = d(\mathrm{id}, x^{-1}y),$$

and, if $e = (v, vg)$ with $v \in I$ and $g \in \mathcal{G}$,

$$w_e = \mu(g)/n, \quad w = 1, \quad 1_{(e\in\gamma_{xy})} = 1_{((x^{-1}v, x^{-1}vg)\in\gamma_{\mathrm{id},x^{-1}y})}.$$

Thus A of Theorem 1 equals

$$\max_{v\in I,\ g\in\mathcal{G}} \frac{1}{\mu(g)} \sum_x \sum_{y\neq x} \tilde{\mu}(x^{-1}y) d(\mathrm{id}, x^{-1}y) 1_{(x^{-1}v, x^{-1}vg\in\gamma_{\mathrm{id},x^{-1}y})},$$

which reduces easily to K of Theorem yyy:14 of Chapter 7. Since π and $\tilde{\pi}$ are both uniform, the extremal characterization

$$\tau_2 = \sup\{\|g\|_2^2/\mathcal{E}(g,g) :\ \sum_i \pi_i g(i) = 0\} \tag{8.17}$$

gives Theorem yyy:14 of Chapter 7.

Theorem 8.1 compares Dirichlet forms. To compare relaxation times using the extremal characterization (8.17), we compare L^2-norms using the same "direct" technique as for Chapter 3 Lemma yyy:26. For any g,

$$\|g\|_2^2 \le \|g\|_2^{\sim 2} \max_i(\pi_i/\tilde{\pi}_i) \tag{8.18}$$

where, as usual, $\pi_i := w_i/w$ and $\tilde{\pi}_i = \tilde{w}_i/\tilde{w}$. So if g has π-mean 0 and $\tilde{\pi}$-mean b, then

$$\frac{\|g\|_2^2}{\mathcal{E}(g,g)} \le \frac{\|g-b\|_2^2}{\mathcal{E}(g-b,g-b)} \le \frac{A\|g-b\|_2^{\sim 2}}{\tilde{\mathcal{E}}(g-b,g-b)} \max_i(\pi_i/\tilde{\pi}_i). \tag{8.19}$$

Thus

Corollary 8.2 (comparison of relaxation times) *In Theorem 1,*

$$\tau_2 \le \frac{A}{a}\tilde{\tau}_2$$

where

$$\begin{aligned} A &:= \frac{w}{\tilde{w}} \max_e \frac{1}{w_e} \sum_x \sum_{y\neq x} \tilde{w}_{xy} |\gamma_{xy}| 1_{(e\in\gamma_{xy})}, \\ a &:= \min_i(\tilde{\pi}_i/\pi_i). \end{aligned}$$

xxx Perhaps restate as

$$\tau_2 \le \tilde{\tau}_2$$

where

$$B := \left(\max_e \frac{1}{w_e} \sum\sum\cdots\right)\left(\max_i \frac{w_i}{\tilde{w}_i}\right)$$

(and similarly for Corollaries 4 and 7)?

xxx Remark about best if $\pi = \tilde{\pi}$?

Here is a simple example, taken from [117], showing the improvement in Corollary 8.2 over Theorem yyy:32 of Chapter 4 provided by the freedom in choice of benchmark chain.

xxx NOTE: After the fact, I realized that the following example was already Example yyy:20 of Chapter 7; must reconcile.

Example 8.3 Consider a card shuffle which transposes the top two cards in the deck, moves the top card to the bottom, or moves the bottom card to the top, each with probability 1/3. This example fits the specialized group framework of Chapter 7 Section yyy:1.9 (see also Remark (b) following Theorem 8.1 above) with I taken to be the symmetric group on m letters and

$$\mathcal{G} := \{(1\ 2), (m\ m-1\ m-2\ \cdots\ 1), (1\ 2\ \cdots\ m)\}$$

in cycle notation. [If the order of the deck is represented by a permutation σ in such a way that $\sigma(i)$ is the position of the card with label i, and if permutations are composed left to right, then $\sigma \cdot (m\ m-1\ m-2\ \cdots\ 1)$ is the order resulting from σ by moving the top card to the bottom.]

We obtain a representation (8.15) for any given permutation x by writing

$$x = h_m h_{m-1} \cdots h_2$$

in such a way that

$$(h_m \cdots h_i)^{-1}(j) = x^{-1}(j) \quad \text{for } i \le j \le m \tag{8.20}$$

(i.e., $h_m \cdots h_i$ and x agree in positions i through m) and each h_i is explicitly represented as a product of generators. To accomplish this, we proceed inductively. Suppose that (8.20) holds for given $i \in \{3, \ldots, m+1\}$, and that $(h_m \cdots h_i)(x^{-1}(i-1)) = l_i = l$, with $1 \le l \le i-1$. Then let

$$\begin{aligned} h_{i-1} \ := \ & (m\ m-1\ m-2\ \cdots\ 1)^{l-1}[(1\ 2)(m\ m-1\ m-2\ \cdots\ 1)]^{i-l-1} \\ & \cdot (m\ m-1\ m-2\ \cdots\ 1)^{m-i+2}. \end{aligned}$$

In words, beginning with $h_m \cdots h_i$, we repeatedly move the top card to the bottom until card $x^{-1}(i-1)$ has risen to the top; then we repeatedly transpose and shift until the top $m-i+2$ cards, in order, are $x^{-1}(i-1), \ldots, x^{-1}(m)$; and finally we cut these $m-i+2$ cards to the bottom.

xxx Either revise Section 1.9 of Chapter 7 to delete requirement of *geodesic* paths, or explain one can erase cycles.

It follows that the diameter Δ of the Cayley graph associated with $\mathcal{G}$ satisfies

$$\Delta \le \sum_{i=2}^{m+1} [(l_i - 1) + 2(i - l_i - 1) + (m - i + 2)] \le 3\binom{m}{2}$$

and so by Chapter 7 Corollary yyy:15 that $\tau_2 \le 27\binom{m}{2}^2 < \frac{27}{4}m^4$.

To improve this bound on the relaxation time we compare the chain of interest to the random transposition chain of Chapter 7 Example yyy:18 and employ Corollary 8.2, or rather its specialization, (yyy:Theorem 14) of Chapter 7.

xxx Continue as in Chapter 7 Example yyy:20 to get

$$\frac{\tau_2}{\tilde{\tau}_2} \le 3\beta^2, \quad \tilde{\tau}_2 = \frac{m}{2}, \quad \tau_2 \le \frac{27}{2}m^3.$$

xxx Test function on page 2139 of [117] shows this is right order.

Corollary 8.2 can be combined with the inequality

$$\hat{\tau} \le \tau_2 \left(\frac{1}{2} \log \frac{1}{\pi_*} + 1 \right) \tag{8.21}$$

from (8.7) to bound the L^2 threshold parameter $\hat{\tau}$ for the chain of interest, but Theorem 8.1 sometimes affords a sharper result. From the Courant–Fischer "min–max" theorem ([183], Theorem 4.2.11) it follows along the same lines as in Chapter 3 Section yyy:6.3 that

$$\lambda^{-1} = \inf \rho(h_1, h_2, \ldots, h_{m-1}), \quad m = 2, \ldots, n, \tag{8.22}$$

where $h_1 \equiv 1$ and xxx Say the conditions better!

$$\begin{aligned} &\rho(h_1, h_2, \ldots, h_{m-1}) \\ &\quad := \sup\{\|g\|_2^2 / \mathcal{E}(g,g) : \sum_i \pi_i h_j(i) g(i) = 0 \quad \text{for } j = 1, \ldots, m-1\} \end{aligned}$$

and the inf in (8.22) is taken over all vectors $h_1, \ldots, h_{m-1}$ that are orthogonal in $L^2(\pi)$ (or, equivalently, that are linearly independent). Using (8.19), Corollary 8.2 now generalizes to

Corollary 8.4 (comparison of eigenvalues) *In Theorem 8.1, the eigenvalues λ_m and $\tilde{\lambda}_m$ in the respective spectral representations satisfy*

$$\lambda_m^{-1} \le \frac{A}{a} \tilde{\lambda}_m^{-1}$$

with A and a as defined in Corollary 8.2.

Here is a simple example [118] not possessing vertex-transitivity:
xxx NOTE: This is a DIRECT comparison!: see Chapter 3 Section 6.4.

Example 8.5 *Random walk on a d-dimensonal grid.*

To keep the notation simple, we let $d = 2$ and consider the grid $I := \{0, \ldots, m_1 - 1\} \times \{0, \ldots, m_2 - 1\}$ as an (unweighted) subgraph of $\mathbf{Z}^2$. The eigenvalues λ_l are not known in closed form. However, if we add self-loops to make a benchmark graph where I is regular with degree 4, then the eigenvalues $\tilde{\lambda}_l$ for the continuous-time walk are

$$1 - \frac{1}{2}\left(\cos\frac{\pi r}{m_1} + \cos\frac{\pi s}{m_2}\right), \quad 0 \le r \le m_1 - 1, \quad 0 \le s \le m_2 - 1.$$

xxx Product chain. Add discussion of *all* eigenvalues to Section yyy:6.2 of Chapter 4?
xxx P.S. See Chapter 5, (66).
In particular, assuming $m_1 \ge m_2$ we have

$$\tilde{\tau}_2 = 2\left(1 - \cos\frac{\pi}{m_1}\right)^{-1}. \tag{8.23}$$

Now we apply Corollary 8.4 to bound the eigenvalues λ_l. In Theorem 8.1, the two graphs agree except for self-loops, so

$$A = w/\tilde{w};$$

furthermore,

$$a = \min_i \frac{\tilde{\pi}_i}{\pi_i} = \frac{w}{\tilde{w}} \min_i \tilde{w}_i w_i,$$

so

$$\frac{A}{a} = \max_i \frac{w_i}{\tilde{w}_i} \le 1.$$

Thus $\lambda_l^{-1} \le \tilde{\lambda}_l^{-1}$ for $1 \le l \le n := m_1 m_2$; in particular,

$$\tau_2 \le \tilde{\tau}_2. \tag{8.24}$$

Comparing the other way around gives

$$\tilde{\lambda}_l^{-1} \le \left(\max_i \frac{\tilde{w}_i}{w_i}\right) \lambda_l^{-1} = 2\lambda_l^{-1}, \quad 1 \le l \le n$$

and in particular

$$\tau_2 \ge \tfrac{1}{2}\tilde{\tau}_2.$$

The result $\frac{1}{2}\tilde{\lambda}_l^{-1} \le \lambda_l^{-1} \le \tilde{\lambda}_l^{-1}$ extends to general d, for which (for example)

$$\tilde{\tau}_2 = d\left(1 - \cos\frac{\pi}{m}\right)^{-1}$$

where $I = \{0, \ldots, m_1 - 1\} \times \cdots \times \{0, \ldots, m_d - 1\}$ and $m := \max_i m_i$.

Example 8.6 *Random walk on a thinned grid.*

As a somewhat more interesting example, suppose we modify the grid in $\mathbf{Z}^2$ in Example 8.5 by deleting at most one edge from each unit square.

xxx Copy picture on page 700 in [118] as example?

Again we can apply Corollary 8.4, using the same benchmark graph as in Example 8.5. In Theorem 8.1, $\tilde{w}_{xy} > 0$ for $x \neq y$ if and only if x and y are neighboring vertices in the (unthinned) grid $\{0, \ldots, m_1-1\} \times \{0, \ldots, m_2-1\}$. We can choose γ_{xy} to have length 1 (if the edge joining x and y has not been deleted) or 3 (if it has). For any directed edge e in the grid, there are at most two paths of length 3 and at most one path of length 1 passing through e. Thus $A \le 7w/\tilde{w}$, and so $A/a \le 7 \max_i(w_i/\tilde{w}_i) \le 7$; comparing the other way around is even easier (all paths have length 1), and we find

$$\tfrac{1}{4}\tilde{\lambda}_l^{-1} \le \lambda_l^{-1} \le 7\tilde{\lambda}_l^{-1}, \quad 2 \le l \le n.$$

xxx REMINDER: NOTES OR ELSEWHERE?: *Mention* exclusion process [149, 118].

Example 8.7 *The n-path with end self-loops.*

The comparison technique does not always provide results as sharp as those in the preceding two examples, even when the two chains are "close." For example, let the chain of interest be the n-path, with self-loops added at each end added to make the graph regular with degree 2, and let the benchmark graph be the n-cycle (Chapter 5, Example yyy:7). Use of Corollary 8.2 gives only $\tau_2 \le n\tilde{\tau}_2$, whereas in fact $\tau_2 = \left(1 - \cos\frac{\pi}{n}\right)^{-1} \sim \frac{2}{\pi^2}n^2$ and $\tilde{\tau}_2 = \left(1 - \cos\frac{2\pi}{n}\right)^{-1} \sim \frac{1}{2\pi^2}n^2$.

It is difficult in general to use Corollary 8.4 to improve upon (8.21). However, when both the chain of interest and the benchmark chain are symmetric reversible chains (as defined in Chapter 7 Section yyy:1.1), it follows from Chapter 4 yyy:(14) by averaging over i that

$$\hat{d}(t) \le \tilde{\hat{d}}\left(\frac{a}{A}t\right), \quad t \ge 0,$$

and hence from (8.1) we obtain

Corollary 8.8 (comparison of L^2 mixing times) *In Theorem* 8.1*, if both the graph of interest and the benchmark graph are vertex-transitive, then the L^2 mixing time parameters $\hat{\tau}$ and $\tilde{\hat{\tau}}$ satisfy*

$$\hat{\tau} \le \frac{A}{a}\tilde{\hat{\tau}}.$$

Example 8.9 Returning to the slow card-shuffling scheme of Example 8.3 with random transpositions benchmark, it is known from group representation methods [112, 122] which make essential use of *all* the eigenvalues $\tilde{\lambda}_r$, not just $\tilde{\lambda}_2$, that

$$\tilde{\hat{\tau}} \sim \tfrac{1}{2} m \log m \quad \text{as } m \to \infty.$$

Since $a = 1$ and A ($= K$ of Chapter 7, Theorem yyy:14) $\le 27m^2$, it follows that

$$\hat{\tau} \le (1 + o(1))\tfrac{27}{2} m^3 \log m. \tag{8.25}$$

This improves upon Example 8.3, which combines with (8.21) to give only

$$\hat{\tau} \le (1 + o(1))\tfrac{27}{4} m^4 \log m.$$

xxx Show truth is $\hat{\tau} = \Theta(m^3 \log m)$?

8.2 Improved bounds on L^2 distance

The central theme of the remainder of this chapter is that norms other than the L^1 norm (and closely related variation distance) and L^2 norm can be used to improve substantially upon the bound

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le \pi^{-1/2} e^{-t/\tau_2}. \tag{8.26}$$

8.2.1 L^q norms and operator norms

Our discussion here of L^q norms will parallel and extend the discussion in Chapter 2 Section yyy:6.2 of L^1 and L^2 norms. Given $1 \le q \le \infty$, both the L^q norm of a function and the L^q norm of a signed measure are defined with respect to some fixed reference probability distribution π on I, which for our purposes will be the stationary distribution of some irreducible but not necessarily reversible chain under consideration. For $1 \le q < \infty$, the L^q norm of a *function* $f : I \to \mathbf{R}$ is

$$\|f\|_q := \left(\sum_i \pi_i |f(i)|^q \right)^{1/q},$$

and we define the L^q norm of a *signed measure* ν on I to be the L^q norm of its density function with respect to π:

$$\|\nu\|_q := \left(\sum_j \pi_j^{1-q} |\nu_j|^q \right)^{1/q}.$$

For $q = \infty$, the corresponding definitions are

$$\|f\|_\infty := \max_i |f(i)|$$

and

$$\|\nu\|_\infty := \max_j (|\nu_j|/\pi_j).$$

Any matrix $\mathbf{A} := (a_{ij} : i, j \in I)$ operates on functions $f : I \to \mathbf{R}$ by left-multiplication:

$$(\mathbf{A}f)(i) = \sum_j a_{ij} f(j), \tag{8.27}$$

and on signed measures ν by right-multiplication:

$$(\nu\mathbf{A})_j = \sum_i \nu_i a_{ij}. \tag{8.28}$$

For (8.27), fix $1 \le q_1 \le \infty$ and $1 \le q_2 \le \infty$ and regard $\mathbf{A}$ as a linear operator mapping L^{q_1} into L^{q_2}. The *operator norm* $\|\mathbf{A}\|_{q_1 \to q_2}$ is defined by

$$\|\mathbf{A}\|_{q_1 \to q_2} := \sup\{\|\mathbf{A}f\|_{q_2} : \|f\|_{q_1} = 1\}. \tag{8.29}$$

The *sup* in (8.29) is always achieved, and there are many equivalent reexpressions, including

$$\begin{aligned} \|\mathbf{A}\|_{q_1 \to q_2} &= \max\{\|\mathbf{A}f\|_{q_2} : \|f\|_{q_1} \le 1\} \\ &= \max\{\|\mathbf{A}f\|_{q_2}/\|f\|_{q_1} : f \ne 0\}. \end{aligned}$$

Note also that

$$\|\mathbf{BA}\|_{q_1 \to q_3} \le \|\mathbf{A}\|_{q_1 \to q_2} \|\mathbf{B}\|_{q_2 \to q_3}, \quad 1 \le q_1, q_2, q_3 \le \infty. \tag{8.30}$$

For (8.28), we may similarly regard $\mathbf{A}$ as a linear operator mapping signed measures ν, measured by $\|\nu\|_{q_1}$, to signed measures $\nu\mathbf{A}$, measured by $\|\nu\mathbf{A}\|_{q_2}$. The corresponding definition of operator norm, call it $|||\mathbf{A}|||_{q_1 \to q_2}$, is then

$$|||\mathbf{A}|||_{q_1 \to q_2} := \sup\{\|\nu\mathbf{A}\|_{q_2} : \|\nu\|_{q_1} = 1\}.$$

A brief calculation shows that

$$|||\mathbf{A}|||_{q_1 \to q_2} = \|\mathbf{A}^*\|_{q_1 \to q_2},$$

where $\mathbf{A}^*$ is the matrix with (i,j) entry $\pi_j a_{ji}/\pi_i$, that is, $\mathbf{A}^*$ is the adjoint operator to $\mathbf{A}$ (with respect to π).

Our applications in this chapter will all have $\mathbf{A} = \mathbf{A}^*$, so we will not need to distinguish between the two operator norms. In fact, all our applications will take $\mathbf{A}$ to be either $\mathbf{P}_t$ or $\mathbf{P}_t - \mathbf{E}$ for some $t \geq 0$, where

$$\mathbf{P}_t := (p_{ij}(t) : i, j \in I)$$

xxx notation $\mathbf{P}_t$ found elsewhere in book?

and $\mathbf{E} = \lim_{t\to\infty} \mathbf{P}_t$ is the transition matrix for the trivial discrete time chain that jumps in one step to stationarity:

$$\mathbf{E} = (\pi_j : i, j \in I),$$

and where we assume that the chain for $(\mathbf{P}_t)$ is *reversible*. Note that $\mathbf{E}$ operates on functions essentially as expectation with respect to π:

$$(\mathbf{E}f)(i) = \sum_j \pi_j f(j), \quad i \in I.$$

The effect of $\mathbf{E}$ on signed measures is to map ν to $(\sum_i \nu_i)\pi$, and

$$\mathbf{P}_t\mathbf{E} = \mathbf{E} = \mathbf{E}\mathbf{P}_t, \quad t \geq 0. \tag{8.31}$$

8.2.2 A more general bound on L^2 distance

The following preliminary result, a close relative to Chapter 3, Lemmas yyy:21 and 23, is used frequently enough in the sequel that we isolate it for reference. It is the simple identity in part (b) that shows why L^2-based techniques are so useful.

Lemma 8.10 *(a) For any function f,*

$$\frac{d}{dt}\|\mathbf{P}_t f\|_2^2 = -2\mathcal{E}(\mathbf{P}_t f, \mathbf{P}_t f) \leq -\frac{2}{\tau_2}\mathrm{var}_\pi \mathbf{P}_t f \leq 0.$$

(b)

$$\|\mathbf{P}_t - \mathbf{E}\|_{2\to 2} = e^{-t/\tau_2}, \quad t \geq 0.$$

Proof. (a) Using the backward equations

$$\frac{d}{dt}p_{ij}(t) = \sum_k q_{ik}\, p_{kj}(t)$$

we find

$$\frac{d}{dt}(\mathbf{P}_t f)(i) = \sum_k q_{ik}[(\mathbf{P}_t f)(k)]$$

and so

$$\begin{aligned}
\frac{d}{dt}\|\mathbf{P}_t f\|_2^2 &= 2\sum_i \sum_k \pi_i[(\mathbf{P}_t f)(i)]q_{ik}[(\mathbf{P}_t f)(k)] \\
&= -2\mathcal{E}(\mathbf{P}_t f, \mathbf{P}_t f) \quad \text{by Chapter 3 yyy:(70)} \\
&\le -\frac{2}{\tau_2}\mathrm{var}_\pi\, \mathbf{P}_t f \quad \text{by the extremal characterization of } \tau_2.
\end{aligned}$$

(b) From (a), for any f we have

$$\frac{d}{dt}\|(\mathbf{P}_t - \mathbf{E})f\|_2^2 = \frac{d}{dt}\|\mathbf{P}_t(f - \mathbf{E}f)\|_2^2 \le -\frac{2}{\tau_2}\|(\mathbf{P}_t - \mathbf{E})f\|_2^2,$$

which yields

$$\begin{aligned}
\|(\mathbf{P}_t - \mathbf{E})f\|_2^2 &\le \|(\mathbf{P}_0 - \mathbf{E})f\|_2^2\, e^{-2t/\tau_2} = (\mathrm{var}_\pi\, f)e^{-2t/\tau_2} \\
&\le \|f\|_2^2\, e^{-2t/\tau_2}.
\end{aligned}$$

Thus $\|\mathbf{P}_t - \mathbf{E}\|_{2\to 2} \le e^{-t/\tau_2}$. Taking f to be the eigenvector

$$f_i := \pi_i^{-1/2} u_{i2}, \quad i \in I,$$

of $\mathbf{P}_t - \mathbf{E}$ corresponding to eigenvalue $\exp(-t/\tau_2)$ demonstrates equality and completes the proof of (b). ∎

The key to all further developments in this chapter is the following result.

Lemma 8.11 *For an irreducible reversible chain with arbitrary initial distribution and any $s, t \ge 0$,*

$$\|P(X_{s+t} \in \cdot) - \pi(\cdot)\|_2 \le \|P(X_s \in \cdot)\|_2 \|\mathbf{P}_t - \mathbf{E}\|_{2\to 2} = \|P(X_s \in \cdot)\|_2\, e^{-t/\tau_2}.$$

Proof. The equality is Lemma 8.10(b), and

$$\|P(X_{s+t} \in \cdot) - \pi(\cdot)\|_2 = \|P(X_s \in \cdot)(\mathbf{P}_t - \mathbf{E})\|_2 \le \|P(X_s \in \cdot)\|_2 \|\mathbf{P}_t - \mathbf{E}\|_{2\to 2}$$

proves the inequality. ■

We have already discussed, in Section 8.1, a technique for bounding τ_2 when (as is usually the case) it cannot be computed exactly. To utilize Lemma 8.11, we must also bound $\|P(X_s \in \cdot)\|_2$. Since

$$\|P(X_s \in \cdot)\|_2 = \|P(X_0 \in \cdot)\mathbf{P}_s\|_2 \tag{8.32}$$

xxx For NOTES: By Jensen's inequality (for $1 \le q < \infty$), any transition matrix contracts L^q for any $1 \le q \le \infty$.

and each $\mathbf{P}_t$ is contractive on L^2, i.e., $\|\mathbf{P}_t\|_{2\to 2} \le 1$ (this follows, for example, from Lemma 8.10(a); and note that $\|\mathbf{P}_t\|_{2\to 2} = 1$ by considering constant functions), it follows that

$$\|P(X_s \in \cdot)\|_2 \text{ decreases monotonically to 1 as } s \uparrow \infty, \tag{8.33}$$

and the decrease is strictly monotone unless $P(X_0 \in \cdot) = \pi(\cdot)$. From (8.32) follows

$$\|P(X_s \in \cdot)\|_2 \le \|P(X_0 \in \cdot)\|_{q^*}\|\mathbf{P}_s\|_{q^*\to 2} \text{ for any } 1 \le q^* \le \infty, \tag{8.34}$$

and again

$$\|\mathbf{P}_s\|_{q^*\to 2} \text{ decreases monotonically to 1 as } s \uparrow \infty. \tag{8.35}$$

The norm $\|\mathbf{P}_s\|_{q^*\to 2}$ decreases in q^* (for fixed s) and is identically 1 when $q^* \ge 2$, but in applications we will want to take $q^* < 2$. The following duality lemma will then often prove useful. Recall that $1 \le q, q^* \le \infty$ are said to be *(Hölder-)conjugate exponents* if

$$\frac{1}{q} + \frac{1}{q^*} = 1. \tag{8.36}$$

Lemma 8.12 *For any operator* $\mathbf{A}$*, let*

$$\mathbf{A}^* = (\pi_j a_{ji}/\pi_i : i, j \in I)$$

denote its adjoint with respect to π*. Then, for any* $1 \le q_1, q_2 \le \infty$*,*

$$\|\mathbf{A}\|_{q_1\to q_2} = \|\mathbf{A}^*\|_{q_2^*\to q_1^*}.$$

In particular, for a reversible chain and any $1 \le q \le \infty$ *and* $s \ge 0$*,*

$$\|\mathbf{P}_s\|_{2\to q} = \|\mathbf{P}_s\|_{q^*\to 2}. \tag{8.37}$$

Proof. Classical duality for L^q spaces (see, e.g., Chapter 6 in [303]) asserts that, given $1 \le q \le \infty$ and g on I,

$$\|g\|_{q^*} = \max\{|\langle f, g\rangle| : \|f\|_q = 1\}$$

where

$$\langle f, g\rangle := \sum_i \pi_i f(i) g(i).$$

Thus

$$\begin{aligned} \|\mathbf{A}^* g\|_{q_1^*} &= \max\{|\langle f, \mathbf{A}^* g\rangle| : \|f\|_{q_1} = 1\} \\ &= \max\{|\langle \mathbf{A} f, g\rangle| : \|f\|_q = 1\}, \end{aligned}$$

and also

$$|\langle \mathbf{A} f, g\rangle| \le \|\mathbf{A} f\|_{q_2} \|g\|_{q_2^*} \le \|\mathbf{A}\|_{q_1 \to q_2} \|f\|_{q_1} \|g\|_{q_2^*},$$

so

$$\|\mathbf{A}^* g\|_{q_1^*} \le \|\mathbf{A}\|_{q_1 \to q_2} \|g\|_{q_2^*}.$$

Since this is true for every g, we conclude $\|\mathbf{A}^*\|_{q_2^* \to q_1^*} \le \|\mathbf{A}\|_{q_1 \to q_2}$. Reverse roles to complete the proof. ∎

As a corollary, if $q^* = 1$ then (8.34) and (8.37) combine to give

$$\|P(X_s \in \cdot)\|_2 \le \|\mathbf{P}_s\|_{1 \to 2} = \|\mathbf{P}_s\|_{2 \to \infty}$$

and then

$$\|P(X_{s+t} \in \cdot) - \pi(\cdot)\|_2 \le \|\mathbf{P}_s\|_{2 \to \infty} \, e^{-t/\tau_2}$$

from Lemma 8.11. Thus

$$\sqrt{\hat{d}(2(s+t))} \le \|\mathbf{P}_s\|_{2 \to \infty} \, e^{-t/\tau_2}. \tag{8.38}$$

Here is a somewhat different derivation of (8.38):

Lemma 8.13 *For $0 \le s \le t$,*

$$\begin{aligned} \sqrt{\hat{d}(2t)} = \|\mathbf{P}_t - \mathbf{E}\|_{2 \to \infty} &\le \|\mathbf{P}_s\|_{2 \to \infty} \|\mathbf{P}_{t-s} - \mathbf{E}\|_{2 \to 2} \\ &= \|\mathbf{P}_s\|_{2 \to \infty} \, e^{-(t-s)/\tau_2}. \end{aligned}$$

Proof. In light of (8.31), (8.30), and Lemma 8.10(b), we need only establish the first equality. Indeed, $\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2$ is the L^2 norm of the *function* $(P_i(X_t \in \cdot)/\pi(\cdot)) - 1$ and so equals

$$\max\left\{\left|\sum_j (p_{ij}(t) - \pi_j) f(j)\right| : \|f\|_2 = 1\right\} = \max\{|((\mathbf{P}_t - \mathbf{E}) f)(i)| : \|f\|_2 = 1\}.$$

Taking the maximum over $i \in I$ we obtain

$$\begin{aligned}\sqrt{\hat{d}(2t)} &= \max\left\{\max_i |((\mathbf{P}_t - \mathbf{E})f)(i)| : \|f\|_2 = 1\right\} \\ &= \max\{\|(\mathbf{P}_t - \mathbf{E})f\|_\infty : \|f\|_2 = 1\} \\ &= \|\mathbf{P}_t - \mathbf{E}\|_{2\to\infty}. \quad \blacksquare\end{aligned}$$

Choosing $s = 0$ in Lemma 8.11 recaptures (8.26), and choosing $s = 0$ in Lemma 8.13 likewise recaptures the consequence (8.5) of (8.26). The central theme for both Nash and log-Sobolev techniques is that one can improve upon these results by more judicious choice of s.

8.2.3 Exact computation of $N(s)$

The proof of Lemma 8.13 can also be used to show that

$$N(s) := \|\mathbf{P}_s\|_{2\to\infty} = \max_i \|P_i(X_s \in \cdot)\|_2 = \max_i \sqrt{\frac{p_{ii}(2s)}{\pi_i}}, \tag{8.39}$$

as at (8.9). In those rare instances when the spectral representation is known explicitly, this gives the formula

xxx Also useful in conjunction with comparison method—see Section 3.

xxx If we can compute this, we can compute $\hat{d}(2t) = N^2(t) - 1$. But the point is to test out Lemma 8.13.

$$N^2(s) = 1 + \max_i \pi_i^{-1} \sum_{m=2}^{n} u_{im}^2 \exp(-2\lambda_m s), \tag{8.40}$$

and the techniques of later sections are not needed to compute $N(s)$. In particular, in the vertex-transitive case

$$N^2(s) = 1 + \sum_{m=2}^{n} \exp(-2\lambda s).$$

The norm $N(s)$ clearly behaves nicely under the formation of products:

$$N(s) = N^{(1)}(s)N^{(2)}(s). \tag{8.41}$$

Example 8.14 *The two-state chain and the d-cube.*

For the two-state chain, the results of Chapter 5 Example yyy:4 show

$$N^2(s) = 1 + \frac{\max(p,q)}{\min(p,q)} e^{-2(p+q)s}.$$

In particular, for the continuized walk on the 2-path,

$$N^2(s) = 1 + e^{-4s}.$$

By the extension of (8.41) to higher-order products, we therefore have

$$N^2(s) = (1 + e^{-4s/d})^d$$

for the continuized walk on the d-cube. This result is also easily derived from the results of Chapter 5 Example yyy:15. For $d \geq 2$ and $t \geq \frac{1}{4} d \log(d-1)$, the optimal choice of s in Lemma 8.13 is therefore

$$s = \tfrac{1}{4} d \log(d-1)$$

and this leads in straightforward fashion to the bound

$$\hat{\tau} \leq \tfrac{1}{4} d(\log d + 3).$$

While this is a significant improvement on the bound [cf. (8.12)]

$$\hat{\tau} \leq \tfrac{1}{4}(\log 2) d^2 + \tfrac{1}{2} d$$

obtained by setting $s = 0$, i.e., obtained using only information about τ_2, it is not

xxx REWRITE!, in light of corrections to my notes.

as sharp as the upper bound

$$\hat{\tau} \leq (1 + o(1)) \tfrac{1}{4} d \log d$$

in (8.13) that will be derived using log-Sobolev techniques.

Example 8.15 *The complete graph.*

For this graph, the results of Chapter 5 Example yyy:9 show

$$N^2(s) = 1 + (n-1) \exp\left(-\frac{2ns}{n-1} \right).$$

It turns out for this example that $s = 0$ is the optimal choice in Lemma 8.13. This is not surprising given the sharpness of the bound in (8.7) in this case. See Example 8.32 below for further details.

Example 8.16 *Product random walk on a d-dimensional grid.*

Consider again the benchmark product chain (i.e., the "tilde chain") in Example 8.5. That chain has relaxation time

$$\tau_2 = d\left(1 - \cos\left(\frac{\pi}{m}\right)\right)^{-1} \le \frac{1}{2}dm^2,$$

so choosing $s = 0$ in Lemma 8.13 gives

$$\begin{aligned} \hat{\tau} &\le \frac{d}{1-\cos(\pi/m)}\left(\frac{1}{2}\log n + 1\right) \\ &\le \tfrac{1}{4}dm^2(\log n + 2). \end{aligned} \tag{8.42}$$

This bound can be improved using $N(\cdot)$. Indeed, if we first consider continuized random walk on the m-path with self-loops added at each end, the stationary distribution is uniform, the eigenvalues are

$$\lambda_l = 1 - \cos(\pi(l-1)/m), \quad 1 \le l \le m,$$

and the eigenvectors are given by

$$u_{il} = (2/m)^{1/2}\cos(\pi(l-1)(i-\tfrac{1}{2})/m), \quad 0 \le i \le m-1, \quad 2 \le l \le m.$$

According to (8.40) and simple estimates, for $s > 0$

$$N^2(s) - 1 \le 2\sum_{l=2}^{m}\exp[-2s(1-\cos(\pi(l-1)/m)] \le 2\sum_{l=1}^{m-1}\exp(-4sl^2/m^2)$$

and

$$\begin{aligned} \sum_{l=2}^{m-1}\exp(-4sl^2/m^2) &\le \int_{x=1}^{\infty}\exp(-4sx^2/m^2)\,dx \\ &= m\left(\frac{\pi}{4s}\right)^{1/2} P\left(Z \ge \frac{2(2s)^{1/2}}{m}\right) \\ &\le \left(\frac{\pi/4}{4s/m^2}\right)^{1/2}\exp(-4s/m^2) \\ &\le (4s/m^2)^{-1/2}\exp(-4s/m^2) \end{aligned}$$

when Z is standard normal; in particular, we have used the well-known (xxx: point to Ross book exercise) bound

$$P(Z \ge z) \le \tfrac{1}{2}e^{-z^2/2}, \quad z \ge 0.$$

Thus

$$N^2(s) \le 1 + 2[1 + (4s/m^2)^{-1/2}] \exp(-4s/m^2), \quad s > 0.$$

Return now to the "tilde chain" of Example 8.5, and assume for simplicity that $m_1 = \cdots = m_d = m$. Since this chain is a slowed-down d-fold product of the path chain, it has

$$N^2(s) \le \left[1 + 2\left(1 + \left(\frac{4s}{dm^2}\right)^{-1/2}\right) \exp\left(-\frac{4s}{dm^2}\right)\right]^d, \quad s > 0. \tag{8.43}$$

In particular, since $\hat{d}(2t) = N^2(t) - 1$, it is now easy to see that

$$\hat{\tau} \le Km^2 d \log d = Kn^{2/d} d \log d \tag{8.44}$$

for a universal constant K.

xxx We've improved on (8.42), which gave order $m^2 d^2 \log m$.

xxx When used to bound $d(t)$ at (8.4), the bound (8.44) is "right": see Theorem 4.1, p. 481, in [120].

The optimal choice of s in Lemma 8.13 cannot be obtained explicitly when (8.43) is used to bound $N(s) = \|\mathbf{P}_s\|_{2\to\infty}$. For this reason, and for later purposes, it is useful to use the simpler, but more restricted, bound

$$N^2(s) \le (4dm^2/s)^{d/2} \quad \text{for } 0 \le s \le dm^2/16. \tag{8.45}$$

To verify this bound, simply notice that $u + 2ue^{-u^2} + 2e^{-u^2} \le 4$ for $0 \le u \le \frac{1}{2}$. When $s = dm^2/16$, (8.45), and $\tau_2 \le dm^2/2$ are used in Lemma 8.13, we find

$$\hat{\tau} \le \tfrac{3}{4} m^2 d^2 (\log 2 + \tfrac{3}{4} d^{-1}).$$

xxx Improvement over (8.42) by factor $\Theta(\log m)$, but display following (8.43) shows still off by factor $\Theta(d/\log d)$.

8.3 Nash inequalities

xxx For NOTES: Nash vs. Sobolev

A Nash inequality for a chain is an inequality of the form

$$\|g\|_2^{2+\frac{1}{D}} \le C\,[\mathcal{E}(g,g) + \tfrac{1}{T}\|g\|_2^2]\,\|g\|_1^{1/D} \tag{8.46}$$

that holds for some positive constants C, D, and T and for all functions g. We connect Nash inequalities to mixing times in Section 8.3.1, and in Section 8.3.2 we discuss a comparison method for establishing such inequalities.

8.3.1 Nash inequalities and mixing times

A Nash inequality implies a useful bound on the quantity

$$N(t) = \|\mathbf{P}_t\|_{2\to\infty} \tag{8.47}$$

appearing in the mixing time Lemma 8.13. This norm is continuous in t and decreases to $\|\mathbf{E}\|_{2\to\infty} = 1$ as $t \uparrow \infty$. Here is the main result:

Theorem 8.17 *If the Nash inequality (8.46) holds for a continuous-time reversible chain, some $C, D, T > 0$, and all g, then the norm $N(s)$ at (8.47) satisfies*

$$N(t) \le e(DC/t)^D \quad \textit{for } 0 < t \le T.$$

Proof. First note $N(t) = \|\mathbf{P}_t\|_{1\to 2}$ by Lemma 8.12. Thus we seek a bound on $h(t) := \|\mathbf{P}_t g\|_2^2$ independent of g satisfying $\|g\|_1 = 1$; the square root of such a bound will also bound $N(t)$.

Substituting $\mathbf{P}_t g$ for g in (8.46) and utilizing the identity in Lemma 8.10(a) and the fact that $\mathbf{P}_t$ is contractive on L^1, we obtain the differential inequality

$$h(t)^{1+\frac{1}{2D}} \le C\,[-\tfrac{1}{2}h'(t) + \tfrac{1}{T}h(t)], \quad t \ge 0.$$

Writing

$$H(t) := \left[\tfrac{1}{2}Ch(t)e^{-2t/T}\right]^{-1/(2D)},$$

the inequality can be equivalently rearranged to

$$H'(t) \ge \left[2D(C/2)^{1+\frac{1}{2D}}\right]^{-1} e^{\frac{t}{DT}}, \quad t \ge 0.$$

Since $H(0) > 0$, it follows that

$$H(t) \ge T\left[2(C/2)^{1+\frac{1}{2D}}\right]^{-1}\left(e^{\frac{t}{DT}} - 1\right), \quad t \ge 0,$$

or equivalently

$$h(t) \le \left[\frac{T}{C}\left(1 - e^{-t/(DT)}\right)\right]^{-2D}, \quad t \ge 0.$$

But

$$e^{-t/(DT)} \le 1 - \frac{t}{T}(1 - e^{-1/D}) \quad \text{for } 0 < t \le T,$$

so for these same values of t we have

$$\begin{aligned} h(t) &\le \left[\frac{t}{C}\left(1-e^{-1/D}\right)\right]^{-2D} \\ &= e^2\left[\frac{t}{C}\left(e^{1/D}-1\right)\right]^{-2D} \le [e(DC/t)^D]^2, \end{aligned}$$

as desired. ∎

We now return to Lemma 8.13 and, for $t \ge T$, set $s = T$. (Indeed, using the bound on $N(s)$ in Theorem 8.17, this is the optimal choice of s if $T < D\tau_2$.) This gives

xxx NOTE: In next theorem, only need *conclusion* of Theorem 8.17, not hypothesis!

Theorem 8.18 *In Theorem 8.17, if $c \ge 1$ and*

$$t \ge T + \tau_2\left(D\log\left(\frac{DC}{T}\right) + c\right),$$

then $\sqrt{\hat{d}(2t)} \le e^{1-c}$; in particular,

$$\hat{\tau} \le T + \tau_2\left(D\log\left(\frac{DC}{T}\right) + 2\right).$$

The following converse of sorts to Theorem 8.17 will be very useful in conjunction with the comparison method.

Theorem 8.19 *If a continuous time reversible chain satisfies*

$$N(t) \le Ct^{-D} \quad \textit{for } 0 < t \le T,$$

then it satisfies the Nash inequality

$$\|g\|_2^{2+\frac{1}{D}} \le C'\left[\mathcal{E}(g,g) + \tfrac{1}{2T}\|g\|_2^2\right]\|g\|_1^{1/D} \quad \textit{for all } g$$

with

$$C' := 2\left(1+\tfrac{1}{2D}\right)[(1+2D)^{1/2}C]^{1/D} \le 2^{2+\frac{1}{2D}}\,C^{1/D}.$$

xxx Detail in proof to be filled in (I have notes): Show $\mathcal{E}(\mathbf{P}_t g, \mathbf{P}_t g) \downarrow$ as $t \uparrow$, or at least that it's maximized at $t = 0$. Stronger of two statements is equivalent to assertion that $\|\mathbf{P}_t f\|_2^2$ is convex in t.

Proof. As in the proof of Theorem 8.17, we note $N(t) = \|\mathbf{P}_t\|_{1\to 2}$. Hence, for any g and any $0 < t \le T$,

$$\begin{aligned}
\|g\|_2^2 &= \|\mathbf{P}_t g\|_2^2 - \int_{s=0}^{t} \tfrac{d}{ds}\|\mathbf{P}_s g\|_2^2\, ds \\
&= \|\mathbf{P}_t g\|_2^2 + 2\int_{s=0}^{t} \mathcal{E}(\mathbf{P}_s g, \mathbf{P}_s g)\, ds \quad \text{by Lemma 8.10(a)} \\
&\le \|\mathbf{P}_t g\|_2^2 + 2\mathcal{E}(g,g)t \quad \text{xxx see above} \\
&\le 2\mathcal{E}(g,g)t + C^2 t^{-2D}\|g\|_1^2.
\end{aligned}$$

This gives

$$\|g\|_2^2 \le t[2\mathcal{E}(g,g) + \tfrac{1}{T}\|g\|_2^2] + C^2 t^{-2D}\|g\|_1^2$$

for *any* $t > 0$. The righthand side here is convex in t and minimized (for $g \neq 0$) at

$$t = \left(\frac{2DC^2\|g\|_1^2}{2\mathcal{E}(g,g) + T^{-1}\|g\|_2^2}\right)^{1/(2D+1)}.$$

Plugging in this value, raising both sides to the power $1+\frac{1}{2D}$ and simplifying yields the desired Nash inequality. The upper bound for C' is derived with a little bit of calculus. ∎

8.3.2 The comparison method for bounding $N(\cdot)$

In Section 8.1 we compared relaxation times for two chains by comparing Dirichlet forms and variances. The point of this subsection is that the comparison can be extended to the norm function $N(\cdot)$ of (8.47) using Nash inequalities. Then results on $N(\cdot)$ like those in Section 8.2.3 can be used to bound mixing times for other chains on the same state space.

xxx For NOTES?: Can even use different spaces. New paragraph:

To see how this goes, suppose that a benchmark chain is known to satisfy

$$\tilde{N}(t) \le \tilde{C}t^{-\tilde{D}} \quad \text{for } 0 < t \le \tilde{T}. \tag{8.48}$$

By Theorem 8.19, it then satisfies a Nash inequality. The L^1- and L^2-norms appearing in this inequality can be compared in the obvious fashion [cf. (8.18)] and the Dirichlet forms can be compared as in Theorem 8.1. This shows that the chain of interest also satisfies a Nash inequality. But then Theorem 8.17 gives a bound like (8.48) for the chain of interest, and Theorem 8.18 can then be used to bound the L^2 threshold time $\hat{\tau}$.

Here is the precise result; the details of the proof are left to the reader.

Theorem 8.20 (comparison of bounds on $N(\cdot)$) *If a reversible benchmark chain satisfies*

$$\tilde{N}(t) \le \tilde{C} t^{-\tilde{D}} \quad \text{for } 0 < t \le \tilde{T}$$

for constants $\tilde{C}, \tilde{D}, \tilde{T} > 0$, then any other reversible chain on the same state space satisfies

$$N(t) \le e(DC/t)^D \quad \text{for } 0 < t \le T,$$

where, with a and A as defined in Corollary 8.2, and with

$$a' := \max_i (\tilde{\pi}_i / \pi_i),$$

we set

$$\begin{aligned} D &= \tilde{D}, \\ C &= a^{-(2+\frac{1}{D})} a'^{1/D} A \times 2(1+\tfrac{1}{2D})[(1+2D)^{1/2}\tilde{C}]^{1/D} \\ &\le a^{-(2+\frac{1}{D})} a'^{1/D} A \times 2^{2+\frac{1}{2D}} \tilde{C}^{1/D}, \\ T &= \frac{2A}{a'^2}\tilde{T}. \end{aligned}$$

xxx Must correct this slightly. Works for *any* A such that $\tilde{\mathcal{E}} \le A\mathcal{E}$, not just minimal one. This is important since we need a *lower* bound on T but generally only have an *upper* bound on $\tilde{\mathcal{E}}/\mathcal{E}$. The same goes for a' (only need *upper bound* on $\tilde{\pi}_i/\pi_i$): we also need an upper bound on T.

Example 8.21 *Random walk on a d-dimensional grid.*

As in Example 8.5, consider the continuized walk on the d-dimensional grid $I = \{0, \dots, m-1\}^d$. In Example 8.5 we compared the Dirichlet form and variance for this walk to the d-fold product of random walk on the m-path with end self-loops to obtain

$$\tau_2 \le \tilde{\tau}_2 = d(1 - \cos\tfrac{\pi}{m})^{-1} \le \tfrac{1}{2} d m^2; \tag{8.49}$$

using the simple bound $\sqrt{\hat{d}(2t)} \le \pi_*^{-1/2} e^{-t/\tau_2} \le (2n)^{-1/2} \exp\left(-\frac{2t}{dm^2}\right)$ we then get

$$\hat{\tau} \le \tfrac{1}{4} m^2 d[\log(2n) + 2], \tag{8.50}$$

which is of order $m^2 d^2 \log m$. Here we will see how comparing $N(\cdot)$, too, gives a bound of order $m^2 d^2 \log d$. In Example 8.43, we will bring log-Sobolev techniques to bear, too, to lower this bound to order $m^2 d \log d$

xxx which is correct, at least for TV. New paragraph:
Recalling (8.45), we may apply Theorem 8.20 with

$$\tilde{D} = d/4, \quad \tilde{C} = (4dm^2)^{d/4}, \quad \tilde{T} = dm^2/16,$$

and, from the considerations in Example 8.5,

$$a \geq \tfrac{1}{2}, \quad A = 1, \quad a' = 2.$$

xxx See xxx following Theorem 8.20. Same paragraph:
This gives

$$D = d/4, \quad C \leq 2^{6+\frac{10}{d}} dm^2 \leq 2^{16} dm^2, \quad T = dm^2/32.$$

Plugging these into Theorem 8.18 yields

$$\hat{\tau} \leq \tfrac{1}{8} m^2 d^2 \log d + (\tfrac{19}{8} \log 2) m^2 d^2 + \tfrac{33}{32} m^2 d, \tag{8.51}$$

which is $\leq 3m^2 d^2 \log d$ for $d \geq 2$.

Other variants of the walk, including the thinned-grid walk of Example 8.6, can be handled in a similar fashion.

xxx Do moderate growth and local Poincaré? Probably *not*, to keep length manageable. Also, will need to rewrite intro a little, since not doing Δ^2-stuff (in any detail).

8.4 Logarithmic Sobolev inequalities

xxx For NOTES: For history and literature, see ([119], first paragraph and end of Section 1).

xxx For NOTES: Somewhere mention relaxing to *nonreversible* chains.

8.4.1 The log-Sobolev time τ_l

Given a probability distribution π on a finite set I, define

xxx For NOTES: Persi's $\mathcal{L}(g)$ is double ours.

$$\mathcal{L}(g) := \sum_i \pi_i g^2(i) \log(|g(i)|/\|g\|_2) \tag{8.52}$$

for $g \not\equiv 0$, recalling $\|g\|_2^2 = \sum_i \pi_i g^2(i)$ and using the convention $0 \log 0 = 0$. By Jensen's inequality,

$$\mathcal{L}(g) \geq 0, \text{ with equality if and only if } |g| \text{ is constant.}$$

Given a finite, irreducible, reversible Markov chain with stationary distribution π, define the *logarithmic Sobolev* (or *log-Sobolev*) *time* by

xxx For NOTES: Persi's α is $1/(2\tau_l)$.

xxx Note $\tau_l < \infty$. (Show?) See also Corollary 8.27.

$$\tau_l := \sup\{\mathcal{L}(g)/\mathcal{E}(g,g) : g \not\equiv \text{constant}\}. \tag{8.53}$$

Notice the similarity between (8.53) and the extremal characterization of τ_2 (Chapter 3, Theorem yyy:22):

$$\tau_2 = \sup\{\|g\|_2^2/\mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0, \quad g \not\equiv 0\}.$$

We discuss exact computation of τ_l in Section 8.4.3, the behavior of τ_l for product chains in Section 8.4.4, and a comparison method for bounding τ_l in Section 8.4.5. In Section 8.4.2 we focus on the connection between τ_l and mixing times. A first such result asserts that the relaxation time does not exceed the log-Sobolev time:

Lemma 8.22 $\tau_2 \le \tau_l$.

xxx Remarks about how "usually" equality?

xxx For NOTES: Proof from [302], via [119].

Proof. Given $g \not\equiv$ constant and ϵ, let $f := 1 + \epsilon g$. Then, writing $\bar{g} = \sum_i \pi_i g(i)$, and with all asymptotics as $\epsilon \to 0$,

$$\begin{aligned}
\log |f|^2 &= 2\epsilon g - \epsilon^2 g^2 + O(\epsilon^3), \\
\log \|f\|_2^2 &= 2\epsilon \bar{g} + \epsilon^2 \|g\|_2^2 - 2\epsilon^2 \bar{g}^2 + O(\epsilon^3), \\
\log \frac{|f|^2}{\|f\|_2^2} &= 2\epsilon(g - \bar{g}) + \epsilon^2(2\bar{g}^2 - \|g\|_2^2 - g^2) + O(\epsilon^3).
\end{aligned}$$

Also,

$$f^2 = 1 + 2\epsilon g + \epsilon^2 g^2;$$

thus

$$f^2 \log(|f|^2/\|f\|_2^2) = 2\epsilon(g - \bar{g}) + \epsilon^2(3g^2 - \|g\|_2^2 - 4g\bar{g} + 2\bar{g}^2) + O(\epsilon^3)$$

and so

$$\mathcal{L}(f) = \epsilon^2(\|g\|_2^2 - \bar{g}^2) + O(\epsilon^3) = \epsilon^2 \mathrm{var}_\pi\, g + O(\epsilon^3).$$

Furthermore, $\mathcal{E}(f,f) = \epsilon^2 \mathcal{E}(g,g)$; therefore

$$\tau_l \ge \frac{\mathcal{L}(f)}{\mathcal{E}(f,f)} = \frac{\mathrm{var}_\pi\, g}{\mathcal{E}(g,g)} + O(\epsilon).$$

Finish by letting $\epsilon \to 0$ and then taking the supremum over g. ∎

8.4.2 τ_l, mixing times, and hypercontractivity

In this subsection we discuss the connection between the L^2 threshold time parameter

$$\hat{\tau} = \inf\{t > 0 : \sqrt{\hat{d}(2t)} = \max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le e^{-1}\} \tag{8.54}$$

and the log-Sobolev time τ_l. As in Section 8.3, we again consider the fundamental quantity

$$N(s) = \|\mathbf{P}_s\|_{2\to\infty}$$

arising in the bound on $\sqrt{\hat{d}(2t)}$ in Lemma 8.13, and recall from Section 8.3.1 that

$$N(s) \text{ decreases strictly monotonically from } \pi_*^{-1/2} \text{ at } s = 0 \text{ to 1 as } s \uparrow \infty.$$

The function N is continuous. It would be nice (especially for use in conjunction with the comparison technique) if we could characterize, in terms of the Dirichlet form $\mathcal{E}$, the value of s, call it s^*, such that $N(s)$ equals 2 (say), but such a characterization is not presently available.

xxx For NOTES?: A partial result is Theorem 3.9 in [119], taking $q = \infty$.

Open Problem 8.23 Characterize s^* in terms of $\mathcal{E}$.

To carry on along these general lines, it turns out to be somewhat more convenient to substitute use of

$$\|P(X_t \in \cdot) - \pi(\cdot)\|_2 \le \|P(X_0 \in \cdot)\|_{\frac{q}{q-1}} \, \|\mathbf{P}_s\|_{2\to q} \, e^{-(t-s)/\tau_2}, \quad 2 \le q < \infty, \tag{8.55}$$

an immediate consequence of Lemmas 8.11 and 8.12 and (8.34), for use of Lemma 8.13. The reason is that, like $N(s)$, $\|\mathbf{P}_s\|_{2\to q}$ decreases monotonically to 1 as $s \uparrow \infty$; but, unlike $N(s)$, it turns out that

$$\text{for each } q \ge 2, \ \|\mathbf{P}_s\|_{2\to q} \textit{ equals } 1 \text{ for all sufficiently large } s. \tag{8.56}$$

The property (8.56) is called *hypercontractivity*, in light of the facts that, for fixed s, $\mathbf{P}_s$ is a contraction on L^2 and $\|\mathbf{P}_s\|_{2\to q}$ is increasing in q. Let

$$s_q := \inf\{s \ge 0 : \|\mathbf{P}_s\|_{2\to q} \le 1\} = \inf\{s : \|\mathbf{P}_s\|_{2\to q} = 1\};$$

then $s_2 = 0 < s_q$, and we will see presently that $s_q < \infty$ for $q \ge 2$. The following theorem affords a connections with the log-Sobolev time τ_l (and hence with the Dirichlet form $\mathcal{E}$).

Theorem 8.24 *For any finite, irreducible, reversible chain,*

$$\tau_l = \sup_{2<q<\infty} \frac{2s_q}{\log(q-1)}.$$

Proof. The theorem is equivalently rephrased as follows:

$$\|\mathbf{P}_t\|_{2\to q} \le 1 \text{ for all } t \ge 0 \text{ and } 2 \le q < \infty \text{ satisfying } e^{2t/u} \ge q-1 \quad (8.57)$$

if and only if $u \ge \tau_l$. The proof will make use of the generalization

$$\mathcal{L}_q(g) := \sum_i \pi_i |g(i)|^q \log(|g(i)|/\|g\|_q)$$

of (8.52). Fixing $0 \not\equiv g \ge 0$ and $u > 0$, we will also employ the notation

$$\begin{aligned} q(t) &:= 1 + e^{2t/u}, \quad G(t) := \|\mathbf{P}_t g\|_{q(t)}^{q(t)}, \\ F(t) &:= \|\mathbf{P}_t g\|_{q(t)} = \exp\left[\frac{1}{q(t)} \log G(t)\right] \end{aligned} \quad (8.58)$$

for $t \ge 0$.

As a preliminary, we compute the derivative of F. To begin, we can proceed as at the start of the proof of Lemma 8.10(a) to derive

$$G'(t) = -q(t)\,\mathcal{E}\left(\mathbf{P}_t g, (\mathbf{P}_t g)^{q(t)-1}\right) + \frac{q'(t)}{q(t)} E_\pi\left[(\mathbf{P}_t g)^{q(t)} \log\left((\mathbf{P}_t g)^{q(t)}\right)\right].$$

Then

$$\begin{aligned} F'(t) &= \left[\frac{G'(t)}{q(t)G(t)} - \frac{q'(t)\log G(t)}{q^2(t)}\right] F(t) \\ &= F(t)^{-(q(t)-1)} \left[\frac{q'(t)}{q(t)} \mathcal{L}_{q(t)}(\mathbf{P}_t g) - \mathcal{E}\left(\mathbf{P}_t g, (\mathbf{P}_t g)^{q(t)-1}\right)\right]. \end{aligned} \quad (8.59)$$

For the first half of the proof we suppose that (8.57) holds and must prove $\tau_l \le u$, that is, we must establish the log-Sobolev inequality

$$\mathcal{L}(g) \le u\,\mathcal{E}(g,g) \text{ for every } g. \quad (8.60)$$

To establish (8.60) it is enough to consider $0 \not\equiv g > 0$,

xxx Do we actually use $g \ge 0$ here?

since for arbitrary g we have

$$\mathcal{L}(g) = \mathcal{L}(|g|) \text{ and } \mathcal{E}(g,g) \ge \mathcal{E}(|g|,|g|). \quad (8.61)$$

Plugging the specific formula (8.58) for $q(t)$ into (8.59) and setting $t = 0$ gives

$$F'(0) = \|g\|_2^{-1}(u^{-1}\mathcal{L}(g) - \mathcal{E}(g,g)). \tag{8.62}$$

Moreover, since

$$\begin{aligned} F(t) = \|\mathbf{P}_t g\|_{q(t)} &\leq \|\mathbf{P}_t\|_{2\to q(t)}\|g\|_2 \leq \|g\|_2 \quad \text{by (8.57)} \\ &= \|\mathbf{P}_0 g\|_2 = F(0), \end{aligned}$$

the (right-hand) derivative of F at 0 must be nonpositive. The inequality (8.60) now follows from (8.62).

For the second half of the proof, we may assume $u = \tau_l$ and must establish (8.57). For $g \geq 0$, (8.53) and Lemma 8.25 (to follow) give

$$\frac{q}{2}\mathcal{L}_q(g) = \mathcal{L}(g^{q/2}) \leq \tau_l\, \mathcal{E}(g^{q/2}, g^{q/2}) \leq \frac{q^2\tau_l}{4(q-1)}\mathcal{E}(g, g^{q-1}) \tag{8.63}$$

for any $1 < q < \infty$. With $q(t) := 1 + e^{2t/\tau_l}$, we have $q'(t) = \frac{2}{\tau_l}(q(t) - 1)$, and replacing g by $\mathbf{P}_t g$ in (8.63) we obtain

$$\frac{q'(t)}{q(t)}\mathcal{L}_{q(t)}(\mathbf{P}_t g) - \mathcal{E}\left(\mathbf{P}_t g, (\mathbf{P}_t g)^{q(t)-1}\right) \leq 0.$$

From (8.59) we then find $F'(t) \leq 0$ for all $t \geq 0$. Since $F(0) = \|g\|_2$, this implies

$$\|\mathbf{P}_t g\|_{q(t)} \leq \|g\|_2. \tag{8.64}$$

We have assumed $g \geq 0$, but (8.64) now extends trivially to general g, and therefore

$$\|\mathbf{P}_t\|_{2\to q(t)} \leq 1.$$

This gives the desired hypercontractivity assertion (8.57). ∎

Here is the technical Dirichlet form lemma that was used in the proof of Theorem 8.24.

Lemma 8.25 *$\mathcal{E}(g, g^{q-1}) \geq \frac{4(q-1)}{q^2}\mathcal{E}(g^{q/2}, g^{q/2})$ for $g \geq 0$ and $1 < q < \infty$.*

xxx Do we somewhere have the following?:

$$\mathcal{E}(f,g) = \tfrac{1}{2}\sum_i \sum_{j\neq i} \pi_i q_{ij}(f(i) - f(j))(g(i) - g(j)). \tag{8.65}$$

Proof. For any $0 \le a < b$

$$\begin{aligned}\left(\frac{b^{q/2}-a^{q/2}}{b-a}\right)^2 &= \left(\frac{q}{2(b-a)}\int_a^b t^{\frac{q}{2}-1}\,dt\right)^2 \\ &\le \frac{q^2}{4(b-a)}\int_a^b t^{q-2}\,dt = \frac{q^2}{4(q-1)}\frac{b^{q-1}-a^{q-1}}{b-a}.\end{aligned}$$

This shows that

$$(b^{q-1}-a^{q-1})(b-a) \ge \frac{4(q-1)}{q^2}(b^{q/2}-a^{q/2})^2$$

and the lemma follows easily from this and (8.65). ∎

Now we are prepared to bound $\hat{\tau}$ in terms of τ_l.

Theorem 8.26 *(a) If $c \ge 0$, then for any state i with $\pi_i \le e^{-1}$,*

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le e^{1-c} \text{ for } t \ge \tfrac{1}{2}\tau_l \log\log \tfrac{1}{\pi_i} + c\tau_2.$$

(b)

$$\hat{\tau} \le \tfrac{1}{2}\tau_l \log\log \tfrac{1}{\pi_*} + 2\tau_2 \le \tau_l(\tfrac{1}{2}\log\log \tfrac{1}{\pi_*} + 2).$$

Proof. Part (b) follows immediately from (8.54), part (a), and Lemma 8.22. To prove part (a), we begin with (8.55):

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le \pi_i^{-1/q}\,\|\mathbf{P}_s\|_{2\to q}\,e^{-(t-s)/\tau_2}.$$

As in the second half of the proof of Theorem 8.24, let $q = q(s) := 1 + e^{2s/\tau_l}$. Then $\|\mathbf{P}_s\|_{2\to q(s)} \le 1$. Thus

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le \pi_i^{-1/q(s)}\,e^{-(t-s)/\tau_2}, \quad 0 \le s \le t.$$

Choosing $s = \frac{1}{2}\tau_l \log\log(\frac{1}{\pi_i})$ we have $q(s) = 1 + \log(\frac{1}{\pi_i})$ and thus

$$\|\mathbf{P}_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le \exp(1 - \tfrac{t-s}{\tau_2}) \text{ for } t \ge s. \quad ∎$$

We have established the upper bound in the following corollary; for the lower bound, see Corollary 3.11 in [119].

Corollary 8.27

$$\tau_l \le \hat{\tau} \le \tau_l(\tfrac{1}{2}\log\log \tfrac{1}{\pi_*} + 2).$$

Examples illustrating the improvement Corollary 8.27 affords over the similar result (8.7) in terms of τ_2 are offered in Examples 8.37 and 8.40.

8.4.3 Exact computation of τ_l

Exact computation of τ_l is exceptionally difficult—so difficult, in fact, that τ_l is known only for a handful of examples. We present some of these examples in this subsection.

Example 8.28 *Trivial two-state chains.*

We consider a discrete-time chain on $\{0,1\}$ that jumps in one step to stationarity (or, since the value of τ_l is unaffected by continuization, the corresponding continuized chain). Thus $(p_{00}, p_{01}, p_{10}, p_{11}) = (\theta, 1-\theta, \theta, 1-\theta)$ with $\theta = \pi_0 = 1 - \pi_1$. We also assume $0 < \theta \le 1/2$. The claim is that

$$\tau_l = \begin{cases} \frac{\log[(1-\theta)/\theta]}{2(1-2\theta)} & \text{if } \theta \neq 1/2 \\ 1 & \text{if } \theta = 1/2. \end{cases} \tag{8.66}$$

Note that this is continuous and decreasing for $\theta \in (0, 1/2]$.

To prove (8.66), we need to show that $\mathcal{L}(g)/\mathcal{E}(g,g) \le \tau(\theta)$ for every nonconstant g on $\{0,1\}$, where $\tau(\theta)$ denotes the righthand side of (8.66), with equality for some g_0. First suppose $\theta \neq 1/2$. For the inequality, as at (8.60)–(8.61) we may suppose $g \ge 0$ and, by homogeneity,

$$E_\pi g = \theta g(0) + (1-\theta) g(1) = 1.$$

We will work in terms of the single variable

$$x := 1/(g(0) - g(1)),$$

so that

$$g(0) = 1 + \frac{1-\theta}{x}, \quad g(1) = 1 - \frac{\theta}{x}$$

and we must consider $x \in (-\infty, -(1-\theta)] \cup [\theta, \infty)$. We calculate

$$\begin{aligned} \mathcal{E}(g,g) &= \theta(1-\theta)(g(0)-g(1))^2 = \theta(1-\theta)/x^2, \\ \|g\|_2^2 &= \theta\left(1 + \frac{1-\theta}{x}\right)^2 + (1-\theta)\left(1 - \frac{\theta}{x}\right)^2 \\ &= [\theta(x+1-\theta)^2 + (1-\theta)(x-\theta)^2]/x^2 = 1 + \frac{\theta(1-\theta)}{x^2}, \\ \ell(x) &:= \mathcal{L}(g) = \theta\left(1 + \frac{1-\theta}{x}\right)^2 \log\left(1 + \frac{1-\theta}{x}\right) \\ &\qquad + (1-\theta)\left(1 - \frac{\theta}{x}\right)^2 \log\left(1 - \frac{\theta}{x}\right) \end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{2}\theta\left(1+\frac{\theta(1-\theta)}{x^2}\right)\log\left(1+\frac{\theta(1-\theta)}{x^2}\right) \\
= & \Big[\theta(x+1-\theta)^2\log(x+1-\theta)^2+(1-\theta)(x-\theta)^2\log(x-\theta)^2 \\
& -(x^2+\theta(1-\theta))\log(x^2+\theta(1-\theta))\Big]/(2x^2),
\end{aligned}
$$

$$
\begin{aligned}
r(x) &:= 2\theta(1-\theta)\frac{\mathcal{L}(g)}{\mathcal{E}(g,g)} = 2x^2\ell(x) \\
&= \theta(x+1-\theta)^2\log(x+1-\theta)^2+(1-\theta)(x-\theta)^2\log(x-\theta)^2 \\
&\quad -(x^2+\theta(1-\theta))\log(x^2+\theta(1-\theta)).
\end{aligned}
$$

¿From here, a straightforward but very tedious calculus exercise shows that r decreases over $(-\infty,-(1-\theta)]$, with $r(-\infty)=2\theta(1-\theta)$, and that r is strictly unimodal over $[\theta,\infty)$, with $r(\theta)=0$ and $r(\infty)=2\theta(1-\theta)$. It follows that $r(x)$ is maximized over $(-\infty,-(1-\theta)]\cup[\theta,\infty)$ by taking x to be the unique root to

$$
\begin{aligned}
0 &= r'(x) = 4\theta(x+(1-\theta))\log(x+(1-\theta)) \\
&\quad +4(1-\theta)(x-\theta)\log(x-\theta)-2x\log(x^2+\theta(1-\theta))
\end{aligned}
\tag{8.67}
$$

over (θ,∞).

There is on hope for solving (8.67) explicitly unless

$$x^2+\theta(1-\theta)=(x+1-\theta)(x-\theta),$$

i.e., $x=2\theta(1-\theta)/(1-2\theta)$. Fortunately, this *is* a solution to (8.67), and it falls in (θ,∞). The corresponding value of r is $\frac{\theta(1-\theta)}{1-2\theta}\log\frac{1-\theta}{\theta}$, so (8.66) follows, and we learn furthermore that the function g maximizing $\mathcal{L}(g)/\mathcal{E}(g,g)$ is g_0, with $g_0(0)=\frac{1}{2\theta}$ and $g_0(1)=\frac{1}{2(1-\theta)}$.

For $\theta=1/2$, the major change is that now r is *increasing*, rather than unimodal, over $[\theta,\infty)$. Thus $r_{\sup}=2\theta(1-\theta)=1/2$, and (8.66) again follows.

Example 8.29 *Two-state chains.*

Now consider *any* irreducible chain (automatically reversible) on $\{0,1\}$, with stationary distribution π. Without loss of generality we may suppose $\pi_0\le\pi_1$. We claim that

$$
\tau_l=\begin{cases}\frac{\pi_1\log(\pi_1/\pi_0)}{2p_{01}(1-2\pi_0)} & \text{if } \pi_0\neq 1/2 \\ \\ 1/(2p_{01}) & \text{if } \pi_0=1/2.\end{cases}
$$

The proof is easy. The functional $\mathcal{L}(g)$ depends only on π and so is unchanged from Example 8.28, and the Dirichlet form changes from $\mathcal{E}(g,g) = \pi_0\pi_1(g(0)-g(1))^2$ in Example 8.28 to $\mathcal{E}(g,g) = p_{01}(g(0)-g(1))^2$ here.

Remark. Recall from Chapter 5, Example yyy:4 that $\tau_2 = 1/(p_{01}+p_{10}) = \pi_1/p_{01}$. It follows that

$$\frac{\tau_l}{\tau_2} = \begin{cases} \frac{\log(\pi_1/\pi_0)}{2(1-2\pi_0)} & \text{if } 0 < \pi_0 < 1/2 \\ \\ 1 & \text{if } \pi_0 = 1/2 \end{cases}$$

is a continuous and decreasing function of π_0. In particular, we have equality in Lemma 8.22 for a two-state chain if and only if $\pi_0 = 1/2$. Moreover,

$$\tau_l/\tau_2 \sim \tfrac{1}{2}\log(1/\pi_0) \to \infty \text{ as } \pi_0 \to 0.$$

Example 8.30 *Trivial chains.*

The proof of Lemma 8.22 and the result of Example 8.28 can be combined to prove the following result: For the "trivial" chain with $p_{ij} \equiv \pi_j$, the log-Sobolev time τ_l is given (when $\pi_* < 1/2$) by

$$\tau_l = \frac{\log(\frac{1}{\pi_*} - 1)}{2(1-2\pi_*)}.$$

We omit the details, referring the reader to Theorem 5.1 of [119].

As an immediate corollary, we get a reverse-inequality complement to Lemma 8.22:

Corollary 8.31 *For any reversible chain (with $\pi_* < 1/2$, which is automatic for $n \geq 3$),*

$$\tau_l \leq \tau_2 \frac{\log(\frac{1}{\pi_*} - 1)}{2(1-2\pi_*)}.$$

Proof. The result of Example 8.30 can be written

$$\mathcal{L}(g) \leq (\mathrm{var}_\pi\, g)\frac{\log(\frac{1}{\pi_*} - 1)}{2(1-2\pi_*)},$$

and $\mathrm{var}_\pi\, g \leq \tau_2\mathcal{E}(g,g)$ by the extremal characterization of τ_2. ∎

Example 8.32 *The complete graph.*

It follows readily from Example 8.30 that the continuized walk of the complete graph has

$$\tau_l = \frac{(n-1)\log(n-1)}{2(n-2)} \sim \frac{1}{2}\log n.$$

Since $\tau_2 = (n-1)/n$, equality holds in Corollary 8.31 for this example.

xxx Move the following warning to follow Corollary 8.27, perhaps?
Warning. Although the ratio of the upper bound on $\hat{\tau}$ to lower bound in Corollary 8.27 is smaller than that in (8.7), the upper bound in Corollary 8.27 is sometimes of larger order of magnitude than the upper bound in (8.7). For the complete graph, (8.7) says

$$\tfrac{n-1}{n} \le \hat{\tau} \le \tfrac{n-1}{n}(\tfrac{1}{2}\log n + 1)$$

and Corollary 8.27 yields

$$(1+o(1))\tfrac{1}{2}\log n \le \hat{\tau} \le (1+o(1))\tfrac{1}{4}(\log n)(\log\log n),$$

while, from Chapter 5, yyy:(33) it follows that

$$\hat{\tau} = \tfrac{1}{2}\log n + O\left(\frac{\log n}{n}\right).$$

As another example, the product chain development in the next subsection together with Example 8.29 will give τ_l exactly for the d-cube. On the other hand, the exact value of τ_l is unknown even for many of the simplest examples in Chapter 5. For instance,

Open Problem 8.33 Calculate τ_l for the n-cycle (Chapter 5 Example yyy:7) when $n \ge 4$.

xxx For NOTES: $n = 3$ is complete graph K_3, covered by Example 8.32. ($\tau_l = \log 2$ for $n = 3$.)

Notwithstanding Open Problem 8.33, the value of τ_l is known *up to multiplicative constants*. Indeed, it is shown in Section 4.2 in [119] that

$$\frac{1}{4\pi^2}n^2 \le \tau_l \le \frac{25}{16\pi^2}n^2.$$

Here is a similar result we will find useful later in dealing with our running example of the grid.

Example 8.34 *The m-path with end self-loops.*

For this example, discussed above in Example 8.16, we claim

$$\frac{2}{\pi^2}m^2 \le \tau_l \le m^2.$$

The lower bound is easy, using Lemma 8.22:

$$\tau_l \ge \tau_2 = (1-\cos(\pi/m))^{-1} \ge \frac{2}{\pi^2}m^2.$$

For the upper bound we use Corollary 8.27 and estimation of $\hat{\tau}$. Indeed, in Example 8.16 it was shown that

$$\hat{d}(2t) = N^2(t) - 1 \le \left[1 + (4t/m^2)^{-1/2}\right] \exp(-4t/m^2), \quad t > 0.$$

Substituting $t = m^2$ gives $\sqrt{\hat{d}(2t)} \le \sqrt{3/2}\, e^{-2} < e^{-1}$, so $\tau_l \le \hat{\tau} \le m^2$.

xxx P.S. Persi (98/07/02) points out that H. T. Yau showed $\tau_l = \Theta(n \log n)$ for random transpositions by combining $\tau_l \ge \tau_2$ (Lemma 8.22) and $\tau_l \le \mathcal{L}(g_0)/\mathcal{E}(g_0, g_0)$ with $g_0 =$ delta function. I have written notes generalizing and discussing this and will incorporate them into a later version.

8.4.4 τ_l and product chains

xxx Remind reader of definition of product chain in continuous time given in Chapter 4 Section yyy:6.2.

xxx Motivate study as providing benchmark chains for comparison method.

xxx Recall from Chapter 4, yyy:(42):

$$\tau_2 = \max(\tau_2^{(1)}, \tau_2^{(2)}). \tag{8.68}$$

xxx Product chain has transition rates equal (off diagonal) to

$$q_{(i_1,i_2),(j_1,j_2)} = \begin{cases} q^{(1)}_{i_1,j_1} & \text{if } i_1 \ne j_1 \text{ and } i_2 = j_2 \\ q^{(2)}_{i_2,j_2} & \text{if } i_1 = j_1 \text{ and } i_2 \ne j_2 \\ 0 & \text{otherwise.} \end{cases} \tag{8.69}$$

xxx Dirichlet form works out very nicely for products:

Lemma 8.35

$$\mathcal{E}(g,g) = \sum_{i_2} \pi^{(2)}_{i_2} \mathcal{E}^{(1)}(g(\cdot,i_2), g(\cdot,i_2)) + \sum_{i_1} \pi^{(1)}_{i_1} \mathcal{E}^{(2)}(g(i_1,\cdot), g(i_1,\cdot)).$$

Proof. This follows easily from (8.69) and the definition of $\mathcal{E}$ in Chapter 3 Section yyy:6.1 (cf. (68)). ∎

The analogue of (8.68) for the log-Sobolev time is also true:

xxx For NOTES?: Can give analagous proof of (8.68): see my notes, page 8.4.24A.

Theorem 8.36 *For a continuous-time product chain,*

$$\tau_l = \max(\tau^{(1)}_l, \tau^{(2)}_l).$$

Proof. The keys to the proof are Lemma 8.35 and the following "law of total $\mathcal{L}$-functional." Given a function $g \not\equiv 0$ on the product state space $I = I_1 \times I_2$, define a function $G_2 \not\equiv 0$ on I_2 by

$$G_2(i_2) := \|g(\cdot,i_2)\|_2 = \left(\sum_{i_1} \pi_{i_1} g^2(i_1,i_2) \right)^{1/2}.$$

Then

$$\begin{aligned} \mathcal{L}(g) &= \sum_{i_1,i_2} \pi_{i_1,i_2}\, g^2(i_1,i_2) \left[\log(|g(i_1,i_2)|/G_2(i_2)) + \log(G_2(i_2)/\|g\|_2) \right] \\ &= \sum_{i_2} \pi^{(2)}_{i_2} \mathcal{L}^{(1)}(g(\cdot,i_2)) + \mathcal{L}^{(2)}(G_2), \end{aligned}$$

where we have used

$$\|G_2\|_2^2 = \|g\|_2^2.$$

Thus, using the extremal characterization (definition) (8.53) of $\tau^{(1)}_l$ and $\tau^{(2)}_l$,

$$\mathcal{L}(g) \le \tau^{(1)}_l \sum_{i_2} \pi^{(2)}_{i_2} \mathcal{E}^{(1)}(g(\cdot,i_2), g(\cdot,i_2)) + \tau^{(2)}_l \mathcal{E}^{(2)}(G_2,G_2). \tag{8.70}$$

But from

$$|G_2(j_2) - G_2(i_2)| = |\|g(\cdot,j_2)\|_2 - \|g(\cdot,i_2)\|_2| \le \|g(\cdot,j_2) - g(\cdot,i_2)\|_2$$

follows

$$\mathcal{E}^{(2)}(G_2,G_2) \le \sum \pi^{(1)}_{i_1} \mathcal{E}^{(2)}(g(i_1,\cdot), g(i_1,\cdot)). \tag{8.71}$$

From (8.70), (8.71), Lemma 8.35, and the extremal characterization of τ_l we conclude $\tau_l \le \max(\tau_l^{(1)}, \tau_l^{(2)})$. Testing on functions that depend only on one of the two variables shows that $\tau_l = \max(\tau_l^{(1)}, \tau_l^{(2)})$. ∎

Theorem 8.36 extends in the obvious fashion to higher-dimensional products.

Example 8.37 *The d-cube.*

The continuized walk on the d-cube (Chapter 5, Example yyy:15) is simply the product of d copies of the continuized walk on the 2-path, each run at rate $1/d$. Therefore, since the log-Sobolev time for the 2-path equals $1/2$ by Example 8.29, the corresponding time for the d-cube is

$$\tau_l = d/2 = \tau_2.$$

¿From this and the upper bound in Corollary 8.27 we can deduce

$$\hat{\tau} \le \tfrac{1}{4} d \log d + (1 - \tfrac{1}{4} \log \tfrac{1}{\log 2}) d.$$

As discussed in this chapter's introduction, this bound is remarkably sharp and improves significantly upon the analogous bound that uses only knowledge of τ_2. xxx Recall corrections marked on pages 8.2.11–12 of my notes.

8.4.5 The comparison method for bounding τ_l

In Section 8.1 we compared relaxation times for two chains by using the extremal characterization and comparing Dirichlet forms and variances. For comparing variances, we used the characterization

$$\mathrm{var}_\pi\, g = \min_{c \in \mathbf{R}} \|g - c\|_2.$$

To extend the comparison method to log-Sobolev times, we need the following similar characterization of $\mathcal{L}$.

xxx For NOTES: Cite [181].

Lemma 8.38 *The functional $\mathcal{L}$ in (8.52) satisfies*

$$\mathcal{L}(g) = \min_{c>0} \sum_i \pi_i L(g(i), c), \quad g \not\equiv 0, \tag{8.72}$$

with

$$L(g(i), c) := g^2(i) \log(|g(i)|/c) - \tfrac{1}{2}(g^2(i) - c^2) \ge 0. \tag{8.73}$$

Proof. We compute

$$\begin{aligned} f(c) &:= 2\sum_i \pi_i L(g(i), c^{1/2}) = E_\pi(g^2 \log |g|^2) - \|g\|_2^2 \log c - \|g\|_2^2 + c, \\ f'(c) &= 1 - c^{-1}\|g\|_2^2, \quad f''(c) = c^{-2}\|g\|_2^2 > 0. \end{aligned}$$

Thus f is strictly convex and minimized by the choice $c = \|g\|_2^2$, and so

$$\min_{c>0} \sum_i \pi_i L(g(i), c) = \tfrac{1}{2} \min_{c>0} f(c) = \tfrac{1}{2} f(\|g\|_2^2) = \mathcal{L}(g).$$

This proves (8.72). Finally, applying the inequality

$$x \log(x/y) - (x - y) \geq 0 \quad \text{for all } x \geq 0,\ y > 0$$

to $x = g^2(i)$ and $y = c^2$ gives the inequality in (8.73). ∎

Now it's easy to see how to compare log-Sobolev times, since, adopting the notation of Section 8.1, Lemma 8.38 immediately yields the analogue

$$\mathcal{L}(g) \leq \tilde{\mathcal{L}}(g) \max_i(\pi_i/\tilde{\pi}_i)$$

of (8.18). In the notation of Corollary 8.2, we therefore have

Corollary 8.39 (comparison of log-Sobolev times)

$$\tau_l \leq \frac{A}{a}\tilde{\tau}_l.$$

Example 8.40 *Random walk on a d-dimensional grid.*

xxx Remarked in Example 8.34 that $\tau_l \leq m^2$ for m-path with end self-loops.

xxx So by Theorem 8.36, benchmark product chain has $\tilde{\tau}_l \leq dm^2$.

Recalling $A \leq 1$ and $a \geq 1/2$ from Example 8.5, we therefore find

$$\tau_l \leq 2m^2 d \tag{8.74}$$

for random walk on the grid. Then Theorem 8.26(b) gives

$$\hat{\tau} \leq m^2 d(\log\log(2n) + 4),$$

which is of order $m^2 d(\log d + \log\log m)$. This is an improvement on the τ_2-only bound $O(m^2 d^2 \log m)$ of (8.50) and may be compared with the Nash-based bound $O(m^2 d^2 \log d)$ of (8.51). In Example 8.43 we will combine Nash-inequality and log-Sobolev techniques to get a bound of order $m^2 d \log d$

xxx right for TV.

8.5 Combining the techniques

To get the maximum power out of the techniques of this chapter, it is sometimes necessary to combine the various techniques. Before proceeding to a general result in this direction, we record a simple fact. Recall (8.36).

Lemma 8.41 *If q and q^* are conjugate exponents with $2 \le q \le \infty$, then*

$$\|f\|_{q^*} \le \|f\|_1^{1-\frac{2}{q}} \|f\|_2^{\frac{2}{q}} \quad \textit{for all } f.$$

Proof. Apply Hölder's inequality

$$\|gh\|_1 \le \|g\|_p \, \|h\|_{p^*}$$

with

$$g = |f|^{(q-2)/(q-1)}, \quad h = |f|^{2/(q-1)}, \quad p = \frac{q-1}{q-2}. \quad \blacksquare$$

Theorem 8.42 *Suppose that a continuous-time reversible chain satisfies*

$$N(t) \le Ct^{-D} \quad \textit{for } 0 < t \le T \tag{8.75}$$

for some constants C, T, D satisfying $CT^{-D} \ge e$. If $c \ge 0$, then

$$\sqrt{\hat{d}(2t)} = \max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le e^{2-c}$$

for

$$t \ge T + \tfrac{1}{2}\tau_l \log\left[\log(CT^{-D}) - 1\right] + c\tau_2,$$

where τ_2 is the relaxation time and τ_l is the log-Sobolev time.

Proof. ¿From Lemma 8.11 and a slight extension of (8.34), for any $s, t, u \ge 0$ and any initial distribution we have

$$\|P(X_{s+t+u} \in \cdot) - \pi(\cdot)\|_2 \le \|P(X_s \in \cdot)\|_{q^*} \, \|\mathbf{P}_t\|_{q^* \to 2} \, e^{-u/\tau_2}$$

for any $1 \le q^* \le \infty$. Choose $q = q(t) = 1 + e^{2t/\tau_l}$ and q^* to be its conjugate. Then, as in the proof of Theorem 8.26(a),

$$\|\mathbf{P}_t\|_{q^* \to 2} = \|\mathbf{P}_t\|_{2 \to q} \le 1.$$

According to Lemma 8.41, (8.39), and (8.75), if $0 < s \le T$ then

$$\|P(X_s \in \cdot)\|_{q^*} \le \|P(X_s \in \cdot)\|_2^{2/q} \le N(s)^{2/q} \le (Cs^{-D})^{2/q}.$$

Now choose $s = T$. Combining everything so far,

$$\sqrt{\hat{d}(2(T+t+u))} \le (CT^{-D})^{2/q(t)}\, e^{-u/\tau_2} \quad \text{for } t, u \ge 0.$$

The final idea is to choose t so that the first factor is bounded by e^2. ¿From the formula for $q(t)$, the smallest such t is

$$\tfrac{1}{2}\tau_l \log\left[\log(CT^{-D}) - 1\right].$$

With this choice, the theorem follows readily. ■

Example 8.43 *Random walk on a d-dimensional grid.*

Return one last time to the walk of interest in Example 8.5. Example 8.21 showed that (8.75) holds with

$$D = d/4, \quad C = e(2^{14}d^2m^2)^{d/4} = e(2^7 dm)^{d/2}, \quad T = dm^2/32.$$

Also recall $\tau_2 \le \frac{1}{2}dm^2$ from (8.49) and $\tau_l \le 2dm^2$ from Example 8.40. Plugging these into Theorem 8.42 with $c = 2$ yields

$$\hat{\tau} \le \tfrac{49}{32}m^2 d \log[\tfrac{1}{4}d\log d + \tfrac{19}{4}d\log 2], \text{ which is } \le 5m^2 d\log d \text{ for } d \ge 2.$$

xxx Finally of right order of magnitude.

8.6 Notes on Chapter 8

xxx The chapter notes go here. Currently, they are interspersed throughout the text.

xxx Also cite and plug [304].

Chapter 9

A Second Look at General Markov Chains (April 21, 1995)

In the spirit of Chapter 2, this is an unsystematic treatment of scattered topics which are related to topics discussed for reversible chains, but where reversibility plays no essential role. Section 9.1 treats constructions of stopping times with various optimality properties. Section 9.2 discusses random spanning trees associated with Markov chains, the probabilistic elaboration of "the matrix-tree theorem". Section 9.3 discusses self-verifying algorithms for sampling from a stationary distribution. Section 9.4 discusses "reversiblizations" of irreversible chains. Section 9.5 gives an example to show that the nonasymptotic interpretation of relaxation time, so useful in the reversible setting, may fail completely in the general case. At first sight these topics may seem entirely unrelated, but we shall see a few subtle connections.

Throughout the chapter, our setting is a finite irreducible discrete-time Markov chain (X_n) with transition matrix $\mathbf{P} = (p_{ij})$.

9.1 Minimal constructions and mixing times

Chapter 4 Theorem yyy involved three mixing time parameters; τ_1 related to variation distance to stationarity, $\tau_1^{(1)}$ related to "separation" from stationarity, and $\tau_1^{(2)}$ related to stationary times (see below). In Chapter 4 these parameters were defined under worst-case initial distributions, and our focus was on "equivalence" of these parameters for reversible chains. Here

we discuss underlying "exact" results. Fix an initial distribution μ. Then associated with each notion of mixing, there is a corresponding construction of a minimal random time T, stated in Theorems 9.1 - 9.3 below.

xxx randomized stopping times

Call a stopping time T a *strong stationary time* if

$$P_\mu(X_t = j, T = t) = \pi_j P_\mu(T = t) \text{ for all } j, t \tag{9.1}$$

i.e. if X_T has distribution π and is independent of T. Call a stopping time T a *stationary time* if

$$P_\mu(X_T = j) = \pi_j \text{ for all } j. \tag{9.2}$$

Call a random time T a *coupling time* if we can construct a joint distribution $((X_t, Y_t); t \geq 0)$ such that (X_t) is the chain with initial distribution μ, (Y_t) is the stationary chain, and $X_t = Y_t, t \geq T$. (A coupling time need not be a stopping time, even w.r.t. the joint process; this is the almost the only instance of a random time which is not a stopping time that we encounter in this book.)

Recall from yyy the notion of *separation* of θ from π:

$$\text{sep}(\theta) \equiv \min\{u : \theta_j \geq (1-u)\pi_j \ \forall j\}.$$

Write $\text{sep}_\mu(t)$ for the separation at time t when the initial distribution was μ:

$$\text{sep}_\mu(t) = \min\{u : P_\mu(X_t = j) \geq (1-u)\pi_j \ \forall j\}.$$

Similarly write $\text{vd}_\mu(t)$ for the variation distance from stationarity at time t:

$$\text{vd}_\mu(t) = \frac{1}{2}\sum_j |P_\mu(X_t = j) - \pi_j|.$$

Theorem 9.1 *Let T be any strong stationary time for the μ-chain. Then*

$$\text{sep}_\mu(t) \leq P_\mu(T > t) \text{ for all } t \geq 0. \tag{9.3}$$

Moreover there exists a minimal *strong stationary time T for which*

$$\text{sep}_\mu(t) = P_\mu(T > t) \text{ for all } t \geq 0. \tag{9.4}$$

Theorem 9.2 *For any coupling time T,*

$$\text{vd}_\mu(t) \leq P_\mu(T > t) \text{ for all } t \geq 0.$$

Moreover there exists a minimal *coupling time T for which*

$$\text{vd}_\mu(t) = P_\mu(T > t) \text{ for all } t \geq 0.$$

Theorem 9.3 *For any stationary time T,*

$$E_\mu T \geq \max_j (E_\mu T_j - E_\pi T_j). \tag{9.5}$$

Moreover there exist mean-minimal *stationary times T for which*

$$E_\mu T = \max_j (E_\mu T_j - E_\pi T_j). \tag{9.6}$$

In each case, the first assertion is immediate from the definitions, and the issue is to carry out a construction of the required T. Despite the similar appearance of the results, attempts to place them all in a common framework have not been fruitful. We will prove Theorems 9.1 and 9.3 below, and illustrate with examples. These two proofs involve only rather simple "greedy" constructions. We won't give the proof of Theorem 9.2 (the construction is usually called the *maximal coupling*: see Lindvall [233]) because the construction is a little more elaborate and the existence of the minimal coupling time is seldom useful, but on the other hand the *coupling inequality* in Theorem 9.2 will be used extensively in Chapter 14. In the context of Theorems 9.1 and 9.2 the minimal times T are clearly unique in distribution, but in Theorem 9.3 there will generically be many mean-minimal stationary times T with different distributions.

9.1.1 Strong stationary times

For any stopping time T, define

$$\theta_j(t) = P_\mu(X_t = j, T \geq t), \quad \sigma_j(t) = P_\mu(X_t = j, T = t). \tag{9.7}$$

Clearly these vectors satisfy

$$0 \leq \sigma(t) \leq \theta(t), \quad (\theta(t) - \sigma(t))\mathbf{P} = \theta(t+1) \ \forall t; \quad \theta_0 = \mu. \tag{9.8}$$

Conversely, given $(\theta(t), \sigma(t); t \geq 0)$ satisfying (9.8), we can construct a randomized stopping time T satisfying (9.7) by declaring that $P(T = t | X_t = j, T \geq t, X_s, s < t) = \sigma_j(t)/\theta_j(t)$. The proofs of Theorems 9.1 and 9.3 use different definitions of vectors satisfying (9.8).

Proof of Theorem 9.1. A particular sequence $(\theta(t), \sigma(t); t \geq 0)$ can be specified inductively by (9.8) and

$$\sigma(t) = r_t \pi \ , \quad \text{where } r_t = \min_j \theta_j(t)/\pi_j. \tag{9.9}$$

The associated stopping time satisfies

$$P_\mu(X_t = j, T = t) = \sigma_j(t) = r_t \pi_j$$

and so is a strong stationary time with $P_\mu(T = t) = r_t$. One can now verify inductively that

$$P_\mu(X_t \in \cdot) = \theta(t) + P_\mu(T \le t-1) \ \cdot pi$$

and so the separation is

$$\mathrm{sep}_\mu(t) = 1 - \min_j \frac{P_\mu(X_t = j)}{\pi_j} = P_\mu(T \ge t) - r_t = P_\mu(T > t).$$

9.1.2 Stopping times attaining a specified distribution

For comparison with the other two results, we stated Theorem 9.3 in terms of stopping times at which the stationary distribution is attained, but the underlying result (amplified as Theorem 9.4) holds for an arbitrary target distribution ρ. So fix ρ as well as the initial distribution μ. Call a stopping time T *admissible* if $P_\mu(X_T \in \cdot) = \rho$. Write $\bar{t}(\mu, \sigma)$ for the *inf* of $E_\mu T$ over all admissible stopping times T.

Theorem 9.4 *(a)* $\bar{t}(\mu, \sigma) = \max_j (E_\mu T_j - E_\rho T_j)$.

(b) The "filling scheme" below defines an admissible stopping time such that $E_\mu T = \bar{t}(\mu, \sigma)$.

(c) Any admissible stopping time T with the property

$$\exists\ k \text{ such that } P_\mu(T \le T_k) = 1. \tag{9.10}$$

satisfies $E_\mu T = \bar{t}(\mu, \sigma)$.

Part (c) is rather remarkable, and can be rephrased as follows. Call a state k with property (9.10) a *halting state* for the stopping time T. In words, the chain must stop if and when it hits a halting state. Then part (c) asserts that, to verify that an admissible time T attains the minimum $\bar{t}(\mu, \rho)$, it suffices to show that there exists some halting state. In the next section we shall see this is very useful in simple examples.

Proof. The greedy construction used here is called a *filling scheme*. Recall from (9.7) the definitions

$$\theta_j(t) = P_\mu(X_t = j, T \ge t), \quad \sigma_j(t) = P_\mu(X_t = j, T = t).$$

Write also $\Sigma_j(t) = P_\mu(X_T = j, T \le t)$. We now define $(\theta(t), \sigma(t); t \ge 0)$ and the associated stopping time $\bar{T}$ inductively via (9.8) and

$$\begin{aligned} \sigma_j(t) &= 0 \text{ if } \Sigma_j(t-1) = \rho_j \\ &= \theta_t \text{ if } \Sigma_j(t-1) + \theta_j(t) \le \rho_j \\ &= \rho_j - \Sigma_j(t-1) \text{ otherwise.} \end{aligned}$$

In words, we stop at the current state (j, say) provided our "quota" ρ_j for the chance of stopping at j has not yet been filled. Clearly

$$\Sigma_j(t) \le \rho_j \ \forall j \ \forall t. \tag{9.11}$$

We now claim that $\bar{T}$ satisfies property (9.10). To see this, consider

$$t_j \equiv \min\{t : \Sigma_j(t) = \rho_j\} \le \infty.$$

Then (9.10) holds by construction for any k such that $t_k = \max_j t_j \le \infty$. In particular, $\bar{T} \le T_k < \infty$ a.s. and then by (9.11) $P_\mu(X_{\bar{T}} \in \cdot) = \lim t \to \infty \Sigma(t) = \rho$. So $\bar{T}$ is an admissible stopping time.

Remark. Generically we expect $t_j = \infty$ for exactly one state j, though other possibilities may occur, e.g. in the presence of symmetry.

Now consider an arbitrary admissible stopping time T, and consider the associated *occupation measure* $\mathbf{x} = (x_j)$:

$$x_j \equiv E_\mu(\text{number of visits to } j \text{ during times } 0, 1, \ldots, T-1).$$

We shall show

$$x_j + \rho_j = \mu_j + \sum_i x_i p_{ij} \ \forall j. \tag{9.12}$$

Indeed, by counting the number of visits during $0, 1, \ldots, T-1, T$ in two ways,

$$x_j + \rho_j = \mu_j + E_\mu(\text{number of visits to } j \text{ during } 1, 2, \ldots, T).$$

Chapter 2 Lemma yyy showed the (intuitively obvious) fact

$$x_i p_{ij} = E_\mu(\text{ number of transitions } i \to j \text{ starting before time } T).$$

So summing over i,

$$\sum_i x_i p_{ij} = E_\mu(\text{number of visits to } j \text{ during } 1, 2, \ldots, T)$$

and (9.12) follows.

Write $\bar{\mathbf{x}}$ for the occupation measure associated with the stopping time $\bar{T}$ produced by the filling scheme. By (9.10), $\min_k \bar{x}_k = 0$. If $\mathbf{x}$ and $\mathbf{x}'$ are solutions of (9.12) then the difference $\mathbf{d} = \mathbf{x} - \mathbf{x}'$ satisfies $\mathbf{d} = \mathbf{dP}$ and so is a multiple of the stationary distribution π. In particular, if $\mathbf{x}$ is the occupation measure for some arbitrary admissible time T, then

$$\mathbf{x} \geq \bar{\mathbf{x}}, \text{ with equality iff } \min_k x_k = 0.$$

Since $E_\mu T = \sum_i x_i$, we have established parts (b) and (c) of the theorem, and

$$\bar{t}(\mu, \sigma) = \sum_i \bar{x}_i.$$

To prove (a), choose a state k such that $\bar{x}_k = 0$, that is such that $\bar{T} \leq T_k$. Then $E_\mu T_k = E_\mu \bar{T} + E_\rho T_k$ and hence $\bar{t}(\mu, \sigma) \leq \max_j (E_\mu T_j - E_\rho T_j)$. But for any admissible stopping time T and any state j

$$E_\mu T_j \leq E_\mu T + E_\rho T_j$$

giving the reverse inequality $\bar{t}(\mu, \sigma) \geq \max_j (E_\mu T_j - E_\rho T_j)$. □

Corollary 9.5 *The minimal strong stationary time has mean $\bar{t}(\mu, \pi)$, i.e. is mean-minimal amongst all not-necessarily-strong stationary times, iff there exists a state k such that*

$$P_\mu(X_t = k)/\pi_k = \min_j P_\mu(X_t = j)/\pi_j \ \forall t.$$

Proof. From the construction of the minimal strong stationary time, this is the condition for k to be a halting state.

9.1.3 Examples

Example 9.6 *Patterns in coin-tossing.*

Recall Chapter 2 Example yyy: (X_t) is the chain on the set $\{H, T\}^n$ of n-tuples $i = (i_1, \ldots, i_n)$. Start at some arbitrary initial state $j = (j_1, \ldots, j_n)$. Here the deterministic stopping time "$T = n$" is a strong stationary time. Now a state $k = (k_1, \ldots, k_n)$ will be a halting state provided it does not overlap j, that is provided there is no $1 \leq u \leq n$ such that $(j_u, \ldots, j_n) = (k_1, \ldots, k_{n+u-1})$. But the number of overlapping states is at most $1 = 2 + 2^2 + \ldots + 2^{n-1} = 2^n - 1$, so there exists a non-overlapping state, i.e. a halting state. So ET attains the minimum ($= n$) of $\bar{t}(j, \pi)$ over all stationary times (and not just over all *strong* stationary times).

Example 9.7 *Top-to-random card shuffling.*

Consider the following scheme for shuffling an n-card deck: the top card is removed, and inserted in one of the n possible positions, chosen uniformly at random. Start in some arbitrary order. Let T be the first time that the card which was originally second-from-bottom has reached the top of the deck. Then it is not hard to show (Diaconis [112] p. 177) that $T+1$ is a strong stationary time. Now any configuration in which the originally-bottom card is the top card will be a halting state, and so $T+1$ is mean-minimal over all stationary times. Here $E(T+1) = 1 + \sum_{m=2}^{n-1} \frac{n}{m} = n(h_n - 1)$.

Example 9.8 *The winning streak chain.*

In a series of games which you win or lose independently with chance $0 < c < 1$, let $\hat{X}_t$ be your current "winning streak", i.e. the number of games won since your last loss. For fixed n, the truncated process $X_t = \min(X_t, n-1)$ is the Markov chain on states $\{0, 1, 2, \ldots, n-1\}$ with transition probabilities

$$p(i,0) = 1-c, \quad p(i, \min(i+1, n-1)) = c; \ 0 \le i \le n-1$$

and stationary distribution

$$\pi_i = (1-c)c^i, \ 0 \le i \le n-2; \quad \pi_{n-1} = c^{n-1}.$$

We present this chain, started at 0, as an example where it is easy to see there are different mean-minimal stationary times T. We'll leave the simplest construction until last – can you guess it now? First consider T_J, where J has the stationary distribution. This is a stationary time, and $n-1$ is a halting state, so it is mean-minimal. Now it is easy to show

$$E_0 T_j = \frac{1}{(1-c)c^j} \ - \frac{1}{1-c}, 1 \le j \le n-1.$$

(Slick proof: in the not-truncated chain, Chapter 2 Lemma yyy says $1 = E_j($ number of visits to j before $T_0) = \pi_j(E_j T_0 + E_0 T_j) = \pi_j(1/(1-c) + E_0 T_j)$.) So

$$\bar{t}(0,\pi) = E_0 T_J = \sum_{j \ge 1} \pi_j E_0 T_j = n - 2 + \frac{\pi_{n-1}}{(1-c)c^{n-1}} - \frac{1-\pi_0}{1-c} = n-1.$$

Here is another stopping time T which is easily checked to attain the stationary distribution, for the chain started at 0. With chance $1-c$ stop at time

0. Otherwise, run the chain until either hitting $n-1$ (in which case, stop) or returning to 0. In the latter case, the return to 0 occurs as a transition to 0 from some state $M \geq 0$. Continue until first hitting $M+1$, then stop. Again $n-1$ is a halting state, so this stationary time also is mean-minimal. Of course, the simplest construction is the deterministic time $T = n-1$. This is a strong stationary time (the winning streak chain is a function of the patterns in coin tossing chain), and again $n-1$ is clearly a halting state. Thus $\bar{t}(0,\pi) = n-1$ without needing the calculation above.

Remark. One could alternatively use Corollary 9.5 to show that the strong stationary times in Examples 9.6 and 9.7 are mean-minimal stationary times. The previous examples are atypical: here is a more typical example in which the hypothesis of Corollary 9.5 is not satisfied and so no mean-optimal stationary time is a strong stationary time.

Example 9.9 *xxx needs a name!*

Chapter 2 Example yyy can be rewritten as follows. Let (U_t) be independent uniform on $\{0,1,\ldots,n-1\}$ and let (A_t) be independent events with $P(A_t) = a$. Define a chain X on $\{0,1,\ldots,n-1\}$ by

$$\begin{aligned} X_{t+1} &= U_{t+1} \text{ on } A_t^c \\ &= X_t + 1 \bmod n \text{ on } A_t. \end{aligned}$$

The stationary distribution is the uniform distribution. Take $X_0 = 0$. Clearly $T \equiv \min\{t : A_t \text{ occurs }\}$ is a strong stationary time, and $ET = 1/(1-a)$, and it is easy to see that T is the minimal strong stationary time. But T is not a mean-minimal stationary time. The occupation measure $\mathbf{x}$ associated with T is such that $\min_j x_j = x_{n-1} = a^{n-1} + a^{2n-1} + \ldots = a^{n-1}/(1-a^n)$, and so the occupation measure $\bar{\mathbf{x}}$ associated with a mean-minimal stationary time is $\bar{\mathbf{x}} = \mathbf{x} - \frac{a^{n-1}}{1-a^n}\pi$, and so $\bar{t}(0,\pi) = \frac{1}{1-a} - \frac{a^{n-1}}{1-a^n}$.

9.2 Markov chains and spanning trees

9.2.1 General Chains and Directed Weighted Graphs

Let's jump into the details and defer the discussion until later. Consider a finite irreducible discrete-time Markov chain (X_n) with transition matrix $\mathbf{P} = (p_{vw})$, and note we are not assuming reversibility. We can identify $\mathbf{P}$ with a weighted directed graph, which has (for each (v,w) with $p_{vw} > 0$) a directed edge (v,w) with weight p_{vw}. A *directed spanning tree* $\mathbf{t}$ is a

spanning tree with one vertex distinguished as the root, and with each edge $e = (v, w)$ of $\mathbf{t}$ regarded as being directed towards the root. Write $\mathcal{T}$ for the set of directed spanning trees. For $\mathbf{t} \in \mathcal{T}$ define

$$\bar{\rho}(\mathbf{t}) \equiv \prod_{(v,w)\in\mathbf{t}} p_{vw}.$$

Normalizing gives a probability distribution ρ on $\mathcal{T}$:

$$\rho(\mathbf{t}) \equiv \frac{\bar{\rho}(\mathbf{t})}{\sum_{\mathbf{t}'\in\mathcal{T}} \bar{\rho}(\mathbf{t}')}.$$

Now fix n and consider the stationary Markov chain $(X_m : -\infty < m \le n)$ run from time minus infinity to time n. We now use the chain to construct a random directed spanning tree $\mathbf{T}_n$. The root of $\mathbf{T}_n$ is X_n. For each $v \ne X_n$ there was a final time, L_v say, before n that the chain visited v:

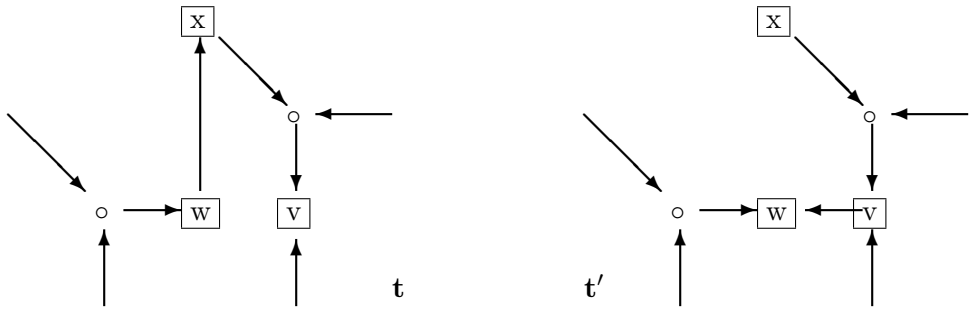
$$L_v \equiv \max\{m \le n : X_m = v\}.$$

Define $\mathbf{T}_n$ to consist of the directed edges

$$(v = X_{L_v}, X_{L_v+1}),\ v \ne X_n.$$

So the edges of $\mathbf{T}_n$ are the last-exit edges from each vertex (other than the root X_n). It is easy to check that $\mathbf{T}_n$ is a directed spanning tree.

Now consider what happens as n changes. Clearly the process $(\mathbf{T}_n : -\infty < n < \infty)$ is a stationary Markov chain on $\mathcal{T}$, with a certain transition matrix $\mathbf{Q} = (q(\mathbf{t}, \mathbf{t}'))$, say. The figure below indicates a typical transition $\mathbf{t} \to \mathbf{t}'$. Here $\mathbf{t}$ was constructed by the chain finishing at its root v, and $\mathbf{t}'$ is the new tree obtained when the chain makes a transition $v \to w$.



Theorem 9.10 (The Markov chain tree theorem) *The stationary distribution of* $(\mathbf{T}_n)$ *is* ρ.

Proof. Fix a directed spanning tree $\mathbf{t}'$. We have to verify

$$\sum_{\mathbf{t}} \bar{\rho}(\mathbf{t}) q(\mathbf{t}, \mathbf{t}') = \bar{\rho}(\mathbf{t}'). \tag{9.13}$$

Write w for the root of $\mathbf{t}'$. For each vertex $x \neq w$ there is a tree $\mathbf{t}_x$ constructed from $\mathbf{t}'$ by adding an edge (w, x) and then deleting from the resulting cycle the edge (v, w) (say, for some $v = v(x)$) leading into w. For $x = w$ set $v(x) = x$. It is easy to see that the only possible transitions into $\mathbf{t}'$ are from the trees $\mathbf{t}_x$, and that

$$\frac{\bar{\rho}(\mathbf{t}_x)}{\bar{\rho}(\mathbf{t}')} = \frac{p_{wx}}{p_{vw}}; \qquad q(\mathbf{t}_x, \mathbf{t}') = p_{vw}.$$

Thus the left side of (9.13) becomes

$$\sum_x \bar{\rho}(\mathbf{t}_x) q(\mathbf{t}_x, \mathbf{t}') = \bar{\rho}(\mathbf{t}') \sum_x p_{wx} = \bar{\rho}(\mathbf{t}').$$

□

The underlying chain X_n can be recovered from the tree-valued chain $\mathbf{T}_n$ via $X_n = \text{root}(\mathbf{T}_n)$, so we can recover the stationary distribution of X from the stationary distribution of T, as follows.

Corollary 9.11 (The Markov chain tree formula) *For each vertex v define*

$$\bar{\pi}(v) \equiv \sum_{\mathbf{t}:\ v = root(\mathbf{t})} \bar{\rho}(\mathbf{t}).$$

$$\pi(v) \equiv \frac{\bar{\pi}(v)}{\sum_w \bar{\pi}(w)}.$$

Then π is the stationary distribution of the original chain (X_n).

See the Notes for comments on this classical result.

Theorem 9.10 and the definition of $\mathbf{T}_0$ come close to specifying an algorithm for constructing a random spanning tree with distribution ρ. Of course the notion of running the chain from time $-\infty$ until time 0 doesn't sound very algorithmic, but we can rephrase this notion using time-reversal. Regarding the stationary distribution π as known, the time-reversed chain X^* has transition matrix $p^*_{vw} \equiv \pi_w p_{wv}/\pi_v$. Here is the restatement of Theorem 9.10 in terms of the time-reversed chain.

Corollary 9.12 *Let $(X^*_m : 0 \le m \le C)$ be the time-reversed chain, run until the cover time C. Define $\mathbf{T}$ to be the directed spanning tree with root X_0 and with edges $(v = X_{T_v}, X_{T_v-1})$, $v \ne X_0$. If X_0 has distribution π then $\mathbf{T}$ has distribution ρ. If X_0 is deterministically v_0, say, then T has distribution ρ conditioned on being rooted at v_0.*

Thus $\mathbf{T}$ consists of the edges by which each vertex is *first* visited, directed backwards.

For a reversible chain, we can of course use the chain itself in Corollary 9.12 above, in place of the time-reversed chain. If the chain is random walk on a unweighted graph G, then

$$\bar{\rho}(\mathbf{t}) = d(\mathrm{root}(\mathbf{t})) \prod_v \frac{1}{d(v)}$$

where $d(v)$ is the degree of v in G. So $\bar{\rho}$, restricted to the set of spanning trees with specified root v_0, is *uniform* on that set. In this setting, Corollary 9.12 specializes as follows.

Corollary 9.13 *Let $(X_m : 0 \le m \le C)$ be random walk on an unweighted graph G, started at v_0 and run until the cover time C. Define $\mathbf{T}$ to be the directed spanning tree with root v_0 and with edges $(v = X_{T_v}, X_{T_v-1})$, $v \ne v_0$. Then $\mathbf{T}$ is uniform on the set of all directed spanning trees of G rooted at v_0.*

We can rephrase this. If we just want "plain" spanning trees without a root and directions, then the $\mathbf{T}$ above, regarded as a plain spanning tree, is uniform on the set of all plain spanning trees. On the other hand, if we want a rooted spanning tree which is uniform on all such trees without prespecified root, the simplest procedure is to construct $\mathbf{T}$ as in Corollary 9.13 with deterministic start v_0, and at the end re-root $\mathbf{T}$ at a uniform random vertex. (This is slightly subtle – we could alternatively start with X_0 uniform, which is typically *not* the stationary distribution π.)

Using the bounds on cover time developed in Chapter 6, we now have an algorithm for generating a uniform spanning tree of a n-vertex graph in $O(n^3)$ steps (and $O(n^2)$ steps on a regular graph). No other known algorithm achieves these bounds.

9.2.2 Electrical network theory

The ideas in this subsection (and much more) are treated in a long but very readable survey paper by Pemantle [279], which we encourage the interested

reader to consult. As observed above, in the reversible setting we have the obvious simplification that we can construct uniform spanning trees using the chain itself. Deeper results can be found using the electrical network analogy. Consider random walk on a weighted graph G. The random spanning tree $\mathbf{T}$ constructed by Corollary 9.12, interpreted as a "plain" spanning tree, has distribution

$$\rho(\mathbf{t}) = c \prod_{e \in \mathbf{t}} w_e$$

where c is the normalizing constant. If an edge e is essential, it must be in every spanning tree, so $P(e \in \mathbf{T}) = 1$. If the edge is inessential, the probability will be strictly between 0 and 1. Intuitively, $P(e \in \mathbf{T})$ should provide a measure of "how nearly essential e is". This should remind the reader of the inessential edge inequality (yyy). Interpreting the weighted graph as an electrical network where an edge $e = (v, x)$ has resistance $1/w_e$, the effective resistance r_{vx} between v and x satisfies

$$r_{vx} \leq 1/w_{vx} \text{ with equality iff } (v, x) \text{ is essential}$$

Proposition 9.14 *For each edge* (v, x),

$$P((v, x) \in \mathbf{T}) = w_{vx} r_{vx}.$$

Note that in a n-vertex graph, $\mathbf{T}$ has exactly $n - 1$ edges, so Proposition 9.14 implies Foster's theorem (Chapter 3 yyy)

$$\sum_{\text{edges } (v,x)} w_{vx} r_{vx} = n - 1.$$

Proof. Consider the random walk started at v and run until the time U of the first return to v after the first visit to x. Let p be the chance that $X_{U-1} = x$, i.e. that the return to x is along the edge (x, v). We can calculate p in two ways. In terms of random walk started at x, p is the chance that the first visit to v is from x, and so by Corollary 9.12 (applied to the walk started at x) $p = P((x, v) \in \mathbf{T})$. On the other hand, consider the walk started at v and let S be the first time that the walk traverses (x, v) in that direction. Then

$$ES = EU/p.$$

But by yyy and yyy

$$ES = w/w_{vx}, \ EU = w r_{vx}$$

and hence $p = w_{vx} r_{vx}$ as required. □

The next result indicates the usefulness of the electrical network analogy.

Proposition 9.15 *For any two edges $e_1 \neq e_2$,*

$$P(e_1 \in \mathbf{T}, e_2 \in \mathbf{T}) \leq P(e_1 \in \mathbf{T})P(e_2 \in \mathbf{T}).$$

Proof. Consider the "shorted" graph G^{short} in which the end-vertices (x_1, x_2) of e_1 are shorted into a single vertex x, with edge-weights $w_{xv} = w_{x_1 v} + w_{x_2 v}$. The natural $1-1$ correspondence $\mathbf{t} \leftrightarrow \mathbf{t} \cup \{e_1\}$ between spanning trees of G^{short} and spanning trees of G containing e_1 maps the distribution ρ^{short} to the conditional distribution $\rho(\cdot | e_1 \in \mathbf{T})$. So, writing $\mathbf{T}^{\text{short}}$ for the random spanning tree associated with G^{short},

$$P(e_2 \in \mathbf{T}^{\text{short}}) = P(e_2 \in \mathbf{T} | e_1 \in \mathbf{T}).$$

But, setting $e_2 = (z_1, z_2)$, Proposition 9.14 shows

$$P(e_2 \in \mathbf{T}^{\text{short}}) = w_{z_1 z_2} r^{\text{short}}_{z_1 z_2}, \quad P(e_2 \in \mathbf{T}) = w_{z_1 z_2} r_{z_1 z_2}.$$

By Rayleigh's monotonicity principle, $r^{\text{short}}_{z_1 z_2} \leq r_{z_1 z - 2}$, and the result follows.

9.3 Self-verifying algorithms for sampling from a stationary distribution

To start with an analogy, we can in principle compute a mean hitting time $E_i T_j$ from the transition matrix $\mathbf{P}$, but we could alternatively estimate $E_i T_j$ by "pure simulation": simulate m times the chain started at i and run until hitting j, and then (roughly speaking) the empirical average of these m hitting times will be $(1 \pm O(m^{-1/2}))E_i T_j$. In particular, for fixed ε we can (roughly speaking) estimate $E_i T_j$ to within a factor $(1 \pm \varepsilon)$ in $O(E_i T_j)$ steps. Analogously, consider some notion of mixing time τ (say τ_1 or τ_2, in the reversible setting). The focus in this book has been on theoretical methods for bounding τ in terms of $\mathbf{P}$, and of theoretical consequences of such bounds. But can we bound τ by pure simulation? More importantly, in the practical context of Markov chain Monte Carlo, can we devise a "self-verifying" algorithm which produces an approximately-stationary sample from a chain in $O(\tau)$ steps without having prior knowledge of τ?

xxx tie up with MCMC discussion.

To say things a little more carefully, a "pure simulation" algorithm is one in which the transition matrix $\mathbf{P}$ is unknown to the algorithm. Instead, there is a list of the states, and at each step the algorithm can obtain, for any state i, a sample from the jump distribution $p(i, \cdot)$, independent of previous samples.

In the MCMC context we typically have an exponentially large state space and seek polynomial-time estimates. The next lemma (which we leave to the reader to state and prove more precisely) shows that no pure simulation algorithm can guarantee to do this.

Lemma 9.16 *Consider a pure simulation algorithm which, given any irreducible n-state chain, eventually outputs a random state whose distribution is guaranteed to be within ε of the stationary distribution in variation distance. Then the algorithm must take $\Omega(n)$ steps for every $\mathbf{P}$.*

Outline of proof. If there is a state k with the property that $1 - p(k,k)$ is extremely small, then the stationary distribution will be almost concentrated on k; an algorithm which has some chance of terminating without sampling a step from every state cannot possibly guarantee that no unvisited state k has this property. □

9.3.1 Exact sampling via the Markov chain tree theorem

Lovasz and Winkler [242] observed that the Markov chain tree theorem (Theorem 9.10) could be used to give a "pure simulation" algorithm for generating exactly from the stationary distribution of an arbitrary n-state chain. The algorithm takes

$$O(\tau_1^* \; n^2 \log n) \tag{9.14}$$

steps, where τ_1^* is the mixing time parameter defined as the smallest t such that

$$P_i(X_{U_\sigma} = j) \geq \frac{1}{2}\pi_j \text{ for all } i, j \in I, \; \sigma \geq t \tag{9.15}$$

where U_σ denotes a random time uniform on $\{0, 1, \ldots, \sigma - 1\}$, independent of the chain.

xxx tie up with Chapter 4 discussion and [241].

The following two facts are the mathematical ingredients of the algorithm. We quote as Lemma 9.17(a) a result of Ross [300] (see also [53] Theorem XIV.37); part (b) is an immediate consequence.

Lemma 9.17 *(a) Let π be a probability distribution on I and let $(F_i; i \in I)$ be independent with distribution π. Fix j, and consider the digraph with edges $\{(i, F_i) : i \neq j\}$. Then with probability (exactly) π_j, the digraph is a tree with edges directed toward the root j.*

(b) So if j is first chosen uniformly at random from I, then the probability above is exactly $1/n$.

As the second ingredient, observe that the Markov chain tree formula (Corollary 9.11) can be rephrased as follows.

Corollary 9.18 *Let π be the stationary distribution for a transition matrix $\mathbf{P}$ on I. Let J be random, uniform on I. Let $(\xi_i; i \in I)$ be independent, with $P(\xi_i = j) = p_{ij}$. Consider the digraph with edges $\{(i, \xi_i) : i \neq J\}$. Then, conditional on the digraph being a tree with edges directed toward the root J, the probability that $J = j$ equals π_j.*

So consider the special case of a chain with the property

$$p^*_{ij} \geq (1/2)^{1/n}\pi_j \ \forall i, j. \tag{9.16}$$

The probability of getting any particular digraph under the procedure of Corollary 9.18 is at least 1/2 the probability of getting that digraph under the procedure of Lemma 9.17, and so the probability of getting some tree is at least $1/2n$, by Lemma 9.17(b). So if the procedure of Corollary 9.18 is repeated $r = \lceil 2n \log 4 \rceil$ times, the chance that some repetition produces a tree is at least $1 - (1 - 1/2n)^{2n \log 4} = 3/4$, and then the root J of the tree has distribution exactly π.

Now for any chain, fix $\sigma > \tau_1^*$. The submultiplicativity (yyy) property of separation, applied to the chain with transition probabilities $\tilde{p}_{ij} = P_i(X_{U_\sigma} = j)$, shows that if V denotes the sum of m independent copies of U_σ, and ξ_i is the state reached after V steps of the chain started at i, then

$$P(\xi_i = j) \equiv P_i(X_V = j) \geq (1 - 2^{-m})\pi_j \ \forall i, j.$$

So putting $m = -\log_2(1 - (1/2)^{1/n}) = \Theta(\log n)$, the set of probabilities $(P(\xi_i = j))$ satisfy (9.16).

Combining these procedures, we have (for fixed $\sigma > \tau_1^*$) an algorithm which, in a mean number $nm\sigma r = O(\sigma n^2 \log n)$ of steps, has chance $\geq 3/4$ to produce an output, and (if so) the output has distribution exactly π. Of course we initially don't know the right σ to use, but we simply try $n, 2n, 4n, 8n, \ldots$ in turn until some output appears, and the mean total number of steps will satisfy the asserted bound (9.14).

9.3.2 Approximate sampling via coalescing paths

A second approach involves the parameter $\tau_0 = \sum_j \pi_j E_i T_j$ arising in the random target lemma (Chapter 2 yyy). Aldous [18] gives an algorithm which, given $\mathbf{P}$ and $\varepsilon > 0$, outputs a random state ξ for which $||P(\xi \in \cdot) - \pi|| \leq \varepsilon$, and such that the mean number of steps is at most

$$81\tau_0/\varepsilon^2. \tag{9.17}$$

The details are messy, so let us just outline the (simple) underlying idea. Suppose we can define a procedure which terminates in some random number Y of steps, where Y is an estimate of τ_0: precisely, suppose that for any $\mathbf{P}$

$$P(Y \le \tau_0) \le \varepsilon; \quad EY \le K\tau_0 \tag{9.18}$$

where K is an absolute constant. We can then define an algorithm as follows.

> Simulate Y; then run the chain for $U_{Y/\varepsilon}$ steps and output the final state ξ

where as above U_σ denotes a random time uniform on $\{0, 1, \ldots, \sigma - 1\}$, independent of the chain. This works because, arguing as at xxx,

$$||P(X_{U_\sigma} \in \cdot) - \pi|| \le \tau_0/\sigma$$

and so

$$||P(\xi \in \cdot) - \pi|| \ \le E \max(1, \frac{\tau_0}{Y/\varepsilon}) \le 2\varepsilon.$$

And the mean number of steps is $(1 + \frac{1}{2\varepsilon})EY$.

So the issue is to define a procedure terminating in Y steps, where Y satisfies (9.18). Label the states $\{1, 2, \ldots, n\}$ and consider the following *coalescing paths* routine.

(i) Pick a uniform random state J.

(ii) Start the chain at state 1, run until hitting state J, and write A_1 for the set of states visited along the path.

(iii) Restart the chain at state $\min\{j : j \notin A_1\}$, run until hitting some state in A_1, and write A_2 for the union of A_1 and the set of states visited by this second path.

(iiii) Restart the chain at state $\min\{j : j \notin A_2\}$, and continue this procedure until every state has been visited. Let Y be the total number of steps.

The random target lemma says that the mean number of steps in (ii) equals τ_0, making this Y a plausible candidate for a quantity satisfying (9.18). A slightly more complicated algorithm is in fact needed – see [18].

9.3.3 Exact sampling via backwards coupling

Write U for a r.v. uniform on $[0, 1]$, and (U_t) for an independent sequence of copies of U. Given a probability distribution on I, we can find a (far from unique!) function $f : [0, 1] \to I$ such that $f(U)$ has the prescribed distribution. So given a transition matrix $\mathbf{P}$ we can find a function $f : I \times [0, 1] \to I$

such that $P(f(i,U) = j) = p_{ij}$. Fix such a function. Simultaneously for each state i, define

$$X_0^{(i)} = i; \quad X_t^{(i)} = f(X_{t-1}^{(i)}, U_t), t = 1, 2, \ldots.$$

xxx tie up with coupling treatment
Consider the (forwards) coupling time

$$C^* = \min\{t : X_t^{(i)} = X_t^{(j)} \ \forall i, j\} \leq \infty.$$

By considering an initial state j chosen according to the stationary distribution π,

$$\max_i ||P_i(X_t \in \cdot) - \pi|| \leq P(C > t).$$

This can be used as the basis for an approximate sampling algorithm. As a simple implementation, repeat k times the procedure defining C^*, suppose we get finite values $C_1^*, \ldots, C_k^*$ each time, then run the chain from an arbitrary initial start for $\max_{1 \leq j \leq k} C_j^*$ steps and output the final state ξ. Then the error $||P(\xi \in \cdot) - \pi||$ is bounded by a function $\delta(k)$ such that $\delta(k) \to 0$ as $k \to \infty$.

Propp and Wilson [286] observed that by using instead a *backwards coupling* method (which has been exploited in other contexts – see Notes) one could make an exact sampling algorithm. Regard our i.i.d. sequence (U_t) as defined for $-\infty < t \leq 0$. For each state i and each time $s < 0$ define

$$X_s^{(i,s)} = i; \quad X_t^{(i,s)} = f(X_{t-1}^{(i,s)}, U_t), t = s+1, s+2, \ldots, 0.$$

Consider the *backwards* coupling time

$$C = \max\{t : X_0^{(i,t)} = X_0^{(j,t)} \ \forall i, j\} \geq -\infty.$$

Lemma 9.19 (Backwards coupling lemma) *If S is a random time such that $-\infty < S \leq C$ a.s. then the random variable $X^{(i,S)}$ does not depend on i and has the stationary distribution π.*

xxx describe algorithm
xxx poset story
xxx analysis in general setting and in poset setting.
xxx compare the 3 methods

9.4 Making reversible chains from irreversible chains

Let $\mathbf{P}$ be an irreducible transition matrix on I with stationary distribution π. The following straightforward lemma records several general ways in which to construct from $\mathbf{P}$ a transition matrix $\mathbf{Q}$ for which the associated chain still has stationary distribution π but is *reversible*. These methods all involve the time-reversed matrix $\mathbf{P}^*$

$$\pi_i p_{ij} = \pi_j p^*_{ji}$$

and so in practice can only be used when we know π explicitly (as we have observed several times previously, in general we cannot write down a useful explicit expression for π in the irreversible setting).

Lemma 9.20 *The following definitions each give a transition matrix* $\mathbf{Q}$ *which is reversible with respect to* π*.*

$$\begin{aligned} \textit{The additive reversiblization:} \quad & \mathbf{Q}^{(1)} = \tfrac{1}{2}(\mathbf{P} + \mathbf{P}^*) \\ \textit{The multiplicative reversiblization:} \quad & \mathbf{Q}^{(2)} = \mathbf{P}\mathbf{P}^* \\ \textit{The Metropolis reversiblization;} \quad & \mathbf{Q}^{(3)}_{i,j} = \min(p_{i,j}, p^*_{j,i}),\ j \neq i. \end{aligned}$$

Of these three construction, only $\mathbf{Q}^{(1)}$ is automatically irreducible. Consider for instance the "patterns in coin tossing" example (Chapter 2 Example yyy). Here are the distributions of a step of the chains from state $(i_1, \ldots, i_n)$.

($\mathbf{Q}^{(1)}$). To $(i_2, \ldots, i_n, 0)$ or $(i_2, \ldots, i_n, 1)$ or $(0, i_1, \ldots, i_{n-1})$ or $(1, i_1, \ldots, i_{n-1})$, with probability 1/4 each.

($\mathbf{Q}^{(2)}$). To $(0, i_2, \ldots, i_n)$ or $(1, i_2, \ldots, i_n)$, with probability 1/2 each. So the state space decomposes into 2-element classes.

($\mathbf{Q}^{(3)}$). Here a "typical" i is isolated.

We shall discuss two aspects of the relationship between irreversible chains and their reversibilizations.

9.4.1 Mixing times

Because the theory of L^2 convergence to stationarity is nicer for reversible chains, a natural strategy to study an irreversible chain (transition matrix $\mathbf{P}$) would be to first study a reversibilization $\mathbf{Q}$ and then seek some general result relating properties of the $\mathbf{P}$-chain to properties of the $\mathbf{Q}$-chain. There are (see Notes) general results relating spectra, but we don't pursue these because (cf. section 9.5) there seems no useful way to derive finite-time results for irreversible chains from spectral gap estimates.

xxx Persi, Fill etc stuff

9.4.2 Hitting times

Here are a matrix-theoretic result and conjecture, whose probabilistic significance (loosely relating to mean hitting times and reversiblization) will be discussed below. As usual $\mathbf{Z}$ is the fundamental matrix associated with $\mathbf{P}$, and $\mathbf{P}^*$ is the time-reversal.

Proposition 9.21 *trace* $\mathbf{Z}(\mathbf{P}^* - \mathbf{P}) \geq 0$.

Conjecture 9.22 *trace* $\mathbf{Z}^2(\mathbf{P}^* - \mathbf{P}) \geq 0$.

Proposition 9.21 is essentially due to Fiedler et al [147]. In fact, what is proved in ([147], p. 91) is that, for a positive matrix $\mathbf{V}$ with largest eigenvalue < 1,

$$\text{trace } (\sum_{m=1}^{\infty} \mathbf{V}^m)(\mathbf{V} - \mathbf{V}^T) \leq 0. \tag{9.19}$$

Applying this to $v_{ij} = s\pi_i^{1/2} p_{ij} \pi_j^{-1/2}$ for $s < 1$ gives

$$\text{trace } (\sum_{m=0}^{\infty} s^m (p_{ij}^{(m)} - \pi_j))(\mathbf{P} - \mathbf{P}^*) = \text{trace } (\sum_{m=0}^{\infty} s^m \mathbf{P}^{(m)})(\mathbf{P} - \mathbf{P}^*)$$

$$= s^{-1} \text{trace } (\sum_{m=0}^{\infty} \mathbf{V}^m)(\mathbf{V} - \mathbf{V}^T) \leq 0.$$

Letting $s \uparrow 1$ gives the Proposition as stated. □

The proof in [147] of (9.19) has no simple probabilistic interpretation, and it would be interesting to find a probabilistic proof. It is not clear to me whether Conjecture 9.22 could be proved in a similar way.

Here is the probabilistic interpretation of Proposition 9.21. Recall the elementary result (yyy) that in a n-state chain

$$\sum_a \sum_b \pi_a p_{ab} E_b T_a = n - 1. \tag{9.20}$$

The next result shows that replacing $E_b T_a$ by $E_a T_b$ gives an inequality. This arose as an ingredient in work of Tetali [325] discussed at xxx.

Corollary 9.23 $\sum_a \sum_b \pi_a p_{ab} E_a T_b \leq n - 1$.

Proof. We argue backwards. By (9.20), the issue is to prove

$$\sum_a \sum_b \pi_a p_{ab} (E_b T_a - E_a T_b) \geq 0.$$

Using Lemma yyy, the quantity in question equals

$$\sum_a \sum_b \pi_a p_{ab} \left(\frac{Z_{aa}}{\pi_a} - \frac{Z_{ba}}{\pi_a} - \frac{Z_{bb}}{\pi_b} + \frac{Z_{ab}}{\pi_b} \right)$$

$$= \text{trace } \mathbf{Z} - \text{trace } \mathbf{PZ} - \text{trace } \mathbf{Z} + \text{trace } \mathbf{P}^*\mathbf{Z} = \text{trace } (\mathbf{P}^* - \mathbf{P})\mathbf{Z} \geq 0.$$

□

Here is the motivation for Conjecture 9.22. For $0 \leq \lambda \leq 1$ let $\mathbf{P}(\lambda) = (1-\lambda)\mathbf{P} + \lambda\mathbf{P}^*$, so that $\mathbf{P}(1/2)$ is the "additive reversiblization" in Lemma 9.20. Consider the average hitting time parameters $\tau_0 = \tau_0(\lambda)$ from Chapter 4.

Corollary 9.24 *Assuming Conjecture 9.22 is true, $\tau_0(\lambda) \leq \tau_0(1/2)$ for all $0 \leq \lambda \leq 1$.*

In other words, making the chain "more reversible" tends to increase mean hitting times.

Proof. This depends on results about differentiating with respect to the transition matrix, which we present as slightly informal calculations. Introduce a "perturbation" matrix $\mathbf{Q}$ such that

$$\sum_j q_{ij} = 0 \ \forall i; \quad q_{ij} = 0 \text{ whenever } p_{ij} = 0. \tag{9.21}$$

Then $\mathbf{P} + \theta\mathbf{Q}$ is a transition matrix, for θ is some neighborhood of 0. Write $\frac{d}{d\theta}$ for the derivative at $\theta = 0$. Then, writing $N_i(t)$ for the number of visits to i before time t,

$$\frac{d}{d\theta} E_a T_b = \sum_i E_a N_i(T_b) \sum_j q_{ij} E_j T_b.$$

This holds because the $\sum_j$ term gives the effect on ET_b of a $\mathbf{Q}$-step from i. Using general identities from Chapter 2 yyy, and (9.21), this becomes

$$\frac{d}{d\theta} E_a T_b = \sum_i \left(\frac{\pi_i(z_{ab} - z_{bb})}{\pi_b} + z_{bi} - z_{ai} \right) \sum_j q_{ij} z_{jb}/\pi_b.$$

Now specialize to the case where π is the stationary distribution for each $\mathbf{P} + \theta\mathbf{Q}$, that is where

$$\sum_i \pi_i q_{ij} = 0 \ \forall j.$$

Then the expression above simplifies to

$$\frac{d}{d\theta}\ E_a T_b = \sum_i (z_{bi} - z_{ai})\ \sum_j q_{ij} z_{jb}/\pi_b.$$

Averaging over a, using $\sum_a \pi_a z_{ai} = 0$,

$$\frac{d}{d\theta}\ E_\pi T_b = \sum_i \sum_j z_{bi} q_{ij} z_{jb}/\pi_b$$

and then averaging over b,

$$\frac{d}{d\theta}\ \tau_0 = \text{ trace } \mathbf{ZQZ} = \text{ trace } \mathbf{Z}^2\mathbf{Q}.$$

So consider $\lambda < 1/2$ in Corollary 9.24. Then

$$\begin{aligned} \frac{d}{d\lambda}\tau_0(\lambda) &= \text{ trace } \mathbf{Z}^2(\lambda)(\mathbf{P}^* - \mathbf{P}) \\ &= (1-2\lambda)^{-1} \text{ trace } \mathbf{Z}^2(\lambda)(\mathbf{P}^*(\lambda) - \mathbf{P}(\lambda)) \end{aligned}$$

and Conjecture 9.22 would imply this is ≥ 0, implying the conclusion of Corollary 9.24.

9.5 An example concerning eigenvalues and mixing times

Here is an example, adapted from Aldous [11]. Let $(\lambda_u : 1 \leq u \leq n)$ be the eigenvalues of $\mathbf{P}$ with $\lambda_1 = 1$, and let

$$\beta = \max\{|\lambda_u| : 2 \leq u \leq n\}.$$

A weak quantification of "mixing" is provided by

$$\alpha(t) \equiv \max_{A,B} |P_\pi(X_0 \in A, X_t \in B) - \pi(A)\pi(B)|.$$

By definition, $\alpha(t)$ is less than the *maximal correlation* $\rho(t)$ discussed in Chapter 4 yyy, and so by yyy

$$\alpha(t) \leq \beta^t \text{ for a reversible chain.} \tag{9.22}$$

The convergence theorem (Chapter 2 yyy) says that $\alpha(t) \to 0$ as $t \to \infty$ provided $\beta < 1$. So one might expect some analog of (9.22) to hold in general. But this is dramatically false: Example 9.26 shows

Lemma 9.25 *There exists a family of n-state chains, with uniform stationary distributions, such that* $\sup_n \beta_n < 1$ *while* $\inf_n \alpha_n(n) > 0$.

Loosely, this implies there is no reasonable hypothesis on the spectrum of a n-state chain which implies an $o(n)$ mixing time. There is a time-asymptotic result

$$\alpha(t) \le \rho(t) \le C\beta^t \ \forall t,$$

for some C depending on the chain. But implicit claims in the literature that bounding the spectrum of a general chain has some consequence for finite-time behavior should be treated with extreme skepticism!

Example 9.26 Let (Y_t) be independent r.v.'s taking values in $\{0, 1, \ldots, n-1\}$ with distribution specified by

$$P(Y \le j) = \frac{j+1}{j+2}, \ 0 \le j \le n-2.$$

Define a Markov chain (X_t) on $\{0, 1, \ldots, n-1\}$ by

$$X_t = \max(X_{t-1} - 1, Y_t).$$

This chain has the property (cf. the "patterns in coin-tossing" chain) of attaining the stationary distribution in finite time. Precisely: for any initial distribution σ, the distribution of X_{n-1} is uniform, and hence X_t is uniform for all $t \ge n-1$. To prove this, we simply observe that for $0 \le j \le n-1$,

$$\begin{aligned} P_\sigma(X_{n-1} \le j) &= P(Y_{n-1} \le j, Y_{n-2} \le j+1, \ldots, Y_0 \le j+n-1) \\ &= \frac{j+1}{j+2} \times \frac{j+2}{j+3} \times \ldots \times \frac{n-1}{n} \times \ 1 \times \ldots 1 \\ &= \frac{j+1}{n}. \end{aligned}$$

If X_0 is either 0 or 1 then X_1 is distributed as Y_1, implying that the vector v with $v_i = 1_{(i=0)} - 1_{(i=1)}$ is an eigenvector of $\mathbf{P}$ with eigenvalue 0. By soft "duality" arguments it can be shown [11] that this is the largest eigenvalue, in the sense that

$$\mathcal{R}(\lambda_u) \le 0 \text{ for all } 2 \le u \le n. \tag{9.23}$$

I believe it is true that

$$\beta_n = \max\{|\lambda_u| : 2 \le u \le n\}$$

is bounded away from 1, but we can avoid proving this by considering the "lazy" chain $\hat{X}_t$ with transition matrix $\hat{\mathbf{P}} = (\mathbf{I} + \mathbf{P})/2$, for which by (9.23)

$$\hat{\beta}_n \leq \sup\{|(1+\lambda)/2| : |\lambda| \leq 1, \mathcal{R}(\lambda) \leq 0\} = \sqrt{1/2}.$$

So the family of lazy chains has the eigenvalue property asserted in Lemma 9.25. But by construction, $X_t \geq X_0 - t$, and so $P(X_0 > 3n/4, X_{n/2} < n/4) = 0$. For the lazy chains we get

$$P_\pi(X_0 > 3n/4, X_n < n/4) \to 0 \text{ as } n \to \infty$$

establishing the (non)-mixing property asserted in the lemma.

9.6 Miscellany

9.6.1 Mixing times for irreversible chains

In Chapter 4 yyy we discussed equivalences between different definitions of "mixing time" in the τ_1 family. Lovasz and Winkler [241] give a detailed treatment of analogous results in the non-reversible case.

xxx state some of this ?

9.6.2 Balanced directed graphs

Any Markov chain can be viewed as random walk on a weighted directed graph, but even on unweighted digraphs it is hard to relate properties on the walk to graph-theoretic properties, because (as we have often observed) it is in general hard to get useful information about the stationary distribution. An exception is the case of a *balanced* digraph, i.e. when the in-degree equals the out-degree ($= r_v$, say) at each vertex v. Random walk on a balanced digraph clearly retains the "undirected" property that the stationary probabilities π_v are proportional to r_v. Now the proofs of Theorems yyy and yyy in Chapter 6 extend unchanged to the balanced digraph setting, showing that the cover-and-return time C^+ satisfies

$$\max_v E_v C^+ \leq n^3 \text{ in general; } \max_v E_v C^+ \leq 6n^2 \text{ on a regular balanced digraph.}$$

(The proofs rely on the edge-commute inequality (Chapter 3 yyy), rather than any "resistance" property).

9.6.3 An absorption time problem

Consider a Markov chain on states $\{1, 2, \ldots, n\}$ for which the only possible transitions are downward, i.e. for $i \geq 2$ we have

$$p(i, j) = 0, \ j \geq i$$

and $p(1, 1) = 1$. The chain is ultimately absorbed in state 1. A question posed by Gil Kalai is whether there is a bound on the mean absorption time involving a parameter similar to that appearing in Cheeger's inequality. For each proper subset A of $\{1, \ldots, n\}$ with $1 \notin A$ define

$$c(A) = \frac{|A||A^c|}{n \ \sum_{i \in A} \sum_{j \in A^c} p(i, j)}$$

and then define

$$\kappa = \max_A c(A).$$

Open Problem 9.27 *Prove that* $\max_i E_i T_1$ *is bounded by a polynomial function of* $\kappa \log n$.

9.7 Notes on Chapter 9

Section 9.1. The idea of a maximal coupling goes back to Goldstein [169]: see Lindvall [233] for further history. Strong stationary times were studied in detail by Diaconis - Fill [115, 114] and Fill [150, 151], with particular attention to the case of one-dimensional stochastically monotone chains where there is some interesting "duality" theory. The special case of random walks on groups had previously been studied in Aldous - Diaconis [21, 22], and the idea is implicit in the regenerative approach to time-asymptotics for general state space chains, discussed at xxx. The theory surrounding Theorem 9.4 goes back to Rost [301]. This is normally regarded as part of the potential theory of Markov chains, which emphasizes analogous results in the transient setting, and the recurrent case is rather a sideline in that setting. See Revuz [289] sec. 2.5 or Dellacherie - Meyer [108] Chapter 9 sec. 3 for textbook treatments in the general-space setting. The observation that the theory applied in simple finite examples such as those in section 9.1.3 was made in Lovasz - Winkler [241], from whom we borrowed the phrase *halting state*. Monotonicity properties like that in the statement of Corollary 9.5 were studied in detail by Brown [73] from the viewpoint of approximate exponentiality of hitting times.

Section 9.2. A slightly more sophisticated and extensive textbook treatment of these topics is in Lyons [250]. The nomenclature reflects my taste: Theorem 9.10 is "the underlying theorem" which implies "the formula" for the stationary distribution in terms of weighted spanning trees. Different textbooks (e.g. [178] p. 340 xxx more refs) give rather different historical citations for the Markov chain tree formula, and in talks I often call it "the most often rediscovered result in probability theory": it would be an interesting project to track down the earliest explicit statement. Of course it can be viewed as part of a circle of ideas (including the matrix-tree theorem for the number of spanning trees in a graph) which is often traced back to Kirchoff. The fact that Theorem 9.10 underlies the formula was undoubtably folklore for many years (Diaconis attributes it to Peter Doyle, and indeed it appears in an undergraduate thesis [317] of one of his students), but was apparently not published until the paper of Anantharam and Tsoucas [30]. The fact that the Markov chain tree theorem can be interpreted as an algorithm for generating uniform random spanning trees was observed by Aldous [13] and Broder [65], both deriving from conversations with Diaconis. [13] initiated study of theoretical properties of uniform random spanning trees, proving e.g. the following bounds on the diameter Δ of the random tree in a regular n-vertex graph.

$$\frac{n^{1/2}}{K_1 \tau_2 \log n} \le E\Delta \le K_2 \tau_2^{1/2} n^{1/2} \log n \tag{9.24}$$

where K_1 and K_2 are absolute constants. Loosely, "in an expander, a random spanning tree has diameter $n^{1/2 \pm o(1)}$". Results on asymptotic Poisson distribution for the degrees in a random spanning tree are given in Aldous [13], Pemantle [279] and Pemantle and Burton [83]. Pemantle [278] discusses the analog of uniform random spanning trees on the *infinite* d-dimensional lattice, and Aldous and Larget [23] give simulation results on quantitative behavior on the d-dimensional torus.

Section 9.2.2. As described in Pemantle [279] and Burton and Pemantle [83], the key to deeper study of random spanning trees is

Theorem 9.28 (Transfer-impedance theorem) *Fix a graph G. There is a symmetric function $H(e_1, e_2)$ on pairs of edges in G such that for any edges $(e_1, \ldots, e_r)$*

$$P(e_i \in \mathbf{T} \text{ for all } 1 \le i \le r) = \det M(e_1, \ldots, e_r)$$

where $M(e_1, \ldots, e_r)$ is the matrix with entries $H(e_i, e_j), 1 \le i, j \le r$.

Section 9.3. The first "pure simulation" algorithm for sampling exactly from the stationary distribution was given by Asmussen et al [35], using a quite different idea, and lacking explicit time bounds.

Section 9.3.1. In our discussion of these algorithms, we are assuming that we have a list of all states. Lovasz - Winkler [242] gave the argument in a slightly different setting, where the algorithm can only "address" a single state, and their bound involved $\max_{ij} E_i T_j$ in place of τ_1^*.

Section 9.3.3. Letac [226] gives a survey of the "backwards coupling" method for establishing convergence of continuous-space chains: it suffices to show there exists a r.x. $X^{-\infty}$ such that $X_0^{(x,s)} \to X^{-\infty}$ a.s. as $s \to -\infty$, for each state x. This method is especially useful in treating matrix-valued chains of the form $X_t = A_t X_{t-1} + B_t$, where $(A_t, B_t), t \geq 1$ are i.i.d. random matrices. See Barnsley and Elton [43] for a popular application.

Section 9.4.1. One result on spectra and reversibilizations is the following. For a transition matrix $\mathbf{P}$ write

$$\tau(\mathbf{P}) = \sup\{\frac{1}{1-|\lambda|} : \ \lambda \neq 1 \text{ an eigenvalue of } \mathbf{P}\}.$$

Then for the additive reversibilization $\mathbf{Q}^{(1)} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^*)$ we have (e.g. [316] Proposition 1)

$$\tau(\mathbf{P}) \leq 2\tau(\mathbf{Q}^{(1)}).$$

Chapter 10

Some Graph Theory and Randomized Algorithms (September 1 1999)

Much of the theory of algorithms deals with algorithms on graphs; conversely, much of the last twenty years of graph theory research pays attention to algorithmic issues. Within these large fields random walks play a comparatively small role, but they do enter in various quite interesting and diverse ways, some of which are described in this chapter. One theme of this chapter is properties of random walks on expander graphs, introduced in sections 10.1.1 and 10.1.2. Some non-probabilistic properties of graphs can be explained naturally (to a probabilist, anyway!) in terms of random walk: see section 10.2. Section 10.3 reviews the general idea of randomized algorithms, and in section 10.4 we treat a diverse sample of randomized algorithms based on random walks. Section 10.5 describes the particular setting of *approximate counting*, giving details of the case of self-avoiding walks. (xxx details not written in this version).

For simplicity let's work in the setting of regular graphs. Except where otherwise stated, G is an n-vertex r-regular connected graph,

$$p_{vw} := r^{-1} 1_{((v,w) \text{ is an edge})}$$

is the transition matrix for discrete-time random walk on G (so $P = r^{-1}A$ for the adjacency matrix A) and $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n \geq -1$ are its eigenvalues, and $\tau_2 = 1/(1 - \lambda_2)$.

10.1 Expanders

10.1.1 Definitions

The Cheeger time constant τ_c discussed in Chapter 4 section 5.1 (yyy 10/11/94 version) becomes, for a r-regular n-vertex graph,

$$\tau_c = \sup_A \frac{r\ |A|\ |A^c|}{n\ |\mathcal{E}(A, A^c)|}$$

where $\mathcal{E}(A, A^c)$ is the set of edges from a proper subset A of vertices to its complement A^c. Our version of Cheeger's inequality is (Chapter 4 Corollary 37 and Theorem 40) (yyy 10/11/94 version)

$$\tau_c \le \tau_2 \le 8\tau_c^2. \tag{10.1}$$

Definition. An *expander family* is a sequence G_n of r-regular graphs (for some fixed $r > 2$), with $n \to \infty$ through some subsequence of integers, such that

$$\sup_n \tau_c(G_n) < \infty$$

or equivalently (by Cheeger's inequality)

$$\sup_n \tau_2(G_n) < \infty.$$

One informally says "expander" for a generic graph G_n in the family. The expander property is stronger than the *rapid mixing* property exemplified by the d-cube (Chapter 5 Example 15) (yyy 4/23/96 version). None of the examples in Chapter 5 is an expander family, and indeed there are no known elementary examples. Certain random constructions of regular graphs yield expanders: see Chapter 30 Proposition 1 (yyy 7/9/96 version). Explicit constructions of expander families, in particular the celebrated *Ramanujan graphs*, depend on group- and number-theoretic ideas outside our scope: see the elegant monograph of Lubotzky [243].

Graph parameters like τ_c are more commonly presented in inverted form (i.e. like $1/\tau_c$) as *coefficients of expansion* such as

$$h := \inf_A \frac{|\mathcal{E}(A, A^c)|}{r\ \min(|A|, |A^c|)}. \tag{10.2}$$

A more familiar version ([93] page 26) of Cheeger's inequality in graph theory becomes, on regular graphs,

$$h^2/2 \le 1 - \lambda_2 \le 2h. \tag{10.3}$$

Since trivially $\tau_c \le 1/h \le 2\tau_c$ the two versions agree up to factors of 2. Inequalities involving coefficients of expansion are often called *isoperimetric inequalities*. Expanders and isoperimetric inequalities have been studied extensively in graph theory and the theory of algorithms, e.g. Chung [93] Chapters 2 and 6, the conference proceedings [156], and the introduction of Lubotzky [243].

One algorithmic motivation for Cheeger-type inequalities concerns computational complexity of calculating parameters like τ_c amd h. Using the definition directly requires exponential (in n) time; but because eigenvalues can be calculated in polynomial time, these general inequalities imply that at least crude bounds can be computed in polynomial time.

10.1.2 Random walk on expanders

If we don't pay attention to numerical constants, then general results about reversible chains easily give us the orders of magnitude of other hitting and mixing time parameters for random walks on expanders.

Theorem 10.1 *For random walk on an expander family, as $n \to \infty$*

$$\tau_1 = \Theta(\log n) \tag{10.4}$$
$$\tau_0 = \Theta(n) \tag{10.5}$$
$$\tau^* = \Theta(n) \tag{10.6}$$
$$\sup_v E_v C = \Theta(n \log n) \tag{10.7}$$

Proof. Recall the general inequality between τ_1 and τ_2 (Chapter 4 Lemma 23) (yyy 10/11/94 version), which on a regular graph becomes

$$\tau_1 \le \tau_2(1 + \tfrac{1}{2}\log n). \tag{10.8}$$

This immediately gives the upper bound $\tau_1 = O(\log n)$. For the lower bound, having bounded degree obviously implies that the diameter Δ of the graph satisfies $\Delta = \Omega(\log n)$. And since the mean distance between an initial vertex v and the position X_T of the walk at a stopping time T is at most ET, the definition of $\tau_1^{(2)}$ implies $d(v,w) \le 2\tau_1^{(2)}$ for any pair of vertices, that is $\tau_1^{(2)} \ge \Delta/2$. This establishes (10.4). The general Markov chain fact $\tau_0 = \Omega(n)$ is Chapter 3 Proposition 14 (yyy 1/26/93 version). Chapter 4 Lemma 25 gives $\tau_0 \le 2n\tau_2$. Combining these and the obvious inequality $\tau_0 \le \tau^*/2$ establishes (10.5,10.6). Finally, the lower bound in (10.7) follows from the general lower bound in Chapter 6 Theorem 31 (yyy 10/31/94 version),

while the upper bound follows from the upper bound on τ^* combined with Chapter 2 Theorem 36 (yyy 8/18/94 version). ∎

In many ways the important aspect of Theorem 10.1 is that τ_1-type mixing times are of order $\log n$. We spell out some implications below. These hold for arbitrary regular graphs, though the virtue of expanders is that τ_2 is bounded.

Proposition 10.2 *There exists constants K_1, K_2 such that the following inequalities hold on any regular graph.*

(i) For each vertex v there exists a stopping time T_v such that $P_v(X(T_v) \in \cdot)$ is uniform and $E_v T_v \le K_1 \tau_2 \log n$.

(ii) For lazy random walk $\widetilde{X}_t$ (with hold-probability $1/2$)

$$P_v(\widetilde{X}_t = w) \ge \tfrac{1}{n}(1 - \tfrac{1}{2^j}) \text{ for all } t \ge jK_2\tau_2 \log n \text{ and all vertices } v, w.$$

Proof. Part (i) is just the definition of $\tau_1^{(2)}$, combined with (10.8) and the fact $\tau_1^{(2)} = O(\tau_1)$.

yyy relate (ii) to Chapter 4 section 3.3 ∎

Repeated use of (i) shows that we can get independent samples from π by sampling at random times $T_1, T_2, T_3, \ldots$ with $E(T_{j+1} - T_j) \le K_1\tau_2 \log n$. Alternatively, repeated use of (ii) shows that we can get almost independent samples from π by examining the lazy chain at deterministic times, as follows.

Corollary 10.3 *Fix j and let $t_0 \ge jK_2\tau_2 \log n$. Write $(Y_1, \ldots, Y_L) = (X_{t_0}, \ldots, X_{Lt_0})$. Then*

$$P(Y_1 = y_1, \ldots, Y_L = y_L) \ge n^{-L}(1 - \tfrac{L}{2^j}), \quad \text{for all } L, y_1, \ldots, y_L$$

the variation distance between $\mathrm{dist}(Y_1, \ldots, Y_L)$ *and $\pi \times \ldots \times \pi$ is at most $L/2^j$.*

Examining the lazy chain at deterministic times means sampling the original walk at random times, but at *bounded* random times. Thus we can get L precisely independent samples using (i) in mean number $K_1 L \tau_2 \log n$ of steps, but without a deterministic upper bound on the number of steps. Using Corollary 10.3 we get almost independent samples (up to variation distance ε) in a number of steps deterministically bounded by $K_2 L \log(L/\varepsilon)\tau_2 \log n$.

10.1.3 Counter-example constructions

Constructions with expanders are often useful in providing counter-examples to conjectures suggested by inspecting properties of random walk on the

elementary examples of graphs in Chapter 5. For example, consider upper bounds on τ_0 in terms of τ_2 and n, in our setting of regular graphs. From general results for reversible chains in Chapter 4 (10/11/94 version: Lemma 24 and below (9))

$$\max(\tau_2, \tfrac{(n-1)^2}{n}) \le \tau_0 \le (n-1)\tau_2.$$

The examples in Chapter 5 are consistent with a conjecture

$$\tau_0 =^{?} O(\max(n, \tau_2) \log n) \tag{10.9}$$

where the $\log n$ term is needed for the 2-dimensional torus. We now outline a counter-example.

Take m copies on the complete graph on m vertices. Distinguish one vertex v_i from each copy i. Add edges to make the (v_i) the vertices of a r-regular expander. For this graph G_m we have, as $m \to \infty$ with fixed r,

$$n = m^2;\ \tau_2 = \Theta(m^2);\ \tau_0 = \Theta(m^3)$$

contradicting conjecture (10.9). We leave the details to the reader: the key point is that random walk on G_m may be decomposed as random walk on the expander, with successive steps in the expander separated by sojourns of times $\Theta(m^2)$ within a clique.

10.2 Eigenvalues and graph theory

Our treatment of the relaxation time τ_2 in Chapter 4 emphasized probabilistic interpretations in the broad setting of reversible Markov chains. Specializing to random walk on unweighted graphs, there are a range of non-probabilistic connections between eigenvalues of the adjacency matrix and other graph-theoretic properties. Such *spectral graph theory* is the subject of Chung [93]: we shall just give a few results with clear probabilistic interpretations.

10.2.1 Diameter of a graph

Implicit in the proof of Theorem 10.1 is that, on a regular graph, the diameter Δ satisfies $\Delta = O(\tau) = O(\tau_2 \log n)$. By being a little careful we can produce numerical constants.

Proposition 10.4 $\Delta/2 \le \left\lceil \frac{1+\frac{1}{2}\log n}{\log \frac{3-\lambda_2}{1+\lambda_2}} \right\rceil$.

Proof. The discrete analog τ_1^{disc} of variation threshold satisfies

$$\tau_1^{\text{disc}} \geq \Delta/2, \tag{10.10}$$

because obviously

$$\text{if } d(v,w) = \Delta \text{ and } t < \Delta/2 \text{ then } ||P_v(X_t \in \cdot) - P_w(X_t \in \cdot)|| = 1. \tag{10.11}$$

Chapter 4 Lemma 26 (yyy 10/11/94 version) specializes to

$$\tau_1^{\text{disc}} \leq \left\lceil \frac{1 + \frac{1}{2}\log n}{\log 1/\beta} \right\rceil, \quad \beta := \max(\lambda_2, -\lambda_n). \tag{10.12}$$

We can remove dependence on λ_n by the trick of introducing artificial holds. (yyy tie up with Chapter 4 section 3.3). The chain with transition matrix $P' := \theta I + (1-\theta)P$ has eigenvalues $\lambda'_i = \theta + (1-\theta)\lambda_i$. Choosing $\theta = (1-\lambda_2)/(3-\lambda_2)$, (this being the value making $\lambda'_2 = -\lambda'_n$ in the worst case $\theta_n = -1$), we have

$$\tfrac{1+\lambda_2}{3-\lambda_2} = \beta' = \lambda'_2 \geq -\lambda'_n.$$

Since (10.10) still holds for the Markov chain P', combining (10.10) and (10.12) with β' establishes the Proposition.

10.2.2 Paths avoiding congestion

Upper bounding τ_2 via the distinguished paths technique (Chapter 4 section 4.3) (yyy 10/11/94 version) is a valuable theoreticial technique: the essence is to choose paths which "avoid congestion". In the opposite direction, one can use upper bounds on mixing times to show existence of paths which avoid congestion. Here's the simplest result of this type. The first part of the proof repeats an idea from Chapter 4 Lemma 21 (yyy new part of Lemma to be added).

Proposition 10.5 *Let $v(1), v(2), \ldots, v(n)$ be any ordering of the vertices $1, 2, \ldots, n$ of a r-regular graph. Then there exists, for each $1 \leq i \leq n$, a path from i to $v(i)$ such that, writing N_{vw} for the number of times the directed edge (v,w) is traversed in all the paths,*

$$\max_{(v,w)} N_{vw} \leq 7 \max\left(e\tau_1^{(2)}/r, \log n\right).$$

So on an expander the bound is $O(\log n)$, using (10.4).

Proof. By definition of $\tau_1^{(2)}$, for each vertex i there is a segment of the chain $i = X_0^{(i)}, X_1^{(i)}, \ldots, X_{U_i}^{(i)}$ such that $EU_i \leq \tau_1^{(2)}$ and $X_{U_i}^{(i)}$ has uniform

distribution. Take these segments independent as i varies. Write $\tilde{N}_{vw}$ for the (random) number of times that (v, w) is traversed by all these random paths. By considering a uniform random start, by (yyy tie up with Chapter 2 Proposition 3)

$$\tfrac{1}{n} E\tilde{N}_{vw} = \tfrac{1}{rn} \sum_i \tfrac{1}{n} EU_i.$$

In particular, $E\tilde{N}_{vw} \le \tau_1^{(2)}/r := \kappa$. By erasing loops we may contract each path to a path in which no directed edge is traversed twice. For fixed (v, w) let p_i be the chance that (v, w) is traversed by the contracted path from vertex i and let N'_{vw} be the total number of traversals. By independence of paths as i varies,

$$\begin{aligned} P(N'_{vw} \ge m) &\le \left(\sum_i p_i\right)^m /m! \quad \text{(expand the sum)} \\ &\le \kappa^m/m! \le (e\kappa/m)^m. \end{aligned}$$

Choosing $m = \lceil 3\max(e\kappa, \log n)\rceil$ makes the bound less that $1/(2n^2)$ and so

$$P\left(\max_{(v,w)} N'_{vw} \ge m\right) < 1/2.$$

Now repeat the entire construction to define another copy $(Y_t^{(i)}, 0 \le t \le V_i)$ of chain segments with traversal counts N''_{vw}. Since $X_{U_i}^{(i)}$ and $Y_{V_i}^{(v(i))}$ have the same uniform distribution, for each i we can construct the chain segments jointly such that $X_{U_i}^{(i)} = Y_{V_i}^{(v(i))}$. Concatenating paths gives a (non-Markov) random path from i to $v(i)$. Then

$$P\left(\max_{(v,w)} N'_{vw} + N''_{vw} \ge 2m\right) < 1$$

and so paths with the maximum $\le 2m - 1$ must exist. ∎

Broder et al. [69] give a more elaborate algorithm for constructing edge-disjoint paths between specified pairs $\{(a_i, b_i), 1 \le i \le k\})$ of distinct vertices on an expander, for $k = n^{1-o(1)}$. The essential idea is to first pick a set S of $4k$ vertices at random, then use a greedy algorithm to construct (as in the proof above) paths from each a_i and b_i to some $\tilde{a}_i$ and $\tilde{b}_i$ in S, then for each i construct a bundle of random walk paths from $\tilde{a}_i$ to $\tilde{b}_i$, and finally show that one path may be selected from each bundle so that the set of paths is edge-disjoint.

10.3 Randomized algorithms

10.3.1 Background

Here we give some background for the mathematician with no knowledge of the theory of algorithms. Typically there are many different possible algorithms for a particular problem; the theory of algorithms seeks "optimal" algorithms according to some notion of "cost". Cost is usually "time", i.e. number of computational steps, but sometimes involves criteria such as (storage) space or simplicity of coding. The phrase *randomized algorithm* refers to settings where the problem itself does not involve randomness but where randomness is introduced into the running of the algorithm. Why this is useful is best seen by example; the textbook of Motwani and Raghavan [265] provides a comprehensive range of examples and classifications of types of problems where randomized algorithms have proved useful. We give three standard examples below (not using Markov chains) and then proceed to talk about algorithms using random walks on graphs.

Example 10.6 *Statistical sampling.*

Consider a population of n individuals. Suppose we wish to know the proportion q of the population with some attribute, i.e. who answer "Yes" to some Yes/No question. To calculate q exactly we need to question all n individuals. But if we can sample uniformly at random from the population, then we can estimate q approximately and can bound the size of error in probability. To do this, we sample independently k random individuals, question them, and calculate the empirical proportion $\bar{q}$ of Yes answers. Use $\bar{q}$ as our estimate of q, and theory gives error probabilities

$$P(|\bar{q} - q| > 2(\bar{q}(1-\bar{q}))^{1/2}k^{-1/2}) \approx 5\%.$$

Such 95% confidence intervals are discussed in every freshman statistics course. Classical statististical sampling is conceptually a bit different from algorithms, in that the "cost" k here refers to real-world costs of interviewing human individuals (or experimenting on individual rats or whatever) rather than to computational cost. However, the key insight in the formula above is that, for prescribed allowable error ε, the cost of this *simple random sampling* is $O(\varepsilon^{-2})$ and this cost does not depend on the "problem size" (i.e. population size) n. The next example is a slightly more subtle use of sampling in a slightly more algorithmic context.

Example 10.7 *Size of a union of sets.*

It's fun to say this as a word problem in the spirit of Chapter 1. Suppose your new cyberpunk novel has been rejected by all publishers, so you have published it privately, and seek to sell copies by mailing advertizements to individuals. So you need to buy mailing lists (from e.g. magazines and specialist bookshops catering to science fiction). Your problem is that such mailing lists might have much overlap. So before buying L lists $A_1, \ldots, A_L$ (where A_i is a set of $|A_i|$ names and addresses) you would like to know roughly the size $|\cup_i A_i|$ of their union. How can you do this without knowing what the sets A_i are (the vendors won't give them to you for free)? Statistical sampling can be used here. Suppose the vendors will allow you to randomly sample a few names (so you can check accuracy) and will allow you to "probe" whether a few specified names are on their list. Then you can sample k times from each list, and for each sampled name X_{ij} probe the other lists to count the number $m(X_{ij}) \geq 1$ of lists containing that name. Consider the identity

$$\begin{aligned} |\cup_i A_i| &= \sum_i |A_i| \times |A_i|^{-1} \sum_{a \in A_i} 1/m(a) \\ &= \sum_i |A_i| \; E(1/M_i) \end{aligned}$$

where M_i is the number of lists containing a uniform random name from A_i. You can estimate $E(1/M_i)$ by $k^{-1} \sum_{j=1}^{k} 1/m(X_{ij})$, and the error has standard deviation $\leq k^{-1/2}$, and the resulting estimate of $|\cup_i A_i|$ has error

$$\pm O((\sum_i |A_i|^2/k)^{1/2}) = \pm O(k^{-1/2} L \max_i |A_i|).$$

As in the previous example, the key point is that the cost of "approximately counting" $\cup_i A_i$ to within a small relative error does not depend on the size of the sets.

Example 10.8 *Solovay-Strassen test of primality [312].*

We can't improve on the concise description given by Babai [37].

> Let $n > 1$ be an odd integer. Call an integer w a Solovay-Strassen *witness* (of compositeness of n) if $1 \leq w \leq n-1$ and either g.c.d.$(w, n) > 1$ or $w^{(n-1)/2} \not\equiv (\frac{w}{n}) \bmod n$, where $(\frac{w}{n})$ is the Jacobi symbol (computed via quadratic reciprocity as easily as g.c.d.'s are computed via Euclid's algorithm). Note that no S-S witness exists if n is prime. On the other hand (this is

the theorem) if n is composite, then *at least half* of the integers $1, 2, \ldots, n-1$ are S-S witnesses.

Suppose now that we want to decide whether or not a given odd 200-digit integer n is prime. Pick k integers $w_1, \ldots, w_k$ independently at random from $\{1, 2, \ldots, n-1\}$. If any one of the w_i turns out to be a witness, we know that n is composite. If none of them are, let us conclude that n is prime. Here we may err, but for any n, the probability that we draw the wrong conclusion is at most $\varepsilon = 2^{-k}$. Setting $k = 500$ is perfectly realistic, so we shall have proven the mathematical statement "n is prime" beyond the shade of doubt.

10.3.2 Overview of randomized algorithms using random walks or Markov chains

Our focus is of course on randomized algorithms using random walks or Markov chains. We will loosely divide these into three categories. *Markov chain Monte Carlo* seeks to simulate a random sample from a (usually non-uniform) given probability distribution on a given set. This is the central topic of Chapter 11. In section 10.4 below we give a selection of miscellaneous graph algorithms. Into this category also falls the idea (Chapter 6 section 8.2) (yyy 10/31/94 version; details to be written) of using random walk as a "undirected graph connectivity" algorithm, and the idea (end of section 10.2.2) of using random walk paths as an ingredient in constructing edge-disjoint paths in an expander graph. A third, intermediate category is the specific topic of *approximate counting via Markov chains*, to be discussed in section 10.5.

10.4 Miscellaneous graph algorithms

10.4.1 Amplification of randomness

In practice, Monte Carlo simulations are done using deterministic pseudo-random number generators. Ideally one would prefer some physical device which generated "truly random" bits. Presumably any such physical random number generator would be rather slow compared to the speed of arithmetical calculations. This thinking has led to an area of theory in which the cost of a randomized algorithm is taken to be the number of truly random bits used.

Recall the Solovay-Strassen test of primality in Example 10.8. Philosophically, there is something unsettling about using a deterministic pseudo-random number generator in this context, so we regard this as a prototype example where one might want to use a hypothetical source of truly random bits. To pick a uniform random integer from $\{1, 2, \ldots, n\}$ requires about $\log_2 n$ random bits, so the cost of the algorithm as presented above is about $k \log_2 n = (\log_2 1/\varepsilon)\ (\log_2 n)$ bits, where ε is the prescribed allowable error probability. But one can use the existence of explicit expanders and results like Lemma 10.12 to devise an algorithm which requires fewer truly random bits. Suppose we have a n-vertex r-regular expander, and label the vertices $\{1, 2, \ldots, n\}$. To simulate a uniform random starting vertex and t steps of the random walk requires about $\log_2 n + t \log_2 r$ bits. The chance that such a walk never hits the set A of witnesses is, by Lemma 10.12, at most $\exp(-\frac{t}{2\tau_2})$. To make this chance $\leq \varepsilon$ we take $t = 2\tau_2\ \log(1/\varepsilon)$, and the cost becomes $\log_2 n + 2\tau_2(\log_2 r)\ \log(1/\varepsilon)$. Thus granted the existence of expanders on which we can efficiently list neighbors of any specified vertex in order to simulate the random walk, the method of simulating (dependent) integers (w_i) via the random walk (instead of independently) reduces the number of truly random bits required from $O((\log n) \times (\log 1/\varepsilon))$ to $O(\max(\log n,\ \log 1/\varepsilon))$.

The idea of using random walks on expanders for such algorithmic purposes is due to Ajtai et al [4]. Following Impagliazzo and Zuckerman [188] one can abstract the idea to rather general randomized algorithms. Suppose we are given a randomized algorithm, intended to show whether an object $x \in \mathcal{X}$ has a property $\mathcal{P}$ by outputting "Yes" or "No", and that for each x the algorithm is correct with probability $\geq 2/3$ and uses at most b random bits. Formally, the algorithm is a function $A : \mathcal{X} \times \{0,1\}^b \to \{\text{Yes,No}\}$ such that

$$\text{if } x \in \mathcal{P} \text{ then } 2^{-b}|\{\mathbf{i} \in \{0,1\}^b : A(x, \mathbf{i}) = \text{YES}\}| \geq 2/3$$

$$\text{if } x \notin \mathcal{P} \text{ then } 2^{-b}|\{\mathbf{i} \in \{0,1\}^b : A(x, \mathbf{i}) = \text{YES}\}| \leq 1/3$$

where $\mathcal{P} \subset \mathcal{X}$ is the subset of all objects possessing the property. To make the probability of incorrect classification be $\leq \varepsilon$ we may simply repeat the algorithm m times, where $m = \Theta(\log 1/\varepsilon)$ is chosen to make

$$P(\text{Binomial}(m, 2/3) \leq m/2) \leq \varepsilon,$$

and output Yes or No according to the majority of the m individual outputs. This requires $bm = \Theta(b \log 1/\varepsilon)$ random bits. But instead we may take $\{0,1\}^b$ as the vertices of a degree-r expander, and simulate a uniform random

starting vertex and m steps of random walk on the expander, using about $b + m \log_2 r$ random bits. For each of the $m+1$ vertices of $\{0,1\}^b$ visited by the walk $(Y_i, 0 \le i \le m)$, compute $A(x, Y_i)$, and output Yes or No according to the majority of the $m+1$ individual outputs. The error probability is at most

$$\max_B P_\pi \left(\tfrac{N_{m+1}(B)}{m+1} - \pi(B) \ge \tfrac{1}{3} \right)$$

where $N_{m+1}(B)$ is the number of visits to B by the walk $(Y_i, 0 \le i \le m)$. By the large deviation bound for occupation measures (Theorem 10.11, yyy to be moved to other chapter) this error probability is at most

$$(1 + c_1 m/\tau_2) \ \exp(-c_2 m/\tau_2)$$

for constants c_1 and c_2. To reduce this below ε requires $m = 0(\tau_2 \log(1/\varepsilon))$. Thus the existence of (bounded-degree) expanders implies that the number of random bits required is only

$$b + m \log_2 r = O(\max(b, \ \log 1/\varepsilon))$$

compared to $O(b \log(1/\varepsilon))$ using independent sampling.

10.4.2 Using random walk to define an objective function

In Chapter 6 section 8.2 (yyy currently at end of this Chapter; to be moved) we gave a standard use of the probabilistic method Here is a less standard use, from Aldous [7], where we use the sample path of a random walk to make a construction.

Consider a function h defined on the vertices of a n-vertex graph G. Constrain h to have no local minima except the global minimum (for simplicity, suppose the values of h are distinct). We seek algorithms to find the vertex v at which $h(v)$ is minimized. Any deterministic "descent" algorithm will work, but it might work slowly. Could there be some more sophisticated algorithm which always works quickly? One idea is *multi-start descent*. Pick $n^{1/2}$ vertices uniformly at random; from these, choose the vertex with minimum h-value, and follow the greedy descent algorithm. On a degree-d graph, the mean time is $O(dn^{1/2})$. Now specialize to the case where G is the d-cube. One can give examples where single-start (from a uniform random start) descent has mean time $\Omega(2^{(1-\varepsilon)d})$, so from a worst-case mean-time viewpoint, multi-start is better. The next theorem shows that (again from a worst-case mean-time viewpoint), one cannot essentially improve on multi-start descent. Consider random walk on the d-cube started

at a uniform random vertex U and let H(v) be the first hitting time on v. Then H is a random function satisfying the constraint, minimized at $v = U$, but

Theorem 10.9 ([6]) *Every algorithm for locating U by examining values $H(v)$ requires examining a mean number $\Omega(2^{d/2-\varepsilon})$ of vertices.*

The argument is simple in outline. As a preliminary calculation, consider random walk on the d-cube of length $t_0 = O(2^{d/2-\varepsilon})$, started at $\mathbf{0}$, and let L_v be the time of the last visit to v, with $L_v = 0$ if v is not visited. Then

$$EL_v \le \sum_{t=1}^{t_0} tP_{\mathbf{0}}(X(t) = v) = O(1) \tag{10.13}$$

where the $O(1)$ bound holds because the worst-case v for the sum is $v = \mathbf{0}$ and, switching to continuous time,

$$\int_0^{2^{d/2}} tP_{\mathbf{0}}(\tilde{X}(t) = \mathbf{0})\, dt = \int_0^{2^{d/2}} t \left(\frac{1 + e^{-2t/d}}{2}\right)^d dt = O(1).$$

Now consider an algorithm which has evaluated $H(v_1), \ldots, H(v_m)$ and write $t_0 = \min_{i \le m} H(v_i) = H(v^*)$ say. It does no harm to suppose $t_0 = O(2^{d/2-\varepsilon})$. Conditional on the information revealed by $H(v_1), \ldots, H(v_m)$, the distribution of the walk $(X(t); 0 \le t \le t_0)$ is specified by
(a) take a random walk from a uniform random start U, and condition on $X(t_0) = v^*$;
(b) condition further on the walk not hitting $\{v_i\}$ before time t_0.
The key point, which of course is technically hard to deal with, is that the conditioning in (b) has little effect. If we ignore the conditioning in (b), then by reversing time we see that the random variables $(H(v^*) - H(v))^+$ have the same distribution as the random variables L_v (up to vertex relabeling). So whatever vertex v the algorithm chooses to evaluate next, inequality (10.13) shows that the mean improvement $E(H(v^*) - H(v))^+$ in objective value is $O(1)$, and so it takes $\Omega(2^{d/2-\varepsilon})$ steps to reduce the objective value from $2^{d/2-\varepsilon}$ to 0.

10.4.3 Embedding trees into the d-cube

Consider again the d-cube $I = \{0,1\}^d$ with Hamming distance $d(\mathbf{i}, \mathbf{j})$. Let $\mathcal{B}$ be the vertices of a M-vertex binary tree. For an *embedding*, i.e. an arbitrary function $\rho : \mathcal{B} \to I$, define

$$\text{load} = \max_{\mathbf{i} \in I} |\{v \in \mathcal{B} : \rho(v) = \mathbf{i}\}|$$

$$\text{dilation} = \max_{\text{edges } (v,w) \text{ of } \mathcal{B}} d(\rho(v), \rho(w)).$$

How can we choose an embedding which makes both load and dilation small? This was studied by Bhatt and Cai [47], as a toy model for parallel computation. In the model I represents the set of processors, $\mathcal{B}$ represents the set of tasks being done at a particular time, the tree structure indicating tasks being split into sub-tasks. To assign tasks to processors, we desire no one processor to have many tasks (small load) and we desire processors working on tasks and their sub-tasks to be close (small dilation) to facilitate communication. As the computation proceeds the tree will undergo local changes, as tasks are completed and new tasks started and split into sub-tasks, and we desire to be able to update the embedding "locally" in response to local changes in the tree. Bhatt and Cai [47] investigated the natural *random walk embedding*, where the root of $\mathcal{B}$ is embedded at $\mathbf{0}$, and recursively each child w of v is embedded at the vertex $\rho(w)$ found at step L (for even L) of a random walk started at $\rho(v)$. So by construction, dilation $\leq L$, and the mathematical issue is to estimate load. As before, the details are technically complicated, but let us outline one calculation. Clearly load $= \Omega(\max(1, M/2^d))$, so we would like the mean number of vertices of $\mathcal{B}$ embedded at any particular vertex $\mathbf{i}$ to be $O(\max(1, M/2^d))$. In bounding this mean, because $p_{\mathbf{0i}}(t) \leq p_{\mathbf{00}}(t)$ for even t (Chapter 7 Corollary 3) (yyy 1/31/94 version) we see that the worst-case $\mathbf{i}$ is $\mathbf{0}$, and then because $p_{\mathbf{00}}(t)$ is decreasing in t we see that the worst-case M-vertex binary tree is a maximally balanced tree. Thus we want

$$\sum_{k=0}^{\log_2 M} 2^k p_{\mathbf{00}}(kL) = O(\max(1, M/2^d)). \tag{10.14}$$

From the analysis of random walk on the d-cube (Chapter 5 Example 15) (yyy 4/23/96 version) one can show

$$p_{\mathbf{00}}(k \log d) = O(\max(d^{-k}, 2^{-d})), \text{ uniformly in } k, d \geq 1.$$

It follows that (10.14) holds if we take $L = \lceil \log d \rceil$.

Of course to bound the load we need to consider the maximally-loaded vertex, rather than a typical vertex. Considering $M = 2^d$ for definiteness, if the M vertices were assigned independently and uniformly, the mean load at a typical vertex would be 1 and classical arguments show the maximal load would be $\Theta(\frac{d}{\log d})$. We have shown that with tree-embedding the mean load at a typical vertex is $O(1)$, so analogously one can show the maximal

load is $O(d/\log d)$. However, [47] shows that by locally redistributing tasks assigned to the same processor, one can reduce the maximal load to $O(1)$ while maintaining the dilation at $O(\log d)$.

10.4.4 Comparing on-line and off-line algorithms

Here we describe work of Coppersmith et al [99]. As in Chapter 3 section 2 (yyy 9/2/94 version) consider a weighted graph on n vertices, but now write the edge-weights as $c_{ij} = c_{ji} > 0$ and regard them as a matrix $\mathbf{C}$ of *costs*. Let $\mathbf{P}$ be the transition matrix of an irreducible Markov chain whose only transitions are along edges of the graph. For each i and j let $m(i,j)$ be the mean cost of the random walk from i to j, when traversing an edge (v,w) incurs cost c_{vw}. Define the *stretch* $s(\mathbf{P},\mathbf{C})$ to be the smallest s such that there exists $a < \infty$ such that, for arbitrary $v_0, v_1, \ldots, v_k$

$$\sum_{i=0}^{k-1} m(v_i, v_{i+1}) \le a + s \sum_{i=0}^{k-1} c_{v_i v_{i+1}}. \tag{10.15}$$

Note that $c(\mathbf{P},\mathbf{C})$ is invariant under scaling of $\mathbf{C}$.

Proposition 10.10 ([99]) *(a)* $s(\mathbf{P},\mathbf{C}) \ge n-1$.

(b) If $\mathbf{P}$ *is reversible and* $\mathbf{C}$ *is the matrix of mean commute times* $E_iT_j + E_jT_i$ *then* $s(\mathbf{P},\mathbf{C}) = n-1$.

(c) For any cost matrix $\tilde{\mathbf{C}}$ *there exists a reversible transition matrix* $\mathbf{P}$ *with matrix* $\mathbf{C}$ *of mean commute times such that, for some constant* α*,*

$$c_{ij} \le \alpha \tilde{c}_{ij}$$

$$c_{ij} = \alpha \tilde{c}_{ij} \text{ when } p_{ij} > 0.$$

So from (b) and invariance under scaling, $s(\mathbf{P},\tilde{\mathbf{C}}) = n-1$.

We shall prove (a) and (b), which are just variations of the standard theory of mean hitting times developed in Chapters 2 and 3. The proof of part (c) involves "convex programming" and is rather outside our scope. The algorithmic interpretations are also rather too lengthy to give in detail, but are easy to say in outline. Imagine a problem where it is required to pick a minimum-cost path, where the cost of a path consists of costs of traversing edges, together with extra costs and constraints. There is some optimal off-line solution, which may be hard to calculate. In such a problem, one may be able to use Proposition 10.10(c) to show that the algorithm which simply picks a random sample path (with transition matrix $\mathbf{P}$ from (c)) has mean cost not more than $n-1$ times the cost of the optimal path.

Proof. Write π for the stationary distribution of $\mathbf{P}$. Write $m^+(v,v)$ for the mean cost of an excursion from v to v, and write $\bar{c} = \sum_v \sum_w \pi_v p_{vw} c_{vw}$. Then $m^+(v,v) = \bar{c}/\pi_v$ by the ergodic argument (Chapter 2 Lemma 30) (yyy 8/18/94 version). and so

$$\begin{aligned} n\bar{c} &= \sum_v \pi_v m^+(v,v) \\ &= \sum_v \pi_v \sum_w p_{vw}(c_{v,w} + m(w,v)) \\ &= \bar{c} + \sum_v \sum_w \pi_v p_{vw} m(w,v). \end{aligned}$$

In other words,

$$\sum_v \sum_w \pi_w p_{wv} m(v,w) = (n-1)\bar{c}. \tag{10.16}$$

Now apply the definition (10.15) of $s(\mathbf{P}, \mathbf{C})$ to the sequence of states visited by the stationary time-reversed chain $\mathbf{P}^*$; by considering the mean of each step,

$$\sum_v \sum_w \pi_v p^*_{vw} m(v,w) \le s(\mathbf{P}, \mathbf{C}) \sum_v \sum_w \pi_v p^*_{vw} c_{vw}. \tag{10.17}$$

But the left sides of (10.16) and (10.17) are equal by definition of $\mathbf{P}^*$, and the sum in the right of (10.17) equals $\bar{c}$ by symmetry of $\mathbf{C}$, establishing (a). For (b), first note that the definition (10.15) of stretch is equivalent to

$$s(\mathbf{P}, \mathbf{C}) = \max_\sigma \frac{\sum_i m(v_i, v_{i+1})}{\sum_i c_{v_i, v_{i+1}}} \tag{10.18}$$

where σ denotes a cycle $(v_1, v_2, \ldots, v_m, v_1)$. Write $t(v,w) = E_v T_w$. Fix a cycle σ and write $\mu = \sum_i t(v_i, v_{i+1})$ for the mean time to complete the cyclic tour. By the ergodic argument (Chapter 2 Lemma 30) (yyy 8/18/94 version). the mean number of traversals of an edge (v,w) during the tour is $\mu \pi_y p_{vw}$, and hence the ratio in (10.18) can be written as

$$\frac{\sum_i t(v_i, v_{i+1})}{\sum_i c_{v_i, v_{i+1}}} \times \sum_v \sum_w \pi_v p_{vw} c_{vw}. \tag{10.19}$$

Now the hypothesis of (b) is that $\mathbf{P}$ is reversible and $c_{vw} = t(v,w) + t(w,v)$. So the second term of (10.19) equals $2(n-1)$ by Chapter 3 Lemma 6 (yyy 9/2/94 version) and the first term equals $1/2$ by the cyclic tour property Chapter 3 Lemma 1 (yyy 9/2/94 version). So for each cycle σ the ratio in (10.18) equals $n-1$, establishing (b).

10.5 Approximate counting via Markov chains

For a finite set S, there is a close connection between

(a) having an explicit formula for the size $|S|$

(b) having a bounded-time algorithm for generating a uniform random element of S.

As an elementary illustration, we all know that there are $n!$ permutations of n objects. From a proof of this fact, we could write down an explicit $1-1$ mapping f between the set of permutations and the set $A = \{(a_1, a_2, \ldots, a_n) : 1 \leq a_i \leq i\}$. Then we could simulate a uniform random permutation by first simulating a uniform random element a of A and then computing $f(a)$. Conversely, given an algorithm which was guaranteed to produce a uniform random permutation after $k(n)$ calls to a random number generator, we could (in principle) analyze the working of the algorithm in order to calculate the chance p of getting the identity permutation. Then we can say that number of permutations equals $1/p$.

A more subtle observation is that, in certain settings, having an algorithm for generating an *approximately* uniform random element of S can be used to estimate *approximately* the size $|S|$. The idea is to estimate successive ratios by sampling. Suppose we can relate S to smaller sets

$$S = S_L \supset S_{L-1} \supset \ldots \supset S_2 \supset S_1 \tag{10.20}$$

where $|S_1|$ is known, the ratios $p_i := |S_{i+1}|/|S_i|$ are bounded away from 0, and where we can sample uniformly from each S_i. Then take k uniform random samples from each S_i and find the sample proportion W_i which fall into S_{i-1}. Because $|S| = |S_1| \ \prod_{i=2}^{L} |S_i|/|S_{i-1}|$, we use

$$\hat{N} := |S_1| \ \prod_{i=2}^{L} W_i^{-1}$$

as an estimate of $|S|$. To study its accuracy, it is simpler to consider $|S|/\hat{N} = \prod_{i=2}^{L} W_i/p_i$. Clearly $E(|S|/\hat{N}) = 1$, and we can calculate the variance by

$$\begin{aligned} \text{var } \left(\frac{|S|}{\hat{N}}\right) &= \text{var } \left(\prod_{i=2}^{L} \frac{W_i}{p_i}\right) \\ &= \prod_{i=2}^{L}(1 + \text{var } (W_i/p_i)) \quad - 1 \\ &= \prod_{i=2}^{L}(1 + \tfrac{1-p_i}{p_i k}) \quad - 1. \end{aligned}$$

The simplest case is where we know a theoretical lower bound p_* for the p_i. Then by taking $k = O(\varepsilon^{-2}L/p_*)$ we get

$$\text{var}\left(\frac{|S|}{\hat{N}}\right) \leq \exp(\tfrac{L}{p_* k}) \ - 1 = O(\varepsilon^2).$$

In other words, with a total number

$$Lk = O(\varepsilon^{-2}L^2/p_*) \tag{10.21}$$

of random samples, we can statistically estimate $|S|$ to within a factor $1 \pm O(\varepsilon)$.

The conceptual point of invoking intermediate sets S_i is that the overall ratio $|S_1|/|S|$ may be exponentially small in some size parameter, so that trying to estimate this ratio directly by sampling from S would involve order $|S|/|S_1|$, i.e. exponentially many, samples. If we can specify the intermediate sets with ratios p_i bounded away from 0 and 1 then $L = O(\log(|S|/|S_1|))$ and so the number of samples required in (10.21) depends on $\log(|S|/|S_1|)$ instead of $|S|/|S_1|$.

The discussion so far has not involved Markov chains. From our viewpoint, the interesting setting is where we cannot directly get uniform random samples from a typical S_i, but instead need to use Markov chain Monte Carlo. That is, on each S_i we set up a reversible Markov chain with uniform stationary distribution (i.e. a chain whose transition matrix is symmetric in the sense $p_{vw} \equiv p_{wv}$) Assume we have a bound τ_1 on the τ_1-values of all these chains. Then as a small modification of Corollary 10.3, one can get m samples from the combined chains whose joint distribution is close (in variation distance) to the the distribution of independent samples from the uniform distributions in $O(\tau_1 m \log m)$ steps. As above, if we can specify the intermediate sets with ratios p_i bounded away from 0 and 1 then we need $m = O(\varepsilon^{-2}\log^2(|S|/|S_1|))$ samples, and so (ignoring dependence on ε) the total number of steps in all the chains is $O(\tau_1 \log^{2+o(1)}(|S|/|S_1|)$.

In summary, to implement this method of approximate counting via Markov chains, one needs

- a way to specify the intermediate sets (10.20)
- a way to specify Markov chains on the S_i whose mixing times can be rigorously bounded.

Two particular examples have been studied in detail, and historically these examples provided major impetus for the development of technical tools to

estimate mixing times. Though the details are too technical for this book, we outline these examples in the next two sections, and then consider in detail the setting of self-avoiding walks.

10.5.1 Volume of a convex set

Given a closed convex set K in R^d, for large d, how can we algorithmically calculate the volume of K? Regard K as being described by an *oracle*, that is for any $x \in R^d$ we can determine in one step whether or not $x \in K$. Perhaps surprisingly, there is no known deterministic algorithm which finds $\text{vol}(K)$ approximately (i.e. to within a factor $1 \pm \varepsilon$) in a polynomial in d number of steps. But this problem is amenable to "approximate counting via Markov chains" technique. This line of research was initiated by Dyer et al [135, 136] who produced an algorithm requiring $O(d^{23+o(1)})$ steps. A long sequence of papers (see the Notes) studied variants of both the Markov chains and the analytic techniques in order to reduce the polynomial degree. Currently the best bound is $O(d^{5+o(1)})$, due to Kannan et al [206].

To outline the procedure in this example, suppose we know $B(1) \subset K \subset B(r)$, where $B(r)$ is the ball of radius $r = r(d)$. (It turns out that one can transform any convex set into one satisfying these constraints with $r = O(d^{3/2})$.) We specify an increasing sequence of convex subsets

$$B(1) = K_0 \subset K_1 \subset \ldots \subset K_L = K$$

by setting $K_i := B(2^{i/d}) \cap K$. This makes the ratios of successive volumes bounded by 2 and requires $L = O(d \log d)$ intermediate sets. So the issue is to design and analyze a chain on a typical convex set K_i whose stationary distribution is uniform. Various Markov chains have been used: simple random walk on a fine discrete lattice restricted to K_i, or spherically symmetric walks. The analysis of the chains has used Cheeger inequalities for chains and the refinement of classical isoperimetric inequalities for convex sets. Recent work of Bubley et al [80] has successfully introduced coupling methods, and it is a challenging problem to refine these coupling methods. There is a suggestive analogy with theoretical study of Brownian motion in a convex set – see Chapter 13 section 1.3 (yyy 7/29/99 version).

10.5.2 Matchings in a graph

For a finite, not necessarily regular, graph G_0 let $\mathcal{M}(G_0)$ be the set of all matchings in G_0, where a *matching* M is a subset of edges such that no vertex is in more than one edge of M. Suppose we want to count $|\mathcal{M}(G_0)|$ (for the

harder setting of counting *perfect* matchings see the Notes). Enumerate the edges of G_0 as $e_1, e_2, \ldots, e_L$, where L is the number of edges of G_0. Write G_i for the graph G with edges $e_1, \ldots, e_i$ deleted. A matching of G_i can be identified with a matching of G_{i-1} which does not contain e_i, so we can write

$$\mathcal{M}(G_{L-1}) \subset \mathcal{M}(G_{L-2}) \subset \ldots \subset \mathcal{M}(G_1) \subset \mathcal{M}(G_0).$$

Since G_{L-1} has one edge, we know $|\mathcal{M}(G_{L-1})| = 2$. The ratio $|\mathcal{M}(G_{i+1})|/|\mathcal{M}(G_i)|$ is the probability that a uniform random matching of G_i does not contain the edge e_{i+1}. So the issue is to design and analyze a chain on a typical set $\mathcal{M}(G_i)$ of matchings whose stationary distribution is uniform. Here is a natural such chain. From a matching M_0, pick a uniform random edge e of G_0, and construct a new matching M_1 from M_0 and e as follows.

If $e \in M_0$ then set $M_1 = M_0 \setminus \{e\}$.

If neither end-vertex of e is in an edge of M_0 then set $M_1 = M_0 \cup \{e\}$.

If exactly one end-vertex of e is in an edge (e' say) of M_0 then set $M_1 = \{e\} \cup M_0 \setminus \{e'\}$.

This construction (the idea goes back to Broder [64]) yields a chain with symmetric transition matrix, because each possible transition has chance $1/L$. An elegant analysis by Jerrum and Sinclair [199], outlined in Jerrum [197] section 5.1, used the distinguished paths technique to prove that on a n-vertex L-edge graph

$$\tau_2 = O(Ln).$$

Since the number of matchings can be bounded crudely by $n!$,

$$\tau_1 = O(\tau_2 \log n!) = O(Ln^2 \log n). \tag{10.22}$$

10.5.3 Simulating self-avoiding walks

xxx to be written

10.6 Notes on Chapter 9

Section 10.1.1. Modern interest in expanders and their algorithmic uses goes back to the early 1980s, e.g. their use in parallel sorting networks by Ajtai et al [3], and was increased by Alon's [26] graph-theoretic formulation of Cheeger's inequality. The conference proceedings [156] provides an overview.

Edge-expansion, measured by parameters like h at (10.2), is more relevant to random walk than vertex-expansion. Walters [334] compares definitions. What we call "expander" is often called *bounded-degree expander*.

Section 10.1.2. Ajtai et al [4], studying the "amplification of randomness" problems in section 10.4.1, was perhaps the first explicit use of random walk on expanders. In Theorem 10.1, the upper bounds on τ^* and EC go back to Chandra et al [85].

Section 10.1.3. With the failure of conjecture (10.9), the next natural conjecture is: on a r-regular graph

$$\tau_0 =^{?} O(\max(n, \tau_2) \max(\log n, r)).$$

It's not clear whether such conjectures are worth pursuing.

Section 10.2. More classical accounts of spectral graph theory are in Cvetkovic et al [107, 106].

On a not-necessarily-regular graph, Chung [93] studies the eigenvalues of the matrix $\mathcal{L}$ defined by

$$\begin{aligned} \mathcal{L}_{vw} &= 1, \quad w = v \\ &= -(d_v d_w)^{-1/2} \text{ for an edge } (v, w) \\ &= 0 \text{ else.} \end{aligned} \tag{10.23}$$

In the regular case, $-\mathcal{L}$ is the Q-matrix of transition rates for the continuous-time random walk, and so Chung's eigenvalues are identical to our continuous-time eigenvalues. In the non-regular case there is no simple probabilistic interpretation of $\mathcal{L}$ and hence no simple probabilistic interpretation of results involving the relaxation time $1/\lambda_2$ associated with $\mathcal{L}$.

Section 10.2.1. Chung [93] Chapter 3 gives more detailed results about diameter and eigenvalues. One can slightly sharpen the argument for Proposition 10.4 by using (10.11) and the analog of Chapter 4 Lemma 26 (yyy 10/11/94 version) in which the threshold for τ_1^{disc} is set at $1 - \varepsilon$. Such arguments give bounds closer to that of [93] Corollary 3.2: if G is not complete then

$$\Delta \leq \left\lceil \frac{\log(n-1)}{\log \frac{3-\lambda_2}{1+\lambda_2}} \right\rceil .$$

Section 10.2.2. Chung [93] section 4.4 analyzes a somewhat related *routing problem*. Broder et al [70] analyze a dynamic version of path selection in expanders.

Section 10.3.1. Example 10.7 (union of sets) and the more general *DNF counting problem* were studied systematically by Karp et al [210]; see also [265] section 11.2.

The Solovay-Strassen test of primality depends on a certain property of the Jacobi symbol: see [265] section 14.6 for a proof of this property.

Section 10.4.1. Several other uses of random walks on expanders can be found in Ajtai et al [4], Cohen and Wigderson [97], Impagliazzo and Zuckerman [188].

Section 10.4.4. Tetali [325] discussions extensions of parts (a,b) of Proposition 10.10 to nonsymmetric cost matrices.

Section 10.5. More extensive treatments of approximate counting are in Sinclair [309] and Motwani and Raghavan [265] Chapter 12.

Jerrum et al [201] formalize a notion of *self-reducibility* and show that, under this condition, approximate counting can be performed in polynomial time iff approximately uniform sampling can. See Sinclair [309] section 1.4 for a nice exposition.

Abstractly, we are studying randomized algorithms which produce a random estimate $\hat{a}(d)$ of a numerical quantity $a(d)$ (where d measures the "size" of the problem) together with a rigorous bound of the form

$$P((1-\varepsilon)a(d) \le \hat{a}(d) \le (1+\varepsilon)a(d)) \ge 1-\delta.$$

Such a scheme is a *FPRAS* (fully polynomial randomized approximation scheme) if the cost of the algorithm is bounded by a polynomial in d, $1/\varepsilon$ and $\log 1/\delta$. Here the conclusion involving $\log 1/\delta$ is what emerges from proofs using large deviation techniques.

Section 10.5.1. Other papers on the volume problem and the related problem of sampling from a log-concave distribution are Lovász and Simonovits [238], Applegate and Kannan [32], Dyer and Frieze [134], Lovász and Simonovits [239] and Frieze et al [158].

Section 10.5.2. In the background is the problem of approximating the *permanent*

$$\mathrm{per}A := \sum_{\sigma} \prod_{i=1}^{n} A_{i\sigma(i)}$$

of a $n \times n$ non-negative matrix, where the sum is over all permutations σ. When A is the adjacency matrix of a $n + n$ bipartite graph, per(A) is the number of *perfect matchings*. Approximate counting of perfect matchings is

in principle similar to approximate counting of all matchings; one seeks to use the chain in section 10.5.2 restricted to $\mathcal{M}_i \cup \mathcal{M}_{i-1}$, where $\mathcal{M}_i$ is the set of matchings with exactly i edges. But successful analysis of this chain requires that we have a *dense* graph, with minimum degree $> n/2$. Jerrum and Sinclair [198] gave the first analysis, using the Cheeger inequality and estimating expansion via distinguished paths. Sinclair [309] Chapter 3 and Motwani and Raghavan [265] Chapter 11 give more detailed expositions. Subsequently it was realized that using the distinguished paths technique directly to bound τ_2 was more efficient. A more general setting is to seek to sample from the non-uniform distribution on matchings M

$$\pi(M) \propto \lambda^{|M|}$$

for a parameter $\lambda > 1$. The distinguished paths technique [199, 197] giving (10.22) works in this setting to give

$$\tau_1 = O(Ln^2\lambda \log(n\lambda)).$$

10.7 Material belonging in other chapters

10.7.1 Large deviation bounds

yyy Somewhere in the book we need to discuss the results on explicit large deviation bounds for occupation measure / empirical averages: [167, 125, 204, 229]. In section 10.4.1 we used the following bound from Gillman [167] Theorem 2.1.

Theorem 10.11

$$P_\mu\left(N_n(B)/n \ - \pi(B) > \gamma\right) \le \left(1 + \frac{\gamma n}{10\tau_2}\right) \ \sqrt{\sum_i \mu_i^2/\pi_i} \ \exp\left(\frac{-\gamma^2 n}{20\tau_2}\right).$$

10.7.2 The probabilistic method in combinatorics

yyy This is to be moved to Chapter 6, where we do the "universal traversal sequences" example.

Suppose one wants to show the existence of a combinatorial object with specified properties. The most natural way is to give an explicit construction of an example. There are a variety of settings where, instead of a giving an explicit construction, it is easier to argue that a randomly-chosen object has a non-zero chance of having the required properties. The monograph by Alon and Spencer [29] is devoted to this topic, under the name *the probabilistic method*. One use of this method is below. Two more example occur later in the book: random construction of expander graphs (Chapter 30 Proposition 1) (yyy 7/9/96 version), and the random construction of an objective function in an optimization problem (Chapter 9 section 4.2) (yyy this version).

10.7.3 copied to Chapter 4 section 6.5

(yyy 10/11/94 version) Combining Corollary 31 with (62) gives the continuous time result below. Recasting the underlying theory in discrete time establishes the discrete-time version.

Lemma 10.12

$$\begin{aligned} \textit{(continuous time)} \quad P_\pi(T_A > t) &\le \exp(-t\pi(A)/\tau_2), \ t \ge 0 \\ \textit{(discrete time)} \quad P_\pi(T_A \ge t) &\le (1 - \pi(A)/\tau_2)^t, \ t \ge 0. \end{aligned}$$

Notes on this section. In studying bounds on T_A such as Lemma 10.12 we usually have in mind that $\pi(A)$ is small. One is sometimes interested

in exit times from a set A with $\pi(A)$ small, i.e. hitting times on A^c where $\pi(A^c)$ is near 1. In this setting one can replace inequalities using τ_2 or τ_c (which parameters involve the whole chain) by inequalities involving analogous parameters for the chain restricted to A and its boundary. See Babai [36] for uses of such bounds.

On several occasions we have remarked that for most properties of random walk, the possibility of an eigenvalue near -1 (i.e. an almost-bipartite graph) is irrelevant. An obvious exception arises when we consider lower bounds for $P_\pi(T_A > t)$ in terms of $|A|$, because in a bipartite graph with bipartition $\{A, A^c\}$ we have $P(T_A > 1) = 0$. It turns out (Alon et al [27] Proposition 2.4) that a corresponding lower bound holds in terms of $\tau_n \equiv 1/(\lambda_n + 1)$.

$$P_\pi(T_A > t) \geq \left(\max(0, 1 - \tfrac{|A|}{n\tau_n})\right)^t .$$

Chapter 11

Markov Chain Monte Carlo (January 8 2001)

This book is intended primarily as "theoretical mathematics", focusing on ideas that can be encapsulated in theorems. *Markov Chain Monte Carlo* (MCMC), which has grown explosively since the early 1990s, is in a sense more of an "engineering mathematics" field – a suite of techniques which attempt to solve applied problems, the design of the techniques being based on intuition and physical analogies, and their analysis being based on experimental evaluation. In such a field, the key insights do not correspond well to theorems.

In section 11.1 we give a verbal overview of the field. Section 11.2 describes the two basic schemes (Metropolis and line-sampling), and section 11.3 describes a few of the many more complex chains which have been suggested. The subsequent sections are fragments of theory, indicating places where MCMC interfaces with topics treated elsewhere in this book. Liu [235] gives a comprehensive textbook treatment of the field.

11.1 Overview of Applied MCMC

11.1.1 Summary

We give a brisk summary here, and expand upon some main ideas (the **boldface phrases**) in section 11.1.2.

Abstractly, we start with the following type of problem.

> Given a probability distribution π on a space S, and a numerical quantity associated with π (for instance, the mean $\bar{g} :=$

> $\sum_x \pi(x)g(x)$ or $:= \int g\, d\pi$, for specified $g : S \to R$), how can one estimate the numerical quantity using Monte Carlo (i.e. randomized algorithm) methods?

Asking such a question implicitly assumes we do not have a solution using mathematical analysis or efficient deterministic numerical methods. *Exact Monte Carlo sampling* presumes the ability to sample exactly from the target distribution π, enabling one to simulate an i.i.d. sequence (X_i) and then use classical statistical estimation, e.g. estimate $\bar{g}$ by $n^{-1}\sum_{i=1}^n g(X_i)$. Where implementable, such exact sampling will typically be the best randomized algorithm. For **one-dimensional distributions** and a host of special distributions on higher-dimensional space or combinatorial structures, exact sampling methods have been devised. But it is unrealistic to expect there to be any exact sampling method which is effective in all settings. *Markov Chain Monte Carlo sampling* is based on the following idea.

> First devise a Markov chain on S whose stationary distribution is π. Simulate n steps $X_1, \ldots, X_n$ of the chain. Treat $X_{\tau^*}, X_{\tau^*+1}, \ldots, X_n$ as dependent samples from π (where τ^* is some estimate of some mixing time) and then use these samples in a statistical estimator of the desired numerical quantity, where the confidence interval takes the dependence into account.

Variations of this basic idea include running multiple chains and introducing auxiliary variables (i.e. defining a chain on some product space $S \times A$). The basic scheme and variations are what make up the field of MCMC. Though there is no a priori reason why one must use *reversible* chains, in practice the need to achieve a target distribution π as stationary distribution makes general constructions using reversibility very useful.

MCMC originated in statistical physics, but mathematical analysis of its uses there are too sophisticated for this book, so let us think instead of Bayesian statistics with high-dimensional data as the prototype setting for MCMC. So imagine a point $x \in R^d$ as recording d numerical characteristics of an individual. So data on n individuals is represented as a $n \times d$ matrix $\mathbf{x} = (x_{ij})$. As a model, we first take a parametric family $\phi(\theta, x)$ of probability densities; that is, $\theta \in R^p$ is a p-dimensional parameter and for each θ the function $x \to \phi(\theta, x)$ is a probability density on R^d. Finally, to make a Bayes model we take θ to have some probability density $h(\theta)$ on R^p. So the probability model for the data is: first choose θ according to $h(\cdot)$, then choose $(x_{i\cdot})$ i.i.d. with density $\phi(\theta, x)$. So there is a posterior distribution

on θ specified by

$$f_{\mathbf{x}}(\theta) := \frac{h(\theta)\ \prod_{i=1}^{n} \phi(\theta, x_{i\cdot})}{z_{\mathbf{x}}} \tag{11.1}$$

where $z_{\mathbf{x}}$ is the normalizing constant. Our goal is to sample from $f_{\mathbf{x}}(\cdot)$, for purposes of e.g. estimating posterior means of real-valued parameters. An explicit instance of (11.1) is the **hierarchical Normal model**, but the general form of (11.1) exhibits features that circumscribe the type of chains it is feasible to implement in MCMC, as follows.

(i) Though the underlying functions $\phi(\cdot,\cdot), h(\cdot)$ which define the model may be mathematically simple, our target distribution $f_{\mathbf{x}}(\cdot)$ depends on actual numerical data (the data matrix $\mathbf{x}$), so it is hard to predict, and dangerous to assume, global regularity properties of $f_{\mathbf{x}}(\cdot)$.

(ii) The normalizing constant $z_{\mathbf{x}}$ is hard to compute, so we want to define chains which can be implemented without calculating $z_{\mathbf{x}}$.

The wide range of issues arising in MCMC can loosely be classified as "design" or "analysis" issues. Here "design" refers to deciding which chain to simulate, and "analysis" involves the interpretation of results. Let us start by discussing design issues. The most famous general-purpose method is the *Metropolis scheme*, of which the following is a simple implementation in setting (11.1). Fix a length scale parameter l. Define a step $\theta \to \theta^{(1)}$ of a chain as follows.

> Pick i uniformly from $\{1, 2, \ldots, p\}$.
> Pick U uniformly from $[\theta_i - l, \theta_i + l]$.
> Let θ' be the p-vector obtained from θ by changing the i'th coordinate to U.
> With probability $\min(1, f_{\mathbf{x}}(\theta')/f_{\mathbf{x}}(\theta))$ set $\theta^{(1)} = \theta'$; else set $\theta^{(1)} = \theta$.

The target density enters the definition only via the ratios $f_{\mathbf{x}}(\theta')/f_{\mathbf{x}}(\theta)$, so the value of $z_{\mathbf{x}}$ is not needed. The essence of a Metropolis scheme is that there is a *proposal chain* which *proposes* a move $\theta \to \theta'$, and then an *acceptance/rejection step* which accepts or rejects the proposed move. See section 11.2.1 for the general definition, and proof that the stationary distribution is indeed the target distribution. There is considerable flexibility in the choice of proposal chain. One might replace the uniform proposal step by a Normal or symmetrized exponential or Cauchy jump; one might instead choose a random (i.e. isotropic) direction and propose to step some random distance in that direction (to make an isotropic Normal step, or a step uniform within a ball, for instance). There is no convincing theory to

say which of these choices is better in general. However, in each proposal chain there is some length scale parameter l: there is a trade-off between making l too small (proposals mostly accepted, but small steps imply slow mixing) and making l too large (proposals rarely accepted), and in section 11.5 we give some theory (admittedly in an artificial setting) which does give guidance on choice of l.

The other well-known general MCMC method is exemplified by the *Gibbs sampler*. In the setting of (11.1), for $\theta = (\theta_1, \ldots, \theta_p)$ and $1 \le j \le p$ write

$$f_{\mathbf{x},j,\theta}(v) = f_{\mathbf{x}}(\theta_1, \ldots, \theta_{j-1}, v, \theta_{j+1}, \ldots, \theta_p).$$

A step $\theta \to \theta^{(1)}$ of the Gibbs sampler is defined as follows.

> Pick j uniformly from $\{1, 2, \ldots, p\}$.
> Pick V from the density on R^1 proportional to $f_{\mathbf{x},j,\theta}(v)$.
> Let $\theta^{(1)}$ be θ with its j'th coordinate replaced by V.

The heuristic appeal of the Gibbs sampler, compared to a Metropolis scheme, is that in the latter one typically considers only small proposal moves (lest proposals be almost always rejected) whereas in the Gibbs sampler one samples over an infinite line, which may permit larger moves. The disadvantage is that sampling along the desired one-dimensional line may not be easy to implement (see section 11.1.2). Closely related to the Gibbs sampler is the *hit-and-run* sampler, where one takes a random (isotropic) direction line instead of a coordinate line; section 11.2.2 abstracts the properties of such *line samplers*, and section 11.3 continues this *design* topic to discuss more complex designs of chains which attain a specified target distribution as their stationary distribution.

We now turn to *analysis* issues, and focus on the simplest type of problem, obtaining an estimate for an expectation $\bar{g} = \sum g(\mathbf{x})\pi(\mathbf{x})$ using an irreducible chain (X_t) designed to have stationary distribution π. How do we obtain an estimate, and how accurate is it? The most straightforward approach is *single-run* estimation. The asymptotic variance rate is

$$\sigma^2 := \lim_{t\to\infty} t^{-1}\mathrm{var}\left(\sum_{s=1}^{t} g(X_s)\right) = \sum_{s=-\infty}^{\infty} \mathrm{cov}_\pi(g(X_0), g(X_s)). \qquad (11.2)$$

So simulate a single run of the chain, from some initial state, for some large number t of steps. Estimate $\bar{g}$ by

$$\hat{g} = \frac{1}{t - t_0} \sum_{i=t_0+1}^{t} g(X_i) \qquad (11.3)$$

and estimate the variance of $\hat{g}$ by $(t-t_0)^{-1}\hat{\sigma}^2$, and report a confidence interval for $\hat{g}$ by assuming $\hat{g}$ has Normal distribution with mean $\bar{g}$ and the estimated variance. Here $\hat{\sigma}^2$ is an estimate of σ^2 obtained by treating the sample covariances $\hat{\gamma}_s$ (i.e. the covariance of the data-set $(g(X_i), g(X_{i+s})); 0 \le i \le t-s$) as estimators of $\gamma_s = \mathrm{cov}_\pi(g(X_0), g(X_s))$. And the *burn-in* time t_0 is chosen as a time after which the $\hat{\gamma}_s$ become small.

Though the **practical relevance of theoretical mixing time parameters** is debatable, one can say loosely that single-run estimates based on t steps will work fine if t is large compared to the relaxation time τ_2. The difficulty is that in practical MCMC problems we do not know, or have reasonable upper bounds on, τ_2, nor can we estimate τ_2 rigorously from simulations. The difficulty in diagnosing convergence from simulations is the possibility of *metastability error* caused by multimodality. Using **statistical physics imagery**, the region around each mode is a *potential well*, and the stationary distribution conditioned to a potential well is a *metastable distribution*. Believing that a simulation reaches the stationary distribution when in fact it only reaches a metastable distribution is the metastability error.

The simplest way to try to guard against metastability error is the *multiple trials diagnostic*. Here we run k independent copies of the chain from different starting states, each for t steps. One diagnostic is to calculate the k sample averages $\hat{g}_j$, and check that the empirical s.d. of these k averages is consistent with the estimated s.d. $(t-t_0)^{-1/2}\hat{\sigma}$. Intuitively, one chooses the initial states to be "overdispersed", i.e. more spread out than we expect the target distribution to be; passing the diagnostic test gives us some reassurance against metastability error (if there were different potential wells, we hope our runs would find more than one well, and that different behavior of g on different wells would be manifest).

Of course, if one intends to perform such diagnostics it makes sense to start out doing the k multiple runs. A more elaborate procedure is to divide $[0,t]$ into L successive blocks, and seek to check whether the kL blocks "look similar". This can be treated as a classical topic in statistics ("analysis of variance"). In brief, we compute the sample mean $\hat{g}_{i,j}$ and sample variance $\hat{\sigma}^2_{i,j}$ for the j'th block of the i'th simulation, and see if this data (perhaps after deleting the first few blocks of each simulation) is consistent with the blocks being i.i.d.. If so, we use the overall average as an estimator of $\bar{g}$, and estimate the accuracy of this estimator by assuming the blocks were independent.

If a multiple-runs diagnostic fails, or if one lacks confidence in one's ability to choose a small number of starting points which might be attracted to

different nodes (if such existed), then one can seek schemes specially adapted to multimodal target densities. Because it is easy to find *local* maxima of a target density f, e.g. by a deterministic hill-climbing algorithm, one can find modes by repeating such an algorithm from many initial states, to try to find an exhaustive list of modes with relatively high f-values. This is *mode-hunting*; one can then design a chain tailored to jump between the wells with non-vanishing probabilities. Such methods are highly problem-specific; more general methods (such as the multi-level or multi-particle schemes of sections 11.3.3 and 11.3.4) seek to automate the search for relevant modes within MCMC instead of having a separate mode-hunting stage.

In seeking theoretical analysis of MCMC one faces an intrinsic difficulty: MCMC is only needed on "hard" problems, but such problems are difficult to study. In comparing effectiveness of different variants of MCMC it is natural to say "forget about theory – just see what works best on real examples". But such experimental evaluation is itself conceptually difficult: **pragmatism is easier in theory than in practice**!

11.1.2 Further aspects of applied MCMC

Sampling from one-dimensional distributions. Consider a probability distribution μ on R^1 with density function f and and distribution function F. In one sense, sampling from μ is easy, because of the elementary result that $F^{-1}(U)$ has distribution μ, where U is uniform on $[0,1]$ and $x = F^{-1}(u)$ is the inverse function of $u = F(x)$. In cases where we have an explicit formula for F^{-1}, we are done. Many other cases can be done using *rejection sampling*. Suppose there is some other density g from which we can sample by the inverse distribution function method, and suppose we know a bound $c \geq \sup_x f(x)/g(x)$. Then the algorithm

> propose a sample x from $g(\cdot)$;
> accept x with probability $\frac{f(x)}{cg(x)}$; else propose a new sample from g

produces an output with density $f(\cdot)$ after mean c steps. By combining these two methods, libraries of algorithms for often-encountered one-dimensional distributions can be built, and indeed exist in statistical software packages.

But what about a general density $f(x)$? If we need to sample many times from the same density, it is natural to use deterministic numerical methods. First probe f at many values of x. Then either
(a) build up a numerical approximation to F and thence to F^{-1}; or

(b) choose from a library a suitable density g and use rejection sampling. The remaining case, which is thus the only "hard" aspect of sampling from one-dimensional distributions, is where we only need one sample from a general distribution. In other words, where we want many samples which are all from different distributions. This is exactly the setting of the Gibbs sampler where the target multidimensional density is complicated, and thus motivates some of the variants we discuss in section 11.3.

Practical relevance of theoretical mixing time parameters. Standard theory from Chapter 4 (yyy cross-refs) relates τ_2 to the asymptotic variance rate $\sigma^2(g)$ at (11.2) for the "worst-case" g:

$$\tau_2 = \frac{1}{1-\lambda_2} \approx \frac{1+\lambda_2}{1-\lambda_2} = \sup_g \frac{\sigma^2(g)}{\text{var}_{\,\pi} g}. \tag{11.4}$$

Moreover Proposition 29 of Chapter 4 (yyy 10/11/94 version) shows that $\sigma^2(g)$ also appears in an upper bound on variances of finite-time averages from the *stationary* chain. So in asking how long to run MCMC simulations, a natural principle (not practical, of course, because we typically don't know τ_2) is

> base estimates on t steps, where t is a reasonable large multiple of τ_2.

But this principle can be attacked from opposite directions. It is sometimes argued that worrying about τ_2 (corresponding to the *worst-case* g) is overly pessimistic in the context of studying some specific g. For instance, Sokal [311] p. 8 remarks that in natural statistical physics models on the infinite lattice near a phase transition in a parameter θ, as θ tends to the critical point the growth exponent of $\sigma^2(g)$ for "interesting" g is typically different from the growth exponent of τ_2. Madras and Slade [252] p. 326 make similar remarks in the context of the pivot algorithm for self-avoiding walk. But we do not know similar examples in the statistical R^d setting. In particular, in the presence of multimodality such counterexamples would require that g be essentially "orthogonal" to the differences between modes, which seems implausible.

Burn-in, the time t_0 excluded from the estimator (11.3) to avoid undue influence of initial state, is conceptually more problematic. Theory says that taking t_0 as a suitable multiple of τ_1 would guarantee reliable estimates. The general fact $\tau_1 \geq \tau_2$ then suggests that allowing sufficient burn-in time is a stronger requirement than allowing enough "mixing" for the stationary chain – so the principle above is overly optimistic. On the other hand, because it

refers to worst-case initial state, requiring a burn-in time of τ_1 seems far too conservative in practice. The bottom line is that one cannot eliminate the possibility of metastability error; in general, all one gets from multiple-runs and diagnostics is confidence that one is sampling from a single potential well, in the imagery below (though section 11.6.2 indicates a special setting where we can do better).

Statistical physics imagery. Any probability distribution π can be written as

$$\pi(x) \propto \exp(-H(x)).$$

One can call H a *potential function*; note that a mode (local maximum) of π is a local minimum of H. One can envisage a realization of a Markov chain as a particle moving under the influence of both a potential function (the particle responds to some "force" pushing it towards lower values of H) and random noise. Associated with each local minimum y of H is a *potential well*, which we envisage as the set of points which under the influence of the potential only (without noise) the particle would move to y (in terms of π, states from which a "steepest ascent" path leads to y).

A fundamental intuitive picture is that the main reason why a reversible chain may relax slowly is that there is more than one potential well, and the chain takes a long time to move from one well to another. In such a case, π conditioned to a single potential well will be a *metastable* (i.e. almost-stationary) distribution. One expects the chain's distribution, from any initial state, to reach fairly quickly one (or a mixture) of these metastable distributions, and then the actual relaxation time to stationarity is dominated by the times taken to move between wells. In more detail, if there are w wells then one can consider, as a coarse-grained approximation, a w-state continuous-time chain where the transition rates $w_1 \to w_2$ are the rates of moving from well w_1 to well w_2. Then τ_2 for the original chain should be closely approximated by τ_2 for the coarse-grained chain.

The hierarchical Normal model. As a very simple instance of (11.1), take $d = 1, p = 2$ and $x \to \phi(\mu, \sigma^2, x)$ the Normal(μ, σ^2) density. Then let (μ, σ) be chosen independently for each individual from some joint density $h(\mu, \sigma)$ on $R \times R^+$. The data is an n-vector $\mathbf{x} = (x_1, \ldots, x_n)$ and the full posterior distribution is

$$f_{\mathbf{x}}(\mu_1, \ldots, \mu_n, \sigma_1, \ldots, \sigma_n) = z_{\mathbf{x}}^{-1} \prod_{i=1}^{n} h(\mu_i, \sigma_i)\phi(\mu_i, \sigma_i^2, x_i).$$

Typically we are interested in a posterior mean of μ_i for fixed i, that is $\bar{g}$ for

$$g(\mu_1, \ldots, \mu_n, \sigma_1, \ldots, \sigma_n) := \mu_i.$$

Pragmatism is easier in theory than in practice. In comparing MCMC methods experimentally, one obvious issue is the choice of example to study. Another issue is that, if we measure "time" as "number of steps", then a step of one chain may not be comparable with a step of another chain. For instance, a Metropolis step is typically easier to implement than a Gibbs step. More subtlely, in combinatorial examples there may be different ways to set up a data structure to represent the current state in a way that permits easy computation of π-values. The alternative of measuring "time" as CPU time introduces different problems – details of coding matter.

11.2 The two basic schemes

We will present general definitions and discussion in the context of finite-state chains on a state space S; translating to continuous state space such as R^d involves slightly different notation without any change of substance.

11.2.1 Metropolis schemes

Write $K = (k_{xy})$ for a *proposal* transition matrix on S. The simplest case is where K is symmetric ($k_{xy} \equiv k_{yx}$). In this case, given π on S we define a step $x \to x'$ of the associated *Metropolis chain* in words by

- pick y from $k(x, \cdot)$ and propose a move to y;
- accept the move (i.e. set $x' = y$) with probability $\min(1, \pi_y/\pi_x)$, otherwise stay ($x' = x$).

This recipe defines the transition matrix P of the Metropolis chain to be

$$p_{xy} = k_{xy} \min(1, \pi_y/\pi_x), \quad y \neq x.$$

Assuming K is irreducible and π strictly positive, then clearly P is irreducible. Then since $\pi_x p_{xy} = k_{xy} \min(\pi_x, \pi_y)$, symmetry of K implies P satisfies the detailed balance equations and so is reversible with stationary distribution π.

The general case is where K is an arbitrary transition matrix, and the acceptance rule becomes

- accept a proposed move $x \to y$ with probability $\min(1, \frac{\pi_y k_{yx}}{\pi_x k_{xy}})$.

The transition matrix of the Metropolis chain becomes

$$p_{xy} = k_{xy} \min\left(1, \frac{\pi_y k_{yx}}{\pi_x k_{xy}}\right), \quad y \neq x. \tag{11.5}$$

To ensure irreducibility, we now need to assume connectivity of the graph on S whose edges are the (x, y) such that $\min(k_{xy}, k_{yx}) > 0$. Again detailed balance holds, because

$$\pi_x p_{xy} = \min(\pi_x k_{xy}, \pi_y k_{yx}), \quad y \neq x.$$

The general case is often called *Metropolis-Hastings* – see Notes for terminological comments.

11.2.2 Line-sampling schemes

The abstract setup described below comes from Diaconis [113]. Think of each S_i as a line, i.e. the set of points in a line.

Suppose we have a collection (S_i) of subsets of state space S, with $\cup_i S_i = S$. Write $I(x) := \{i : x \in S_i\}$. Suppose for each $x \in S$ we are given a probability distribution $i \to w(i, x)$ on $I(x)$, and suppose

$$\text{if } x, y \in S_i \text{ then } w(i, x) = w(i, y). \tag{11.6}$$

Write $\pi^{[i]}(\cdot) = \pi(\cdot|S_i)$. Define a step $x \to y$ of the *line-sampling chain* in words by

- choose i from $w(\cdot, x)$;
- then choose y from $\pi^{[i]}$.

So the chain has transition matrix

$$p_{xy} = \sum_{i \in I(x)} w(i, x) \pi^{[i]}_y, \quad y \neq x.$$

We can rewrite this as

$$p_{xy} = \sum_{i \in I(x) \cap I(y)} w(i, x) \pi_y / \pi(S_i)$$

and then (11.6) makes it clear that $\pi_x p_{xy} = \pi_y p_{yx}$. For irreducibility, we need the condition

the union over i of the edges in the complete graphs

on S_i form a connected graph on S. (11.7)

Note in particular we want the S_i to be overlapping, rather than a partition.

This setting includes many examples of random walks on combinatorial sets. For instance, card shuffling by random transpositions (yyy cross-ref) is essentially the case where the collection of subsets consists of all 2-card subsets. In the R^d setting, with target density f, the Gibbs sampler is the case where the collection consists of all lines parallel to some axis. Taking instead all lines in all directions gives the *hit-and-run* sampler, for which a step from x is defined as follows.

- Pick a direction uniformly at random, i.e. a point y on the surface on the unit ball.
- Step from x to $x + Uy$, where $-\infty < U < \infty$ is chosen with density proportional to
$$u^{d-1} f(x + uy).$$

The term u^{d-1} here arises as a Jacobean; see Liu [235] Chapter 8 for explanation and more examples in R^d.

11.3 Variants of basic MCMC

11.3.1 Metropolized line sampling

Within the Gibbs or hit-and-run scheme, at each step one needs to sample from a one-dimensional distribution, but a different one-dimensional distribution each time. As mentioned in section 11.1.2, this is in general not easy to implement efficiently. An alternative is *Metropolized line sampling*, where one instead takes a single step of a Metropolis (i.e. propose/accept) chain with the correct stationary distribution. To say the idea abstractly, in the general "line sampling" setting of section 11.2.2, assume also:

for each i we have an irreducible transition matrix K^i on S_i whose stationary distribution is $\pi^{[i]}$.

Then define a step $x \to y$ of the Metropolized line sampler as

- choose i from $w(\cdot, x)$;

- then choose y from $k^i(x, \cdot)$.

It is easy to check that the chain has stationary distribution π, and is reversible if the K^i are reversible, so in particular if the K^i are defined by a Metropolis-type propose-accept scheme. In the simplest setting where the line sampler is the Gibbs sampler and we use the same one-dimensional proposal step distribution each time, this scheme is *Metropolis-within-Gibbs*. In that context is seems intuitively natural to use a long-tailed proposal distribution such as the Cauchy distribution. Because we might encounter wildly different one-dimensional target densities, e.g. one density with s.d. $1/10$ and another with two modes separated by 10, and using a $U(-L, L)$ step proposal would be inefficient in the latter case if L is small, and inefficient in the former case if L is large. Intuitively, a long-tailed distribution avoids these worst cases, at the cost of having the acceptance rate be smaller in good cases.

11.3.2 Multiple-try Metropolis

In the setting (section 11.2.1) of the Metropolis scheme, one might consider making several draws from the proposal distribution and choosing one of them to be the proposed move. Here is one way, suggested by Liu et al [236], to implement this idea. It turns out that to ensure the stationary distribution is the target distribution π, we need extra samples which are used only to adjust the acceptance probability of the proposed step.

For simplicity, we take the case of a symmetric proposal matrix K. Fix $m \geq 2$. Define a step from x of the *multiple-try Metropolis* (MTM) chain as follows.

- Choose $y_1, \ldots, y_m$ independently from $k(x, \cdot)$;
- Choose y_i with probability proportional to $\pi(y_i)$;
- Choose $x_1, \ldots, x_{m-1}$ independently from $k(y_i, \cdot)$, and set $x_m = x$;
- Accept the proposed move $x \to y_i$ with probability $\min\left(1, \frac{\sum_i \pi(y_i)}{\sum_i \pi(x_i)}\right)$.

Irreducibility follows from irreducibility of K. To check detailed balance, write the acceptance probability as $\min(1, q)$. Then

$$p_{xy} = m k_{xy} \sum \prod_{i=1}^{m-1} k_{x,y_i} \prod_{i=1}^{m-1} k_{y,x_i} \frac{\pi_y}{\sum_i \pi_{y_i}} \min(1, q)$$

where the first sum is over ordered $(2m-2)$-tuples $(y_1, \ldots, y_{m-1}, x_1, \ldots, x_{m-1})$. So we can write

$$\pi_x p_{xy} = m k_{xy} \pi_x \pi_y \sum \prod_{i=1}^{m-1} k_{x,y_i} \prod_{i=1}^{m-1} k_{y,x_i} \min\left(\frac{1}{\sum_i \pi_{y_i}}, \frac{q}{\sum_i \pi_{y_i}}\right).$$

The choice of q makes the final term become $\min(\frac{1}{\sum_i \pi_{y_i}}, \frac{1}{\sum_i \pi_{x_i}})$. One can now check $\pi_x p_{xy} = \pi_y p_{yx}$, by switching the roles of x_j and y_j.

To compare MTM with single-try Metropolis, consider the $m \to \infty$ limit, in which the empirical distribution of $y_1, \ldots, y_m$ will approach $k(x, \cdot)$, and so the distribution of the chosen y_i will approach $k(x, \cdot)\pi(\cdot)/a_x$ for $a_x := \sum_y k_{xy}\pi_y$. Thus for large m the transition matrix of MTM will approximate

$$p^{\infty}_{xy} = \frac{k_{xy}\pi_y}{a_x} \min(1, a_x/a_y), \quad y \neq x.$$

To compare with single-try Metropolis P, rewrite both as

$$\begin{aligned} p^{\infty}_{xy} &= k_{xy}\pi_y \min\left(\tfrac{1}{a_x}, \tfrac{1}{a_y}\right), \quad y \neq x \\ p_{xy} &= k_{xy}\pi_y \min\left(\tfrac{1}{\pi_x}, \tfrac{1}{\pi_y}\right), \quad y \neq x. \end{aligned}$$

Thinking of a step of the proposal chain as being in a random direction unrelated to the behavior of π, from a π-typical state x we expect a proposed move to tend to make π decrease, so we expect $a_x < \pi_x$ for π-typical x. In this sense, the equations above show that MTM is an improvement. Of course, if we judge "cost" in terms of the number of evaluations of π_x, then a step of MTM costs $2m-1$ times the cost of single-step Metropolis. By this criterion it seems implausible that MTM would be cheaper than single-step. On the other hand one can envisage settings where there is substantial cost in updating a data structure associated with the current state x, and in such a setting MTM may be more appealing.

11.3.3 Multilevel sampling

Writing $\pi(x) \propto \exp(-H(x))$, as in the statistical physics imagery (section 11.1.2), suggests defining a one-parameter family of probability distributions by

$$\pi_\theta(x) \propto \exp(-\theta H(x)).$$

(In the physics analogy, θ corresponds to 1/temperature). If π is multimodal we picture π_θ, as θ increases from 0 to 1, interpolating between the uniform

distribution and π by making the potential wells grow deeper. Fix a proposal matrix K, and let P_θ be the transition matrix for the Metropolized chain (11.5) associated with K and π_θ. Now fix L and values $0 = \theta_1 < \theta_2 < \ldots < \theta_L = 1$. The idea is that for small θ the P_θ-chain should have less difficulty moving between wells; for $\theta = 1$ we get the correct distribution within each well; so by varying θ we can somehow sample accurately from all wells. There are several ways to implement this idea. *Simulated tempering* [254] defines a chain on state space $S \times \{1, \ldots, L\}$, where state (x, i) represents configuration x and parameter θ_i, and where each step is either of the form

- $(x, i) \to (x', i)$; $x \to x'$ a step of P_{θ_i}

or of the form

- $(x, i) \to (x, i')$; where $i \to i'$ is a proposed step of simple random walk on $\{1, 2, \ldots, L\}$.

However, implementing this idea is slightly intricate, because normalizing constants z_θ enter into the desired acceptance probabilities. A more elegant variation is the *multilevel exchange chain* suggested by Geyer [163] and implemented in statistical physics by Hukushima and Nemoto [185]. First consider L independent chains, where the i'th chain $X_t^{(i)}$ has transition matrix P_{θ_i}. Then introduce an interaction; propose to switch configurations $X^{(i)}$ and $X^{(i+1)}$, and accept with the appropriate probability. Precisely, take state space S^L with states $\mathbf{x} = (x_1, \ldots, x_L)$. Fix a (small) number $0 < \alpha < 1$.

- With probability $1 - \alpha$ pick i uniformly from $\{1, \ldots, L\}$, pick x_i' according to $P_{\theta_i}(x_i, \cdot)$ and update $\mathbf{x}$ by changing x_i to x_i'.
- With probability α, pick uniformly an adjacent pair $(i, i+1)$, and propose to update $\mathbf{x}$ by replacing (x_i, x_{i+1}) by (x_{i+1}, x_i). Accept this proposed move with probability

$$\min\left(1, \frac{\pi_{\theta_i}(x_{i+1})\pi_{\theta_{i+1}}(x_i)}{\pi_{\theta_i}(x_i)\pi_{\theta_{i+1}}(x_{i+1})}\right).$$

To check that the product $\pi = \pi_{\theta_1} \times \ldots \times \pi_{\theta_L}$ is indeed a stationary distribution, write the acceptance probability as $\min(1, q)$. If $\mathbf{x}$ and $\mathbf{x}'$ differ only by interchange of (x_i, x_{i+1}) then

$$\frac{\pi(\mathbf{x})}{\pi(\mathbf{x}')}\frac{p(\mathbf{x}, \mathbf{x}')}{p(\mathbf{x}', \mathbf{x})} = \frac{\pi(\mathbf{x})}{\pi(\mathbf{x}')}\frac{\frac{\alpha}{L-1}\min(1, q)}{\frac{\alpha}{L-1}\min(1, q^{-1})} = \frac{\pi(\mathbf{x})}{\pi(\mathbf{x}')}q$$

and the definition of q makes the expression $= 1$. The case of steps where only one component changes is easier to check.

11.3.4 Multiparticle MCMC

Consider the setting of section 11.2.2. There is a target distribution π on S and a collection of subsets (S_i). Write $\pi^{[i]} = \pi(\cdot|S_i)$ and $I(x) = \{i : x \in S_i\}$. Now fix $m \geq 2$. We can use the line-sampling scheme of section 11.2.2 to define (recall Chapter 4 section 6.2) (yyy 10/11/94 version) a product chain on S^m with stationary distribution $\pi \times \pi \times \ldots \times \pi = \pi^k$. For this product chain, picture m particles, at each step picking a random particle and making it move as a step from the line-sampling chain. Now let us introduce an interaction: the line along which a particle moves may depend on the positions of the other particles.

Here is a precise construction. Suppose that for each $(x, \hat{\mathbf{x}}) \in S \times S^{m-1}$ we are given a probability distribution $w(\cdot, x, \hat{\mathbf{x}})$ on $I(x)$ satisfying the following analog of (11.6):

$$\text{if } x, y \in S_i \text{ then } w(i, x, \hat{\mathbf{x}}) = w(i, y, \hat{\mathbf{x}}). \tag{11.8}$$

A step of the chain from (x_i) is defined by

- Pick k uniformly from $\{1, 2, \ldots, m\}$
- Pick i from $w(\cdot, x_k, (x_i, i \neq k))$
- Pick x'_k from $\pi^{[i]}(\cdot)$
- Update (X_i) by replacing x_k by x'_k.

It is easy to check that π^m is indeed a stationary distribution; and the chain is irreducible under condition (11.7). Of course we could, as in section 11.3.1, use a Metropolis step instead of sampling from $\pi^{[K]}$.

Constructions of this type in statistical applications on R^d go back to Gilks et al [166], under the name *adaptive directional sampling*. In particular they suggested picking a distinct pair (j, k) of the "particles" and taking the straight line through x_j and x_k as the line to sample x'_k from. Liu et al [236] suggest combining this idea with mode-hunting. Again pick a distinct pair (j, k) of "particles"; but now use some algorithm to find a local maximum $m(x_j)$ of the target density starting from x_j, and sample x'_k from the line through x_k and $m(x_j)$.

11.4 A little theory

The chains designed for MCMC in previous sections are reversible, and therefore the theory of reversible chains developed in this book is available.

Unfortunately there is very little extra to say – in that sense, there is no "theory of MCMC". What follows is rather fragmentary observations.

11.4.1 Comparison methods

Consider the Metropolis chain

$$p_{xy}^{\text{Metro}} = k_{xy} \min\left(1, \frac{\pi_y k_{yx}}{\pi_x k_{xy}}\right), \quad y \neq x.$$

The requirement that a step of a chain be constructible as a proposal from K followed by acceptance/rejection, is the requirement that $p_{xy} \leq k_{xy},\ y \neq x$. Recall the asymptotic variance rate

$$\sigma^2(P, f) := \lim_t t^{-1} \text{var} \sum_{s=1}^{t} f(X_s).$$

Lemma 11.1 (Peskun's Theorem [280]) *Given K and π, let P be a reversible chain with $p_{xy} \leq k_{xy},\ y \neq x$ and with stationary distribution π. Then $\sigma^2(P, f) \geq \sigma^2(P^{\text{Metro}}, f)\ \forall f$.*

Proof. Reversibility of P implies

$$p_{xy} = \frac{\pi_y p_{yx}}{\pi_x} \leq \frac{\pi_y k_{yx}}{\pi_x} = k_{xy} \frac{\pi_y k_{yx}}{\pi_x k_{xy}}$$

and hence

$$p_{xy} = p_{xy}^{\text{Metro}} \beta_{xy}$$

where $\beta_{xy} = \beta_{yx} \leq 1,\ y \neq x$. So the result follows directly from Peskun's lemma (yyy Lemma 11.5, to be moved elsewhere). □

This result can be interpreted as saying that the Metropolis rates (11.5) are the optimal way of implementing a proposal-rejection scheme. Loosely speaking, a similar result holds in any natural Metropolis-like construction of a reversible chain using a $\max(1, \cdot)$ acceptance probability.

It is important to notice that Lemma 11.1 does not answer the following question, which (except for highly symmetric graphs) seems intractable.

Question. Given a connected graph and a probability distribution π on its vertices, consider the class of reversible chains with stationary distribution π and with transitions only across edges of the graph. Within that class, which chain has smallest relaxation time?

Unfortunately, standard comparison theorems don't take us much further in comparing MCMC methods. To see why, consider Metropolis on R^d with isotropic Normal$(0, \sigma^2 \mathbf{I}_d)$ proposal steps. This has some relaxation time $\tau_2(f, \sigma)$, where f is the target density. For $\sigma_1 < \sigma_2$, the normal densities $g_\sigma(x)$ satisfy $g_{\sigma_2}(x)/g_{\sigma_1}(x) \geq (\sigma_1/\sigma_2)^d$. So the comparison theorem (Chapter 3 Lemma 29) (yyy 9/2/94 version) shows

$$\tau_2(f, \sigma_2) \geq (\sigma_1/\sigma_2)^d \tau_2(f, \sigma_2), \quad \sigma_1 < \sigma_2.$$

But this is no help in determining the optimal σ.

11.4.2 Metropolis with independent proposals

Though unrealistic in practical settings, the specialization of the Metropolis chain to the case where the proposal chain is i.i.d., that is where $k_{xy} = k_y$, is mathematically a natural object of study. In this setting the transition matrix (11.5) becomes

$$p_{xy} = k_y \min(1, w_y/w_x), \quad y \neq x$$

where $w_x := \pi_w/k_x$. It turns out there is a simple and sharp coupling analysis, based on the trick of labeling states as $1, 2, \ldots, n$ so that $w_1 \geq w_2 \geq \ldots \geq w_n$ (Liu [234] used this trick to give an eigenvalue analysis, extending part (b) below). Let ρ be the chance that a proposed step from state 1 is rejected (count a proposed step from state 1 to 1 as always accepted). So

$$\rho = \sum_{i=1}^{n} k_i (1 - \tfrac{w_i}{w_1}) < 1.$$

Proposition 11.2 *For the Metropolis chain over independent proposals, with states ordered as above,*
(a) $\bar{d}(t) \leq \rho^t$
(b) The relaxation time $\tau_2 = (1 - \rho)^{-1}$.

Proof. For the chain started at state 1, the time T of the first acceptance of a proposed step satisfies

$$P(T > t) = \rho^t.$$

Recall from (yyy Chapter 4-3 section 1; 10/11/99 version) the notion of *coupling*. For this chain a natural coupling is obtained by using the same $U(0, 1)$ random variable to implement the accept/reject step (accept if $U <$ $P(\text{accept})$) in two versions of the chain. It is easy to check this coupling

(X_t, X'_t) respects the ordering: if $X_0 \le X'_0$ then $X_t \le X'_t$. At time T the fact that a proposed jump from 1 is accepted implies that a jump from any other state must be accepted. So T is a coupling time, and the coupling inequality (yyy Chapter 4-3 section 1.1; 10/11/99 version) implies $\bar{d}(t) \le P(T > t)$. This establishes (a), and the general inequality $\bar{d}(t) = \Omega(\lambda_2^t)$ implies $\lambda_2 \le \rho$. On the other hand, for the chain started at state 1, on $\{T = 1\}$ the time-1 distribution is π; in other words

$$P_1(X_1 \in \cdot) = \rho\delta_1(\cdot) + (1 - \rho)\pi(\cdot).$$

But this says that ρ is an eigenvalue of P (corresponding to the eigenvector $\delta_1 - \pi$), establishing (b). □

In the continuous-space setting, with a proposal distribution uniform on $[0, 1]$ and target density f with $f^* := \max_x f(x)$, part (b) implies the relaxation time τ_2 equals f^*. So (unsurprisingly) Metropolis-over-independent is comparable to the basic rejection sampling scheme (section 11.1.2), which gives an exact sample in mean f^* steps.

11.5 The diffusion heuristic for optimal scaling of high dimensional Metropolis

In any Metropolis scheme for sampling from a target distribution on R^d, there arises the question of how large to take the steps of the proposal chain. One can answer this for isotropic-proposal schemes in high dimensions, in the setting where the target is a product distribution, and the result in this (very artificial) setting provides a heuristic for more realistic settings exemplified by (11.1).

11.5.1 Optimal scaling for high-dimensional product distribution sampling

Fix a probability density function f on R^1. For large d consider the i.i.d. product distribution $\pi_f(d\mathbf{x}) = \prod_{i=1}^d f(x_i)\, dx_i$ for $\mathbf{x} = (x_i) \in R^d$. Suppose we want to sample from π_f using Metropolis or Gibbs; what is the optimal scaling (as a function of d) for the step size of the proposal chain, and how does the relaxation time scale?

For the Gibbs sampler this question is straightforward. Consider the one-dimensional case, and take the proposal step increments to be Normal$(0, \sigma^2)$. Then (under technical conditions on f – we omit technical conditions here and in Theorem 11.3) the Gibbs chain will have some finite relaxation time

depending on f and σ, and choosing the optimal σ^* gives a relaxation time $\tau_2(f)$, say. The Gibbs sampler chain in which we choose a random coordinate and propose changing only that coordinate (using the optimal σ^* above) is a product chain in the sense of Chapter 4 section 6.2 (yyy 10/11/94 version), and so the relaxation time of this product chain is $\tau_2^{\text{Gibbs}}(f) = \tau_2(f)\ d$.

Though the argument above is very simple, it is unsatisfactory because there is no simple expression for relaxation time as a function of σ or for the optimal σ^*. It turns out that this difficulty is eliminated in the isotropic-proposal Metropolis chain. In the Gibbs sampler above, the variance of the length of a proposed step is σ^2, so we retain this property by specifying the steps of the proposal chain to have Normal$(0, \sigma^2 d^{-1}\mathbf{I}_d)$ distribution. One expects the relaxation time to grow linearly in d in this setting also. The following result of Roberts et al [293] almost proves this, and has other useful corollaries.

Theorem 11.3 *Fix $\sigma > 0$. Let $(\mathbf{X}(t), t = 0, 1, 2, \ldots)$ be the Metropolis chain for sampling from product measure π_f on R^d based on a proposal random walk with step distribution Normal$(0, \sigma^2 d^{-1}\mathbf{I}_d)$. Write $X^{(1)}(t)$ for the first coordinate of $\mathbf{X}(t)$, and let $Y_d(t) := X^{(1)}(\lfloor td \rfloor)$ be this coordinate process speeded up by a factor d, for continuous $0 \le t < \infty$. Suppose $\mathbf{X}(0)$ has the stationary distribution π_f. Then*

$$(Y_d(t), 0 \le t < \infty) \stackrel{d}{\to} (Y(t), 0 \le t < \infty) \text{ as } d \to \infty \tag{11.9}$$

where the limit process is the stationary one-dimensional diffusion

$$dY_t = \theta^{1/2} dW_t + \theta \mu(Y_t) dt \tag{11.10}$$

for standard Brownian motion W_t, where

$$\begin{aligned} \mu(y) &:= \tfrac{f'(y)}{2f(y)} \\ \theta &:= 2\sigma^2 \Phi(-\sigma\kappa/2) \text{ where } \Phi \text{ is the Normal distribution function} \\ \kappa &:= \left(\int \tfrac{(f'(x))^2}{f(x)}\ dx \right)^{1/2}. \end{aligned}$$

Moreover, as $d \to \infty$ the proportion of accepted proposals in the stationary chain tends to $2\Phi(-\sigma\kappa/2)$.

We outline the proof in section 11.5.3. The result may look complicated, so one piece of background may be helpful. Given a probability distribution on the integers, there is a Metropolis chain for sampling from it based on the

simple random walk proposal chain. As a continuous-space analog, given a density f on R^1 there is a "Metropolis diffusion" with stationary density f based on $\theta^{1/2}W_t$ (for arbitrary constant θ) as "proposal diffusion", and this Metropolis diffusion is exactly the diffusion (11.10): see Notes to (yyy final Chapter).

Thus the appearance of the limit diffusion Y is not unexpected; what is important is the explicit formula for θ in terms of σ and f. Note that the parameter θ affects the process (Y_t) only as a speed parameter. That is, if Y^*_t is the process (11.10) with $\theta = 1$ then the general process can be represented as $Y_t = Y^*_{\theta t}$. In particular, the relaxation time scales as $\tau_2(Y) = \theta^{-1}\tau_2(Y^*)$. Thus we seek to maximize θ as a function of the underlying step variance σ, and a simple numerical calculation shows this is maximized by taking $\sigma = 2.38/\kappa$, giving $\theta = 1.3/\kappa^2$.

Thus Theorem 11.3 suggests that for the Metropolis chain $\mathbf{X}$, the optimal variance is $2.38^2\kappa^{-2}d^{-1}\mathbf{I}_d$, and suggests that the relaxation time $\tau_2(f,d)$ scales as

$$\tau_2(f,d) \sim \frac{d\kappa^2}{1.3}\ \tau_2(Y^*). \tag{11.11}$$

In writing (11.11) we are pretending that the Metropolis chain is a product chain (so that its relaxation time is the relaxation time of its individual components) and that relaxation time can be passed to the limit in (11.9). Making a rigorous proof of (11.11) seems hard.

11.5.2 The diffusion heuristic.

Continuing the discussion above, Theorem 11.3 says that the long-run proportion of proposed moves which are accepted is $2\Phi(-\kappa\sigma/2)$. At the optimal value $\sigma = 2.38/\kappa$ we find this proportion is a "pure number" 0.23, which does not depend on f. To quote [293]

> This result gives rise to the useful heuristic for random walk Metropolis in practice:
>
> *Tune the proposal variance so that the average acceptance rate is roughly* 1/4.

We call this the *diffusion heuristic* for proposal-step scaling. Intuitively one might hope that the heuristic would be effective for fairly general *unimodel* target densities on R^d, though it clearly has nothing to say about the problem of passage between modes in a multimodal target. Note also that to invoke the diffusion heuristic in a combinatorial setting, where the proposal

chain is random walk on a graph, one needs to assume that the target distribution is "smooth" in the sense that $\pi(v)/\pi(w) \approx 1$ for a typical edge (v, w). In this case one can make a Metropolis chain in which the proposal chain jumps σ edges in one step, and seek to optimize σ. See Roberts [292] for some analysis in the context of smooth distributions on the d-cube. However, such smoothness assumptions seem inapplicable to most practical combinatorial MCMC problems.

11.5.3 Sketch proof of Theorem

Write a typical step of the proposal chain as

$$(x_1, x_2, \ldots, x_d) \to (x_1 + \xi_1, x_2 + \xi_2, \ldots, x_d + \xi_d).$$

Write

$$J = \log \frac{f(x_i + \xi_1)}{f(x_1)}; \quad S = \log \prod_{i=2}^{d} \frac{f(x_i + \xi_i)}{f(x_i)}.$$

The step is accepted with probability $\min(1, \prod_{i=1}^{d} \frac{f(x_i+\xi_i)}{f(x_i)}) = \min(1, e^{J+S})$. So the increment of the first coordinate of the Metropolis chain has mean and mean-square $E\xi_1 \ \min(1, e^{J+S})$ and $E\xi_1^2 \ \min(1, e^{J+S})$. The essential issue in the proof is to show that, for "typical" values of $(x_2, \ldots, x_n)$,

$$E\xi_1 \ \min(1, e^{J+S}) \ \sim \ \theta\mu(x_1)/d \tag{11.12}$$
$$E\xi_1^2 \ \min(1, e^{J+S}) \ \sim \ \theta/d. \tag{11.13}$$

This identifies the asymptotic drift and variance rates of $Y_d(t)$ with those of $Y(t)$.

Write $h(u) := E \min(1, e^{u+S})$. Since

$$J \approx \log\left(1 + \tfrac{f'(x_1)}{f(x_1)}\xi_1\right) \approx \tfrac{f'(x_1)}{f(x_1)}\xi_1 = 2\mu(x_1)\xi_1,$$

the desired estimates (11.12,11.13) can be rewritten as

$$EJh(J) \ \sim \ 2\theta\mu^2(x_1)/d \tag{11.14}$$
$$EJ^2h(J) \ \sim \ 4\theta\mu^2(x_1)/d. \tag{11.15}$$

Now if J has Normal$(0, \beta^2)$ distribution then for sufficiently regular $h(\cdot)$ we have

$$EJh(J) \sim \beta^2 h'(0); \quad EJ^2h(J) \sim \beta^2 h(0) \text{ as } \beta \to 0.$$

Since J has approximately Normal$(0, 4\mu^2(x_1)$var $\xi_1 = 4\mu^2(x_1)\sigma^2/d)$ distribution, proving (11.14,11.15) reduces to proving

$$h'(0) \rightarrow \frac{\theta}{2\sigma^2} \tag{11.16}$$

$$h(0) \rightarrow \frac{\theta}{\sigma^2}. \tag{11.17}$$

We shall argue

$$\text{dist}(S) \text{ is approximately Normal}(-\kappa^2\sigma^2/2, \kappa^2\sigma^2). \tag{11.18}$$

Taking the first two terms in the expansion of $\log(1+u)$ gives

$$\log \tfrac{f(x_i+\xi_i)}{f(x_i)} \approx \tfrac{f'(x_i)}{f(x_i)}\ \xi_i - \tfrac{1}{2}\left(\tfrac{f'(x_i)}{f(x_i)}\right)^2 \xi_i^2.$$

Write $K(\mathbf{x}) = d^{-1}\sum_{i=2}^{d}(\frac{f'(x_i)}{f(x_i)})^2$. Summing the previous approximation over i, the first sum on the right has approximately Normal$(0, \sigma^2 K(\mathbf{x}))$ distribution, and (using the weighted law of large numbers) the second term is approximately $-\frac{1}{2}\sigma^2 K(\mathbf{x})$. So the distribution of S is approximately Normal$(-K(\mathbf{x})\sigma^2/2, K(\mathbf{x})\sigma^2)$. But by the law of large numbers, for a typical $\mathbf{x}$ drawn from the product distribution π_f we have $K(\mathbf{x}) \approx \kappa^2$, giving (11.18).

To argue (11.17) we pretend S has exactly the Normal distribution at (11.18). By a standard formula, if S has Normal(α, β^2) distribution then

$$E\max(1, e^S) = \Phi(\alpha/\beta) + e^{\alpha+\beta^2/2}\Phi(-\beta - \alpha/\beta).$$

This leads to

$$h(0) = 2\Phi(-\kappa\sigma/2)$$

which verifies (11.17). From the definition of $h(u)$ we see

$$h'(0) = Ee^S 1_{(S\le 0)} = h(0) - P(S \ge 0) = h(0) - \Phi(-\kappa\sigma/2) = \Phi(-\kappa\sigma/2)$$

which verifies (11.16).

11.6 Other theory

11.6.1 Sampling from log-concave densities

As mentioned in Chapter 9 section 5.1 (yyy version 9/1/99) there has been intense theoretical study of the problem of sampling uniformly from a convex

set in R^d, in the $d \to \infty$ limit. This problem turns out to be essentially equivalent to the problem of sampling from a *log-concave* density f, that is a density of the form $f(x) \propto \exp(-H(x))$ for convex H. The results are not easy to state; see Bubley et al [80] for discussion.

11.6.2 Combining MCMC with slow exact sampling

Here is a special setting in which one can make rigorous inferences from MCMC without rigorous bounds on mixing times. Suppose we have a guess $\hat{\tau}$ at the relaxation time of a Markov sampler from a traget distribution π; suppose we have some separate method of sampling exactly from π, but where the cost of one exact sample is larger than the cost of $\hat{\tau}$ steps of the Markov sampler. In this setting it is natural to take m exact samples and use them as initial states of m multiple runs of the Markov sampler. It turns out (see [19] for precise statement) that one can obtain confidence intervals for a mean $\bar{g}$ which are always rigorously correct (without assumptions on τ_2) and which, if $\hat{\tau}$ is indeed approximately τ_2, will have optimal length, that is the length which would be implied by this value of τ_2.

11.7 Notes on Chapter MCMC

Liu [235] provides a nice combination of examples and carefully-described methodology in MCMC, emphasizing statistical applications but also covering some statistical physics. Other statistically-oriented books include [91, 165, 291]. We should reiterate that most MCMC "design" ideas originated in statistical physics; see the extensive discussion by Sokal [311]. Neal [267] focuses on neural nets but contains useful discussion of MCMC variants.

Section 11.1.1. In the single-run setting, the variance of sample means (11.3) could be estimated by classical methods of time series [63].

The phrase *metastability error* is our coinage – though the idea is standard, there seems no standard phrase.

Elaborations of the multiple-runs method are discussed by Gelman and Rubin [161]. The applied literature has paid much attention to diagnostics: for reviews see Cowles and Carlin [102] or Robert [290].

Section 11.1.2. Devroye [109] gives the classical theory of sampling from one-dimensional and other specific distributions.

Section 11.2.1. The phrase "Metropolis algorithm" is useful shorthand for "MCMC sampling, where the Markov chain is based on a proposal-acceptance scheme like those in section 11.2.1". The idea comes from the

1953 paper by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller [262] in the context of statistical physics, and the variant with general proposal matrix is from the 1970 paper of Hastings [180]. Of course the word "algorithm" means a *definite* rule for attaining some goal; the arbitraryness of proposal matrix, and vagueness about when to stop, makes it an extreme stretch to use the word for the Metropolis scheme.

The map $K \to P$ in the Metropolis-Hastings construction (11.5) has an interpretation as a minimum-length projection in a certain L^1 space of matrices – see Billera and Diaconis [50].

Section 11.2.2. The Gibbs sampler was popularized in 1984 by Geman and Geman [162] in the context of Bayesian image analysis. The idea is older in statistical physics, under the name *heat bath*. Hit-and-run was introduced in 1984 by Smith [310]. General line-sampling schemes go back to Goodman and Sokal [170].

Section 11.3.1. Terminology for this type of construction is not standard. What we call "Metropolized line sampling" is what Besag and Greene [46] call an auxiliary variable construction, and this type of construction goes back to Edwards and Sokal [139] in statistical physics.

Section 11.3.2. One can also define MTM using a general proposal matrix K [236], though (in contrast to Metropolis) the specialization of the general case to the symmetric case is different from the symmetric case described in the text. Liu et al [236] discuss the use of MTM as an ingredient in other variations of MCMC.

Other MCMC variations. In statistical physics, it is natural to think of particles in R^d having position and velocity. This suggests MCMC schemes in which velocity is introduced as an auxiliary variable. In particular one can use deterministic equations of motion to generate proposal steps for Metropolis, an idea called *hybrid Monte Carlo* – see Neal [268].

Section 11.4. The survey by Diaconis and Saloff-Coste [121] has further pieces of theory, emphasizing the low-dimensional discrete setting. For target densities on R^d one needs some regularity conditions to ensure τ_2 is finite; see Roberts and Tweedie [294] for results of this type.

Section 11.4.1. As background to Peskun's theorem, one might think (by vague physical analogy) that it would be desirable to have acceptance probabilities behave as some "smooth" function; e.g. in the symmetric-proposal case, instead of $\min(1, \pi_y/\pi_x)$ take $\frac{\pi_y}{\pi_x+\pi_y}$. Lemma 11.1 shows this intuition is wrong, at least using asymptotic variance rate or relaxation time as a criterion. Liu [235] section 12.3 gives further instances where Peskun's Theorem can be applied. As usual, it is hard to do such comparison arguments for τ_1.

Section 11.4.2. The coupling here is an instance of a one-dimensional *monotone coupling*, which exists for any stochastically monotone chain.

Section 11.5.2. Discussion of practical aspects of the diffusion heuristic can be found in Roberts et al [160], and discussion in the more complicated setting of Gibbs distributions of $(X_v; v \in Z^d)$ is in Breyer and Roberts [60].

11.8 Belongs in other chapters

yyy: add to what's currently sec. 10.2 of Chapter 2, version 9/10/99, but which may get moved to the new Chapter 8.

Where π does not vary with the parameter α we get a simple expression for $\frac{d}{d\alpha}\mathbf{Z}$.

Lemma 11.4 *In the setting of (yyy Chapter 2 Lemma 37), suppose π does not depend on α. Then*

$$\tfrac{d}{d\alpha}\mathbf{Z} = \mathbf{ZRZ}.$$

xxx JF: I see this from the series expansion for $\mathbf{Z}$ – what to do about a proof, I delegate to you!

11.8.1 Pointwise ordered transition matrices

yyy: belongs somewhere in Chapter 3.

Recall from Chapter 2 section 3 (yyy 9/10/99 version) that for a function $f : S \to R$ with $\sum_i \pi_i f_i = 0$, the asymptotic variance rate is

$$\sigma^2(\mathbf{P}, f) := \lim_t t^{-1}\text{var} \sum_{s=1}^{t} f(X_s) = f\Gamma f \tag{11.19}$$

where $\Gamma_{ij} = \pi_i Z_{ij} + \pi_j Z_{ji} + \pi_i\pi_j - \pi_i\delta_{ij}$. These individual-function variance rates can be compared between chains with the same stationary distribution, under a very strong "coordinatewise ordering" of transition matrices.

Lemma 11.5 (Peskun's Lemma [280]) *Let $\mathbf{P}$ and $\mathbf{Q}$ be reversible with the same stationary distribution π. Suppose $p_{ij} \le q_{ij}$ $\forall j \ne i$. Then $\sigma^2(\mathbf{P}, f) \ge \sigma^2(\mathbf{Q}, f)$ for all f with $\sum_i \pi_i f_i = 0$.*

Proof. Introduce a parameter $0 \le \alpha \le 1$ and write $\mathbf{P}^\alpha = (1-\alpha)\mathbf{P} + \alpha\mathbf{Q}$. Write $(\cdot)'$ for $\frac{d}{d\alpha}(\cdot)$ at $\alpha = 0$. It is enough to show

$$(\sigma^2(\mathbf{P}, f))' \le 0.$$

By (11.19)

$$(\sigma^2(\mathbf{P}, f))' = f\Gamma' f = 2\sum_i \sum_j f_i \pi_i z'_{ij} f_j.$$

By (yyy Lemma 11.4 above) $\mathbf{Z}' = \mathbf{Z}\mathbf{P}'\mathbf{Z}$. By setting

$$g_i = \pi_i f_i; \quad a_{ij} = z_{ij}/\pi_j; \quad w_{ij} = \pi_i p_{ij}$$

we can rewrite the equality above as

$$(\sigma^2(\mathbf{P}, f))' = 2\ g\mathbf{A}\mathbf{W}'\mathbf{A}g.$$

Since $\mathbf{A}$ is symmetric wsith row-sums equal to zero, it is enough to show that $\mathbf{W}'$ is non-negative definite. By hypothesis $\mathbf{W}'$ is symmetric and $w'_{ij} \geq 0$ for $j \neq i$. These properties imply that, ordering states arbitrarily, we may write

$$\mathbf{W}' = \sum \sum_{i<j} w'_{ij} \mathbf{M}^{ij}$$

where $\mathbf{M}^{ij}$ is the matrix whose only non-zero entries are $m(i,i) = m(j,j) = -1$; $m(i,j) = m(j,i) = 1$. Plainly $\mathbf{M}^{ij}$ is non-negative definite, hence so is $\mathbf{W}'$.

Chapter 12

Coupling Theory and Examples (October 11, 1999)

xxx This is intended as a section in a Chapter near Chapter 4; maybe a new Chapter consisting of this and another section on bounding τ_2 via distinguished paths. Need some preliminary discussion, e.g. on
relations between τ_1 and τ_2;
observe that τ_2 is tied to reversibility whereas coupling isn't.

12.1 Using coupling to bound variation distance

Recall from Chapter 2 section 4.1 (yyy 9/10/99 version) several ways in which variation distance is used to measure the deviation from stationarity of the time-t distribution of a Markov chain:

$$\begin{aligned} d_i(t) : &= ||P_i(X_t = \cdot) - \pi(\cdot)|| \\ d(t) : &= \max_i d_i(t) \\ \bar{d}(t) : &= \max_{ij} ||P_i(X_t = \cdot) - P_j(X_t = \cdot)||. \end{aligned}$$

Recall also from yyy the definition of *variation threshold time*

$$\tau_1 := \min\{t : \bar{d}(t) \le e^{-1}\}.$$

Since τ_1 is affected by continuization, when using the above definition with a discrete-time chain we write τ_1^{disc} for emphasis.

Coupling provides a methodology for seeking to upper bound $\bar{d}(t)$ and hence τ_1. After giving the (very simple) theory in section 12.1.1 and some

discussion in section 12.1.2, we proceed to give a variety of examples of its use. We shall say more about theory and applications of coupling in Chapter 8 (where we discuss the related idea of *coupling from the past*) and in Chapter 10 (on interacting random walks).

12.1.1 The coupling inequality

Consider a finite-state chain in discrete or continuous time. Fix states i, j. Suppose we construct a *coupling*, that is a joint process $((X_t^{(i)}, X_t^{(j)}),\ t \geq 0)$ such that

$$\begin{aligned} &(X_t^{(i)}, t \geq 0) \text{ is distributed as the chain started at } i \\ &(X_t^{(j)}, t \geq 0) \text{ is distributed as the chain started at } j. \end{aligned} \tag{12.1}$$

And suppose there is a random time $T^{ij} \leq \infty$ such that

$$X_t^{(i)} = X_t^{(j)},\ T^{ij} \leq t < \infty. \tag{12.2}$$

Call such a T^{ij} a *coupling time*. Then the *coupling inequality* is

$$||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)|| \leq P(T^{ij} > t),\ 0 \leq t < \infty. \tag{12.3}$$

The inequality holds because

$$\begin{aligned} ||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)|| &= ||P(X_t^{(i)} \in \cdot) - P(X_t^{(j)} \in \cdot)|| \\ &\leq P(X_t^{(i)} \neq X_t^{(j)}) \\ &\leq P(T^{ij} > t). \end{aligned}$$

12.1.2 Comments on coupling methodology

The coupling inequality provides a method of bounding the variation distance $\bar{d}(t)$, because if we can construct a coupling for an arbitrary pair (i, j) of initial states then

$$\bar{d}(t) \leq \max_{i,j} P(T^{ij} > t).$$

The reader may wish to look at a few of the examples before reading this section in detail.

In applying coupling methodology there are two issues. First we need to specify the coupling, then we need to analyze the coupling time. The most common strategy for constructing couplings is via Markov couplings, as follows. Suppose the underlying chain has state space I and (to take

the continuous-time case) transition rate matrix $\mathbf{Q} = (q(i,k))$. Consider a transition rate matrix $\tilde{\mathbf{Q}}$ on the product space $I \times I$. Write the entries of $\tilde{\mathbf{Q}}$ as $\tilde{q}(i,j;k,l)$ instead of the logical-but-fussy $\tilde{q}((i,j),(k,l))$. Suppose that, for each pair (i,j) with $j \neq i$,

$$\tilde{q}(i,j;\cdot,\cdot) \text{ has marginals } q(i,\cdot) \text{ and } q(j,\cdot) \tag{12.4}$$

in other words $\sum_l \tilde{q}(i,j;k,l) = q(i,k)$ and $\sum_k \tilde{q}(i,j;k,l) = q(j,l)$. And suppose that

$$\begin{aligned} \tilde{q}(i,i;k,k) &= q(i,k) \text{ for all } k \\ \tilde{q}(i,i;k,l) &= 0 \text{ for } l \neq k. \end{aligned}$$

Take $(X_t^{(i)}, X_t^{(j)})$ to be the chain on $I \times I$ with transition rate matrix $\widetilde{\mathbf{Q}}$ and initial position (i,j), Then (12.1) must hold, and $T^{ij} := \min\{t : X_t^{(i)} = X_t^{(j)}\}$ is a coupling time. This construction gives a *natural Markov coupling*, and all the examples where we use the coupling inequality will be of this form. In practice it is much more understandable to define the joint process in words, and we usually do so.

In constructing and analyzing couplings, we often exploit (explicitly or implicitly) some integer-valued metric $\rho(i,j)$ on the state space I. Then with a Markovian coupling,

$$P(T^{ij} > t) = P(\rho(X_t^i, X_t^{(j)}) \geq 1)$$

and it is enough to study the integer-valued process $Z_t := \rho(X_t^{(i)}, X_t^{(j)})$. Typically (Z_t) is *not* Markov, but one can try to compare it with some integer-valued Markov process (Z_t^*). Indeed, in defining the coupling one has in mind trying to make such a comparison possible. Often one shows that for any initial (i,j) the random time T^{ij} is stochastically smaller than the hitting time T_{a0}^* for the comparison chain (Z_t^*) to reach 0 starting from $a := \max_{i,j} \rho(i,j)$. This would imply

$$\bar{d}(t) \leq P(T_{a0}^* > t).$$

Finally, one does calculations with the integer-valued chain (Z_t^*), either bounding the tail probability $P(T_{a0}^* > t)$ directly or (what is often simpler) just bounding the expectation ET_{a0}^*, so that by Markov's inequality and the submultiplicativity property (Chapter 2 Lemma 20) (yyy 9/10/99 version) we have in continuous time

$$\tau_1 \leq eET_{a0}^*; \quad \bar{d}(t) \leq \exp(1 - \tfrac{t}{\tau_1}).$$

Here is perhaps the simplest comparison lemma, whose proof is left to the reader.

Lemma 12.1 (Decreasing functional lemma) *Let (Y_t) be a Markov chain on S and $f : S \to \{0, 1, 2, \ldots, \Delta\}$ a function. Suppose that for each $1 \le i \le \Delta$ and each initial state y with $f(y) = i$,*
(i) $f(Y_1) \le i$;
(ii) $P(f(Y_1) \le i - 1) \ge a_i > 0$.
Then

$$\max_{y \in S} E_y T_A \le \sum_{i=1}^{\Delta} 1/a_i$$

where $A := \{y : f(y) = 0\}$.

We now start a series of examples. Note that when presenting a coupling proof we don't need to explicitly check irreducibility, because the conclusion of a bound on coupling time obviously implies irreducibility.

12.1.3 Random walk on a dense regular graph

(Chapter 5 Example 16).

Consider an r-regular n-vertex graph. Write $\mathcal{N}(v)$ for the set of neighbors of v. For any pair v, w we can define a $1-1$ map $\theta_{v,w} : \mathcal{N}(v) \to \mathcal{N}(w)$ such that $\theta_{v,w}(x) = x$ for $x \in \mathcal{N}(v) \cap \mathcal{N}(w)$. Consider discrete-time random walk on the graph. We define a "greedy coupling" by specifying the joint transition matrix

$$\tilde{p}(v, w; x, \theta_{v,w}(x)) = 1/r, \quad x \in \mathcal{N}(v).$$

That is, from vertices v and w, if the first chain jumps to x then the second chain jumps to $\theta_{v,w}(x)$, and we maximize the chance of the two chains meeting after a single step. In general one cannot get useful bounds on the coupling time. But consider the dense case, where $r > n/2$. As observed in Chapter 5 Example 16, here $|\mathcal{N}(v) \cap \mathcal{N}(w)| \ge 2r - n$ and so the coupled process (X_t, Y_t) has the property that for $w \ne v$

$$P(X_{t+1} = Y_{t+1} | X_t = v, Y_t = w) = \frac{|\mathcal{N}(v) \cap \mathcal{N}(w)|}{r} \ge \frac{2r - n}{r}$$

implying that the coupling time T (for any initial pair of states) satisfies

$$P(T > t) \le \left(\frac{n - r}{r}\right)^t.$$

So the coupling inequality implies $\bar{d}(t) \le (\frac{n-r}{r})^t$. In particular the variation threshold satisfies

$$\tau_1^{\text{disc}} = O(1) \text{ as } n \to \infty,\ r/n \to \alpha > 1/2.$$

12.1.4 Continuous-time random walk on the d-cube

(Chapter 5 Example 15).

For $\mathbf{i} = (i_1, \ldots, i_d)$ and $\mathbf{j} = (j_1, \ldots, j_d)$ in $I = \{0,1\}^d$, let $\mathcal{D}(\mathbf{i}, \mathbf{j})$ be the set of coordinates u where $\mathbf{i}$ and $\mathbf{j}$ differ. Write $\mathbf{i}^u$ for the state obtained by changing the u'th coordinate of $\mathbf{i}$. Recall that in continuous time the components move independently as 2-state chains with transition rates $1/d$. Define a coupling in words as "run unmatched coordinates independently until they match, and then run them together". Formally, the non-zero transitions of the joint process are

$$\begin{aligned}
\tilde{q}(\mathbf{i}, \mathbf{j}; \mathbf{i}^u, \mathbf{j}^u) &= 1/d \text{ if } u \notin \mathcal{D}(\mathbf{i}, \mathbf{j}) \\
\tilde{q}(\mathbf{i}, \mathbf{j}; \mathbf{i}^u, \mathbf{j}) &= 1/d \text{ if } u \in \mathcal{D}(\mathbf{i}, \mathbf{j}) \\
\tilde{q}(\mathbf{i}, \mathbf{j}; \mathbf{i}, \mathbf{j}^u) &= 1/d \text{ if } u \in \mathcal{D}(\mathbf{i}, \mathbf{j}).
\end{aligned}$$

For each coordinate which is initially unmatched, it takes exponential (rate $2/d$) time until it is matched, and so the coupling time $T = T^{\mathbf{ij}}$ satisfies

$$T \stackrel{d}{=} \max(\xi_1, \ldots, \xi_{d_0})$$

where the (ξ_u) are independent exponential (rate $2/d$) and $d_0 = |\mathcal{D}(\mathbf{i}, \mathbf{j})|$ is the initial number of unmatched coordinates. So

$$P(T \le t) = (1 - \exp(-2t/d))^{d_0}$$

and the coupling inequality bounds variation distance as

$$\bar{d}(t) \le 1 - (1 - \exp(-2t/d))^d.$$

This leads to an upper bound on the variation threshold time

$$\tau_1 \le (\tfrac{1}{2} + o(1)) d \log d \text{ as } d \to \infty.$$

This example is discussed in more detail in Chapter 5 Example 15 (yyy 4/23/96 version) where it is shown that

$$\tau_1 \sim \tfrac{1}{4} d \log d \text{ as } d \to \infty$$

so the coupling bound is off by a factor of 2.

12.1.5 The graph-coloring chain

Fix a n-vertex graph with maximal degree r. Fix an integer $c \geq r+2$ and consider $[c] := \{1, 2, \ldots, c\}$ as a set of c colors. Let $\mathrm{col}(G, c)$ be the set of c-colorings of G, where a *c-coloring* is an assignment of a color to each vertex, in such a way that no two adjacent vertices have the same color. One can put a natural "product graph" structure on $\mathrm{col}(G, c)$, in which two colorings are adjacent if they differ at only one vertex. It is not hard to check that the condition $c \geq r+2$ ensures that $\mathrm{col}(G, c)$ is non-empty and the associated graph is connected. There is a natural discrete-time Markov chain on $\mathrm{col}(G, c)$:

> Pich a vertex v of G uniformly at random, pick a color γ uniformly at random, assign color γ to vertex v if feasible (i.e. if no neighbor of v has color γ), else retain existing color of v.

Under certain conditions a simple coupling analysis succeeds in bounding the mixing time. (The bound is far from sharp – see Notes).

Proposition 12.2 *If $c > 4r$ then $\bar{d}(t) \leq n \exp(-\frac{(c-4r)t}{cn})$ and so $\tau_1^{\mathrm{disc}} \leq 1 + \frac{cn}{c-4r}(1 + \log n)$.*

Proof. We couple two versions of the chain by simply using the same v and γ in both chains at each step. Write D_t for the number of vertices at which the colors in the two chains differ. Then $D_{t+1} - D_t \in \{-1, 0, 1\}$ and the key estimate is the following.

Lemma 12.3 *Conditional on the state of the coupled process at time t,*

$$P(D_{t+1} = D_t + 1) \leq \frac{2rD_t}{cn} \tag{12.5}$$

$$P(D_{t+1} = D_t - 1) \geq \frac{(c-2r)D_t}{cn} \tag{12.6}$$

Proof. In order that $D_{t+1} = D_t + 1$ it is necessary that the chosen pair (v, γ) is such that

> (*) there exists a neighbor (w, say) of v such that w has color γ in one chain but not in the other chain.

But the total number of pairs (v, γ) equals nc while the number of pairs satisfying (*) is at most $D_t \cdot 2r$. This establishes (12.5). Similarly, for $D_{t+1} = D_t - 1$ it is sufficient that v is currently unmatched and that no

neighbor of v in either chain has color γ; the number of such pairs (v, γ) is at least $D_t \cdot (c - 2r)$. □

Lemma 12.3 implies $E(D_{t+1} - D_t | D_t) \le -(c - 4r)D_t/(cn)$ and so

$$ED_{t+1} \le \kappa ED_t; \quad \kappa := 1 - \tfrac{c-4r}{cn}.$$

Since $D_0 \le n$ we have, for any initial pair of states,

$$P(D_t \ge 1) \le ED_t \le \kappa^t n \le n \exp(-\tfrac{(c-4r)t}{cn})$$

and the coupling lemma establishes the Proposition.

12.1.6 Permutations and words

The examples in sections 12.1.4 and 12.1.5 were simple prototypes of *interacting particle systems*, more examples of which appear in Chapter 10, whose characteristic property is that a step of the chain involves only "local" change. Chains making "global" changes are often hard to analyze, but here is a simple example.

Fix a finite alphabet A of size $|A|$. Fix m, and consider the set A^m of "words" $\mathbf{x} = (x_1, \ldots, x_m)$ with each $x_i \in A$. Consider the discrete-time Markov chain on A^m in which a step $\mathbf{x} \to \mathbf{y}$ is specified by the following two-stage procedure.

Stage 1. Pick a permutation σ of $\{1, 2, \ldots, m\}$ uniformly at random from the set of permutations σ satisfying $x_{\sigma(i)} = x_i \forall i$.

Stage 2. Let $(c_j(\sigma); j \ge 1)$ be the cycles of σ. For each j, and independently as j varies, pick uniformly an element α_j of A, and define $y_i = \alpha_j$ for every $i \in c_j(\sigma)$.

Here is an alternative description. Write Π for the set of permutations of $\{1, \ldots, m\}$. Consider the bipartite graph on vertices $A^m \cup \Pi$ with edge-set $\{(\mathbf{x}, \sigma) : x_{\sigma(i)} = x_i \forall i\}$. Then the chain is random walk on this bipartite graph, watched every second step (that is, when it is in A^m).

From the second description, it is clear that the stationary probabilities $\pi(\mathbf{x})$ are proportional to the degree of $\mathbf{x}$ in the bipartite graph, giving

$$\pi(\mathbf{x}) \propto \prod_a n_a(\mathbf{x})!$$

where $n_a(\mathbf{x}) = |\{i : x_i = a\}|$. We shall use a coupling argument to establish the following bound on variation distance:

$$\bar{d}(t) \le m \left(1 - \frac{1}{|A|}\right)^t \tag{12.7}$$

implying that the variation threshold satisfies

$$\tau_1^{\text{disc}} \le 1 + \frac{1 + \log m}{-\log(1 - \frac{1}{|A|})} \le 1 + (1 + \log m)|A|.$$

The construction of the coupling depends on the following lemma.

Lemma 12.4 *Given finite sets F^1, F^2 we can construct (for $u = 1, 2$) a uniform random permutation σ^u of F^u with cycles $(C_j^u; j \ge 1)$, where the cycles are labeled such that*

$$C_j^1 \cap F^1 \cap F^2 = C_j^2 \cap F^1 \cap F^2 \text{ for all } j.$$

In the equality we interpret the C_j^u as sets.

Proof. Given a permutation σ of a finite set G, there is an induced permutation on a subset G' obtained by deleting from the cycle representation of σ those elements not in G'. It is easy to check that, for a uniform random permutation of G, the induced random permutation of G' is also uniform. In the setting of the lemma, take a uniform random permutation σ of $F^1 \cup F^2$, and let σ^u be the induced random permutations of F^u. Then the equality holds because each side is representing the cycles of the induced permutation on $F^1 \cap F^2$. □

We construct a step $(\mathbf{x}^1, \mathbf{x}^2) \to (\mathbf{Y}^1, \mathbf{Y}^2)$ of the coupled processes as follows. For each $a \in A$, set $F^{1,a} = \{i : x_i^1 = a\}$, $F^{2,a} = \{i : x_i^2 = a\}$. Take random permutations $\sigma^{1,a}$, $\sigma^{2,a}$ as in the lemma, with cycles $C_j^{1,a}, C_j^{2,a}$. Then $(\sigma^{1,a}, a \in A)$ define a uniform random permutation σ^1 of $\{1, \ldots, m\}$, and similarly for σ^2. This completes stage 1. For stage 2, for each pair (a, j) pick a uniform random element α_j^a of A and set

$$Y_i^1 = \alpha_j^a \text{ for every } i \in C_j^{1,a}$$

$$Y_i^2 = \alpha_j^a \text{ for every } i \in C_j^{2,a}.$$

This specifies a Markov coupling. By construction

$$\begin{array}{ll} \text{if } x_i^1 = x_i^2 & \text{then } Y_i^1 = Y_i^2 \\ \text{if } x_i^1 \ne x_i^2 & \text{then } P(Y_i^1 = Y_i^2) = 1/|A| \end{array}$$

because Y_i^1 and Y_i^2 are independent uniform choices from A. So the coupled processes $(\mathbf{X}^1(t), \mathbf{X}^2(t))$ satisfy

$$P(X_i^1(t) \ne X_i^2(t)) = \left(1 - \frac{1}{|A|}\right)^t P(X_i^1(0) \ne X_i^2(0)).$$

In particular $P(\mathbf{X}^1(t) \ne \mathbf{X}^2(t)) \le m(1 - 1/|A|)^t$ and the coupling inequality (12.3) gives (12.7).

12.1.7 Card-shuffling by random transpositions

We mentioned in Chapter 1 section 1.4 (yyy 7/20/99 version) that card-shuffling questions provided a natural extrinsic motivation for the study of mixing times. The example here and in section 12.1.9 give a first study of mathematically (if not physically) simple random shuffles, and these discrete-time chains are prototypes for more complex chains arising in other contexts.

Consider a d-card deck. The *random transpositions* shuffle is:

> Make two independent uniform choices of cards, and interchange them.

With chance $1/d$ the two choices are the same card, so no change results. To make a coupling analysis, we first give an equivalent reformulation.

> Pick a label a and a position i uniformly at random; interchange the label-a card with the card in position i.

This reformulation suggests the coupling in which the same choice of (a, i) is used for each chain. In the coupled process (with two arbitrary starting states) let D_t be the number of unmatched cards (that is, cards whose positions in the two decks are different) after t steps. Then
(i) $D_{t+1} \le D_t$.
(ii) $P(D_{t+1} \le j - 1 | D_t = j) \ge j^2/d^2$.

Here (i) is clear, and (ii) holds because whenever the card labeled a and the card in position i are both unmatched, the step of the coupled chain creates at least one new match (of the card labeled a).

Noting that D_t cannot take value 1, we can use the decreasing functional lemma (Lemma 12.1) to show that the coupling time $T := \min\{t : D_t = 0\}$ satisfies

$$ET \le \sum_{j=2}^{d} d^2/j^2 \le d^2(\tfrac{\pi^2}{6} - 1).$$

In particular, the coupling inequality implies $\tau_1^{\text{disc}} = O(d^2)$.

We revisit this example in Chapter 7 Example 18 (yyy 1/31/94 version) where it is observed that in fact

$$\tau \sim \tfrac{1}{2} d \log d \tag{12.8}$$

An analogous continuous-space chain on the simplex is studied in Chapter 13-4 Example 3 (yyy 7/29/99 version)

12.1.8 Reflection coupling on the n-cycle

Consider continuous-time random walk on the n-cycle $I = \{0, 1, 2, \ldots, n-1\}$. That is, the transition rates are

$$i \xrightarrow{1/2} i+1; \quad i \xrightarrow{1/2} i-1$$

where here and below ± 1 is interpreted modulo n. One can define a coupling by specifying the following transition rates for the bivariate process.

$$\begin{array}{lll} & (i,i) \xrightarrow{1/2} (i+1, i+1); & (i,i) \xrightarrow{1/2} (i-1, i-1) \\ (\text{if } |j-i| > 1) & (i,j) \xrightarrow{1/2} (i+1, j-1); & (i,j) \xrightarrow{1/2} (i-1, j+1) \\ (i, i+1) \xrightarrow{1/2} (i,i); & (i,i+1) \xrightarrow{1/2} (i+1, i+1); & (i,i+1) \xrightarrow{1/2} (i-1, i+2) \end{array} \quad (12.9)$$

and symmetrically for $(i+1, i)$. The joint process $((X_t^{(0)}, X_t^{(k)}), t \geq 0)$ started at $(0, k)$ can be visualized as follows. Let $\phi(i) := k - i \bmod n$. Picture the operation of ϕ as reflection in a mirror which passes through the points $\{x_1, x_2\} = \{k/2, k/2 + n/2 \bmod n\}$ each of which is either a vertex or the middle of an edge. In the simplest case, where x_1 and x_2 are vertices, let (X_t^0) be the chain started at vertex 0, let $T^{0k} = \min\{t : X_t \in \{x_1, x_2\}\}$ and define

$$\begin{aligned} X_t^{(k)} &= \phi(X_t^{(0)}), \ t \leq T^{0k} \\ &= X_t^{(0)}, \ t > T^{0k}. \end{aligned}$$

This constructs a bivariate process with the transition rates specified above, with coupling time T^{0k}, and the pre-T^{0k} path of $X^{(k)}$ is just the reflection of the pre-T^{0k} path of $X^{(0)}$. In the case where a mirror point is the middle of an edge $(j, j+1)$ and the two moving particles are at j and $j+1$, we don't want simultaneous jumps across that edge; instead (12.9) specifies that attempted jumps occur at *independent* times, and the process is coupled at the time of the first such jump.

It's noteworthy that in this example the coupling inequality

$$||P_0(X_t \in \cdot) - P_k(X_t \in \cdot)|| \leq P(X_t^{(0)} \neq X_t^{(k)})$$

is in fact an equality. Indeed this assertion, at a given time t, is equivalent to the assertion

$$P(X_t^{(0)} \in \cdot, T > t) \text{ and } P(X_t^{(k)} \in \cdot, T > t) \text{ have disjoint support.}$$

But the support A^0 of the first measure is the set of vertices which can be reached from 0 without meeting or crossing any mirror point (and similarly for A^k); and A^0 and A^k are indeed disjoint.

It is intuitively clear that the minimum over k of T^{0k} is attained by $k = \lfloor n/2 \rfloor$: we leave the reader to find the simple non-computational proof. It follows, taking e.g. the simplest case where n is multiple of 4, that we can write

$$\bar{d}(t) = P(T_{\{-n/4,n/4\}} > t) \tag{12.10}$$

where $T_{\{-n/4,n/4\}}$ is the hitting time for continuous-time random walk on the integers.

Parallel results hold in discrete time but only when the chains are suitably *lazy*. The point is that (12.9) isn't allowable as transition *probabilities*. However, if we fix $0 < a \leq 1/3$ then the chain with transition probabilities

$$i \xrightarrow{a} i+1; \quad i \xrightarrow{a} i-1$$

(and which holds with the remaining probability) permits a coupling of the form (12.9) with all transition probabilities being a instead of $1/2$. The analysis goes through as above, leading to (12.10) where T refers to the discrete-time lazy walk on the integers.

Similar results hold for random walk on the n-path (Chapter 5 Example 8) (yyy 4/23/96 version). and we call couplings of this form *reflection couplings*. They are simpler in the context of continuous-path Brownian motion – see Chapter 13-4 section 1 (yyy 7/29/99 version).

12.1.9 Card-shuffling by random adjacent transpositions

As in section 12.1.7 we take a d-card deck; here we define a (lazy) shuffle by

> With probability 1/2 make no change; else pick a uniform random position $i \in \{1, 2, \ldots, d\}$ and interchange the cards in positions i and $i+1$ (interpret $d+1$ as 1).

To study this by coupling, consider two decks. In some positions i the decks match (the label on the card in position i is the same in both decks). Write $\mathcal{D}$ for the set of i such that either position i or position $i+1$ or both match. Specify a step of the coupled chain by:

$$\begin{aligned}
P(\text{interchange } i \text{ and } i+1 \text{ in each deck}) &= \tfrac{1}{2d}, \quad i \in \mathcal{D} \\
P(\text{interchange } i \text{ and } i+1 \text{ in first deck, no change in second deck}) &= \tfrac{1}{2d}, \quad i \notin \mathcal{D} \\
P(\text{interchange } i \text{ and } i+1 \text{ in second deck, no change in first deck}) &= \tfrac{1}{2d}, \quad i \notin \mathcal{D} \\
P(\text{no change in either deck}) &= \tfrac{|\mathcal{D}|}{2d}.
\end{aligned}$$

Consider a particular card a. From the coupling description we see
(a) if the card gets matched then it stays matched;
(b) while unmatched, at each step the card can move in at most one of the decks.
It follows that the "clockwise" distance $D(t) := X_a^1(t) - X_a^2(t) \bmod d$ between the positions of card a in the two decks behaves exactly as a lazy random walk on the d-cycle:

$$p_{j,j+1} = p_{j,j-1} = 1/d, \quad 1 \le j \le d$$

until $D(t)$ hits 0. By the elementary formula for mean hitting times on the cycle (Chapter 5 eq. (24)) (yyy 4/23/96 version), the mean time $T^{(a)}$ until card a becomes matched satisfies

$$ET^{(a)} \le \tfrac{d}{2}\ \tfrac{d^2}{4}$$

uniformly over initial configurations. By submultiplicativity (Chapter 2 section 4.3) (yyy 9/10/99 version)

$$P(T^{(a)} > md^3/4) \le 2^{-m},\ m = 1, 2, \ldots.$$

The chains couple at time $T := \max_a T^{(a)}$ and so

$$\bar{d}(md^3/4) \le P(T > md^3/4) \le d2^{-m}.$$

In particular

$$\tau_1^{\text{disc}} = O(d^3 \log d).$$

In this example it turns out that coupling does give the correct order of magnitude; the corresponding lower bound

$$\tau_1^{\text{disc}} = \Omega(d^3 \log d)$$

was proved by Wilson [338]. Different generalizations of this example appear in section 12.1.13 and in Chapter 14 section 5 (yyy 3/10/94 version), where we discuss relaxation times.

A generalization of this example, the *interchange process*, is studied in Chapter 14 section 5 (yyy 3/10/94 version).

12.1.10 Independent sets

Fix a graph G on n vertices with maximal degree r, An *independent set* is a set of vertices which does not contain any adjacent vertices. Fix m

and consider the space of all independent sets of size m in G. Picture an independent set $\mathbf{x}$ as a configuration of m particles at distinct vertices, with no two particles at adjacent vertices. A natural discrete-time chain $(\mathbf{X}_t)$ on I is

> pick a uniform random particle a and a uniform random vertex v; move particle a to vertex v if feasible, else make no move.

To study mixing times, we can define a coupling $(\mathbf{X}_t, \mathbf{Y}_t)$ by simply making the same choice of (a, v) in each of the two coupled chains, where at each time we invoke a matching of particles in the two realizations which is arbitrary except for matching particles at the same vertex. To analyze the coupling, let ρ be the natural metric on I: $\rho(\mathbf{x}, \mathbf{y})$ = number of vertices occupied by particles of $\mathbf{x}$ but not by particles of $\mathbf{y}$. Clearly $D_t := \rho(\mathbf{X}_t, \mathbf{Y}_t)$ can change by at most 1 on each step. Let us show that, for initial states with $\rho(\mathbf{x}, \mathbf{y}) = d > 0$,

$$P_{(\mathbf{x},\mathbf{y})}(D_1 = d+1) \le \frac{m-d}{m} \frac{2d(r+1)}{n} \tag{12.11}$$

$$P_{(\mathbf{x},\mathbf{y})}(D_1 = d-1) \ge \frac{d}{m} \frac{n-(m+d-2)(r+1)}{n}. \tag{12.12}$$

For in order that $D_1 = d+1$ we must first choose a matched particle a (chance $(m-d)/d$) and then choose a vertex v which is a neighbor of (or the same as) some vertex v' which is in exactly one of $\{\mathbf{x}, \mathbf{y}\}$: there are $2d$ such vertices v' and hence at most $2d(r+1)$ possibilities for v. This establishes (12.11). Similarly, in order that $D_t = d-1$ it is sufficient that we pick an unmatched particle a (chance d/m) and then choose a vertex v which is not a neighbor of (or the same as) any vertex v' which is occupied in one or both realizations by some particle other than a: there are $m+d-2$ such forbidden vertices v' and hence at most $(m+d-2)(r+1)$ forbidden positions for v. This establishes (12.12).

From (12.11,12.12) a brief calculation gives

$$\begin{aligned} E_{(\mathbf{x},\mathbf{y})}(D_1 - d) &\le \tfrac{-d}{mn}(n - (3m-d-2)(r+1)) \\ &\le \tfrac{-d}{mn}(n - 3(m-1)(r+1)). \end{aligned}$$

In other words

$$E_{(\mathbf{x},\mathbf{y})} D_1 \le \kappa d; \quad \kappa := 1 - \frac{n-3(m-1)(r+1)}{mn}.$$

If $m < 1 + \frac{n}{3(r+1)}$ then $\kappa < 1$. In this case, by copying the end of the analysis of the graph-coloring chain (section 12.1.5)

$$\bar{d}(t) \le m\kappa^t; \quad \tau_1 = O\left(\tfrac{\log m}{1-\kappa}\right).$$

To clarify the size-asymptotics, suppose $m, n \to \infty$ with $m/n \to \rho < \frac{1}{3(r+1)}$. Then for fixed ρ

$$\tau_1 = O(n \log n).$$

12.1.11 Two base chains for genetic algorithms

One way of motivating study of Markov chains on combinatorial sets with uniform stationary distributions is as "base chains" on which to base Markov chain Monte Carlo, that is to create other chains designed to have some specified distribution as their stationary distributions. Here is a typical base chain underlying genetic algorithms.

Fix integers $K, L \ge 1$ with K even. A state of the chain is a family of *words* $(\mathbf{x}^k, 1 \le k \le K)$, where each word is a binary L-tuple $\mathbf{x}^k = (x^k_l, 1 \le l \le L)$. A step of the chain is defined as follows.

> Use a uniform random permutation π of $\{1, 2, \ldots, K\}$ to partition the words into $K/2$ pairs $\{\mathbf{x}^{\pi(1)}, \mathbf{x}^{\pi(2)}\}, \quad \{\mathbf{x}^{\pi(3)}, \mathbf{x}^{\pi(4)}\} \ldots$. Create a new pair $\{\mathbf{y}^1, \mathbf{y}^2\}$ from $\{\mathbf{x}^{\pi(1)}, \mathbf{x}^{\pi(2)}\}$ by setting, independently for each $1 \le l \le L$
>
> $$P((y^1_l, y^2_l) = (x^{\pi(1)}_l, x^{\pi(2)}_l)) = P((y^1_l, y^2_l) = (x^{\pi(2)}_l, x^{\pi(1)}_l)) = 1/2. \tag{12.13}$$
>
> Repeat independently for $1 \le i \le K/2$ to create new pairs $\{\mathbf{y}^{2i-1}, \mathbf{y}^{2i}\}$ from $\{\mathbf{x}^{\pi(2i-1)}, \mathbf{x}^{\pi(2i)}\}$. The new state is the family of words $\mathbf{y}^k$.

Associated with an initial state $(\mathbf{x}^k)$ is a vector of column sums $\mathbf{m} = (m_l, 1 \le l \le L)$ where $m_l = \sum_k x^k_l$. These sums are preserved by the chain, so the proper state space is the space $I_{\mathbf{m}}$ of families with column-sums $\mathbf{m}$. The transition matrix is symmetric and so the chain is reversible with uniform stationary distribution on $I_{\mathbf{m}}$.

To describe the coupling, first rephrase (12.13) in words as "(y^1_l, y^2_l) is $\{x^{\pi(1)}_l, x^{\pi(2)}_l\}$ in random order, either forwards or backwards". Now specify the coupling as follows.
(i) Use the same random permutation π for both chains.
(ii) For each i and each l, in creating the new words (y^{2i-1}_l, y^{2i}_l) from the old

words $\{x_l^{\pi(2i-1)}, x_l^{\pi(2i)})\}$ use the same choice (forwards or backwards) in both chains, except when $(x_l^{\pi(2i-1)}, x_l^{\pi(2i)}) = (1,0)$ for one chain and $= (0,1)$ for the other chain, in which case use opposite choices of (forwards, backwards) in the two chains.

To study the coupled processes $(\mathbf{X}(t), \hat{\mathbf{X}}(t))$, fix l and consider the number $W(t) := \sum_{k=1}^{K} |X_l^k(t) - \hat{X}_l^k(t)|$ of words in which the l'th letter is not matched in the two realizations. Suppose $W(0) = w$. Consider the creation of the first two new words in each chain. The only way that the number of matches changes is when we use opposite choices of (forwards, backwards) in the two chains, in which case two new matches are created. The chance that the l'th letter in the two chains is 1 and 0 in the $\pi(1)$'th word and is 0 and 1 in the $\pi(2)$'th word equals $\frac{w/2}{K} \times \frac{w/2}{K-1}$, and so (taking into account the symmetric case) the mean number of new matches at l in these two words equals $\frac{w^2}{K(K-1)}$. Summing over the $K/2$ pairs,

$$E(W(1)|W(0) = w) = w - \tfrac{w^2}{2(K-1)}.$$

We can now apply a comparison lemma (Chapter 2 Lemma 32) (yyy 9/10/99 version) which concludes that the hitting time T^l of $W(t)$ to 0 satisfies

$$ET^l \le \sum_{w=2}^{K} \tfrac{2(K-1)}{w^2} \le 2K.$$

Since $T := \max_l T^l$ is a coupling time, a now-familiar argument shows that for $u = 1, 2, \ldots$

$$\bar{d}(4uK) \le P(T > 4uK) \le LP(T^l > u \cdot 4K) \le L\ 2^{-u}$$

and so

$$\tau_1^{\text{disc}} = O(K \log L).$$

Open Problem 12.5 *Show* $\tau_1^{\text{disc}} = O(\log K \times \log L)$.

We expect this bound by analogy with the "random transpositions" shuffle (section 12.1.7). Loosely speaking, the action of the chain on a single position in words is like the random transpositions chain speeded up by a factor $K/2$, so from (12.8) we expect its mixing time to be $\Theta(\log K)$. It would be interesting to study this example via the group representation or strong stationary time techniques which have proved successful for the random transpositions chain.

To make a metaphor involving biological genetics, the letters represent chromosomes and the words represent the chromosomal structure of a gamete; the process is "sexual reproduction from the viewpoint of gametes". If instead we want a word to represent a particular chromosome and the letters to represent genes within that chromosome, then instead of flipping bits independently it is more natural to model *crossover*. That is, consider a chain in which the rule for creating a new pair $\{\mathbf{y}^{2i-1}, \mathbf{y}^{2i}\}$ from $\{\mathbf{x}^{\pi(2i-1)}, \mathbf{x}^{\pi(2i)}\}$ becomes

Take U_i uniform on $\{1, 2, \ldots, L, L+1\}$. Define

$$\begin{aligned}(y_l^{2i-1}, y_l^{2i}) &= (x_l^{\pi(2i-1)}, x_l^{\pi(2i)}),\ l < U \\ (y_l^{2i-1}, y_l^{2i}) &= (x_l^{\pi(2i)}, x_l^{\pi(2i-1)}),\ l \geq U.\end{aligned}$$

As an exercise (hint in Notes), find a coupling argument to show that for this chain

$$\tau_1^{\text{disc}} = O(KL^2). \tag{12.14}$$

12.1.12 Path coupling

In certain complicated settings it is useful to know that it is enough to couple versions of the chain which start in "nearby" states. To say this carefully, let I be finite and consider a $\{0, 1, 2, \ldots\}$-valued function $\rho(i, j)$ defined on some symmetric subset $\mathcal{E} \subset I \times I$. Call ρ a *pre-metric* if
(i) $\rho(i, j) = 0$ iff $i = j$.
(ii) $\rho(i, j) = \rho(j, i)$.
(iii) $\rho(i_0, i_k) \leq \sum_{u=0}^{k-1} \rho(i_u, i_{u+1})$, whenever (i_0, i_k) and each (i_u, i_{u+1}) are in $\mathcal{E}$.

Clearly a pre-metric extends to a metrric $\bar{\rho}$ by defining

$$\bar{\rho}(i, j) := \min \left\{ \sum_u \rho(i_u, i_{u+1}) \right\} \tag{12.15}$$

the minimum over all paths $i = i_0, i_1, \ldots, i_k = j$ with each $(i_u, i_{u+1}) \in \mathcal{E}$. Note $\bar{\rho}(i, j) \leq \rho(i, j)$ for $(i, j) \in \mathcal{E}$.

Lemma 12.6 *Let S be a state space. Let $(\mu_{i,i+1},\ 0 \leq i \leq d-1)$ be probability distributions on $S \times S$ such that the second marginal of $\mu_{i,i+1}$ coincides with the first marginal of $\mu_{i+1,i+2}$ for $0 \leq i \leq d-2$. Then there exists a S-valued random sequence $(V_i, 0 \leq i \leq d)$ such that $\mu_{i,i+1} = \text{dist}(V_i, V_{i+1})$ for $0 \leq i \leq d-1$.*

Proof. Just take (V_i) to be the non-homogeneous Markov chain whose transition probabilities $P(V_{i+1} \in \cdot | V_i = v)$ are the conditional probabilities determined by the specified joint distribution $\mu_{i,i+1}$.

Lemma 12.7 (Path-coupling lemma) *Take a discrete-time Markov chain (X_t) with finite state space I. Write $X_1^{(i)}$ for the time-1 value of the chain started at state i. Let ρ be a pre-metric defined on some subset $\mathcal{E} \subset I \times I$. Suppose that for each pair (i,j) in $\mathcal{E}$ we can construct a joint law $(X_1^{(i)}, X_1^{(j)})$ such that*

$$E\bar{\rho}(X_1^{(i)}, X_1^{(j)}) \le \kappa \rho(i,j) \tag{12.16}$$

for some constant $0 < \kappa < 1$. Then

$$\bar{d}(t) \le \Delta_\rho \kappa^t \quad (\kappa < 1) \tag{12.17}$$

where $\Delta_\rho := \max_{i,j \in I} \bar{\rho}(i,j)$.

See the Notes for comments on the case $\kappa = 1$.

Proof. Fix states i, j and consider a path (i_u) attaining the minimum in (12.15). For each u let $(X_1^{(i_u)}, X_1^{(i_{u+1})})$ have a joint distribution satisfying (12.16). By Lemma 12.6 there exists a random sequence $(X_1^{(i)} = X_1^{(i_0)}, X_1^{(i_1)}, \ldots, X_1^{(j)})$ consistent with these bivariate distributions. In particular, there is a joint distribution $(X_1^{(i)}, X_1^{(j)})$ such that

$$E\bar{\rho}(X_1^{(i)}, X_1^{(j)}) \le E \sum_u \bar{\rho}(X_1^{(i_u)}, X_1^{(i_{u+1})}) \le \kappa \sum_u \rho(i_u, i_{u+1}) = \kappa \bar{\rho}(i,j).$$

This construction gives one step of a coupling of two copies of the chain started at arbitrary states, and so extends to a coupling $((X_t^{(i)}, X_t^{(j)}),\ t = 0, 1, 2, \ldots)$ of two copies of the entire processes. The inequality above implies

$$E(\bar{\rho}(X_{t+1}^{(i)}, X_{t+1}^{(j)}) | X_t^{(i)}, X_t^{(j)}) \le \kappa \bar{\rho}(X_t^{(i)}, X_t^{(j)})$$

and hence

$$P(X_t^{(i)} \ne X_t^{(j)}) \le E\bar{\rho}(X_t^{(i)}, X_t^{(j)}) \le \kappa^t \bar{\rho}(i,j) \le \kappa^t \Delta_\rho$$

establishing (12.17). □

Bubley and Dyer [77] introduced Lemma 12.7 and the name *path-coupling*. It has proved useful in extending the range of applicability of coupling methods in settings such as graph-coloring (Bubley et al [79] Vigoda [333]) and independent sets (Luby and Vigoda [244]). These are too intricate for presentation here, but the following example will serve to illustrate the use of path-coupling.

12.1.13 Extensions of a partial order

Fix a partial order $\preceq$ on an n-element set, and let I_m be the set of *linear extensions* of $\preceq$, that is to say total orders consistent with the given partial order. We can define a discrete-time Markov chain on I_m by re-using the idea in the "random adjacent transpositions" example (section 12.1.9). Let $w(\cdot)$ be a probability distribution on $\{1, 2, \ldots, n-1\}$. Define a step of the chain as follows.

> Pick position i with probability $w(i)$, and independently pick one of { stay, move } with probability 1/2 each. If pick "move" then interchange the elements in positions i and $i+1$ if feasible (i.e. if consistent with the partial order); else make no change.

The transition matrix is symmetric, so the stationary distribution is uniform on I_m.

To analyze by coupling, define one step of a bivariate coupled process as follows.

> Make the same choice of i in both chains. Also make the same choice of { move, stay }, except in the case where the elements in positions i and $i+1$ are the same elements in opposite order in the two realizations, in which case use the opposite choices of { stay, move }.

The coupling is similar (but not identical) to that in section 12.1.9, where the underlying chain is that corresponding to the "null" partial order. For a general partial order, the coupling started from an arbitrary pair of states seems hard to analyze directly. For instance, an element in the same position in both realizations at time t may not remain so at time $t+1$. Instead we use path-coupling, following an argument of Bubley and Dyer [78]. Call two states $\mathbf{x}$ and $\mathbf{y}$ *adjacent* if they differ by only one (not necessarily adjacent) transposition; if the transposed cards are in positions $i < j$ then let $\rho(\mathbf{x}, \mathbf{y}) = j - i$. We want to study the increment $\Phi := \rho(X_1, Y_1) - \rho(\mathbf{x}, \mathbf{y})$ where (X_1, Y_1) is the coupled chain after one step from $(\mathbf{x}, \mathbf{y})$. The diagram shows a typical pair of adjacent states.

	a	b	c	α	d	e	f	β	g	h
	a	b	c	β	d	e	f	α	g	h
position	$\cdot$	$\cdot$	$\cdot$	i	$\cdot$	$\cdot$	$\cdot$	j	$\cdot$	$\cdot$

Observe first that any choice of position other than $i-1, i, j-1, j$ will have no effect on Φ. If position i and "move" are chosen, then $\{\alpha, d\}$ are interchanged in the first chain and $\{\beta, d\}$ in the second; both lead to feasible

configurations by examining the relative orders in the other chain's previous configuration. This has chance $w(i)/2$ and leads to $\Phi = -1$. If position $i-1$ and "move" are chosen (chance $w(i-1)/2$), then if either or both moves are feasible $\Phi = 1$, while if neither are feasible then $\Phi = 0$. Arguing similarly for choices $j-1, j$ leads to

$$E(\Phi) \le \tfrac{1}{2}(w(i-1) - w(i) - w(j-1) + w(j)).$$

This estimate remains true if $j = i+1$ because in that case choosing position i (chance $w(i)$) always creates a match. Now specify

$$w(i) := \tfrac{i(n-i)}{w_n}, \quad w_n := \sum_{j=1}^{n-1} j(n-j)$$

and then $E\Phi \le -\frac{j-i}{w_n}$. This leads to

$$E_{(\mathbf{x},\mathbf{y})}\rho(X_1, Y_1) \le \left(1 - \tfrac{1}{w_n}\right) \rho(\mathbf{x}, \mathbf{y})$$

for adjacent $(\mathbf{x}, \mathbf{y})$. We are thus in the setting of Lemma 12.7, which shows

$$\bar{d}(t) \le \Delta_n \exp(-t/w_n).$$

Since $\Delta_n = O(n^2)$ and $w_n \sim n^3/6$ we obtain

$$\tau_1^{\text{disc}} = (\tfrac{1}{3} + o(1))\ n^3 \log n.$$

12.2 Notes on Chapter 4-3

Coupling has become a standard tool in probability theory. The monograph of Lindvall [233] contains an extensive treatment, and history. In brief, Doeblin [126] used the idea of running two copies of the chain independently until they meet, in order to prove the convergence theorem (Chapter 2 Theorem 2) for finite-state chains, and this is now the textbook proof ([270] Theorem 1.8.3) of the convergence theorem for countable-state chains. The first wide-ranging applications were in the context of infinite-site interacting particles in the 1970s, where (e.g. Liggett [230]) couplings were used to study uniqueness of invariant distributions and convergence thereto. Theory connecting couplings and variation distance is implicit in Griffeath [171] and Pitman [282], though the first systematic use to bound variation distance in finite-state chains was perhaps Aldous [9], where examples including those in sections 12.1.4, 12.1.7 and 12.1.9 were given.

Section 12.1.2. There may exist Markov couplings which are not of the *natural* form (12.4), but examples typically rely on very special symmetry properties. For the theoretically-interesting notion of (non-Markov) *maximal coupling* see Chapter 9 section 1 (yyy 4/21/95 version).

The coupling inequality is often presented using a first chain started from an arbitrary point and a second chain started with the stationary distribution, leading to a bound on $d(t)$ instead of $\bar{d}(t)$. See Chapter 13-4 yyy for an example where this is used in order to exploit distributional properties of the stationary chain.

Section 12.1.5. This chain was first studied by Jerrum [200], who proved rapid mixing under the weaker assumption $c \geq 2r$. His proof involved a somewhat more careful analysis of the coupling, exploiting the fact that "bad" configurations for the inequalities (12.5,12.6) are different. This problem attracted interest because the same constraint $c \geq 2r$ appears in proofs of the absence of phase transition in the zero-temperature anti-ferromagnetic Potts model in statistical physics. Proving rapid mixing under weaker hypotheses was first done by Bubley et al [79] in special settings and using computer assistance. Vigoda [333] then showed that rapid mixing still holds when $c > \frac{11}{6}r$: the proof first studies a different chain (still reversible with uniform stationary distribution) and then uses a comparison theorem.

Section 12.1.6. The chain here was suggested by Jerrum [196] in the context of a general question of counting the number of orbits of a permutation group acting on words. More general cases (using a subgroup of permutations instead of the whole permutation group) remain unanalyzed.

Section 12.1.10. See Luby and Vigoda [244] for more detailed study and references.

Section 12.1.11. Conceptually, the states in these examples are *unordered* families of words. In genetic algorithms for optimization one has an objective function $f : \{0,1\}^L \to R$ and accepts or rejects offspring words with probabilities depending on their f-values.

Interesting discussion of some different approaches to genetics and computation is in Rabani et al [287].

Hint for (12.14). First match the L'th letters in each word, using the occasions when $U_i = L$ or $L+1$. This takes $O(LK)$ time.

Section 12.1.12. Another setting where path-coupling has been used is contingency tables: Dyer and Greenhill [137].

In the case where (12.16) holds for $\kappa = 1$, one might expect a bound of the form

$$\bar{d}(t) = O(\Delta_\rho^2/\alpha) \tag{12.18}$$

$$\alpha := \min_{(i,j)\in\mathcal{E}} P(\rho(X_1^{(i)}, X_1^{(j)}) \le \rho(i,j) - 1).$$

by arguing that, for arbitrary (i,k), the process $\rho(X_t^{(i)}, X_t^{(k)})$ can be compared to a mean-zero random walk with chance α of making a negative step. Formalizing this idea seems subtle. Consider three states i, j, k with $(i,j) \in \mathcal{E}$ and $(j,k) \in \mathcal{E}$. Suppose

$$P(\rho(X_1^{(i)}, X_1^{(j)}) = \rho(i,j) + 1) = P(\rho(X_1^{(i)}, X_1^{(j)}) = \rho(i,j) - 1) = \alpha$$

and otherwise $\rho(\cdot,\cdot)$ is unchanged; similarly for (j,k). The changes for the (i,j) process and for the (j,k) process will typically be dependent, and in the extreme case we might have

$$\rho(X_1^{(i)}, X_1^{(j)}) = \rho(i,j) + 1 \text{ iff } \rho(X_1^{(j)}, X_1^{(k)}) = \rho(j,k) - 1$$

and symmetrically, in which case $\rho(X_t^{(i)}, X_t^{(k)})$ might not change at all. Thus proving a result like (12.18) must require further assumptions.

Section 12.1.13. The Markov chain here (with uniform weights) was first studied by Karzanov and Khachiyan [211].

Chapter 13

Continuous State, Infinite State and Random Environment (June 23, 2001)

13.1 Continuous state space

We have said several times that the theory in this book is fundamentally a theory of *inequalities*. "Universal" or "a priori" inequalities for reversible chains on finite state space, such as those in Chapter 4, should extend unchanged to the continuous space setting. Giving proofs of this, or giving the rigorous setup for continuous-space chains, is outside the scope of our intermediate-level treatment. Instead we just mention a few specific processes which parallel or give insight into topics treated earlier.

13.1.1 One-dimensional Brownian motion and variants

Let $(B_t, 0 \le t < \infty)$ be one-dimensional *standard Brownian motion* (BM). Mentally picture a particle moving along an erratic continuous random trajectory. Briefly, for $s < t$ the increment $B_t - B_s$ has Normal$(0, t-s)$ distribution, and for non-overlapping intervals (s_i, t_i) the increments $B_{t_i} - B_{s_i}$ are independent. See Norris [270] section 4.4, Karlin and Taylor [208] Chapter 7, or Durrett [133] Chapter 7 for successively more detailed introductions. One can do explicit calculations, directly in the continuous setting, of distributions of many random quantities associated with BM. A particular

calculation we need ([133] equation 7.8.12) is

$$G(t) := P\left(\sup_{0\le s\le t} |B_s| < 1\right) = \frac{4}{\pi}\sum_{m=0}^{\infty} \frac{(-1)^m}{2m+1} \exp(-(2m+1)^2\pi^2 t/8) \quad (13.1)$$

where

$$G^{-1}(1/e) = 1.006. \quad (13.2)$$

One can also regard BM as a limit of rescaled random walk, a result which generalizes the classical central limit theorem. If $(X_m, m = 0, 1, 2, \ldots)$ is simple symmetric random walk on Z, then the central limit theorem implies $m^{-1/2}X_m \stackrel{d}{\to} B_1$ and the generalized result is

$$(m^{-1/2}X_{\lfloor mt\rfloor}, 0 \le t < \infty) \stackrel{d}{\to} (B_t, 0 \le t < \infty) \quad (13.3)$$

where the convergence here is *weak convergence of processes* (see e.g. Ethier and Kurtz [141] for detailed treatment). For more general random flights on Z, that is $X_m = \sum_{j=1}^{m} \xi_j$ with $\xi_1, \xi_2, \ldots$ independent and $E\xi = 0$ and var $\xi = \sigma^2 < \infty$, we have *Donsker's theorem* ([133] Theorem 7.6.6)

$$(m^{-1/2}X_{\lfloor mt\rfloor}, 0 \le t < \infty) \stackrel{d}{\to} (\sigma B_t, 0 \le t < \infty). \quad (13.4)$$

Many asymptotic results for random walk on the integers or on the n-cycle or on the n-path, and their d-dimensional counterparts, can be explained in terms of Brownian motion or its variants. The variants of interest to us take values in compact sets and have uniform stationary distributions.

Brownian motion on the circle can be defined by

$$B^\circ_t := B_t \bmod 1$$

and then random walk $(X^{(n)}_m, m = 0, 1, 2, \ldots)$ on the n-cycle $\{0, 1, 2, \ldots, n-1\}$ satisfies, by (13.3),

$$n^{-1}(X^{(n)}_{\lfloor n^2 t\rfloor}, 0 \le t < \infty) \stackrel{d}{\to} (B^\circ_t, 0 \le t < \infty) \text{ as } n \to \infty. \quad (13.5)$$

The process B° has eigenvalues $\{2\pi^2 j^2, \quad 0 \le j < \infty\}$ with eigenfunction $\equiv 1$ for $j = 0$ and two eigenfunctions $\cos(2\pi jx)$ and $\sin(2\pi jx)$ for $j \ge 1$. In particular the relaxation time is

$$\tau_2 = \tfrac{1}{2\pi^2}.$$

The result for random walk on the n-cycle (Chapter 5 Example 7)

$$\tau_2 \sim \tfrac{n^2}{2\pi^2} \text{ as } n \to \infty$$

can therefore be viewed as a consequence of the n^2 time-rescaling in (13.5) which takes random walk on the n-cycle to Brownian motion on the circle. This argument is a prototype for the *weak convergence paradigm*: proving size-asymptotic results for discrete structures in terms of some limiting continuous structure.

Variation distance can be studied via coupling. Construct two Brownian motions on R started from 0 and $x > 0$ as follows. Let $B^{(1)}$ be standard Brownian motion, and let

$$T_{x/2} := \inf\{t : B^{(1)}_t = x/2\}.$$

Then $T_{x/2} < \infty$ a.s. and we can define $B^{(2)}$ by

$$\begin{aligned} B^{(2)}_t &= x - B^{(1)}_t, \quad 0 \le t \le T_{x/2} \\ &= B^{(1)}_t, \quad T_{x/2} \le t < \infty. \end{aligned}$$

That is, the segment of $B^{(2)}$ over $0 \le t \le T_{x/2}$ is the image of the corresponding segment of $B^{(1)}$ under the reflection which takes 0 to x. It is easy to see that $B^{(2)}$ is indeed Brownian motion started at x. This is the *reflection coupling* for Brownian motion. We shall study analogous couplings for variant processes. Given Brownian motion on the circle $B^{\circ 1}$ started at 0, we can construct another Brownian motion on the circle $B^{\circ 2}$ started at $0 < x \le 1/2$ via

$$\begin{aligned} B^{\circ 2}_t &= x - B^{\circ 1}_t \bmod 1, \quad 0 \le t \le T_{\{\frac{x}{2}, \frac{x}{2}+\frac{1}{2}\}} \\ &= B^{\circ 1}_t, \quad T_{\{\frac{x}{2}, \frac{x}{2}+\frac{1}{2}\}} \le t < \infty \end{aligned}$$

where

$$T_{\{\frac{x}{2}, \frac{x}{2}+\frac{1}{2}\}} := \inf\{t : B^{\circ 1}_t = \tfrac{x}{2} \text{ or } \tfrac{x}{2} + \tfrac{1}{2}\}.$$

Again, the segment of $B^{\circ 2}$ over $0 \le t \le T_{\{\frac{x}{2}, \frac{x}{2}+\frac{1}{2}\}}$ is the image of the corresponding segment of $B^{\circ 1}$ under the reflection of the circle which takes 0 to x, so we call it the reflection coupling for Brownian motion on the circle. Because sample paths cannot cross without meeting, it is easy to see that the general coupling inequality (Chapter 4-3 section 1.1) becomes an equality:

$$||P_0(B^\circ_t \in \cdot) - P_x(B^\circ_t \in \cdot)|| = P(T_{\{\frac{x}{2}, \frac{x}{2}+\frac{1}{2}\}} > t).$$

The worst starting point is $x = 1/2$, and the hitting time in question can be written as the hitting time $T_{\{-1/4,1/4\}}$ for standard Brownian motion, so

$$\bar{d}(t) = P(T_{\{-1/4,1/4\}} > t) = G(16t) \tag{13.6}$$

by *Brownian scaling*, that is the property

$$(B_{c^2 t}, 0 \le t < \infty) \stackrel{d}{=} (cB(t), 0 \le t < \infty). \tag{13.7}$$

See the Notes for an alternative formula. Thus for Brownian motion on the circle

$$\tau_1 = \tfrac{1}{16} G^{-1}(1/e) = 0.063. \tag{13.8}$$

If simple random walk is replaced by aperiodic random flight with step variance σ^2 then the asymptotic values of τ_2 and τ_1 are replaced by τ_2/σ^2 and τ_1/σ^2; this may be deduced using the local central limit theorem ([133] Theorem 2.5.2).

Reflecting Brownian motion $\bar{B}$ on the interval $[0, 1]$ is very similar. Intuitively, imagine that upon hitting an endpoint 0 or 1 the particle is instantaneously inserted an infinitesimal distance into the interval. Formally one can construct $\bar{B}_t$ as $\bar{B}_t := \phi(B_t)$ for the concertina map

$$\phi(2j+x) = x, \quad \phi(2j+1+x) = 1-x; \quad 0 \le x \le 1,\ j = \ldots -2, -1, 0, 1, 2, \ldots.$$

The process $\bar{B}$ has eigenvalues $\{\pi^2 j^2/2, \quad 0 \le j < \infty\}$ with eigenfunctions $\cos(\pi j x)$. In particular the relaxation time is

$$\tau_2 = \tfrac{2}{\pi^2}.$$

The result for random walk on the n-path (Chapter 5 Example 8)

$$\tau_2 \sim \tfrac{2n^2}{\pi^2} \text{ as } n \to \infty$$

is another instance of the weak convergence paradigm, a consequence of the n^2 time-rescaling which takes random walk on the n-path to reflecting Brownian motion on the interval. The variation distance function $\bar{d}(t)$ for $\bar{B}$ can be expressed in terms of the corresponding quantity (write as $d^\circ(t)$) for B°. Briefly, it is easy to check

$$(\bar{B}_t, 0 \le t < \infty) \stackrel{d}{=} (2\min(B^\circ_{t/4}, 1 - B^\circ_{t/4}), 0 \le t < \infty)$$

and then to deduce $\bar{d}(t) = d^\circ(t/4)$. Then using (13.8)

$$\tau_1 = \tfrac{1}{4} G^{-1}(1/e) = 0.252. \tag{13.9}$$

13.1.2 d-dimensional Brownian motion

Standard d-dimensional Brownian motion can be written as

$$\mathbf{B}_t = (B_t^{(1)}, \ldots, B_t^{(d)})$$

where the component processes $(B_t^{(i)}, i = 1, \ldots, d)$ are independent one-dimensional standard Brownian motions. A useful property of $\mathbf{B}$ is *isotropy*: its distribution is invariant under rotations of R^d. In approximating simple random walk $(X_m, m = 0, 1, 2, \ldots)$ on Z^d one needs to be a little careful with scaling constants. The analog of (13.3) is

$$(m^{-1/2} X_{\lfloor mt \rfloor}, 0 \le t < \infty) \stackrel{d}{\to} (d^{-1/2}\mathbf{B}_t, 0 \le t < \infty) \tag{13.10}$$

where the factor $d^{-1/2}$ arises because the components of the random walk have variance $1/d$ — see (13.4). Analogous to (13.5), random walk $(X_m^{(n)}, m = 0, 1, 2, \ldots)$ on the discrete torus Z_n^d converges to Brownian motion $\mathbf{B}^\circ$ on the continuous torus $[0, 1)^d$:

$$n^{-1}(X^{(n)}_{\lfloor n^2 t \rfloor}, 0 \le t < \infty) \stackrel{d}{\to} (d^{-1/2}\mathbf{B}^\circ_t, 0 \le t < \infty) \text{ as } n \to \infty. \tag{13.11}$$

13.1.3 Brownian motion in a convex set

Fix a convex polyhedron $K \subset R^d$. One can define reflecting Brownian motion in K; heuristically, when the particle hits a face it is replaced an infinitesimal distance inside K, orthogonal to the face. As in the previous examples, the stationary distribution is uniform on K. We will outline a proof of

Proposition 13.1 *For Brownian motion $\mathbf{B}$ in a convex polyhedron K which is a subset of the ball of radius r,*
(i) $\tau_1 \le G^{-1}(1/e)\ r^2$
(ii) $\tau_2 \le 8\pi^{-2} r^2$.

Proof. By the d-dimensional version of Brownian scaling (13.7) we can reduce to the case $r = 1$. The essential fact is

Lemma 13.2 *Let $\bar{B}_t$ be reflecting Brownian motion on $[0, 1]$ started at 1, and let T_0 be its hitting time on 0. Versions $\mathbf{B}^{(1)}, \mathbf{B}^{(2)}$ of Brownian motion in K started from arbitrary points of K can be constructed jointly with $\bar{B}$ such that*

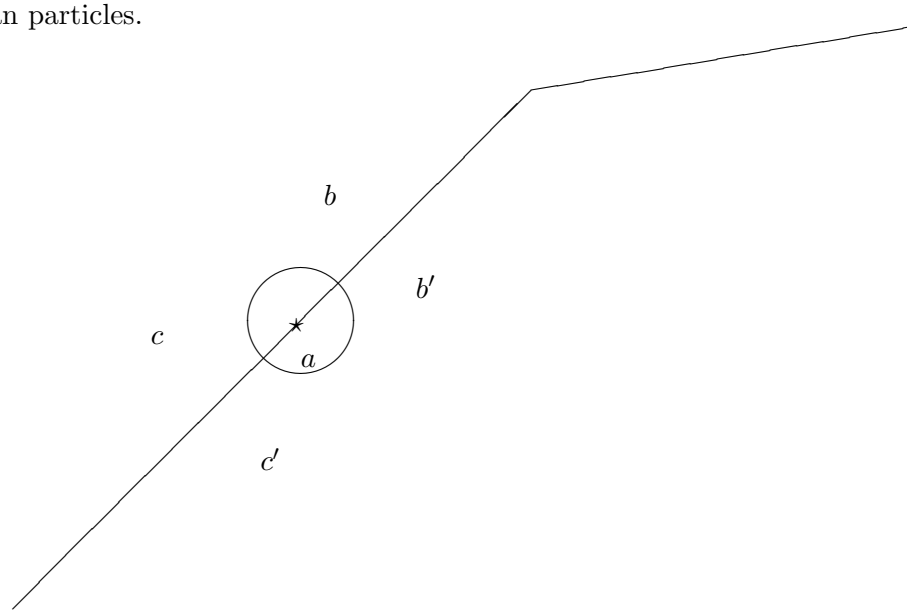
$$|\mathbf{B}_t^{(1)} - \mathbf{B}_t^{(2)}| \le 2\bar{B}_{\min(t, T_0)}, \quad 0 \le t < \infty. \tag{13.12}$$

Granted this fact, $\bar{d}(t)$ for Brownian motion on K satisfies

$$\bar{d}(t) = \max_{\text{startingpoints}} P(\mathbf{B}_t^{(1)} \neq \mathbf{B}_t^{(2)}) \leq P(T_0 > t) = G(t)$$

where the final equality holds because T_0 has the same distribution as the time for Brownian motion stared at 1 to exit the interval $(0, 2)$. This establishes (i) for $r = 1$. Then from the $t \to \infty$ asymptotics of $G(t)$ in (13.1) we have $\bar{d}(t) = O(\exp(-\pi^2 t/8))$, implying $\tau_2 \leq 8/\pi^2$ by Lemma **??** and establishing (ii).

Sketch proof of Lemma. Details require familiarity with stochastic calculus, but this outline provides the idea. For two Brownian motions in R^d started from $(0, 0, \ldots, 0)$ and from $(x, 0, \ldots, 0)$, one can define the reflection coupling by making the first coordinates evolve as the one-dimensional reflection coupling, and making the other coordinate processes be identical in the two motions. Use isotropy to entend the definition of reflection coupling to arbitrary starting points. Note that the distance between the processes evolves as 2 times one-dimensional Brownian motion, until they meet. The desired joint distribution of $((\mathbf{B}_t^{(1)}, \mathbf{B}_t^{(2)}), \quad 0 \leq t < \infty)$ is obtained by specifying that while both processes are in the interior of K, they evolve as the reflection coupling (and each process reflects orthogonally at faces). As the figure illustrates, the effect of reflection can only be to decrease distance between the two Brownian particles.



For a motion hitting the boundary at a, if the unreflected process is at b or c an infinitesimal time later then the reflected process is at b' or c'.

By convexity, for any $x \in K$ we have $|b' - x| \leq |b - x|$; so reflection can only decrease distance between coupled particles. To argue the inequality carefully, let α be the vector normal to the face. The projection $\mathcal{P}_\alpha$ satisfies $|\mathcal{P}_\alpha(b' - x)| \leq |\mathcal{P}_\alpha(b - x)|$. Further, $b - b' \perp \alpha$, implying $P_{\alpha^\perp}(b' - x) = P_{\alpha^\perp}(b - x)$. Therefore by Pythagoras $|b' - x| \leq |b - x|$.

We can therefore write, in stochastic calculus notation,

$$d|\mathbf{B}_t^{(1)} - \mathbf{B}_t^{(2)}| = d(2B_t) - dA_t$$

where B_t is a one-dimensional Brownian motion and A_t is an increasing process (representing the contribution from reflections off faces) which increases only when one process is at a face. But we can construct reflecting Brownian motion $\bar{B}$ in terms of the same underlying B_t by

$$d(2\bar{B}_t) = d(2B_t) - dC_t$$

where C_t (representing the contribution from reflections off the endpoint 1) is increasing until T_0. At time 0 we have (because $r = 1$)

$$|\mathbf{B}_0^{(1)} - \mathbf{B}_0^{(2)}| \leq 2 = 2\bar{B}_0.$$

We have shown

$$d(|\mathbf{B}_t^{(1)} - \mathbf{B}_t^{(2)}| - 2\bar{B}_t) = -dA_t + dC_t.$$

If the desired inequality (13.12 fails then it fails at some first time t, which can only be a time when dC_t is increasing, that is when $\bar{B}_t = 1$, at which times the inequality holds a priori. ■.

Proposition 13.1 suggests an approach to the algorithmic question of simulating a uniform random point in a convex set $K \subset R^d$ where d is large, discussed in Chapter 9 section 5.1. If we could simulate the discrete-time chain defined as reflecting Brownian motion $\mathbf{B}$ on K examined at time intervals of h^2/d for some small h (so that the length of a typical step is of order $\sqrt{(h^2/d) \times d} = h$), then Proposition 13.1 implies that $O(d/h^2)$ steps are enough to approach the stationary distribution. Since the convex set is available only via an oracle, one can attempt to do the simulation via acceptance/rejection. That is, from $\mathbf{x}$ we propose a move to $\mathbf{x}' = \mathbf{x} + \sqrt{h^2/d}\ Z$ where Z has standard d-variate Normal distribution, and accept the move iff $\mathbf{x}' \in K$. While this leads to a plausible heuristic argument, the rigorous difficulty is that it is not clear how close an acceptance/rejection step is to the true step of reflecting Brownian motion. No rigorous argument based directly on Brownian motion has yet been found, though the work of Bubley et al [80] on coupling of random walks has elements in common with reflection coupling.

13.1.4 Discrete-time chains: an example on the simplex

Discrete-time, continuous-space chains arise in many settings, in particular (Chapter MCMC) in Markov Chain Monte Carlo sampling from a target distribution on R^d. As discussed in that chapter, estimating mixing times for such chains with general target distributions is extremely difficult. The techniques in this book are more directly applicable to chains with (roughly) uniform stationary distribution. The next example is intended to give the flavor of how techniques might be adapted to the continuous setting: we will work through the details of a coupling argument.

Example 13.3 *A random walk on the simplex.*

Fix d and consider the simplex $\Delta = \{\mathbf{x} = (x_1, \ldots, x_d\} : x_i \geq 0, \sum_i x_i = 1\}$. Consider the discrete-time Markov chain $(\mathbf{X}(t), t = 0, 1, 2, \ldots)$ on Δ with steps:

from state $\mathbf{x}$, pick 2 distinct coordinates $\{i, j\}$ uniformly at random, and replace the 2 entries $\{x_i, x_j\}$ by $\{U, x_i + x_j - U\}$ where U is uniform on $(0, x_i + x_j)$.

The stationary distribution π is the uniform distribution on Δ. We will show that the mixing time τ_1 satisfies

$$\tau_1 = O(d^2 \log d) \text{ as } d \to \infty. \tag{13.13}$$

The process is somewhat reminiscent of card shuffling by random transpositions (Chapter 7 Example 18), so by analogy with that example we expect that in fact $\tau_1 = \Theta(d \log d)$. What we show here is that the coupling analysis of that example (Chapter 4-3 section 1.7) extends fairly easily to the present example.

As a preliminary, let us specify two distinct couplings (A, B) of the uniform$(0, a)$ and the uniform$(0, b)$ distributions. In the *scaling coupling* we take $(A, B) = (aU, bU)$ for U with uniform$(0, 1)$ distribution. In the *greedy coupling* we make $P(A = B)$ have its maximal value, which is $\min(a, b)/\max(a, b)$, and we say the coupling *works* if $A = B$.

Fix $\mathbf{x}(0) \in \Delta$. We now specify a coupling $(\mathbf{X}(t), \mathbf{Y}(t))$ of the chains started with $\mathbf{X}(0) = \mathbf{x}(0)$ and with $\mathbf{Y}(0)$ having the uniform distribution. (This is an atypical couplig argument, in that it matters that one version is the stationary version).

> From state $(\mathbf{x}, \mathbf{y})$, choose the same random pair $\{i, j\}$ for each process, and link the new values x_i' and y_i' (which are uniform

on different intervals) via the scaling coupling for the first $t_1 = 3d^2 \log d$ steps, then via the greedy coupling for the next $t_2 = Cd^2$ steps.

We shall show that, for any fixed constant $C > 0$,

$$P(\mathbf{X}(t_1 + t_2) = \mathbf{Y}(t_1 + t_2)) \geq 1 - C^{-1} - o(1) \text{ as } d \to \infty \tag{13.14}$$

establishing (13.13).

Consider the effect on l_1 distance $||\mathbf{x} - \mathbf{y}|| := \sum_i |x_i - y_i|$ of a step of the scaling coupling using coordinates $\{i, j\}$. The change is

$$|U(x_i+x_j)-U(y_i+y_j)|+|(1-U)(x_i+x_j)-(1-U)(y_i+y_j)|-|x_i-y_i|-|x_j-y_j|$$

$$= |(x_i + x_j) - (y_i + y_j)| - |x_i - y_i| - |x_j - y_j|$$

$$= \begin{cases} 0 & \text{if sgn } (x_i - y_i) = \text{sgn } (x_j - y_j) \\ -2\min(|x_i - y_i|, |x_j - y_j|) & \text{if not .} \end{cases}$$

Thus

$$E_{(\mathbf{x},\mathbf{y})} \left(||\mathbf{X}(1) - \mathbf{Y}(1)|| - ||\mathbf{x} - \mathbf{y}|| \right)$$

$$= \frac{-2}{d(d-1)} \sum_i \sum_{j \neq i} \min(|x_i - y_i|, |x_j - y_j|) 1_{(\text{sgn } (x_i - y_i) \neq \text{sgn } (x_j - y_j))}$$

$$= \frac{-4}{d(d-1)} \sum_{i \in A} \sum_{j \in B} \min(c_i, d_j)$$

(where $c_i := x_i - y_i$ on $A := \{i : x_i > y_i\}$; $d_j := y_j - x_j$ on $B := \{j : y_j > x_j\}$)

$$\begin{aligned} &= \frac{-4}{d(d-1)} \sum_{i \in A} \sum_{j \in B} \frac{c_i d_j}{\max(c_i, d_j)} \\ &\leq \frac{-4}{d(d-1)} \sum_{i \in A} \sum_{j \in B} \frac{c_i d_j}{||\mathbf{x} - \mathbf{y}||/2} \\ &= \frac{-2}{d(d-1)} ||\mathbf{x} - \mathbf{y}|| \end{aligned}$$

because $\sum_{i \in A} c_i = \sum_{j \in B} d_j = ||\mathbf{x} - \mathbf{y}||/2$. So

$$E_{(\mathbf{x},\mathbf{y})} ||\mathbf{X}(1) - \mathbf{Y}(1)|| \leq \left(1 - \frac{2}{d(d-1)}\right) ||\mathbf{x} - \mathbf{y}||.$$

Because $||\mathbf{X}(0) - \mathbf{Y}(0)|| \le 2$, it follows that after t steps using the scaling coupling,

$$E||\mathbf{X}(t) - \mathbf{Y}(t)|| \le 2\left(1 - \frac{2}{d(d-1)}\right)^t.$$

So by taking $t_1 \sim 3d^2 \log d$, after t_1 steps we have

$$P(||\mathbf{X}(t_1) - \mathbf{Y}(t_1)|| \le d^{-5}) = 1 - o(1). \tag{13.15}$$

Now consider the greedy coupling. If a step works, the l_1 distance $||\mathbf{X}(t) - \mathbf{Y}(t)||$ cannot increase. The chance that a step from $(\mathbf{x}, \mathbf{y})$ involving coordinates $\{i, j\}$ works is

$$\begin{aligned}
\frac{\min(x_i + x_j, y_i + y_j)}{\max(x_i + x_j, y_i + y_j)} &\ge \frac{y_i + y_j - ||\mathbf{x} - \mathbf{y}||}{\max(x_i + x_j, y_i + y_j)} \\
&\ge \frac{y_i + y_j - ||\mathbf{x} - \mathbf{y}||}{y_i + y_j + ||\mathbf{x} - \mathbf{y}||} \\
&\ge \frac{y_i + y_j - 2||\mathbf{x} - \mathbf{y}||}{y_i + y_j} \\
&\ge 1 - \frac{||\mathbf{x} - \mathbf{y}||}{\min(y_i, y_j)}.
\end{aligned}$$

So unconditionally

$$P_{(\mathbf{x},\mathbf{y})}(\text{greedy coupling works on first step}) \ge 1 - \frac{||\mathbf{x} - \mathbf{y}||}{\min_k y_k}. \tag{13.16}$$

Now the uniform distribution $(Y_1^{(d)}, \ldots, Y_d^{(d)})$ on the simplex has the property (use [133] Exercise 2.6.10 and the fact that the uniform distribution on the simplex is the joint distribution of spacings between $d-1$ uniform$(0,1)$ variables and the endpoint 1)

$$\text{if constants } a_d > 0 \text{ satisfy } da_d \to 0 \text{ then } P(Y_1^{(d)} \le a_d) \sim da_d.$$

Since $(\mathbf{Y}(t))$ is the stationary chain and $Y_i^{(d)} \stackrel{d}{=} Y_1^{(d)}$,

$$P(\min_{1 \le k \le d} Y_k(t) \le d^{-4.5} \text{ for some } t_1 < t \le t_1 + t_2) \le t_2 d P(Y_1^{(d)} < d^{-4.5})$$

and since $t_2 = O(d^2)$ this bound is $o(1)$. In other words

$$P(\min_{1 \le k \le d} Y_k(t) \ge d^{-4.5} \text{ for all } t_1 < t \le t_1 + t_2) = 1 - o(1) \text{ as } d \to \infty.$$

Combining this with (13.15,13.16) and the non-increase of l_1 distance, we deduce

$$P(\text{greedy coupling works for all} t_1 < t \le t_1 + t_2) = 1 - o(1). \qquad (13.17)$$

Now consider the number $M(t)$ of unmatched coordinates i at time $t \ge t_1$, that is, the number of i with $X_i(t) \ne Y_i(t)$. Provided the greedy coupling works, this number $M(t)$ cannot increase, and decreases by at least 1 each time two unmatched coordinates are chosen. So we can compare $(M(t_1 + t), t \ge 0)$ with the chain $(N(t), t \ge 0)$ with $N(0) = d$ and

$$P(N(t+1) = m-1 | N(t) = m) = \frac{m(m-1)}{d(d-1)} = 1 - P(N(t+1) = m | N(t) = m).$$

As in the analysis of the shuffling example, the time $T = \min\{t : N(t) = 1\}$ has $ET = \sum_{m=2}^{d} \frac{d(d-1)}{m(m-1)} \le d^2$. When the number $M(t)$ goes strictly below 2 it must become 0, and so

$$\begin{aligned} &P(\text{greedy coupling works for all } t_1 < t \le t_1 + t_2, \mathbf{X}(t_1 + t_2) \ne \mathbf{Y}(t_1 + t_2)) \\ &= P(\text{greedy coupling works for all } t_1 < t \le t_1 + t_2, M(t_1 + t_2) > 1) \\ &\le P(T > t_2) \le 1/C_2. \end{aligned}$$

This and (13.17) establish (13.14).

13.1.5 Compact groups

Parallel to random flights on finite groups one can discuss discrete-time random flights on classical (continuous) compact groups such as the *orthogonal group* $O(d)$ of $d \times d$ real orthogonal matrices. For instance, specify a *reflection* to be an automorphism which fixes the points in some hyperplane, so that a reflection matrix can be written as

$$A = I - 2xx^T$$

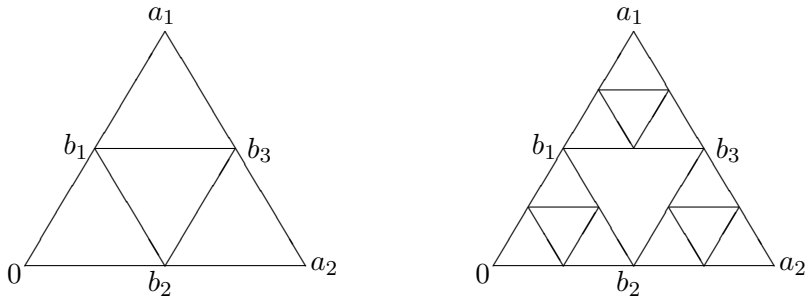
where I is the $d \times d$ identity matrix and x is a unit-length vector in R^d. Assigning to x the Haar measure on the $(d-1)$-sphere creates a *uniform random reflection*, and a sequence of uniform random reflections define a random flight on $O(d)$. Porod [285] shows that the variation threshold satisfies

$$\tau_1 \sim \tfrac{1}{2} d \log d$$

and that the cut-off phenomenon occurs. The result, and its proof via group representation theory, are reminiscent of card-shuffling via random transpositions (Chapter 7 Example 18).

13.1.6 Brownian motion on a fractal set

Constructions and properties of analogs of Brownian motion taking values in fractal subsets of R^d have been studied in great detail over the last 15 years. Since these processes are most easily viewed as limits of random walks on graphs, we shall say a little about the simplest example. The figure illustrates the first two stages of the construction of the well-known *Sierpinski gasket*.



Graph G_1 Graph G_2

In the topology setting one may regard G_d as a closed subset of R^2, that is as a set of line-segments, and then the closure of $\cup_{d=1}^{\infty} G_d$ is the Sierpinski gasket G (this is equivalent to the usual construction by "cutting out middle triangles"). In the graph setting, regard G_d as a graph and write $(X_t^{(d)}, t = 0, 1, 2, \ldots)$ for discrete-time random walk on G_d started at point 0. Let M_d be the number of steps of $X^{(d)}$ until first hitting point a_1 or a_2. Using symmetry properties of the graphs, there is a simple relationship between the distributions of M_1 and M_2. For the walk on G_2, the length of the time segment until first hitting b_1 or b_2 is distributed as M_1; successive segments (periods until next hitting one of $\{0, a_1, a_2, b_1, b_2, b_3\}$ other than the current one) are like successive steps of the walk on G_1, so the number of segments is distributed as M_1. Using the same argument for general d gives

> M_d is distributed as the d'th generation size in a Galton-Watson branching process with 1 individual in generation 0 and offspring distributed as M_1.

It is easy to calculate $EM_1 = 5$; indeed the distribution of M_1 is determined by its generating function, which can be calculated to be $Ez^{M_1} = z^2/(4-3z)$. So $EM_d = 5^d$. This suggests the existence of a limit process on G after

rescaling time, that is a limit

$$(X^{(d)}_{\lfloor 5^{-d}t \rfloor}, 0 \le t < \infty) \ \stackrel{d}{\to} \ (X^{(\infty)}_t, 0 \le t < \infty).$$

In fact we can be more constructive. Branching process theory ([133] Example 4.4.1) shows that $M_d/5^d \ \stackrel{d}{\to} \ W$ where $EW = 1$ and where W has the self-consistency property

$$\sum_{i=1}^{M} W_i \ \stackrel{d}{=} \ 5W \tag{13.18}$$

where $(M; W_1, W_2, \ldots)$ are independent, $M \ \stackrel{d}{=} \ M_1$ and $W_i \ \stackrel{d}{=} \ W$. Now in the topological setting, the vertices of G_d are a subset of G. Let $\tilde{X}^{(d)}_t$ be the process on $G_d \subset G$ whose sequence of jumps is as the jumps of the discrete-time walk $X^{(d)}$ but where the times between jumps are independent with distribution $5^{-d}W$. Using (13.18) we can construct the processes $\tilde{X}^{(d)}$ jointly for all d such that the process $\tilde{X}^{(d)}$, watched only at the times of hitting (successively distinct) points of G_{d-1}, is exactly the process $\tilde{X}^{(d-1)}$. These coupled processes specify a process $X^{(\infty)}_t$ on G at a random subset of times t. It can be shown that this random subset is dense and that sample paths extend continuously to all t, and it is natural to call $X^{(\infty)}$ *Brownian motion on the Sierpinski gasket*.

13.2 Infinite graphs

There is a huge research literature concerning random walks on infinite discrete groups, and more generally on infinite graphs, and the recent monograph of Woess [339] provides an in-depth treatment. This section focuses narrowly on two aspects of an issue not emphasized in [339]: what does study of random walk on infinite graphs tell us about random walks on finite graphs? One aspect of this issue is that random walks on certain specific infinite graphs may be used to get approximations or inequalities for random walks on specific finite graphs. We treat three examples.

- The infinite lattice Z^d as an approximation to the discrete torus Z^d_N for large N (section 13.2.4).
- The infinite degree-r tree $\mathbf{T}^r$ and bounds for r-regular expander graphs of large size (section 13.2.6).

- The hierarchical tree $\mathbf{T}^r_{\text{hier}}$ as an approximation to balanced $(r-1)$-ary trees (section 13.2.9).

The second aspect concerns properties such as transience, non-trivial boundary, and "spectral radius < 1", which have been well-studied as qualitative properties which an infinite-state chain either possesses or does not possess. What are the quantitative finite-state analogs of such properties? Here actual theorems are scarce; we present conceptual discussion in sections 13.2.3 and 13.2.10 as a spur to future research.

13.2.1 Set-up

We assume the reader has some acquaintance with classical theory (e.g., [133] Chapter 5) for a countable-state irreducible Markov chain, which emphasizes the trichotomy *transient* or *null-recurrent* or *positive-recurrent*. We use the phrase *general chain* to refer to the case of an arbitrary irreducible transition matrix $\mathbf{P}$, without any reversibility assumption.

Recall from Chapter 3 section 2 the identification, in the finite-state setting, of reversible chains and random walks on weighted graphs. Given a reversible chain we defined edge-weights $w_{ij} = \pi_i p_{ij} = \pi_j p_{ji}$; conversely, given edge-weights we defined random walk as the reversible chain

$$p_{vx} = w_{vx}/w_v;\ w_v = \sum_x w_{vx}. \tag{13.19}$$

In the infinite setting it is convenient (for reasons explained below) to take the "weighted graph" viewpoint. Thus the setting of this section is that we are given a connected weighted graph satisfying

$$w_v \equiv \sum_x w_{vx} < \infty\ \forall x, \quad \sum_v w_v = \infty \tag{13.20}$$

and we study the associated *random walk* (X_t), i.e., the discrete-time chain with $p_{vx} = w_{vx}/w_v$. So in the unweighted setting ($w_e \equiv 1$), we have nearest-neighbor random walk on a locally finite, infinite graph.

To explain why we adopt this set-up, say π is *invariant* for $\mathbf{P}$ if

$$\sum_i \pi_i p_{ij} = \pi_j\ \forall j; \quad \pi_j > 0\ \forall j.$$

Consider asymmetric random walk on Z, say

$$p_{i,i+1} = 2/3,\ p_{i,i-1} = 1/3;\ -\infty < i < \infty. \tag{13.21}$$

One easily verifies that each of the two measures $\pi_i = 1$ and $\pi_i = 2^i$ is invariant. Such nonuniqueness makes it awkward to seek to define reversibility of $\mathbf{P}$ via the detailed balance equations

$$\pi_i p_{ij} = \pi_j p_{ji} \ \forall i, j \tag{13.22}$$

without a prior definition of π. Stating definitions via weighted graphs avoids this difficulty.

The second assumption in (13.20), that $\sum_v w_v = \infty$, excludes the positive-recurrent case (see Theorem 13.4 below); because in that case the questions one asks, such as whether the relaxation time τ_2 is finite, can be analyzed by the same techniques as in the finite-state setting.

Our intuitive interpretation of "reversible" in Chapter 3 was "a movie of the chain looks the same run forwards or run backwards". But the chain corresponding to the weighted graph with weights $w_{i,i+1} = 2^i$, which is the chain (13.21) with $\pi_i = 2^i$, has a particle moving towards $+\infty$ and so certainly doesn't satisfy this intuitive notion. On the other hand, a probabilistic interpretation of an infinite invariant measure π is that if we start at time 0 with independent Poisson(π_v) numbers of particles at vertices v, and let the particles move independently according to $\mathbf{P}$, then the particle process is stationary in time. So the detailed balance equations (13.22) correspond to the intuitive "movie" notion of reversible *for the infinite particle process*, rather than for a single chain.

13.2.2 Recurrence and Transience

The next Theorem summarizes parts of the standard theory of general chains (e.g., [133] Chapter 5). Write $\rho_v := P_v(T_v^+ < \infty)$ and let $N_v(\infty)$ be the total number of visits (including time 0) to v.

Theorem 13.4 *For a general chain, one of the following alternatives holds.*
Recurrent. $\rho_v = 1$ *and* $E_v N_v(\infty) = \infty$ *and* $P_v(N_w(\infty) = \infty) = 1$ *for all* v, w.
Transient. $\rho_v < 1$ *and* $E_v N_v(\infty) < \infty$ *and* $P_v(N_w(\infty) < \infty) = 1$ *for all* v, w.
In the recurrent case there exists an invariant measure π*, unique up to constant multiples, and the chain is either*
positive-recurrent*:* $E_v T_v^+ < \infty \ \forall v$ *and* $\sum_v \pi_v < \infty$*; or*
null-recurrent*:* $E_v T_v^+ = \infty \ \forall v$ *and* $\sum_v \pi_v = \infty$.
In the transient and null-recurrent cases, $P_v(X_t = w) \to 0$ *as* $t \to \infty$ *for all* v, w.

Specializing to random walk on a weighted graph, the measure (w_v) is invariant, and the second assumption in (13.20) implies that the walk cannot be positive-recurrent. By a natural abuse of language we call the weighted graph *recurrent* or *transient*. Because $E_v N_v(\infty) = \sum_t p^{(t)}_{vv}$, Theorem 13.4 contains the "classical" method to establish transience or recurrence by considering the $t \to \infty$ behavior of $p^{(t)}_{vv}$. This method works easily for random walk on Z^d (section 13.2.4).

Some of the "electrical network" story from Chapter 3 extends immediately to the infinite setting. Recall the notion of a flow $\mathbf{f}$, and the net flow $f_{(x)}$ out of a vertex x. Say $\mathbf{f}$ is a *unit flow from x to infinity* if $f_{(x)} = 1$ and $f_{(v)} = 0\ \forall v \neq x$. Thompson's principle (Chapter 3 Proposition 35) extends to the infinite setting, by considering subsets $A_n \downarrow \phi$ (the empty set) with A_n^c finite.

Theorem 13.5 *Consider a weighted graph satisfying (13.20). For each v,*

$$\inf \left\{ \tfrac{1}{2} \sum_e f_e^2 / w_e : \mathbf{f} \text{ a unit flow from } v \text{ to infinity} \right\} = \frac{1}{w_v(1-\rho_v)}.$$

In particular, the random walk is transient iff for some (all) v there exists a unit flow $\mathbf{f}$ from v to infinity such that $\sum_e f_e^2/w_e < \infty$.

By analogy with the finite setting, we can regard the *inf* as the effective resistance between v and infinity, although (see section **??**) we shall not attempt an axiomatic treatment of infinite electrical networks.

Theorem 13.5 has the following immediate corollary: of course (a) and (b) are logically equivalent.

Corollary 13.6 *(a) If a weighted graph is recurrent, then so is any subgraph.*
(b) To show that a weighted graph is transient, it suffices to find a transient subgraph.

Thus the classical fact that Z^2 is recurrent implies that a subgraph of Z^2 is recurrent, a fact which is hard to prove by bounding t-step transition probabilities. In the other direction, it is possible (but not trivial) to prove that Z^3 is transient by exhibiting a flow: indeed Doyle and Snell [131] construct a transient tree-like subgraph of Z^3.

Here is a different formulation of the same idea.

Corollary 13.7 *The return probability $\rho_v = P_v(T_v^+ < \infty)$ cannot increase if a new edge (not incident at v) is added, or the weight of an existing edge (not incident at v) is increased.*

13.2.3 The finite analog of transience

Recall the mean hitting time parameter τ_0 from Chapter 4. For a sequence of n-state reversible chains, consider the property

$$n^{-1}\tau_0(n) \text{ is bounded as } n \to \infty. \tag{13.23}$$

We assert, as a conceptual paradigm, that property (13.23) is the analog of the "transient" property for a single infinite-state chain. The connection is easy to see algebraically for symmetric chains (Chapter 7), where $\tau_0 = E_\pi T_v$ for each v, so that by Chapter 2 Lemma 10

$$n^{-1}\tau_0 = z_{vv} = \sum_{t=0}^{\infty}(p_v v^{(t)} - n^{-1}).$$

The boundedness (in n) of this sum is a natural analog of the transience condition

$$\sum_{t=0}^{\infty} p_{vv}^{(t)} < \infty$$

for a single infinite-state chain. So in principle the methods used to determine transience or recurrence in the infinite-state case ([339] Chapter 1) should be usable to determine whether property (13.23) holds for finite families, and indeed Proposition 37 of Chapter 3 provides a tool for this purpose. In practice these extremal methods haven't yet proved very successful; early papers [85] proved (13.23) for expanders in this way, but other methods are easier (see our proof of Chapter 9 Theorem 1). There is well-developed theory ([339] section 6) which establishes recurrence for infinite *planar* graphs under mild assumptions. It is natural to conjecture that under similar assumptions, a planar n-vertex graph has $\tau_0 = \Theta(n \log n)$, as in the case of Z^2 in Proposition 13.8 below.

13.2.4 Random walk on Z^d

We consider the lattice Z^d as an infinite $2d$-regular unweighted graph. Write X_t for simple random walk on Z^d, and write $\widetilde{X}_t$ for the continuized random walk. Of course, general random flights (i.e. "random walks", in everyone's terminology except ours) and their numerous variations comprise a well-studied classical topic in probability theory. See Hughes [184] for a wide-ranging intermediate-level treatment, emphasizing physics applications. Our discussion here is very narrow, relating to topics treated elsewhere in this book.

To start some calculations, for $d = 1$ consider

$$\begin{aligned}
\tilde{p}(t) &\equiv P_0(\widetilde{X}_t = 0) \\
&= P(J_t^+ = J_t^-), \text{ where } J_t^+ \text{ and } J_t^- \text{ are the} \\
&\quad \text{independent Poisson}(t/2) \text{ numbers of } +1 \text{ and } -1 \text{ jumps} \\
&= \sum_{n=0}^{\infty} \left(\frac{e^{-t/2}(t/2)^n}{n!} \right)^2 \\
&= e^{-t} I_0(t)
\end{aligned}$$

where $I_0(t) := \sum_{n=0}^{\infty} \frac{(t/2)^{2n}}{(n!)^2}$ is the *modified Bessel function of the first kind of order* 0. Now var $\widetilde{X}_t = t$, and as a consequence of the local CLT (or by quoting asymptotics of the Bessel function I_0) we have

$$\tilde{p}(t) \sim (2\pi t)^{-1/2} \text{ as } t \to \infty. \tag{13.24}$$

As discussed in Chapter 4 section 6.2 and Chapter 5 Example 17, a great advantage of working in continuous time in dimensions $d \geq 2$ is that the coordinate processes are independent copies of slowed-down one-dimensional processes, so that $\tilde{p}^{(d)}(t) \equiv P_0(\widetilde{X}_t = 0)$ in dimension d satisfies

$$\tilde{p}^{(d)}(t) = (\tilde{p}(t/d))^d = e^{-t}(I_0(t/d))^d. \tag{13.25}$$

In particular, from (13.24),

$$\tilde{p}^{(d)}(t) \sim (\tfrac{d}{2\pi})^{d/2} t^{-d/2} \text{ as } t \to \infty. \tag{13.26}$$

One can do a similar analysis in the discrete time case. In dimension $d = 1$,

$$\begin{aligned}
p(t) &\equiv P_0(X_t = 0) \\
&= 2^{-t} \binom{t}{t/2}, t \text{ even} \\
&\sim 2\ (2\pi t)^{-1/2} \text{ as } t \to \infty,\ t \text{ even}.
\end{aligned} \tag{13.27}$$

This agrees with (13.26) but with an extra "artificial" factor of 2 arising from periodicity. A more tedious argument gives the analog of (13.26) in discrete time for general d:

$$p^{(d)}(t) \sim 2(\tfrac{d}{2\pi})^{d/2} t^{-d/2} \text{ as } t \to \infty,\ t \text{ even}. \tag{13.28}$$

From the viewpoint of classical probability, one can regard (13.26,13.28) as the special case $j = 0$ of the local CLT: in continuous time in dimension d,

$$\sup_j \left| P_0(\widetilde{X}_t = j) - (\tfrac{d}{2\pi})^{d/2} t^{-d/2} \exp(-d|j|^2/(2t)) \right| = o(t^{-d/2}) \text{ as } t \to \infty$$

where $|j|$ denotes Euclidean norm.

The occupation time $N_0(t)$ satisfies $E_0 N_0(t) = \int_0^t \tilde{p}(s)\ ds$ (continuous time) and $= \sum_{s=0}^{t-1} p(s)$ (discrete time). In either case, as $t \to \infty$,

$$(d=1)\quad E_0 N_0(t) \sim \sqrt{\tfrac{2}{\pi}}\ t^{1/2} \tag{13.29}$$

$$(d=2)\quad E_0 N_0(t) \sim \tfrac{1}{\pi} \log t \tag{13.30}$$

$$(d \geq 3)\quad E_0 N_0(t) \to R_d \equiv \int_0^\infty \tilde{p}^{(d)}(t) dt$$
$$= \int_0^\infty e^{-t} (I_0(t/d))^d\ dt \tag{13.31}$$

where $R_d < \infty$ for $d \geq 3$ by (13.26). This is the classical argument for establishing transience in $d \geq 3$ and recurrence in $d \leq 2$, by applying Theorem 13.4. Note that the return probability $\rho^{(d)} := P_0(T_0^+ < \infty)$ is related to $E_0 N_0(\infty)$ by $E_0 N_0(\infty) = \frac{1}{1-\rho^{(d)}}$; in other words

$$\rho^{(d)} = \frac{R_d - 1}{R_d},\ d \geq 3.$$

Textbooks sometimes give the impression that calculating $\rho^{(d)}$ is hard, but one can just calculate numerically the integral (13.31). Or see [174] for a table.

The quantity $\rho^{(d)}$ has the following sample path interpretation. Let V_t be the number of distinct vertices visited by the walk before time t. Then

$$t^{-1} V_t \to 1 - \rho^{(d)} \text{ a.s. },\ d \geq 3. \tag{13.32}$$

The proof of this result is a textbook application of the ergodic theorem for stationary processes: see [133] Theorem 6.3.1.

13.2.5 The torus Z_m^d

We now discuss how random walk on Z^d relates to $m \to \infty$ asymptotics for random walk on the finite torus Z_m^d, discussed in Chapter 5. We now use superscript $\cdot^{(m)}$ to denote the length parameter. From Chapter 5 Example 17 we have

$$\tau_2^{(m)} = \frac{d}{1 - \cos(2\pi/m)} \sim \frac{dm^2}{2\pi^2}$$
$$\tau_1^{(m)} = \Theta(m^2) \tag{13.33}$$

where asymptotics are as $m \to \infty$ for fixed d. One can interpret this as a consequence of the dN^2 time rescaling in the wweak convergence of rescaled

random walk to Brownian motion of the d-dimensional torus, for which (cf. sections 13.1.1 and 13.1.2) $\tau_2 = 2\pi^{-2}$. At (74)–(75) of Chapter 5 we saw that the eigentime identity gave an exact formula for the mean hitting time parameter $\tau_0^{(m)}$, whose asymptotics are, for $d \geq 3$,

$$m^{-d}\tau_0^{(m)} \to \hat{R}_d \equiv \int_0^1 \ldots \int_0^1 \frac{1}{\frac{1}{d}\sum_{u=1}^d (1 - \cos(2\pi x_u))}\, dx_1 \ldots dx_d < \infty. \tag{13.34}$$

Here we give an independent analysis of this result, and the case $d = 2$.

Proposition 13.8

$$(d = 1) \qquad \tau_0^{(n)} \sim \tfrac{1}{6}n^2 \tag{13.35}$$

$$(d = 2) \qquad \tau_0^{(m)} \sim 2\pi^{-1}m^2 \log m \tag{13.36}$$

$$(d \geq 3) \qquad \tau_0^{(m)} \sim R_d m^d \tag{13.37}$$

for R_d defined by (13.31). In particular, the expressions for R_d and $\hat{R}_d$ at (13.31) and (13.34) are equal, for $d \geq 3$.

The $d = 1$ result is from Chapter 5 (26). We now prove the other cases.

Proof. We may construct continuized random walk $\widetilde{X}_t^{(m)}$ on Z_m^d from continuized random walk $\widetilde{X}_t$ on Z^d by

$$\widetilde{X}_t^{(m)} = \widetilde{X}_t \bmod m \tag{13.38}$$

and then $P_0(\widetilde{X}_t^{(m)} = 0) \geq P_0(\widetilde{X}_t = 0)$. So

$$\begin{aligned} m^{-d}\tau_0^{(m)} &= \int_0^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-d}\right)\, dt \\ &\quad \text{(Chapter 2, Corollary 12 and (8))} \\ &= \int_0^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-d}\right)^+ dt \text{ by complete monotonicity} \\ &\geq \int_0^\infty \left(P_0(\widetilde{X}_t = 0) - m^{-d}\right)^+\, dt \qquad (13.39) \\ &\to \int_0^\infty P_0(\widetilde{X}_t = 0)\, dt \quad = R_d. \end{aligned}$$

Consider the case $d \geq 3$. To complete the proof, we need the corresponding upper bound, for which it is sufficient to show

$$\int_0^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-d} - P_0(\widetilde{X}_t = 0)\right)^+ dt \to 0 \text{ as } m \to \infty. \tag{13.40}$$

To verify (13.40) without detailed calculations, we first establish a 1-dimensional bound

$$(d = 1) \qquad \tilde{p}^{(m)}(t) \le \tfrac{1}{m} + \tilde{p}(t). \tag{13.41}$$

To obtain (13.41) we appeal to a coupling construction (the *reflection coupling*, described in continuous-space in section 13.1.3 – the discrete-space setting here is similar) which shows that continuized random walks $\widetilde{X}^{(m)}, \widetilde{Y}^{(m)}$ on Z_m with $\widetilde{X}_0^{(m)} = 0$ and $\widetilde{Y}_0^{(m)}$ distributed uniformly can be coupled so that

$$\widetilde{Y}_t^{(m)} = 0 \text{ on the event } \{\widetilde{X}_t^{(m)} = 0, T \le t\}$$

where T is the first time that $\widetilde{X}^{(m)}$ goes distance $\lfloor m/2 \rfloor$ from 0. And by considering the construction (13.38)

$$P(\widetilde{X}_t^{(m)} = 0) \le P(\widetilde{X}_t = 0) + P(\widetilde{X}_t^{(m)} = 0, T \le t)$$

and (13.41) follows, since $P(\widetilde{Y}_t^{(m)} = 0) = 1/m$.

Since the d-dimensional probabilities relate to the 1-dimensional probabilities via $P_0(\widetilde{X}_t^{(m)} = 0) = \left(\tilde{p}^{(m)}(t/d)\right)^d$ and similarly on the infinite lattice, we can use inequality (13.41) to bound the integrand in (13.40) as follows.

$$\begin{aligned}
&P_0(\widetilde{X}_t^{(m)} = 0) - m^{-d} - P_0(\widetilde{X}_t = 0) \\
&\quad\le \left(\tfrac{1}{m} + \tilde{p}(t/d)\right)^d - m^{-d} - (\tilde{p}(t/d))^d \\
&\quad= \sum_{j=1}^{d-1} \binom{d}{j} (\tilde{p}(t/d))^j (\tfrac{1}{m})^{d-j} \\
&\quad= \frac{\tilde{p}(t/d)}{m} \sum_{j=1}^{d-1} \binom{d}{j} (\tilde{p}(t/d))^{j-1} (\tfrac{1}{m})^{d-1-j} \\
&\quad\le \frac{\tilde{p}(t/d)}{m} \sum_{j=1} d-1 \binom{d}{j} \left[\max((\tilde{p}(t/d))^{d-2}, \tfrac{1}{m})\right]^{d-2} \\
&\quad= (2^d - 2) \frac{\tilde{p}(t/d)}{m} \left[\max((\tilde{p}(t/d))^{d-2}, (\tfrac{1}{m})^{d-2})\right] \\
&\quad\le (2^d - 2) \frac{\tilde{p}(t/d)}{m} \left[(\tilde{p}(t/d))^{d-2} + \tfrac{1}{m})^{d-2}\right] \\
&\quad= (2^d - 2) \left[\frac{(\tilde{p}(t/d))^{d-1}}{m} + \frac{\tilde{p}(t/d)}{m^{d-1}}\right].
\end{aligned}$$

The fact (13.24) that $\tilde{p}(t) = \Theta(t^{-1/2})$ for large t easily implies that the integral in (13.40) over $0 \le t \le m^3$ tends to zero. But by (13.33) and

submultiplicativity of $\bar{d}(t)$,

$$0 \le P_0(\widetilde{X}_t^{(m)} = 0) - m^{-d} \le d(t) \le \bar{d}(t) \le B_1 \exp(-\tfrac{t}{B_2 m^2}) \qquad (13.42)$$

where B_1, B_2 depend only on d. This easily implies that the integral in (13.40) over $m^3 \le t < \infty$ tends to zero, completing the proof of (13.37).

In the case $d = 2$, we fix $b > 0$ and truncate the integral in (13.39) at bm^2 to get

$$\begin{aligned} m^{-2}\tau_0^{(m)} &\ge -b + \int_0^{bm^2} P_0(\widetilde{X}_t = 0)\ dt \\ &= -b + (1+o(1))\tfrac{2}{\pi}\log(bm^2) \text{ by (13.30)} \\ &= (1+o(1))\tfrac{2}{\pi}\log m. \end{aligned}$$

Therefore

$$\tau_0^{(m)} \ge (1+o(1))\tfrac{2}{\pi} m^2 \log m.$$

For the corresponding upper bound, since $\int_0^{m^2} P_0(\widetilde{X}_t = 0)dt \sim \frac{2}{\pi}\log m$ by (13.30), and $m^{-2}\tau_0^{(m)} = \int_0^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2}\right)\ dt$, it suffices to show that

$$\begin{aligned} &\int_0^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2} - P_0(\widetilde{X}_t = 0)\right)^+\ dt \\ &+ \int_{m^2}^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2}\right)^+ dt = o(\log N). \end{aligned} \qquad (13.43)$$

To bound the first of these two integrals, we observe from (13.41) that $P_0(\widetilde{X}_t^{(m)} = 0) \le (m^{-1} + \tilde{p}(t/2))^2$, and so the integrand is bounded by $\frac{2}{m}\tilde{p}(t/2)$. Using (13.24), the first integral is $O(1) = o(\log m)$. To analyze the second integral in (13.43) we consider separately the ranges $m^2 \le t \le m^2 \log^{3/2} m$ and $m^2 \log^{3/2} m \le t < \infty$. Over the first range, we again use (13.41) to bound the integrand by $\frac{2}{m}\tilde{p}(t/2) + (\tilde{p}(t/2))^2$. Again using (13.24), the integral is bounded by

$$\begin{aligned} &(1+o(1))\frac{2}{\pi^{1/2}m}\int_{m^2}^{m^2\log^{3/2} m} t^{-1/2}dt \ +\ (1+o(1))\pi^{-1}\int_{m^2}^{m^2\log^{3/2} m} t^{-1}dt \\ &= \Theta(\log^{3/4} m) + \Theta(\log\log m) = o(\log m). \end{aligned}$$

To bound the integral over the second range, we use (13.42) and find

$$\begin{aligned} \int_{m^2\log^{3/2} m}^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2}\right) dt &\le B_1 B_2 m^2 \exp(-\tfrac{\log^{3/2} m}{B_2}) \\ &= o(1) = o(\log m). \end{aligned}$$

□

13.2.6 The infinite degree-r tree

Fix $r \geq 3$ and write $\mathbf{T}^r$ for the infinite tree of degree r. We picture $\mathbf{T}^r$ as a "family tree", where the root ϕ has r children, and each other vertex has one parent and $r-1$ children. Being a vertex-transitive graph (recall Chapter 7 section 1.1; for r even, $\mathbf{T}^r$ is the Cayley graph of the free group on $r/2$ generators), one can study many more general "random flights" on $\mathbf{T}^r$ (see Notes), but we shall consider only the simple random walk (X_t).

We can get some information about the walk without resorting to calculations. The "depth" process $d(X_t, \phi)$ is clearly the "reflecting asymmetric random walk" on $Z^+ := \{0, 1, 2, \ldots\}$ with

$$p_{0,1} = 1;\ p_{i,i-1} = 1/r;\ p_{i,i+1} = (r-1)/r,\ i \geq 1.$$

By comparison with asymmetric random walk on all Z, which has drift $(r-2)/r$, we see that

$$t^{-1} d(X_t, \phi) \to \frac{r-2}{r} \text{ a.s. as } t \to \infty. \tag{13.44}$$

In particular, the number of returns to ϕ is finite and so the walk is transient. Now consider the return probability $\rho = P_\phi(T_\phi^+ < \infty)$ and note that (by considering the first step) $\rho = P_\phi(T_c < \infty)$ where c is a child of ϕ. Considering the first two steps, we obtain the equation $\rho = \frac{1}{r} + \frac{r-1}{r}\rho^2$, and since by transience $\rho < 1$, we see that

$$\rho := P_\phi(T_\phi^+ < \infty) = P_\phi(T_c < \infty) = \frac{1}{r-1}. \tag{13.45}$$

So

$$E_\phi N_\phi(\infty) = \frac{1}{1-\rho} = \frac{r-1}{r-2}. \tag{13.46}$$

As at (13.32), ρ has a sample path interpretation: the number V_t of distinct vertices visited by the walk before time t satisfies

$$t^{-1} V_t \to 1 - \rho = \tfrac{r-2}{r-1} \text{ a.s. as } t \to \infty.$$

By transience, amongst the children of ϕ there is some vertex L_1 which is visited last by the walk; then amongst the children of L_1 there is some vertex L_2 which is visited last by the walk; and so on, to define a "path to infinity" $\phi = L_0, L_1, L_2, \ldots$. By symmetry, given $L_1, L_2, \ldots, L_{i-1}$ the conditional distribution of L_i is uniform over the children of L_{i-1}, so in the natural sense we can describe (L_i) as the uniform random path to infinity.

13.2.7 Generating function arguments

While the general qualitative behavior of random walk on $\mathbf{T}^r$ is clear from the arguments above, more precise quantitative estimates are most naturally obtained via generating function arguments. For any state i of a Markov chain, the generating functions $G_i(z) := \sum_{t=0}^{\infty} P_i(X_t = i)z^t$ and $F_i(z) := \sum_{t=1}^{\infty} P_i(T_i^+ = t)z^t$ are related by

$$G_i = 1 + F_i G_i \tag{13.47}$$

(this is a small variation on Chapter 2 Lemma 19). Consider simple *symmetric* reflecting random walk on Z^+. Clearly

$$G_0(z) = \sum_{t=0}^{\infty} \binom{2t}{t} 2^{-2t} z^{2t} = (1 - z^2)^{-1/2},$$

the latter identity being the series expansion of $(1 - x)^{-1/2}$. So by (13.47)

$$F_0(z) := \sum_{t=0 \text{ or } 1}^{\infty} P_0(T_0^+ = 2t)z^{2t} = 1 - (1 - z^2)^{1/2}.$$

Consider an excursion of length $2t$, that is, a path $(0 = i_0, i_1, \ldots, i_{2t-1}, i_{2t} = 0)$ with $i_j > 0, 1 \le j \le 2t - 1$. This excursion has chance 2^{1-2t} for the symmetric walk on Z^+, and has chance $((r-1)/r)^{t-1}(1/r)^t$ for the asymmetric walk $d(X_t, \phi)$. So

$$\frac{P_\phi(T_\phi^+ = 2t)}{P_0(T_0^+ = 2t)} = \frac{r}{2(r-1)} \left(\frac{4(r-1)}{r^2}\right)^t$$

where the numerator refers to simple RW on the tree, and the denominator refers to simple symmetric reflecting RW on Z^+. So on the tree,

$$F_\phi(z) = \frac{r}{2(r-1)} F_0\left(z\sqrt{\tfrac{4(r-1)}{r^2}}\right) = \frac{r}{2(r-1)} \left(1 - \left(1 - \frac{4(r-1)z^2}{r^2}\right)^{1/2}\right).$$

Then (13.47) gives an expression for $G_\phi(z)$ which simplifies to

$$G_\phi(z) = \frac{2(r-1)}{r - 2 + \sqrt{r^2 - 4(r-1)z^2}}. \tag{13.48}$$

In particular, G_ϕ has radius of convergence $1/\beta$, where

$$\beta = 2r^{-1}\sqrt{r-1} < 1. \tag{13.49}$$

Without going into details, one can now use standard Tauberian arguments to show

$$P_\phi(X_t = \phi) \sim \alpha t^{-3/2}\beta^t, \quad t \text{ even} \tag{13.50}$$

for a computable constant α, and this format (for different values of α and β) remains true for more general radially symmetric random flights on $\mathbf{T}^r$ ([339] Theorem 19.30). One can also in principle expand (13.48) as a power series to obtain $P_\phi(X_t = \phi)$. Again we shall not give details, but according to Giacometti [164] one obtains

$$\begin{aligned} P_\phi(X_t = \phi) &= \frac{r-1}{r}\left(\frac{\sqrt{r-1}}{r}\right)^t \frac{\Gamma(1+t)}{\Gamma(2+t/2)\Gamma(1+t/2)} \\ &\quad \times {}_2F_1(\tfrac{t+1}{2}, 1, 2+\tfrac{t}{2}, \tfrac{4(r-1)}{r^2}),\ t \text{ even} \end{aligned} \tag{13.51}$$

where ${}_2F_1$ is the generalized hypergeometric function.

Finally, the β at (13.49) can be interpreted as an eigenvalue for the infinite transition matrix (p_{ij}), so we anticipate a corresponding eigenfunction f_2 with

$$\sum_j p_{ij} f_2(j) = \beta f_2(i)\ \forall i, \tag{13.52}$$

and one can verify this holds for

$$f_2(i) := (1 + \tfrac{r-2}{r} i)(r-1)^{-i/2}. \tag{13.53}$$

13.2.8 Comparison arguments

Fix $r \geq 3$ and consider a sequence (G_n) of n-vertex r-regular graphs with $n \to \infty$. Write (X_t^n) for the random walk on G_n. We can compare these random walks with the random walk (X_t^∞) on $\mathbf{T}^r$ via the obvious inequality

$$P_v(X_t^n = v) \geq P_\phi(X_t^\infty = \phi),\ t \geq 0. \tag{13.54}$$

To spell this out, there is a *universal cover* map $\gamma : \mathbf{T}^r \to G_n$ with $\gamma(\phi) = v$ and such that for each vertex w of $\mathbf{T}^r$ the r edges at w are mapped to the r edges of G_n at $\gamma(w)$. Given the random walk X^∞ on $\mathbf{T}^r$, the definition $X_t^n = \gamma(X_t^\infty)$ constructs random walk on G_n, and (13.54) holds because $\{X_t^n = v\} \supseteq \{X_t^\infty = \phi\}$.

It is easy to use (13.54) to obtain asymptotic lower bounds on the fundamental parameters discussed in Chapter 4. Instead of the relaxation time τ_2, it is more natural here to deal directly with the second eigenvalue λ_2.

Lemma 13.9 *For random walk on n-vertex r-regular graphs, with $r \geq 3$ fixed and $n \to \infty$*
(a) $\liminf n^{-1}\tau_0(n) \geq \frac{r-1}{r-2}$;
(b) $\liminf \frac{\tau_1(n)}{\log n} \geq \frac{r}{(r-2)\log(r-1)}$;
(c) $\liminf \lambda_2(n) \geq \beta := 2r^{-1}\sqrt{r-1}$.

Theory concerning *expanders* (Chapter 9 section 1) shows there exist graphs where the limits above are finite constants (depending on r), so Lemma 13.9 gives the optimal order of magnitude bound.

Proof. For (a), switch to the continuous-time walk, consider an arbitrary vertex v in G_n, and take $t_0(n) \to \infty$ with $t_0(n)/n \to 0$. Then we repeat the argument around (13.39) in the torus setting:

$$\begin{aligned} n^{-1}E_\pi T_v &= \int_0^\infty (P_v(X_t^n = v) - \tfrac{1}{n})\, dt \\ &\geq \int_0^{t_0} (P_v(X_t^n = v) - \tfrac{1}{n})\, dt \\ &\geq -\frac{t_0}{n} + \int_0^{t_0} P_\phi(X_t^\infty = \phi)\, dt \text{ by (13.54)} \\ &\to \int_0^\infty P_\phi(X_t^\infty = \phi)\, dt \\ &= E_\phi N_\phi(\infty) = \frac{r-1}{r-2}, \end{aligned}$$

which is somewhat stronger than assertion (a). Next, the discrete-time spectral representation implies

$$P_v(X_t^n = v) \leq \tfrac{1}{n} + n\beta^t(n).$$

Using (13.54) and (13.50), for any $n \to \infty, t \to \infty$ with t even,

$$t^{-3/2}\beta^t(\alpha - o(1)) \leq \tfrac{1}{n} + n\beta^t(n). \tag{13.55}$$

For (b), the argument for (13.54) gives a coupling between the process X^n started at v and the process X^∞ started at ϕ such that

$$d^n(X_t^n, v) \leq d^\infty(X_t^\infty, \phi)$$

where d^n and d^∞ denote graph distance. Fix $\varepsilon > 0$ and write $\gamma = \frac{r-2}{r} + \varepsilon$. By the coupling and (13.44), $P(d^n(X_t^n, v) \geq \gamma t) \to 0$ as $n, t \to \infty$. This

remains true in continuous time. Clearly $\tau_1(n) \to \infty$, and so by definition of τ_1 we have

$$\limsup \pi\{w : d^n(w,v) \geq \gamma\tau_1(n)\} \leq e^{-1}.$$

But by counting vertices,

$$\begin{aligned} \pi\{w : d^n(v,w) \leq d\} &\leq \frac{1 + r + r(r-1) + \cdots + r(r-1)^{d-1}}{n} \\ &\to 0 \text{ if } d \sim (1-\varepsilon)\frac{\log n}{\log(r-1)}. \end{aligned}$$

For these two limit results to be consistent we must have $\gamma\tau_1(n) \geq (1-\varepsilon)\frac{\log n}{\log(r-1)}$ for all large n, establishing (b).

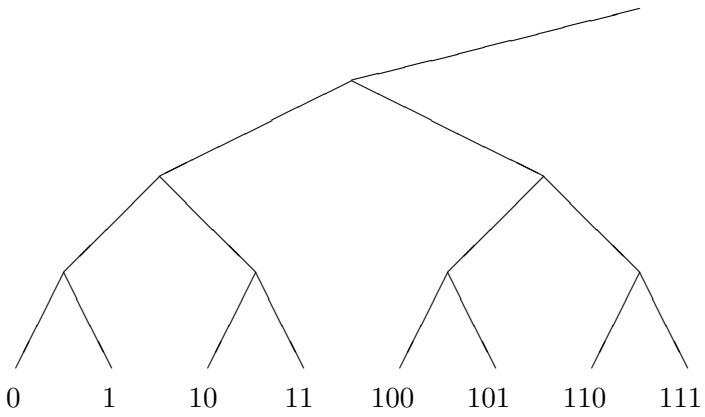
For (c), fix a vertex v_0 of G_n and use the function f_2 at (13.53) to define $f(v) := f_2(d(v, v_0))$ for all vertices v of G_n. The equality (13.52) for f_2 on the infinite tree easily implies the inequality $\mathbf{P}f \geq \beta f$ on G_n. Set $\bar{f} := n^{-1}\sum_v f(v)$ and write 1 for the unit function. By the Rayleigh-Ritz characterization (Chapter 4 eq. (73)), writing $\langle g, h\rangle := \sum_{ij} \pi_i g_i p_{ij} h_j$,

$$\begin{aligned} \lambda_2(n) &\geq \frac{\langle f - \bar{f}1, \mathbf{P}(f - \bar{f}1)\rangle}{||f - \bar{f}1||_2^2} \\ &= \frac{\langle f, \mathbf{P}f\rangle - \bar{f}^2}{||f||_2^2 - \bar{f}^2} \\ &\geq \frac{\beta||f||_2^2 - \bar{f}^2}{||f||_2^2 - \bar{f}^2}. \end{aligned}$$

As $n \to \infty$ we have $\bar{f} \to 0$ while $||f||_2$ tends to a non-zero limit, establishing (c).

13.2.9 The hierarchical tree

Fix $r \geq 2$. There is an infinite tree (illustrated for $r = 2$ in the figure) specified as follows. Each vertex is at some height $0, 1, 2, \ldots$. A vertex at height h has one parent vertex at height $h + 1$ and (if $h \geq 1$) r child vertices at height $h - 1$. The height-0 vertices are leaves, and the set L of leaves has a natural labeling by finite r-ary strings. The figure illustrates the binary ($r = 2$) case, where $L = \{0, 1, 10, 11, 100, 101, \ldots\}$. L forms an Abelian group under entrywise addition modulo r, e.g. for $r = 2$ we have $1101 + 110 = 1101 + 0110 = 1011$. Adopting a name used for generalizations of this construction in statistical physics, we call L the *hierarchical lattice* and the tree $\mathbf{T}^r_{\text{hier}}$ the *hierarchical tree*.



Fix a parameter $0 < \lambda < r$. Consider biased random walk X_t on the tree $\mathbf{T}^r_{\rm hier}$, where from each non-leaf vertex the transition goes to the parent with probability $\lambda/(\lambda + r)$ and to each child with probability $1/(\lambda + r)$. Then consider $Y =$ "X watched only on L", that is the sequence of (not-necessarily distinct) successive leaves visited by X. The group L is distance-transitive (for Hamming distance on L) and Y is a certain isotropic random flight on L. A nice feature of this example is that *without calculation* we can see that Y is recurrent if and only if $\lambda \leq 1$. For consider the path of ancestors of 0. The chain X must spend an infinite time on that path (side-branches are finite); on that path X behaves as asymmetric simple random walk on Z^+, which is recurrent if and only if $\lambda \leq 1$; so X and thence Y visits 0 infinitely often if and only if $\lambda \leq 1$.

Another nice feature is that we can give a fairly explicit expression for the t-step transition probabilities of Y. Writing H for the maximum height reached by X in an excursion from the leaves, then

$$P(H \geq h) = P_1(\hat{T}_h < \hat{T}_0) = \frac{\frac{r}{\lambda} - 1}{(\frac{r}{\lambda})^h - 1}, \quad h \geq 1$$

where $\hat{T}$ denotes hitting time for the height process. Writing M_t for the maximum height reached in t excursions,

$$P(M_t < h) = (P(H < h))^t = \left(1 - \frac{\frac{r}{\lambda} - 1}{(\frac{r}{\lambda})^h - 1}\right)^t.$$

It is clear by symmetry that the distribution of Y_t is conditionally uniform on the leaves which are descendants of the maximal-height vertex previously

visited by X. So for leaves v, x with branchpoint at height d,

$$P_v(Y_t = x) = \sum_{h \geq d} r^{-h} P(M_t = h).$$

Since $P(M_t = h) = P(M_t < h+1) - P(M_t < h)$, we have found the "fairly explicit expression" promised above. A brief calculation gives the following time-asymptotics. Fix $s > 0$ and consider $t \sim s(\frac{r}{\lambda})^j$ with $j \to \infty$; then

$$P_v(Y_t = v) \sim r^{-j} f(s); \quad \text{where}$$

$$f(s) = \sum_{i=-\infty}^{\infty} r^{-i} \left(\exp(-s(\tfrac{r}{\lambda} - 1)(\tfrac{r}{\lambda})^{-i-1}) - \exp(-s(\tfrac{r}{\lambda} - 1)(\tfrac{r}{\lambda})^{-i}) \right).$$

In particular,

$$P_v(Y_t = v) = \Theta(t^{-d/2}) \text{ as } t \to \infty, \ d = \frac{2 \log r}{\log r - \log \lambda}. \tag{13.56}$$

Comparing with (13.26), this gives a sense in which Y mimics simple random walk on Z^d, for d defined above. Note that d increases continuously from 0 to ∞ as λ increases from 0 to r, and that Y is recurrent if and only if $d \leq 2$.

Though we don't go into details, random walk on the hierarchical lattice is a natural infinite-state analog of biased random walk on the balanced finite tree (Chapter 5 section 2.1). In particular, results in the latter context showed that, writing n for number of vertices, $\tau_0(n) = O(n)$ if and only if $\lambda/r > 1/r$, that is if and only if $d > 2$. This is the condition for transience of the infinite-state walk, confirming the paradigm of section 13.2.3.

13.2.10 Towards a classification theory for sequences of finite chains

Three chapters of Woess [339] treat in detail three properties that random walk on an infinite graph may or may not possess:

- transience
- spectral radius < 1
- non-trivial boundary.

Can these be related to properties for sequences of finite chains? We already mentioned (section 13.2.3) that the property $\tau_0(n) = O(n)$ seems to be the analog of transience. In this speculative section we propose definitions of three other properties for sequences of finite chains, which we name

- compactness
- infinite-dimensionality
- expander-like.

Future research will show whether these are useful definitions! Intuitively we expect that every reasonably "natural" sequence should fall into one of these three classes.

For simplicity we consider reversible random walks on Cayley graphs. It is also convenient to continuize. The resulting chains are special cases of (reversible) *Lévy processes*. We define the general Lévy process to be a continuous-time process with stationary independent increments on a (continuous or discrete) group. Thus the setting for the rest of this section is a sequence $(X_t^{(n)})$ of reversible Lé—vy processes on finite groups $G^{(n)}$ of size $n \to \infty$ through some subsequence. Because we work in continuous time, the eigenvalues satisfy $0 = \lambda_1^{(n)} < \lambda_2^{(n)} \leq \cdots$.

(A): Compactness. Say the sequence $(X_t^{(n)})$ is *compact* if there exists a (discrete or continuous) compact set S and a reversible Lévy process $\widetilde{X}_t$ on S such that

(i) $\tilde{d}(t) \equiv ||P_\pi(\widetilde{X}_t \in \cdot) \to \pi|| \to 0$ as $t \to \infty$;

(ii) $\frac{\lambda_j(n)}{\lambda_2(n)} \to \tilde{\lambda}_j$ as $n \to \infty,\ j \geq 2$; where $1 = \tilde{\lambda}_2 \leq \tilde{\lambda}_3 \leq \cdots$ are the eigenvalues of $(\widetilde{X}_t)$;

(iii) $d^{(n)}(t\ \tau_2(n)) \to \tilde{d}(t)$ as $n \to \infty;\ t > 0$.

These properties formalize the idea that the sequence of random walks form discrete approximations to a limit Lévy process on a compact group, at least as far as mixing times are concerned. Simple random walk on Z_m^d, and the limit Brownian motion on R^d (section 13.1.2) form the obvious example. Properties (i) and (iii) imply, in particular, that

$$\tau_1(n)/\tau_2(n) \text{ is bounded as } n \to \infty. \tag{13.57}$$

One might hope that a converse is true:

> Does every sequence satisfying (13.57) have a compact subsequence?

Unfortunately, we are convinced that the answer is "no", for the following reason. Take (X_t^n) which is compact, where the limit Lévy process has function $\tilde{d}(t)$ as at (i). Now consider a product chain $(X_t^{(n)}, Y_t^{(n)})$, where components run independently, and where $Y^{(n)}$ has the cut-off property

(Chapter 7) and $\tau_1^Y(n) \sim \tau_2^X(n)$. Note that by Chapter 7-1 Lemma 1 we have $\tau_2^Y(n) = o(\tau_1^Y(n))$. If the product chain had a subsequential limit, then its total variation function at (i), say $d'(t)$, must satisfy

$$\begin{aligned} d'(t) &= \tilde{d}(t), \quad t > 1 \\ &= 1. \quad t < 1. \end{aligned}$$

But it seems intuitively clear (though we do not know a proof) that every Lévy process on a compact set has continuous $d(\cdot)$. This suggests the following conjecture.

Conjecture 13.10 *For any sequence of reversible Lévy processes satisfying (13.57), there exists a subsequence satisfying the definition of* compact *except that condition (ii) is replaced by*
(iv): $\exists t_0 \geq 0$:

$$\begin{aligned} d^{(n)}(t\ \tau_2(n)) &\to 1; \quad t < t_0 \\ &\to \tilde{d}(t); \quad t > t_0. \end{aligned}$$

Before describing the other two classes of chains, we need a definition and some motivating background. In the present setting, the property "trivial boundary" is equivalent (see Notes) to the property

$$\lim_{t\to\infty} ||P_v(X_t \in \cdot) - P_w(X_t \in \cdot)|| = 0, \quad \forall v, w. \tag{13.58}$$

This suggests that an analogous finite-state property might involve whether the variation distance for nearby starts becomes small before time τ_1. Say that a sequence $(L_n(\varepsilon))$ of subsets is an *asymptotic ε-neighborhood* if

$$||P_\phi(X_{\varepsilon\tau_1} \in \cdot) - P_v(X_{\varepsilon\tau_1} \in \cdot)|| \to 0 \text{ as } n \to \infty$$

uniformly over $v \in L_n(\varepsilon)$; here ϕ is an arbitrary reference vertex. From Chapter 7-1 Lemma 1(b) we can deduce that, if the cut-off property holds, such a neighborhood must have size $|L_n(\varepsilon)| = o(n)$.

(B): Infinite dimensional. Say the sequence $(X_t^{(n)})$ is *infinite-dimensional* if the following three properties hold.

(i) $\tau_1(n) = \Theta(\tau_2 \log\log n)$

(ii) The cut-off property holds

(iii) there exists some $\delta(\varepsilon)$, increasing from 0 to 1 as ε increases from 0 to 1, such that a maximal-size asymptotic ε-neighborhood $(L_n(\varepsilon))$ has

$$\log|L_n(\varepsilon)| = (\log n)^{\delta(\varepsilon)+o(1)} \text{ as } n \to \infty.$$

This definition is an attempt to abstract the essential properties of random walk on the d-cube (Chapter 5 Example 15), where properties (i) and (ii) were already shown. We outline below a proof of property (iii) in that example. Another fundamental example where (i) and (ii) hold is card-shuffling by random transpositions (Chapter 7 Example 18)), and we conjecture that property (iii) also holds there. Conceptually, this class *infinite-dimensional* of sequences is intended (cf. (13.58)) as the analog of a single random walk with trivial boundary on an infinite-dimensional graph.

Property (iii) for the d-cube. Let $(X(t))$ be continuous-time random walk on the d-cube, and $(X^*(t))$ continuous-time random walk on the b-cube, where $b \le d$. The natural coupling shows

if $d(v,w) = b$ then
$$||P_v(X(t) \in \cdot) - P_w(X(t) \in \cdot)|| = ||P_{\mathbf{0}}(X^*(tb/d) \in \cdot) - P_{\mathbf{1}}(X^*(tb/d) \in \cdot)||.$$

Take $d \to \infty$ with
$$b(d) \sim d^\alpha, \quad t(d) \sim \tfrac{1}{4}\varepsilon d \log d$$
for some $0 < \alpha, \varepsilon < 1$, so that
$$\lim_d \frac{t(d)b(d)/d}{\frac{1}{4}b(d)\log b(d)} \to \tfrac{\varepsilon}{\alpha}.$$

Since the variation cut-off for the b-cube is at $\frac{1}{4}b \log b$, we see that for vertices v and w at distance $b(d)$,
$$\begin{aligned} ||P_v(X(t(d)) \in \cdot) - P_w(X(t(d)) \in \cdot)|| &\to 1, \quad \varepsilon > \alpha \\ &\to 0, \quad \varepsilon < \alpha. \end{aligned}$$

So a maximal-size asymptotic ε-neighborhood $(L_n(\varepsilon))$ of $\mathbf{0}$ must be of the form $\{w : d(w,\mathbf{0}) \le d^{\varepsilon + o(1)}\}$. So
$$\log |L_n(\varepsilon)| = \log \binom{d}{d^{\varepsilon+o(1)}} = d^{\varepsilon+o(1)} = (\log n)^{\varepsilon+o(1)}$$
as required.

Finally, we want an analog of a random walk with non-trivial boundary, expressed using property (ii) below.

(C): Expander-like. Say the sequence $(X_t^{(n)})$ is *expander-like* if
(i) $\tau_1 = \Theta(\tau_2 \log n)$

(ii) every asymptotic ε-neighborhood ($L_n(\varepsilon)$) has

$$\log |L_n(\varepsilon)| = (\log n)^{o(1)} \text{ as } n \to \infty$$

(iii) The cut-off property holds.

Recall from Chapter 9 section 1 that, for symmetric graphs which are r-regular expanders for fixed r, we have $\tau_2(n) = \Theta(1)$ and $\tau_1(n) = \Theta(\log n)$. But it is not known whether properties (ii) and (iii) always hold in this setting.

13.3 Random Walks in Random Environments

In talking about random walk on a weighted graph, we have been assuming the graph is fixed. It is conceptually only a minor modification to consider the case where the "environment" (the graph or the edge-weights) is itself first given in some specified random manner. This has been studied in several rather different contexts, and we will give a brief description of known results without going into many details.

Quantities like our mixing time parameters τ from Chapter 4 are now random quantities $\boldsymbol{\tau}$. In general we shall use **boldface** for quantities depending on the realization of the environment but not depending on a realization of the walk.

13.3.1 Mixing times for some random regular graphs

There is a body of work on estimating mixing times for various models of random regular graph. We shall prove two simple results which illustrate two basic techniques, and record some of the history in the Notes.

The first result is Proposition 1.2.1 of Lubotzky [243]. This illustrates the technique of proving expansion (i.e., upper-bounding the Cheeger time constant τ_c) by direct counting arguments in the random graph.

Proposition 13.11 *Let $G_{k,n}$ be the $2k$-regular random graph on vertices $\{1, 2, \ldots, n\}$ with edges $\{(i, \pi_j(i)) : 1 \le i \le n, 1 \le j \le k\}$, where $(\pi_j, 1 \le i \le k)$ are independent uniform random permutations of $\{1, 2, \ldots, n\}$. Write $\boldsymbol{\tau}_c(k, n)$ for the Cheeger time constant for random walk on $G_{k,n}$. Then for fixed $k \ge 7$,*

$$P(\boldsymbol{\tau}_c(k, n) > 2k) \to 0 \text{ as } n \to \infty.$$

Note that a realization of $G_{k,n}$ may be disconnected (in which case $\tau_c = \infty$) and have self-loops and multiple edges.

Outline of proof. Suppose a realization of the graph has the property

$$|A| \le n/2 \Rightarrow |\partial A| \ge |A|/2 \tag{13.59}$$

where $\partial A := \{\text{edges}(i,j) : i \in A, j \in A^c\}$. Then

$$\tau_c = \sup_{A:1\le|A|\le n/2} \frac{k|A|(n-|A|)}{n|\partial A|} \le \sup_{A:1\le|A|\le n/2} \frac{k|A|(n-|A|)}{n|A|/2} \le 2k.$$

So we want to show that (13.59) holds with probability $\to 1$ as $n \to \infty$. If (13.59) fails for some A with $|A| = a$, then there exists B with $|B| = \lfloor \frac{3}{2}a \rfloor = b$ such that

$$\pi_j(A) \subseteq B, \ 1 \le j \le k \tag{13.60}$$

(just take $B = \cup_j \pi_j(A_j)$ plus, if necessary, arbitrary extra vertices). For given A and B, the chance of (13.60) equals $((b)_a/(n)_a)^k$, where $(n)_a := \prod_{r=0}^{a-1}(n-r)$. So the chance that (13.59) fails is at most

$$\sum_{1\le a\le n/2} q(a), \text{ where } q(a) = \binom{n}{a}\binom{n}{b}\left((b)_a/(n)_a\right)^k.$$

So it suffices to verify $\sum_{1\le a\le n/2} q(a) \to 0$. And this is a routine but tedious verification (see Notes). □

Of course the bound on τ_c gives, via Cheeger's inequality, a bound on τ_2, and thence a bound on τ_1 via $\tau_1 = O(\tau_2 \log n)$. But Proposition 13.11 is unsatisfactory in that these bounds get *worse* as k increases, whereas intuitively they should get better. For bounds on τ_1 which improve with k we turn to the second technique, which uses the "$L^1 \le L^2$" inequality to bound the variation threshold time τ_1. Specifically, recall (Chapter 3 Lemma 8b) that for an n-state reversible chain with uniform stationary distribution, the variation distance $d(t)$ satisfies

$$d(t) \le 2 \max_i (np_{ii}(2t) - 1)^{1/2}. \tag{13.61}$$

This is simplest to use for random walk on a group, as illustrated by the following result of Roichman [296].

Proposition 13.12 *Fix $\alpha > 1$. Given a group G, let S be a random set of $k = \lfloor \log^\alpha |G| \rfloor$ distinct elements of G, and consider random walk on the associated Cayley graph with edges $\{(g, gs) : g \in G, s \in S \cup S^{-1}\}$. For any sequence of groups with $|G| \to \infty$,*

$$P(\boldsymbol{\tau}_1 > t_1) \to 0, \text{ where } t_1 = \lceil \tfrac{\alpha}{\alpha-1} \tfrac{\log|G|}{\log k} \rceil.$$

Proof. We first give a construction of the random walk jointly with the random set S. Write $A = \{a, b, \ldots\}$ for a set of k symbols, and write $\bar{A} = \{a, a^{-1}, b, b^{-1}, \ldots\}$. Fix $t \geq 1$ and let $(\xi_s, 1 \leq s \leq t)$ be independent uniform on $\bar{A}$. Choose $(g(a), a \in A)$ by uniform sampling without replacement from G, and set $g(a^{-1}) = (g(a))^{-1}$. Then the process $(X_s; 1 \leq s \leq t)$ constructed via $X_s = g(\xi_1)g(\xi_2)\ldots g(\xi_s)$ is distributed as the random walk on the random Cayley graph, started at the identity ι. So $P(X_t = \iota) = E\mathbf{p}_{\iota\iota}(t)$ where $\mathbf{p}_{\iota\iota}(t)$ is the t-step transition probability in the random environment, and by (13.61) it suffices to take $t = t_1$ (for t_1 defined in the statement of the Proposition) and show

$$|G|P(X_{2t} = \iota) - 1 \to 0. \tag{13.62}$$

To start the argument, let $J(2t)$ be the number of distinct values taken by $(\langle\xi_s\rangle, 1 \leq s \leq 2t)$, where we define $\langle a\rangle = \langle a^{-1}\rangle = a$. Fix $j \leq t$ and $1 \leq s_1 < s_2 < \ldots < s_j \leq 2t$. Then

$$P(J(2t) = j | \langle\xi_{s_i}\rangle \text{ distinct for } 1 \leq i \leq j) = (j/k)^{2t-j} \leq (t/k)^t.$$

By considering the possible choices of (s_i),

$$P(J(2t) = j) \leq \binom{2t}{j}(t/k)^t.$$

Since $\sum_j \binom{2t}{j} = 2^{2t}$ we deduce

$$P(J(2t) \leq t) \leq (4t/k)^t. \tag{13.63}$$

Now consider the construction of X_{2t} given above. We claim

$$P(X_{2t} = \iota | \xi_s, 1 \leq s \leq 2t) \leq \frac{1}{|G| - 2t} \text{ on } \{J(2t) > t\}. \tag{13.64}$$

For if $J(2t) > t$ then there exists some $b \in A$ such that $\langle\xi_s\rangle = b$ for exactly one value of s in $1 \leq s \leq 2t$. So if we condition also on $\{g(a); a \in A, a \neq b\}$, then $X_{2t} = g_1 g(b) g_2$ or $g_1 g(b)^{-1} g_2$ where g_1 and g_2 are determined by the conditioning, and then the conditional probability that $P(X_{2t} = \iota)$ is the conditional probability of $g(b)$ taking a particular value, which is at most $1/(|G| - 2t)$.

Combining (13.64) and (13.63),

$$P(X_{2t} = \iota) \leq (4t/k)^t + \tfrac{1}{|G|-2t} \leq (4t/k)^t + \tfrac{1}{|G|} + O(\tfrac{t}{|G|^2}).$$

So proving (13.62) reduces to proving

$$|G|(4t/k)^t + t/|G| \to 0$$

and the definition of t was made to ensure this.

13.3.2 Randomizing infinite trees

Simple random walk on the infinite regular tree is a fundamental process, already discussed in section 13.2.6. There are several natural ways to randomize the environment; we could take an infinite regular tree and attach random edge-weights; or we could consider a Galton–Watson tree, in which numbers of children are random. Let us start by considering these possibilities simultaneously. Fix a distribution $(\xi; W_1, W_2, \ldots, W_\xi)$ where

$$\xi \geq 1;\ P(\xi \geq 2) > 0;\ W_i > 0, i \leq \xi. \tag{13.65}$$

Note the (W_i) may be dependent. Construct a tree via:

> the root ϕ has ξ children, and the edge (ϕ, i) to the ith child has weight W_i; repeat recursively for each child, taking independent realizations of the distribution (13.65).

So the case $\xi_i \equiv r - 1$ gives the randomly-weighted r-ary tree (precisely, the modification where the root has degree $r - 1$ instead of r), and the case $W_i \equiv 1$ gives a Galton–Watson tree. As in Chapter 3 section 2 to each realization of a weighted graph we associate a random walk with transition probabilities proportional to edge-weights. Since random walk on the unweighted r-ary tree is transient, a natural first issue is prove transience in this "random environment" setting. In terms of the electrical network analogy (see comment below Theorem 13.5), interpreting W as conductance, we want to know whether the (random) resistance $\mathbf{R}$ between ϕ and ∞ is a.s. finite. By considering the children of ϕ, it is clear that the distribution of $\mathbf{R}$ satisfies

$$\mathbf{R} \stackrel{d}{=} \left(\sum_{i=1}^{\xi} (\mathbf{R}_i + W_i^{-1})^{-1} \right)^{-1} \tag{13.66}$$

where the $(\mathbf{R}_i)$ are independent of each other and of $(\xi; W_1, W_2, \ldots, W_\xi)$, and $\mathbf{R}_i \stackrel{d}{=} \mathbf{R}$. But $\hat{\mathbf{R}} \equiv \infty$ is a solution of (13.66), so we need some work to actually prove that $\mathbf{R} < \infty$.

Proposition 13.13 *The resistance* $\mathbf{R}$ *between* ϕ *and* ∞ *satisfies* $\mathbf{R} < \infty$ *a.s..*

Proof. Write $\mathbf{R}^{(k)}$ for the resistance between ϕ and height k (i.e. the height-k vertices, all shorted together). Clearly $\mathbf{R}^{(k)} \uparrow \mathbf{R}$ as $k \to \infty$, and analogously to (13.66)

$$\mathbf{R}^{(k+1)} \stackrel{d}{=} \left(\sum_{i=1}^{\xi} (\mathbf{R}_i^{(k)} + W_i^{-1})^{-1} \right)^{-1}$$

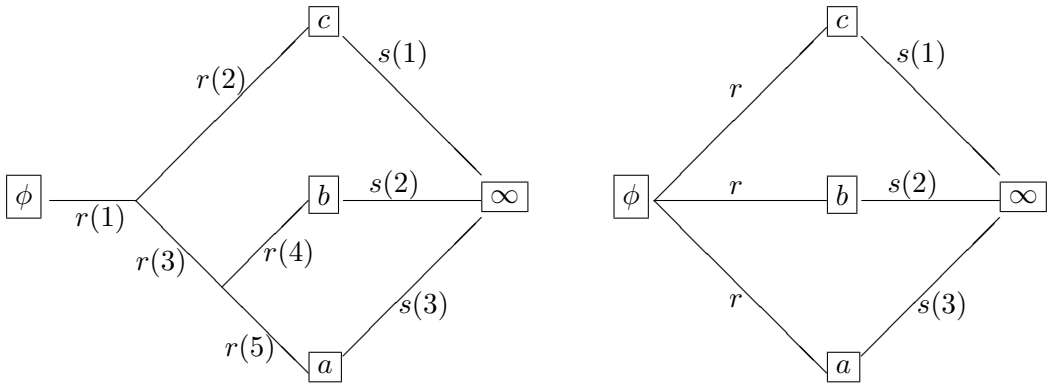
where the $(\mathbf{R}_i^{(k)})$ are independent of each other and of $(\xi; W_1, W_2, \ldots, W_\xi)$, and $\mathbf{R}_i^{(k)} \stackrel{d}{=} \mathbf{R}^{(k)}$.

Consider first the special case $\xi \equiv 3$. Choose x such that $P(W_i^{-1} > x \text{ for some } i) \le 1/16$. Suppose inductively that $P(\mathbf{R}^{(k)} > x) \le 1/4$ (which holds for $k = 0$ since $\mathbf{R}^{(0)} = 0$). Then

$$P(\mathbf{R}_i^{(k)} + W_i^{-1} > 2x \text{ for at least 2 } i\text{'s }) \le \tfrac{1}{16} + 3(\tfrac{1}{4})^2 \le \tfrac{1}{4}.$$

This implies $P(\mathbf{R}^{(k+1)} > x) \le 1/4$, and the induction goes through. Thus $P(\mathbf{R} > x) \le 1/4$. By (13.66) $p := P(\mathbf{R} = \infty)$ satisfies $p = p^3$, so $p = 0$ or 1, and we just eliminated the possibility $p = 1$. So $\mathbf{R} < \infty$ a.s..

Reducing the general case to the special case involves a comparison idea, illustrated by the figure.



Here the edge-weights are *resistances*. In the left network, ϕ is linked to $\{a, b, c\}$ via an arbitrary tree, and in the right network, this tree is replaced by three direct edges, each with resistance $r = 3(r(1) + r(2) + \ldots + r(5))$. We claim that this replacement can only *increase* the resistance between ϕ and ∞. This is a nice illustration of Thompson's principle (Chapter 3 section 7.1) which says that in a realization of either graph, writing $r^*(e)$ for resistance and suming over undirected edges e,

$$R_{\phi\infty} = \inf_{\mathbf{f}} \sum_e r^*(e) f^2(e)$$

where $\mathbf{f} = (f(e))$ is a unit flow from ϕ to ∞. Let $\mathbf{f}$ be the minimizing flow in the right network; use $\mathbf{f}$ to define a flow $\mathbf{g}$ in the left network by specifying that the flow through a (resp. b, c) is the same in the left network and the right network. It is easy to check

$$(\text{left network}) \sum_e r(e) g^2(e) \le (\text{right network}) \sum_e r(e) f^2(e)$$

and hence the resistance $R_{\phi\infty}$ can indeed only be less in the left network.

In the general case, the fact $P(\xi \geq 2) > 0$ implies that the number of individuals in generation g tends to ∞ a.s. as $g \to \infty$. So in particular we can find 3 distinct individuals $\{A, B, C\}$ in some generation G. Retain the edges linking ϕ with these 3 individuals, and cut all other edges within the first G generations. Repeat recursively for descendants of $\{A, B, C\}$. This procedure constructs an infinite subtree, and it suffices to show that the resistance between ϕ and ∞ in the subtree is a.s. finite. By the comparison argument above, we may replace the network linking ϕ to $\{A, B, C\}$ by three direct edges with the same (random) resistance, and similarly for each stage of the construction of the subtree; this gives another tree $\mathcal{T}$, and it suffices to shows its resistance is finite a.s.. But $\mathcal{T}$ fits the special case $\xi \equiv 3$. ∎

It is not difficult (we won't give details) to show that the distribution of $\mathbf{R}$ is the *unique* distribution on $(0, \infty)$ satisfying (13.66). It does seem difficult to say anything explicit about the distribution of $\mathbf{R}$ in Proposition 13.13. One can get a little from comparison arguments. On the binary tree ($\xi \equiv 2$), by using the exact potential function and the exact flows from the unweighted case as "test functions" in the Dirichlet principle and Thompson's principle, one obtains

$$E\mathbf{R} \leq EW^{-1}; \quad E\mathbf{R}^{-1} \leq EW.$$

13.3.3 Bias and speed

Lyons et al [246, 247, 248], summarized in [250] Chapter 10, have studied in detail questions concerning a certain model of biased random walk on deterministic and random infinite trees. Much of their focus is on topics too sophisticated (boundary theory, dimension) to recount here, but let us give one simple result.

Consider the unweighted Galton–Watson tree with offspring distribution $\mu = \text{dist}\ (\xi)$, i.e., the case $W_i \equiv 1$ of (13.65). Fix a parameter $0 \leq \lambda < \infty$. In the *biased random walk* X_t, from a vertex with r children the walker goes to any particular child with probability $1/(\lambda + r)$, and to the parent with probability $\lambda/(\lambda + r)$. It turns out [248] that the biased random walk is recurrent for $\lambda \geq E\xi$ and transient for $\lambda < E\xi$. We will just prove one half of that result.

Proposition 13.14 *The biased random walk is a.s. recurrent for $\lambda \geq E\xi$.*

Proof. We use a "method of fictitious roots". That is, to the root ϕ of the Galton-Watson tree we append an extra edge to a "fictitious" root ϕ^*,

and we consider random walk on this extended tree (rooted at ϕ^*). Write $\mathbf{q}$ for the probability (conditional on the realization of the tree) that the walk started at ϕ never hits ϕ^*. It will suffice to prove $P(\mathbf{q} = 0) = 1$. Fix a realization of the tree, in which ϕ has z children. Then

$$q = \sum_{i=1}^{z} \frac{1}{\lambda + z}(q_i + (1 - q_i)q)$$

where q_i is the probability (on this realization) that the walk started at the i'th child of ϕ never hits ϕ. Rearrange to see $q = (\sum_i q_i)/(\lambda + \sum_i q_i)$. So on the random tree we have

$$\mathbf{q} \stackrel{d}{=} \frac{\sum_{i=1}^{\xi} \mathbf{q}_i}{\lambda + \sum_{i=1}^{\xi} \mathbf{q}_i}$$

where the $(\mathbf{q}_i)$ are independent of each other and ξ, and $\mathbf{q}_i \stackrel{d}{=} \mathbf{q}$. Applying Jensen's inequality to the concave function $x \to \frac{x}{x+\lambda}$ shows

$$E\mathbf{q} \le \frac{(E\xi)(E\mathbf{q})}{\lambda + (E\xi)(E\mathbf{q})}.$$

By considering the relevant quadratic equation, one sees that for $\lambda \ge E\xi$ this inequality has no solution with $E\mathbf{q} > 0$. So $E\mathbf{q} = 0$, as required. ∎

In the transient case, we expect there to exist a non-random *speed* $s(\lambda, \mu) \le 1$ such that

$$t^{-1} d(X_t, \phi) \to s(\lambda, \mu) \text{ a.s. as } t \to \infty. \tag{13.67}$$

Lyons et al [248] show that, when $E\xi < \infty$, (13.67) is indeed true and that $s(\lambda, \mu) > 0$ for all $1 \le \lambda < E\xi$. Moreover in the unbiased ($\lambda = 1$) case there is a simple formula [247]

$$s(1, \mu) = E\frac{\xi - 1}{\xi + 1}.$$

There is apparently no such simple formula for $s(\lambda, \mu)$ in general. See Lyons et al [249] for several open problems in this area.

13.3.4 Finite random trees

Cayley's formula ([313] p. 25) says there are n^{n-2} different trees on $n \ge 2$ labeled vertices $\{1, 2, \ldots, n\}$. Assuming each such tree to be equally likely gives one tractable definition (there are others) of *random n-tree* $\mathbf{T}_n$. One can combine the formulas from Chapter 5 section 3 for random walks on

general trees with known distributional properties of $\mathbf{T}_n$ to get a variety of formulas for random walk on $\mathbf{T}_n$, an idea going back to Moon [264].

As an illustration it is known [264] that the distance $\mathbf{d}(1,2)$ between vertex 1 and vertex 2 in $\mathbf{T}_n$ has distribution

$$P(\mathbf{d}(1,2) = k) = (k+1)n^{-k}(n-2)_{k-1},\ 1 \le k \le n-1$$

where $(m)_s = m(m-1)\cdots(m-s+1)$. Routine calculus gives

$$E\mathbf{d}(1,2) \sim \sqrt{\pi/2}\ n^{1/2}. \tag{13.68}$$

Now on any n-vertex tree, the mean hitting time $t(i,j) = E_i T_j$ satisfies

$$t(i,j) + t(j,i) = 2(n-1)d(i,j) \tag{13.69}$$

(Chapter 5 (84)), and so

$$E\mathbf{t}(1,2) = (n-1)E\mathbf{d}(1,2).$$

Combining with (13.68),

$$E\mathbf{t}(1,2) \sim \sqrt{\pi/2}\ n^{3/2}. \tag{13.70}$$

Instead of deriving more formulas of this type for random walk on $\mathbf{T}_n$, let's jump to the bottom line. It turns out that all the mixing and hitting time parameters $\boldsymbol{\tau}_u^{(n)}$ of Chapter 4, and the analogous "mean cover time" parameters of Chapter 6, are of order $n^{3/2}$ but are *random* to first order: that is,

$$n^{-3/2}\boldsymbol{\tau}_u^{(n)} \ \xrightarrow{d}\ \boldsymbol{\tau}_u^{(\infty)} \text{ as } n \to \infty \tag{13.71}$$

for non-deterministic limits $\boldsymbol{\tau}_u^{(\infty)}$. The fact that all these parameters have the same order is of course reminiscent of the cases of the n-cycle and n-path (Chapter 5 Examples 7 and 8), where all the parameters are $\Theta(n^2)$. And the sophisticated explanation is the same: one can use the weak convergence paradigm (section 13.1.1). In the present context, the random tree $\mathbf{T}_n$ rescales to a limit *continuum random tree* $\mathbf{T}_\infty$, and the random walk converges (with time rescaled by $n^{3/2}$ and space rescaled by $n^{1/2}$) to Brownian motion on $\mathbf{T}_\infty$, and (analogously to section 13.1.1) the rescaled limits of the parameters are just the corresponding parameters for the Brownian motion. See the Notes for further comments.

13.3.5 Randomly-weighted random graphs

Fix a distribution W on $(0, \infty)$ with $EW < \infty$. For each n consider the random graph $G(n, p(n))$, that is the graph on n vertices where each possible edge has chance $p(n)$ to be present. Attach independent random conductances, distributed as W, to the edges. Aspects of this model were studied by Grimmett and Kesten [176]. As they observe, much of the behavior is intuitively rather clear, but technically difficult to prove: we shall just give the intuition.

Case (i): $p(n) = \mu/n$ for fixed $1 < \mu < \infty$. Here the number of edges at vertex 1 is asymptotically Poisson(μ), and the part of the graph within a fixed distance d of vertex 1 is asymptotically like the first d generations in the random family tree $\mathcal{T}^\infty$ of a Galton–Watson branching process with Poisson(μ) offspring distribution, with independent edge-weights attached. This tree essentially fits the setting of Proposition 13.13, except that the number of offspring may be zero and so the tree may be finite, but it is not hard to show (modifying the proof of Proposition 13.13) that the resistance $\mathbf{R}$ in $\mathcal{T}^\infty$ between the root and ∞ satisfies $\{\mathbf{R} < \infty\} = \{\mathcal{T}^\infty \text{ is infinite}\}$ and its distribution is characterized by the analog of (
refRdef). The asymptotic approximation implies that, for $d(n) \to \infty$ slowly, the resistance $\mathbf{R}_{n,d(n)}$ between vertex 1 and the depth-$d(n)$ vertices of $G(n, p(n))$ satisfies $\mathbf{R}_{n,d(n)} \stackrel{d}{\to} \mathbf{R}^{(1)} \stackrel{d}{=} \mathbf{R}$. We claim that the resistance $\mathbf{R}^{(n)}_{1,2}$ between vertices 1 and 2 of $G(n, p(n))$ satisfies

$$\mathbf{R}^{(n)}_{1,2} \stackrel{d}{\to} \mathbf{R}^{(1)} + \mathbf{R}^{(2)}; \text{ where } \mathbf{R}^1 \text{ and } \mathbf{R}^2 \text{ are independent copies of } \mathbf{R} .$$

The lower bound is clear by shorting, but the upper bound requires a complicated construction to connect the two sets of vertices at distances $d(n)$ from vertices 1 and 2 in such a way that the effective resistance of this connecting network tends to zero.

The number of edges of the random graph is asymptotic to $n\mu/2$. So the total edge weight $\sum_i \sum_j W_{ij}$ is asymptotic to $n\mu EW$, and by the commute interpretation of resistance the mean commute time $\mathbf{C}^{(n)}_{1,2}$ for random walk on a realization of the graph satisfies

$$n^{-1}\mathbf{C}^{(n)}_{1,2} \stackrel{d}{\to} \mu EW(\mathbf{R}^{(1)} + \mathbf{R}^{(2)}).$$

Case (ii): $p(n) = o(1) = \Omega(n^{\varepsilon-1})$, some $\varepsilon > 0$. Here the degree of vertex 1 tends to ∞, and it is easy to see that the (random) stationary probability π_1 and the (random) transition probabilities and stationary

distribution the random walk satisfy

$$\max_j \mathbf{p}_{1,j} \stackrel{p}{\to} 0, \; n\boldsymbol{\pi}_1 \stackrel{p}{\to} 1 \text{ as } n \to \infty.$$

So for fixed $k \geq 1$, the k-step transition probabilities satisfy $p_{11}^{(k)} \stackrel{p}{\to} 0$ as $n \to \infty$. This suggests, but it is technically hard to *prove*, that the (random) fundamental matrix $\mathbf{Z}$ satisfies

$$\mathbf{Z}_{11} \stackrel{p}{\to} 1 \text{ as } n \to \infty. \tag{13.72}$$

Granted (13.72), we can apply Lemma 11 of Chapter 2 and deduce that the mean hitting times $\mathbf{t}(\pi, 1) = E_\pi T_1$ on a realization of the random graph satisfies

$$n^{-1}\mathbf{t}(\pi, 1) = \tfrac{\mathbf{Z}_{11}}{n\boldsymbol{\pi}_1} \stackrel{p}{\to} 1, \text{ as } n \to \infty. \tag{13.73}$$

13.3.6 Random environments in d dimensions

The phrase *random walk in random environment* (RWRE) is mostly used to denote variations of the classical "random flight in d dimensions" model. Such variations have been studied extensively in mathematical physics as well as theoretical probability, and the monograph of Hughes [184] provides thorough coverage. To give the flavor of the subject we quote one result, due to Boivin [52].

Theorem 13.15 *Assign random conductances (w_e) to the edges of the two-dimensional lattice $\mathbf{Z}^2$, where*
(i) the process (w_e) is stationary ergodic.
(ii) $c_1 \leq w_e \leq c_2$ a.s., for some constants $0 < c_1 < c_2 < \infty$.
Let $(X_t; t \geq 0)$ be the associated random walk on this weighted graph, $X_0 = 0$. Then $t^{-1/2}X_t \stackrel{d}{\to} Z$ where Z is a certain two-dimensional Normal distribution, and moreover this convergence holds for the conditional distribution of X_t given the environment, for almost all environments.

13.4 Notes on Chapter 13

Section 13.1. Rigorous setup for discrete-time continuous-space Markov chains is given concisely in Durrett [133] section 5.6 and in detail in Meyn and Tweedie [263]. For the more sophisticated continuous-time setting see e.g. Rogers and Williams [295]. Aldous et al [24] prove some of the Chapter 4 mixing time inequalities in the discrete-time continuous-space setting.

The central limit theorem (for sums of functions of a Markov chain) does not automatically extend from the finite-space setting (Chapter 2 Theorem 17) to the continuous-space setting: regularity conditions are required. See [263] Chapter 17. But a remarkable result of Kipnis - Varadhan [217] shows that for *stationary* reversible chains the central limit theorem remains true under very weak hypotheses.

Sections 13.1.1 - 13.1.3. The eigenvalue analysis is classical. The reflection coupling goes back to folklore; see e.g. Lindvall [233] Chapter 6 for applications to multidimensional diffusions and Matthews [259] for Brownian motion in a polyhedron. Burdzy and Kendall [82] give a careful study of coupling for Brownian motion in a triangle. Chen [89] surveys use of coupling to estimate spectral gaps for diffusions on manifolds.

Here is a more concise though less explicit expression for $\bar{d}(t)$ at (13.6) (and hence for $G(t)$at (13.1)). Consider Brownian motions B° on the circle started at 0 and at 1/2. At any time t, the former distribution dominates the latter on the interval $(-1/4, 1/4)$ only, and so

$$\begin{aligned}
\bar{d}(t) &= P_0(B^\circ_t \in (-1/4, 1/4)) - P_{1/2}(B^\circ_t \in (-1/4, 1/4)) \\
&= P_0(B^\circ_t \in (-1/4, 1/4)) - P_0(B^\circ_t \in (1/4, 3/4)) \\
&= 2P_0(B^\circ_t \in (-1/4, 1/4)) \ - 1 \\
&= 2P((t^{1/2}Z) \bmod 1 \in (-1/4, 1/4)) - 1
\end{aligned}$$

where Z has Normal$(0, 1)$ law. We quoted this expression in the analysis of Chapter 5 Example 7

Section 13.1.5. Janvresse [195], Porod [285] and Rosenthal [298] study mixing times for other flights on matrix groups involving rotations and reflections; Porod [284] also discusses more general Lie groups.

Section 13.1.6. The mathematical theory has mostly been developed for classes of nested fractals, of which the Sierpinski gasket is the simplest. See Barlow [40], Lindstrøm [232], Barlow [41] for successively more detailed treatments. Closely related is Brownian motion on the continuum random tree, mentioned in section 13.3.4.

One-dimensional diffusions. The continuous-space analog of a birth-and-death process is a one-dimensional *diffusion* (X_t), described by a stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$$

where B_t is standard Brownian motion and $\mu(\cdot)$ and $\sigma(\cdot)$ are suitably regular specified functions. See Karlin and Taylor [209] for non-technical introduction. Theoretical treatments standardize (via a one-to-one transformation $R \to R$) to the case $\mu(\cdot) = 0$, though for our purposes the standardization to $\sigma(\cdot) = 1$ is perhaps more natural. In this case, if the formula

$$f(x) \propto \exp\left(\int^x 2\mu(y)dy\right)$$

can give a *density* function $f(x)$ then f is the stationary density. Such diffusions relate to two of our topics.

(i) For MCMC, to estimate a density $f(x) \propto \exp(-H(x))$, one can in principle simulate the diffusion with $\sigma(x) = 1$ and $\mu(x) = -H'(x)/2$. This idea was used in Chapter MCMC section 5.

(ii) Techniques for bounding the relaxation time for one-dimensional diffusions parallel techniques for birth-and-death chains [90].

Section 13.2. We again refer to Woess [339] for systematic treatment of random walks on infinite graphs.

Our general theme of using the infinite case to obtain limits for finite chains goes back at least to [8], in the case of Z^d; similar ideas occur in the study of interacting particle systems, relating properties of finite and infinite-site models.

Section 13.2.2. There is a remarkable connection between recurrence of reversible chains and a topic in Bayesian statistics: see Eaton [138]. Properties of random walk on fractal-like infinite subsets of Z^d are studied by Telcs [322, 323].

Section 13.2.9. One view of (Y_t) is as one of several "toy models" for the notion of random walk on fractional-dimensional lattice. Also, when we seek to study complicated variations of random walk, it is often simpler to use the hierarchical lattice than Z^d itself. See for instance the sophisticated study of self-avoiding walks by Brydges et al [76]; it would be interesting to see whether direct combinatorial methods could reproduce their results.

Section 13.2.10. Another class of sequences could be defined as follows. There are certain continuous-time, continuous-space reversible processes on compact spaces which "hit points" and for which $\tau_0 < \infty$; for example

(i) Brownian motion on the circle (section 13.1.1)
(ii) Brownian motion on certain fractals (section 13.1.6)
(iii) Brownian motion on the continuum random tree (section 13.3.4).
So for a sequence of finite-state chains one can define the property

$$\tau_0(n)/\tau_2(n) \text{ is bounded}$$

as the finite analog of "diffusions which hit points". This property holds for the discrete approximations to the examples above: (i) random walk on the n-cycle
(i) random walk on graphs approximating fractals (section 13.1.6)
(iii) random walk on random n-vertex trees (section 13.3.4).

Equivalence (13.58) is hard to find in textbooks. The property "trivial boundary" is equivalent to "no non-constant bounded harmonic functions" ([339] Corollary 24.13), which is equivalent ([328] Theorem 6.5.1) to existence of successful *shift-coupling* of two versions of the chain started at arbitrary points. The property (13.58) is equivalent ([328] Theorem 4.9.4) to existence of successful *couplings*. In the setting of interest to us (continuized chains on countable space), existence of a shift-coupling (a priori weaker than existence of a coupling) for the discrete-time chain implies existence of a coupling for the continuous-time chain, by using independence of jump chain and hold times.

Section 13.3. Grimmett [175] surveys "random graphical networks" from a somewhat different viewpoint, emphasising connections with statistical physics models.

Section 13.3.1. More precise variants of Proposition 3.35 were developed in the 1970s, e.g. [281, 92]. Lubotzky [243], who attributes this method of proof of Proposition 13.11 to Sarnak [305], asserts the result for $k \geq 5$ but our own calculations give only $k \geq 7$. Note that Proposition 13.11 uses the *permutation model* of a $2k$-regular random graph. In the alternative *uniform* model we put $2k$ balls labeled 1, $2k$ balls labeled 2, ... and $2k$ balls labeled n into a box; then draw without replacement two balls at a time, and put an edge between the two vertices. In both models the graphs may be improper (multiple edges or self-loops) and unconnected, but are in fact proper with probability $\Omega(1)$ and connected with probability $1 - o(1)$ as $n \to \infty$ for fixed k. Behavior of τ_c in the uniform model is implicitly studied in Bollobás [54]. The L^2 ideas underlying the proof of Proposition 13.12 were used by Broder and Shamir [68], Friedman [155] and Kahn and Szemerédi [157] in the setting of the permutation model of random r-regular graphs. One result is that $\beta \equiv \max(\lambda_2, -\lambda_n) = O(\frac{2\sqrt{2r-1}}{r})$ with probability

$1-o(1)$. Further results in the "random Cayley graph" spirit of Proposition 13.12 can be found in [28, 130, 271].

Section 13.3.2. The monograph of Lyons and Peres [250] contains many more results concerning random walks on infinite deterministic and Galton–Watson trees. A challenging open problem noted in [249] is to prove that $\mathbf{R}$ has absolutely continuous distribution when ξ is non-constant. The method of fictitious roots used in Proposition 13.14 is also an ingredient in the analysis of cover times on trees [16].

Section 13.3.4. Moon [264] gives further results in the spirit of (13.70), e.g. for variances of hitting times. The fact that random walk on $\mathbf{T}_n$ rescales to Brownian motion on a "continuum random tree" $\mathbf{T}_\infty$ was outlined in Aldous [14] section 5 and proved in Krebs [218]. While this makes the "order $n^{3/2}$" property (13.71) of the parameters essentially obvious, it is still difficult to get explicit information about the limit distributions $\boldsymbol{\tau}^{(\infty)}$. What's known [14] is

(a) $E\boldsymbol{\tau}_0^{(\infty)} = \sqrt{\pi/2}$, as suggested by (13.70);

(b) $\boldsymbol{\tau}^{(\infty)*} = \frac{8}{3}\sqrt{2\pi}$, from (13.69) and the known asymptotics for the diameter of $\mathbf{T}_n$;

(c) The "cover and return" time C_n^+ appearing in Chapter 6 satisfies $n^{-3/2}EC_n^+ \to 6\sqrt{2\pi}$, modulo some technical issues.

Section 13.3.5. Grimmett and Kesten [176] present their results in terms of resistances, without explicitly mentioning random walk, so that results like (13.73) are only implicit in their work.

Chapter 14

Interacting Particles on Finite Graphs (March 10, 1994)

There is a well-established topic "interacting particle systems", treated in the books by Griffeath [172], Liggett [231], and Durrett [132], which studies different models for particles on the infinite lattice Z^d. All these models make sense, but mostly have not been systematically studied, in the context of finite graphs. Some of these models – the voter model, the antivoter model, and the exclusion process – are related (either directly or "via duality") to interacting random walks, and setting down some basic results for these models on finite graphs (sections 14.3 - 14.5) is the main purpose of this chapter. Our focus is on applying results developed earlier in the book. With the important exception of *duality*, we do not use the deeper theory developed in the infinite setting. As usual, whether the deeper theory is applicable to the type of questions we ask in the finite setting is an interesting open question. These models are most naturally presented in continuous time, so our default convention is to work with continuous-time random walk.

We have already encountered results whose natural proofs were "by coupling", and this is a convenient place to discuss couplings in general.

14.1 Coupling

If X and Y are random variables with Binomial (n, p_1) and (n, p_2) distributions respectively, and if $p_1 \le p_2$, then it is intuitively obvious that

$$P(X \ge x) \le P(Y \ge x) \text{ for all } x. \tag{14.1}$$

One could verify this from the exact formulas, but there is a more elegant non-computational proof. For $1 \le i \le n$ define events (A_i, B_i, C_i), independent as i varies, with $P(A_i) = p_1, P(B_i) = p_2 - p_1, P(C_i) = 1 - p_2$. And define

$$\begin{aligned} X' = \sum_i 1_{A_i} &= \text{number of A's which occur} \\ Y' = \sum_i 1_{A_i \cup B_i} &= \text{number of A's and B's which occur.} \end{aligned}$$

Then $X' \le Y'$, so (14.1) holds for X' and Y', but then because $X' \stackrel{d}{=} X$ and $Y' \stackrel{d}{=} Y$ we have proved that (14.1) holds for X and Y. This is the prototype of a *coupling argument*, which (in its wide sense) means

> to prove some *distributional* inequality relating two random processes X, Y by constructing versions X', Y' which satisfy some *sample path* inequality.

Our first "process" example is a somewhat analogous proof of part (a) of the following result, which abstracts slightly a result stated for random walk on distance-regular graphs (Chapter 7 Proposition yyy).

Proposition 14.1 *Let (X_t) be an irreducible continuous-time birth-and-death chain on states $\{0, 1, \ldots, \Delta\}$.*

(a) $\frac{P_0(X_t=i)}{\pi_i}$ is non-increasing in i, for fixed t

(b) $\frac{P_0(X_t=i)}{P_0(X_t=0)}$ is non-decreasing in t, for fixed i

Proof. Fix $i_1 \le i_2$. Suppose we can construct processes Y_t and Z_t, distributed as the given chain started at i_1 and i_2 respectively, such that

$$Y_t \le Z_t \text{ for all } t. \tag{14.2}$$

Then

$$P_{i_1}(X_t = 0) = P(Y_t = 0) \ge P(Z_t = 0) = P_{i_2}(X_t = 0).$$

But by reversibility

$$P_{i_1}(X_t = 0) = \frac{\pi_0}{\pi_{i_1}} \; P_0(X_t = i_1)$$

and similarly for i_2, establishing (a).

Existence of processes satisfying (14.2) is a consequence of the Doeblin coupling discussed below. The proof of part (b) involves a different technique and is deferred to section 14.1.3.

14.1.1 The coupling inequality

Consider a finite-state chain in discrete or continuous time. Fix states i, j. Suppose we construct a joint process $(X_t^{(i)}, X_t^{(j)}; t \geq 0)$ such that

$$\begin{aligned} &(X_t^{(i)}, t \geq 0) \text{ is distributed as the chain started at } i \\ &(X_t^{(j)}, t \geq 0) \text{ is distributed as the chain started at } j. \end{aligned} \tag{14.3}$$

And suppose there is a random time $T \leq \infty$ such that

$$X_t^{(i)} = X_t^{(j)}, \; T \leq t < \infty. \tag{14.4}$$

Call such a T a *coupling time*. Then the *coupling inequality* is

$$||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)|| \leq P(T > t), \; 0 \leq t < \infty. \tag{14.5}$$

The inequality is clear once we observe $P(X_t^{(i)} \in \cdot, T \leq t) = P(X_t^{(j)} \in \cdot, T \leq t)$. The coupling inequality provides a method of bounding the variation distance $\bar{d}(t)$ of Chapter 2 section yyy.

The most common strategy for constructing a coupling satisfying (14.3) is via Markov couplings, as follows. Suppose the underlying chain has state space I and (to take the continuous-time case) transition rate matrix $\mathbf{Q} = (q(i,j))$. Consider a transition rate matrix $\tilde{\mathbf{Q}}$ on the product space $I \times I$. Write the entries of $\tilde{\mathbf{Q}}$ as $\tilde{q}(i,j;k,l)$ instead of the logical-but-fussy $\tilde{q}((i,j),(k,l))$. Suppose that, for each pair (i,j) with $j \neq i$,

$$\tilde{q}(i,j;\cdot,\cdot) \text{ has marginals } q(i,\cdot) \text{ and } q(j,\cdot) \tag{14.6}$$

in other words $\sum_l \tilde{q}(i,j;k,l) = q(i,k)$ and $\sum_k \tilde{q}(i,j;k,l) = q(j,l)$. And suppose that

$$\begin{aligned} \tilde{q}(i,i;k,k) &= q(i,k) \text{ for all } k \\ \tilde{q}(i,i;k,l) &= 0 \text{ for } l \neq k. \end{aligned}$$

Take $(X_t^{(i)}, X_t^{(j)})$ to be the chain on $I \times I$ with transition rate matrix $\tilde{\mathbf{Q}}$ and initial position (i,j), Then (14.3) must hold, and $T \equiv \min\{t : X_t^{(i)} = X_t^{(j)}\}$ is a coupling time. This construction gives a *Markov coupling*, and all the examples where we use the coupling inequality will be of this form. In practice it is much more understandable to define the joint process in words

xxx red and black particles.

A particular choice of $\tilde{\mathbf{Q}}$ is

$$\tilde{q}(i,j;k,l) = q(i,k)q(j,l), \ j \neq i \tag{14.7}$$

in which case the joint process is called to *Doeblin coupling*. In words, the Doeblin coupling consists of starting one particle at i and the other particle at j, and letting the two particles move independently until they meet, at time $M_{i,j}$ say, and thereafter letting them stick together. In the particular case of a birth-and-death process, the particles cannot cross without meeting (in continuous time), and so if $i < j$ then $X_t^{(i)} \leq X_t^{(j)}$ for all t, the property we used at (14.2).

14.1.2 Examples using the coupling inequality

Use of the coupling inequality has nothing to do with reversibility. In fact it finds more use in the irreversible setting, where fewer alternative methods are available for quantifying convergence to stationarity. In the reversible setting, coupling provides a quick way to get bounds which usually (but not always) can be improved by other methods. Here are two examples we have seen before.

Example 14.2 *Random walk on the d-cube (Chapter 5 Example yyy).*

For $\mathbf{i} = (i_1, \ldots, i_d)$ and $\mathbf{j} = (j_1, \ldots, j_d)$ in $I = \{0,1\}^d$, let $D(\mathbf{i},\mathbf{j})$ be the set of coordinates u where $\mathbf{i}$ and $\mathbf{j}$ differ. Write $\mathbf{i}^u$ for the state obtained by changing the i'th coordinate of $\mathbf{i}$. Recall that in continuous time the components move independently as 2-state chains with transition rates $1/d$. In words, the coupling is "run unmatched coordinates independently until they match, and then run them together". Formally, the non-zero transitions of the joint process are

$$\begin{aligned}
\tilde{q}(\mathbf{i},\mathbf{j};\mathbf{i}^u,\mathbf{j}^u) &= 1/d \text{ if } i_u = j_u \\
\tilde{q}(\mathbf{i},\mathbf{j};\mathbf{i}^u,\mathbf{j}) &= 1/d \text{ if } i_u \neq j_u \\
\tilde{q}(\mathbf{i},\mathbf{j};\mathbf{i},\mathbf{j}^u) &= 1/d \text{ if } i_u \neq j_u.
\end{aligned}$$

For each coordinate which is initially unmatched, it takes exponential (rate $2/d$) time until it is matched, and so the coupling time T satisfies

$$T \stackrel{d}{=} \max(\xi_1, \ldots, \xi_{d_0})$$

where the (ξ_u) are independent exponential (rate $2/d$) and $d_0 = d(\mathbf{i}, \mathbf{j})$ is the initial number of unmatched coordinates. So

$$P(T \le t) = (1 - \exp(-2t/d))^{d_0}$$

and the coupling inequality bounds variation distance as

$$\bar{d}(t) \le (1 - \exp(-2t/d))^d.$$

This leads to an upper bound on the variation threshold time

$$\tau_1 \le (\frac{1}{2} + o(1)) d \log d \text{ as } d \to \infty.$$

In this example we saw in Chapter 5 that in fact

$$\tau_1 \sim \frac{1}{4} d \log d \text{ as } d \to \infty$$

so the coupling bound is off by a factor of 2.

Example 14.3 *Random walk on a dense regular graph (Chapter 5 Example yyy).*

Consider a r-regular n-vertex graph. Write $\mathcal{N}(v)$ for the set of neighbors of v. For any pair v, w we can define a $1-1$ map $\theta_{v,w} : \mathcal{N}(v) \to \mathcal{N}(w)$ such that $\theta_{v,w}(x) = x$ for $x \in \mathcal{N}(v) \cap \mathcal{N}(w)$. We can now define a "greedy coupling" by

$$\tilde{q}(v, w; x, \theta_{v,w}(x)) = 1/r, \quad x \in \mathcal{N}(v).$$

In general one cannot get useful bounds on the coupling time T. But consider the dense case, where $r > n/2$. As observed in Chapter 5 Example yyy, here $|\mathcal{N}(v) \cap \mathcal{N}(w)| \ge 2r - n$ and so the coupled processes (X_t, Y_t) have the property that for $w \ne v$

$$P(X_{t+dt} = Y_{t+dt} | X_t = v, Y_t = w) = \frac{|\mathcal{N}(v) \cap \mathcal{N}(w)|}{r} dt \ge \frac{2r - n}{r} dt$$

implying that T satisfies

$$P(T > t) \le \exp(-(2r - n)t/r).$$

So the coupling inequality implies $\bar{d}(t) \le \exp(-(2r-n)t/r)$, and in particular the variation threshold satisfies

$$\tau_1 \le \frac{r}{2r - n}.$$

14.1.3 Comparisons via couplings

We now give two examples of coupling in the wide sense, to compare different processes. The first is a technical result (inequality (14.8) below) which we needed in Chapter 6 yyy. The second is the proof of Proposition 14.1(b).

Example 14.4 *Exit times for constrained random walk.*

Let (X_t) be discrete-time random walk on a graph G, let A be a subset of the vertices of G and let (Y_t) be random walk on the subgraph induced by A. Given $B \subset A$, let S be the first hitting time of (Y_t) on B, and let T be the first hitting time of (X_t) on $B \cup A^c$. Then

$$E_i T \le E_i S, \ i \in A. \tag{14.8}$$

This is "obvious", and the reason it's obvious is by coupling. We can construct coupled processes (X', Y') with the property that, if both particles are at the same position a in A, and if X jumps to another state b in A, then Y jumps to the same state b. This property immediately implies that, for the coupled processes started at the same state in A, we have $T' \le S'$ and hence (14.8).

In words, here is the coupling (X', Y'). When the particles are at different positions they jump independently. When they are at the same position, first let X' jump; if X' jumps to a vertex in A let Y' jump to the same vertex, and otherwise let Y' jump to a uniform random neighbor in A. Formally, the coupled process moves according to the transition matrix $\tilde{\mathbf{P}}$ on $G \times A$ defined by

$$\tilde{p}(x,a;y,b) = p_G(x,y)\ p_A(a,b) \text{ if } x \notin A \text{ or } x \ne a$$
$$\tilde{p}(a,a;b,b) = p_G(a,b), \ b \in A$$
$$\tilde{p}(a,a;y,b) = p_G(a,y) p_A(a,b), \ b \in A, y \in A^c$$

where p_A and p_G refer to transition probabilities for the original random walks on A and G.

Proof of Proposition 14.1(b). Fix $i \ge 1$. By reversibility it is sufficient to prove

$$\frac{P_0(X_t = i)}{P_0(X_t = 0)} \text{ is non-decreasing in } t \ .$$

Consider the Doeblin coupling $(X_t^{(0)}, X_t^{(i)})$ of the processes started at 0 and at i, with coupling time T. Since $X_t^{(0)} \le X_t^{(i)}$ we have

$$P(X_t^{(i)} = 0) = P(X_t^{(0)} = 0, T \le t)$$

and so we have to prove

$$P(T \le t | X_t^{(0)} = 0) \text{ is non-decreasing in } t \ .$$

It suffices to show that, for $t_1 > t$,

$$P(T \le t | X_{t_1}^{(0)} = 0) \ge P(T \le t | X_t^{(0)} = 0)$$

and thus, by considering the conditional distribution of $X_t^{(0)}$ given $X_{t_1}^{(0)} = 0$, it suffices to show that

$$P(T \le t | X_t^{(0)} = j) \ge P(T \le t | X_t^{(0)} = 0) \tag{14.9}$$

for $j \ge 0$. So fix j and t. Write $(X_s^{(0,j)}, 0 \le s \le t)$ for the process conditioned on $X_0 = 0, X_t = j$. By considering time running backwards from t to 0, the processes $X^{(0,0)}$ and $X^{(0,j)}$ are the same non-homogeneous Markov chain started at the different states 0 and j, and we can use the Doeblin coupling in this non-homogeneous setting to construct versions of these processes with

$$X_s^{(0,0)} \le X_s^{(0,j)}, \ 0 \le s \le t.$$

Now introduce an independent copy of the original process, started at time 0 in state i. If this process meets $X^{(0,0)}$ before time t then it must also meet $X^{(0,j)}$ before time t, establishing (14.9).

14.2 Meeting times

Given a Markov chain, the *meeting time* $M_{i,j}$ is the time at which independent copies of the chain started at i and at j first meet. Meeting times arose in the Doeblin coupling and arise in several other contexts later, so deserve a little study. It is natural to try to relate meeting times to properties such as hitting times for a single copy of the chain. One case is rather simple. Consider a distribution dist(ξ) on a group G such that

$$\xi \stackrel{d}{=} \xi^{-1}; \qquad g\xi \stackrel{d}{=} \xi g \text{ for all } g \in G.$$

Now let X_t and Y_t be independent copies of the continuization of random flight on G with step-distribution ξ. Then if we define $Z_t = X_t^{-1} Y_t$, it is easy to check that Z is itself the continuization of the random flight, but run at twice the speed, i.e. with transition rates

$$q_Z(g,h) = 2P(g\xi = h).$$

It follows that $EM_{i,j} = \frac{1}{2}E_iT_j$. The next result shows this equality holds under less symmetry, and (more importantly) that an inequality holds without any symmetry.

Proposition 14.5 *For a continuous-time reversible Markov chain, let T_j be the usual first hitting time and let $M_{i,j}$ be the meeting time of independent copies of the chain started at i and j. Then* $\max_{i,j} EM_{i,j} \le \max_{i,j} E_iT_j$. *If moreover the chain is symmetric (recall the definition from Chapter 7 yyy) then $EM_{i,j} = \frac{1}{2}E_iT_j$.*

Proof. This is really just a special case of the cat and mouse game of Chapter 3 section yyy, where the player is using a random strategy to decide which animal to move. Write X_t and Y_t for the chains started at i and j. Write $f(x,y) = E_xT_y - E_\pi T_y$. Follow the argument in Chapter 3 yyy to verify

$$S_t \equiv (2t + f(X_t, Y_t); 0 \le t \le M_{ij}) \text{ is a martingale.}$$

Then

$$\begin{aligned} E_iT_j - E_\pi T_j &= ES_0 \\ &= ES_{M_{i,j}} \text{ by the optional sampling theorem} \\ &= 2EM_{ij} + Ef(X_{M_{ij}}, Y_{M_{ij}}) \\ &= 2EM_{i,j} - E\bar{t}(X_{M_{i,j}}), \text{ where } \bar{t}(k) = E_\pi T_j. \end{aligned}$$

In the symmetric case we have $\bar{t}(k) = \tau_0$ for all k, establishing the desired equality. In general we have $\bar{t}(k) \le \max_{i,j} E_iT_j$ and the stated inequality follows.

Remarks. Intuitively the bound in Proposition 14.5 should be reasonable for "not too asymmetric" graphs. But on the n-star (Chapter 5 yyy), for example, we have $\max_{i,j} EM_{i,j} = \Theta(1)$ while $\max_{i,j} E_iT_j = \Theta(n)$. The "$\Theta(1)$" in that example comes from concentration of the stationary distribution, and on a regular graph we can use Chapter 3 yyy to obtain

$$\sum_i \sum_j \pi_i \pi_j EM_{i,j} \ge \frac{(n-1)^2}{2n}.$$

But we can construct regular graphs which mimic the n-star in the sense that $\max_{i,j} EM_{i,j} = o(\tau_0)$. A more elaborate result, which gives the correct order of magnitude on the n-star, was given in Aldous [15].

Proposition 14.6 *For a continuous-time reversible chain,*

$$\max_{i,j} EM_{i,j} \le K \left(\sum_i \frac{\pi_i}{\max(E_\pi T_i, \tau_1)} \right)^{-1}$$

for an absolute constant K.

The proof is too lengthy to reproduce, but let us observe as a corollary that we can replace the $\max_{i,j} E_i T_j$ bound in Proposition 14.5 by the *a priori* smaller quantity τ_0, at the expense of some multiplicative constant.

Corollary 14.7 *For a continuous-time reversible chain,*

$$\max_{i,j} EM_{i,j} \le K\tau_0$$

for an absolute constant K.

Proof of Corollary 14.7. First recall from Chapter 4 yyy the inequality

$$\tau_1 \le 66\tau_0. \tag{14.10}$$

"Harmonic mean $\le$ arithmetic mean" gives the first inequality in

$$\begin{aligned} \left(\sum_i \frac{\pi_i}{\max(E_\pi T_i, \tau_1)} \right)^{-1} &\le \sum_i \pi_i \max(E_\pi T_i, \tau_1) \\ &\le \sum_i \pi_i (E_\pi T_i + \tau_1) \\ &\le \tau_0 + \tau_1 \\ &\le 67\tau_0 \text{ by (14.10)} \end{aligned}$$

and so the result is indeed a corollary of Proposition 14.6.

Two interesting open problems remain. First, does Proposition 14.6 always give the right order of magnitude, i.e.

Open Problem 14.8 *In the setting of Proposition 14.6, does there exist an absolute constant K such that*

$$K \max_{i,j} EM_{i,j} \ge \left(\sum_i \frac{\pi_i}{\max(E_\pi T_i, \tau_1)} \right)^{-1}$$

The other open problem is whether some modification of the proof of Proposition 14.5 would give a *small* constant K in Corollary 14.7. To motivate this question, note that the coupling inequality applied to the Doeblin coupling shows that for any chain $\bar{d}(t) \leq \max_{i,j} P(M_{i,j} > t)$. Then Markov's inequality shows that the variation threshold satisfies $\tau_1 \leq e \max_{i,j} EM_{i,j}$. In the reversible setting, Proposition 14.5 now implies $\tau_1 \leq eK\tau_0$ where K is the constant in Corollary 14.7. So a direct proof of Corollary 14.7 with small K would improve the numerical constant in inequality (14.10).

14.3 Coalescing random walks and the voter model

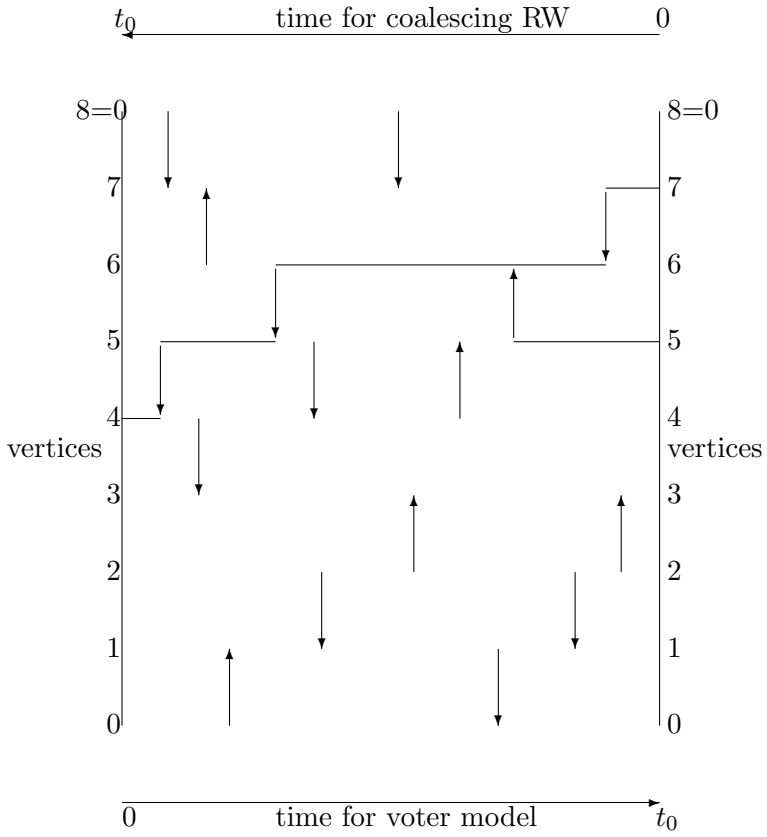
Sections 14.3 and 14.4 treat some models whose behavior relates "by duality" to random-walk-type processes. It is possible (see Notes) to fit all our examples into an abstract duality framework, but for the sake of concreteness I haven't done so. Note that for simplicity we work in the setting of *regular* graphs, though the structural results go over to general graphs and indeed to weighted graphs.

Fix a r-regular n-vertex graph G. In the *voter model* we envisage a person at each vertex. Initially each person has a different opinion (person i has opinion i, say). As time passes, opinions change according to the following rule. For each person i and each time interval $[t, t + dt]$, with chance dt the person chooses uniformly at random a neighbor (j, say) and changes (if necessary) their opinion to the current opinion of person j. Note that the total number of existing opinions can only decrease with time, and at some random time $C_{\mathbf{vm}}$ there will be only one "consensus" opinion.

In the *coalescing random walk* process, at time 0 there is one particle at each vertex. These particles perform independent continuous-time random walks on the graph, but when particles meet they coalesce into clusters and the cluster thereafter sticks together and moves as a single random walk. So at time t there are clusters, composed of one or more particles, at distinct vertices, and during $[t, t+dt]$ each cluster has chance dt to move to a random neighbor and (if that neighbor is occupied by another cluster) to coalesce with that other cluster. Note that the total number of clusters can only decrease with time, and at some random time $C_{\mathbf{crw}}$ the particles will have all coalesced into a single cluster.

Remarkably, the two random variables $C_{\mathbf{vm}}$ and $C_{\mathbf{crw}}$ associated with the two models turn out to have the same distribution, depending only on the graph G. The explanation is that the two processes can be obtained by looking at the same picture in two different ways. Here's the picture. For

each edge e and each direction on e, create a Poisson process of rate $1/r$. In the figure, G is the 8-cycle, "time" is horizontal and an event of the Poisson process for edge (i, j) at time t is indicated by a vertical arrow $i \to j$ at time t.



In the voter model, we interpret time as increasing left-to-right from 0 to t_0, and we interpret an arrow $j \to i$ at time t as meaning that person j adopts i's opinion a time t. In the coalescing random walk model, we interpret time as increasing right-to-left from 0 to t_0, and we interpret an arrow $j \to i$ at time t as meaning that the cluster (if any) at state j at time t jumps to state i, and coalesces with the cluster at i (if any).

So for fixed t_0, we can regard both processes as constructed from the same Poisson process of "arrows". For any vertices i, j, k the event (for the voter model)

The opinions of persons i and j at time t_0 are both the opinion

initially held by k

is exactly the same as the event (for the coalescing random walk process)

The particles starting at i and at j have coalesced before time t_0 and their cluster is at vertex k at time t_0.

The horizontal lines in the figure indicate part of the trajectories. In terms of the coalescing random walks, the particles starting at 5 and 7 coalesce, and the cluster is at 4 at time t_0. In terms of the voter model, the opinion initially held by person 4 is held by persons 5 and 7 at time t_0. The reader may (provided this is not a library book) draw in the remaining trajectories, and will find that exactly 3 of the initial opinions survive, i.e. that the random walks coalesce into 3 clusters.

In particular, the event (for the voter model)

By time t_0 everyone's opinion is the opinion initially held by person k

is exactly the same as the event (for the coalescing random walk process)

All particles have coalesced by time t_0, and the cluster is at k at time t_0.

So $P(C_{\mathbf{vm}} \le t_0) = P(C_{\mathbf{crw}} \le t_0)$, and these two times (which we shall now call just C) do indeed have the same distribution.

We now discuss bounds on EC. It is interesting that the two models give us quite different ways to prove bounds. Bounding EC here is somewhat analogous to the problem of bounding mean cover time, discussed in Chapter 6.

14.3.1 A bound using the voter model

Recall from Chapter 4 yyy the definition of the Cheeger time constant τ_c. In the present setting of a r-regular graph, the definition implies that for any subset A of vertices

$$\text{number of edges linking } A \text{ and } A^c \ \ge \frac{r|A|(n-|A|)}{n\tau_c}. \tag{14.11}$$

Proposition 14.9 *(a) If G is s-edge-connected then $EC \le \frac{rn^2}{4s}$.*

(b) $EC \le 2 \log 2 \ \tau_c n$.

Proof. The proof uses two ideas. The first is a straightforward comparison lemma.

Lemma 14.10 *Let (X_t) be a continuous-time chain on states I. Let $f : I \to \{0, 1, \ldots, n\}$ be such that $f(X_t)$ never jumps by more than 1, and such that there exist strictly positive constants $\gamma, a(1), \ldots, a(n-1)$ such that, for each $1 \le i \le n-1$ and each state x with $f(x) = i$,*

$$\frac{P(f(X_{t+dt}) = i+1 | X_t = x)}{dt} = \frac{P(f(X_{t+dt}) = i-1 | X_t = x)}{dt} \ge \gamma a(i).$$

Then

$$E_x T_{\{f^{-1}(0), f^{-1}(n)\}} \le \gamma^{-1} E^*_{f(x)} T^*_{\{0,n\}}$$

where E^*T^* refers to mean hitting time for the chain X^* on states $\{0, 1, \ldots, n\}$ with transition rates

$$q_{i,i+1} = q_{i,i-1} = a(i).$$

The second idea is that our voter model can be used to define a less-informative "two-party" model. Fix an initial set B of vertices, and group the opinions of the individuals in B into one political party ("Blues") and group the remaining opinions into a second party ("Reds"). Let N^B_t be the number of Blues at time t and let $C^B \le C$ be the first time at which everyone belongs to the same party. Then

$$\begin{aligned} &P(N^B_{t+dt} = N^B_t + 1 | \text{ configuration at time } t) \\ &= P(N^B_{t+dt} = N^B_t - 1 | \text{ configuration at time } t) \\ &= \frac{\text{number of edges linking Blue - Red vertices at time } t}{r}\, dt. \end{aligned} \quad (14.12)$$

Cases (a) and (b) now use Lemma 14.10 with different comparison chains. For (a), while both parties coexist, the number of edges being counted in (14.12) is at least s. To see this, fix two vertices v, x of different parties, and consider (c.f. Chapter 6 yyy) a collection of s edge-disjoint paths from v to x. Each path must contain at least one edge linking Blue to Red. Thus the quantity (14.12) is at least $\frac{s}{r}\, dt$. If that quantity were $\frac{1}{2}\, dt$ then N^B_t would be continuous time random walk on $\{0, \ldots, n\}$ and the quantity EC^B would be the mean time, starting at $|B|$, for simple random walk to hit 0 or n, which by Chapter 5 yyy we know equals $|B|(n - |B|)$. So using Lemma 14.10

$$EC^B \le \frac{r}{2s}|B|(n - |B|) \le \frac{rn^2}{8s}. \quad (14.13)$$

For (b), use (14.11) to see that the quantity (14.12) must be at least $\frac{N_t^B(n-N_t^B)}{n\tau_c}\,dt$. Consider for comparison the chain on $\{0,\ldots,n\}$ with transition rates $q_{i,i+1} = q_{i,i-1} = i(n-i)/n$. For this chain

$$\begin{aligned} E_i^* T^*_{\{0,n\}} &= \sum_{j=1}^{n-1} E_i(\text{ time spent in } j \text{ before } T^*_{\{0,n\}} \\ &= \sum_{j=1}^{n-1} m_i(j) \frac{\frac{1}{2}}{j(n-j)/n} \end{aligned}$$

where $m_i(j)$ is the mean occupation time for simple symmetric random walk and the second term is the speed-up factor for the comparison chain under consideration. Using the formula for $m_i(j)$ from Chapter 5 yyy,

$$E_i^* T^*_{\{0,n\}} = i \sum_{j=i}^{n-1} \frac{1}{j} + (n-i) \sum_{j=1}^{i-1} \frac{1}{n-j} \le n \log 2.$$

So using Lemma 14.10

$$EC^B \le \tau_c n \log 2. \tag{14.14}$$

Finally, imagine choosing $\mathbf{B}$ at random by letting each individual initially be Blue or Red with probability 1/2 each, independently for different vertices. Then by considering some two individuals with different opinions at time t,

$$P(C^{\mathbf{B}} > t) \ge \frac{1}{2} P(C > t).$$

Integrating over t gives $EC \le 2EC^{\mathbf{B}}$. But $EC^{\mathbf{B}} \le \max_B EC^B$, so the Proposition follows from (14.13) and (14.14).

14.3.2 A bound using the coalescing random walk model

The following result bounds the mean coalescing time in terms of mean hitting times of a single random walk.

Proposition 14.11 $EC \le e(\log n + 2) \max_{i,j} E_i T_j$

Proof. We can construct the coalescing random walk process in two steps. Order the vertices arbitrarily as $i_1, \ldots, i_n$. First let the n particles perform independent random walks for ever, with the particles starting at i, j first meeting at time $M_{i,j}$, say. Then when two particles meet, let them cluster and follow the future path of the lower-labeled particle. Similarly, when

two clusters meet, let them cluster and follow the future path of the lowest-labeled particle in the combined cluster. Using this construction, we see

$$C_{\mathbf{crw}} \le \max_j M_{i_1,j}. \tag{14.15}$$

Now let $m^* \equiv \max_{i,j} EM_{i,j}$. Using subexponentiality as in Chapter 2 section yyy,

$$P(M_{i,j} > t) \le \exp(-\lfloor \frac{t}{em^*} \rfloor). \tag{14.16}$$

and so

$$\begin{aligned} EC &= \int_0^\infty P(C > t)dt \\ &\le \int_0^\infty \min(1, \sum_j P(M_{i_1,j} > t))\ dt \text{ by (14.15)} \\ &\le \int_0^\infty \min(1, ne\exp(-\frac{t}{em^*}))\ dt \text{ by (14.16)} \\ &= em^*(2 + \log n) \end{aligned}$$

where the final equality is the calculus fact

$$\int_0^\infty \min(1, Ae^{-at})\ dt = a^{-1}(1 + \log A), \quad A \ge 1.$$

The result now follows from Proposition 14.5.

14.3.3 Conjectures and examples

The complete graph. On the complete graph, the number K_t of clusters at time t in the coalescing random walk model is itself the continuous-time chain with transition rates

$$q_{k,k-1} = k(k-1)/(n-1); \quad n \ge k \ge 2.$$

Since $C_{\mathbf{crw}}$ is the time taken for K_t to reach state 1,

$$EC = \sum_{k=2}^{m} \frac{n-1}{k(k-1)} = \frac{(n-1)^2}{n} \sim n.$$

Recall from Chapter 7 yyy that in a vertex-transitive graph with τ_2/τ_0 small, the first hitting time to a typical vertex has approximately exponential distribution with mean τ_0. Similarly, the meeting time $M_{i,j}$ for typical

i, j has approximately exponential distribution with mean $\tau_0/2$. It seems intuitively clear that, for fixed small k, when the number of clusters first reaches k these clusters should be approximately uniformly distributed, so that the mean further time until one of the $k(k-1)/2$ pairs coalesce should be about $\frac{\tau_0}{k(k-1)}$. Repeating the analysis of the complete graph suggests

Open Problem 14.12 *Prove that for a sequence of vertex-transitive graphs with $\tau_2/\tau_0 \to 0$, we have $EC \sim \tau_0$.*

In the general setting, there is good reason to believe that the *log* term in Proposition 14.11 can be removed.

Open Problem 14.13 *Prove there exists an absolute constant K such that on any graph*

$$EC \le K \max_{v,w} E_v T_w.$$

The assertion of Open Problem 14.12 in the case of the torus Z_m^d for $d \ge 2$ was proved by Cox [103]. A detailed outline is given in [132] Chapter 10b, so we will not repeat it here, but see the remark in section 14.3.5 below.

xxx discuss $d = 1$?

14.3.4 Voter model with new opinions

For a simple variation of the voter model, fix a parameter $0 < \lambda < \infty$ and suppose that each individual independently decides at rate λ (i.e. with chance λdt in each time interval $[t, t + dt]$) to adopt a new opinion, not previously held by anyone. For this process we may take as state space the set of partitions $\mathbf{A} = \{A_1, A_2, \ldots\}$ of the vertex-set of the underlying graph G, where two individuals have the same opinion iff they are in the same component A of $\mathbf{A}$. The duality relationship holds with the following modification. In the dual process of coalescing random walks, each cluster "dies" at rate λ. Thus in the dual process run forever, each "death" of a cluster involves particles started at some set A of vertices, and this partition $\mathbf{A} = \{A_i\}$ of vertices into components is (by duality) distributed as the stationary distribution of the voter model with new opinions. This is the *unique* stationary distribution, even though (e.g. on the n-cycle) the Markov chain may not be irreducible because of the existence of transient states.

The time to approach stationarity in this model is controlled by the time $\tilde{C}$ for the dual process to die out completely. Clearly $E\tilde{C} \le EC + 1/\lambda$, where C is the coalescing time discussed in previous sections, and we do not have anything new to say beyond what is implied by previous results. Instead,

we study properties of the stationary distribution $\mathbf{A} = \{A_i\}$. A natural parameter is the chance, γ say, that two random individuals have the same opinion, i.e.

$$\gamma \equiv E \sum_i \frac{|A_i|^2}{n^2}. \tag{14.17}$$

Lemma 14.14

$$\gamma = \frac{2E\mathcal{E}}{\lambda r n^2} + \frac{1}{n},$$

where $\mathcal{E}$ is the number of edges with endpoints in different components, under the stationary distribution.

Proof. Run the stationary process, and let $\mathbf{A}(t)$ and $\mathcal{E}(t)$ be the partition and the number of edges linking distinct components, at time t, and let $S(t) = \sum_i |A_i(t)|^2$. Then

$$\frac{E(S^2(t+dt) - S^2(t)| \text{ configuration at time } t)}{dt}$$

$$= \frac{4}{r}\mathcal{E}(t) + 2\lambda \sum_i |A_i(t)|(1 - |A_i(t)|). \tag{14.18}$$

The first term arises from the "voter" dynamics. If an opinion change involves an edge linking components of sizes a and b, then the change in S^2 has expectation

$$\frac{(a+1)^2 + (a-1)^2 + (b+1)^2 - (b-1)^2}{2} - (a^2 + b^2) = 2$$

and for each of the $\mathcal{E}(t)$ edges linking distinct components, opinion changes occur at rate $2/r$. The second term arises from new opinions. A new opinion occurs in a component of size a at rate λa, and the resulting change in S^2 is

$$(a-1)^2 + 1^2 - a^2 = 2(1-a).$$

Stationarity implies that the expectation of (14.18) equals zero, and so

$$\frac{4}{r}E\mathcal{E} = 2\lambda \sum_i E|A_i|(|A_i| - 1) = 2\lambda(n^2\gamma - n)$$

and the lemma follows.

Corollary 14.15 $\frac{1+\lambda\tau_c}{1+\lambda\tau_c n} \le \gamma \le \frac{\lambda+1}{\lambda n}$.

Proof. Clearly $E\mathcal{E}$ is at most the total number of edges, $nr/2$, so the upper bound follows from the lemma. For the lower bound, (14.11) implies

$$\xi \geq \frac{r \sum_i |A_i|(n - |A_i|)}{2n\tau_c}$$

and hence

$$E\mathcal{E} \geq \frac{r}{2n\tau_c}(n^2 - n^2\gamma)$$

and the bound follows from the lemma after brief manipulation.

We now consider bounds on γ obtainable by working with the dual process. Consider the meeting time M of two independent random walks started with the stationary distribution. Then by duality (xxx explain)

$$\gamma = P(M < \xi_{(2\lambda)})$$

where $\xi_{(2\lambda)}$ denotes a random variable with exponential (2λ) distribution independent of the random walks. Now M is the hitting time of the stationary "product chain" (i.e. two independent continuous-time random walks) on the diagonal $A = \{(v, v)\}$, so by Chapter 3 yyy M has completely monotone distribution, and we shall use properties of complete monotonicity to get

Corollary 14.16

$$\frac{1}{1 + 2\lambda EM} \leq \gamma \leq \frac{1}{1 + 2\lambda EM} + \frac{\tau_2}{EM}.$$

Proof. We can write $M \stackrel{d}{=} R\xi_{(1)}$, where $\xi_{(1)}$ has exponential(1) distribution and R is independent of $\xi_{(1)}$. Then

$$\begin{aligned}
\gamma &= P(R\xi_{(1)} < \xi_{(2\lambda)} \\
&= E\ P(R\xi_{(1)} < \xi_{(2\lambda)}|R) \\
&= E\frac{1}{1 + 2\lambda R} \\
&\geq \frac{1}{1 + 2\lambda ER} \text{ by Jensen's inequality} \\
&= \frac{1}{1 + 2\lambda EM}.
\end{aligned}$$

For the upper bound, apply Chapter 3 yyy to the product chain to obtain

$$P(M > t) \geq \exp(-t/EM) - \tau_2/EM$$

(recall that τ_2 is the same for the product chain as for the underlying random walk). So

$$\begin{aligned} 1-\gamma &= P(M \geq \xi_{(2\lambda)}) \\ &= \int_0^\infty P(M \geq t)\ 2\lambda e^{-2\lambda t} dt \\ &\geq \frac{2\lambda EM}{1 = 2\lambda EM} - \frac{\tau_2}{EM} \end{aligned}$$

and the upper bound follows after rearrangement.

Note that on a vertex-transitive graph Proposition 14.5 implies $EM = \tau_0/2$. So on a sequence of vertex-transitive graphs with $\tau_2/\tau_0 \to 0$ and with $\lambda\tau_0 \to \theta$, say, Corollary 14.16 implies $\gamma \to \frac{1}{1+\theta}$. But in this setting we can say much more, as the next section will show.

14.3.5 Large component sizes in the voter model with new opinions

xxx discuss coalescent, GEM and population genetics.

xxx genetics already implicit in xxx

Fix $0 < \theta < \infty$. take independent random variables (ξ_i) with distribution

$$P(\xi > x) = (1-x)^\theta, \quad 0 < x < 1$$

and define

$$(X_1^{(\theta)}, X_2^{(\theta)}, X_3^{(\theta)}, \ldots) = (\xi_1, (1-\xi_1)\xi_2, (1-\xi_1)(1-\xi_2)\xi_3, \ldots)$$

so that $\sum_i X_i^{(\theta)} = 1$.

Proposition 14.17 *Consider a sequence of vertex-transitive graphs for which $\tau_2/\tau_0 \to 0$. Consider the stationary distribution* $\mathbf{A}$ *of the voter model with new opinions, presented in size-biased order. If $\lambda\tau_0 \to \theta$ then*

$$\left(\frac{|A_1|}{n}, \ldots, \frac{|A_k|}{n}\right) \ \stackrel{d}{\to}\ (X_1^{(\theta)}, \ldots, X_k^{(\theta)}) \text{ for all fixed } k.$$

xxx proof

Remark. The same argument goes halfway to proving Open Problem 14.12, by showing

Corollary 14.18 *Consider a sequence of vertex-transitive graphs for which $\tau_2/\tau_0 \to 0$. Let $C^{(k)}$ be the coalescing time for k walks started at independent uniform positions. Then, for fixed k, $EC^{(k)} \sim \tau_0(1-k^{-1})$.*

xxx argument similar (?) to part of the proof in Cox [103] for the torus.

14.3.6 Number of components in the voter model with new opinions

xxx τ_c result

14.4 The antivoter model

Recall from section 14.3 the definition of the voter model on a r-regular n-vertex graph. We now change this in two ways. First, we suppose there are only two different opinions, which it is convenient to call ± 1. Second, the evolution rule is

> For each person i and each time interval $[t, t+dt]$, with chance dt the person chooses uniformly at random a neighbor (j, say) and changes (if necessary) their opinion to the opposite of the opinion of person j.

The essential difference from the voter model is that opinions don't disappear. Writing $\eta_v(t)$ for the opinion of individual v at time t, the process $\eta(t) = (\eta_v(t), v \in G)$ is a continuous-time Markov chain on state-space $\{-1, 1\}^G$. So, provided this chain is irreducible, there is a unique stationary distribution $(\eta_v, v \in G)$ for the antivoter model.

This model on infinite lattices was studied in the "interacting particle systems" literature [172, 231], and again the key idea is duality. In this model the dual process consists of *annihilating* random walks. We will not go into details about the duality relation, beyond the following definition we need later. For vertices v, w, consider independent continuous-time random walks started at v and at w. We have previously studied $M_{v,w}$, the time at which the two walks first meet, but now we define $N_{v,w}$ to be the total number of jumps made by the two walks, up to and including the time $M_{v,w}$. Set $N_{v,v} = 0$.

Donnelly and Welsh [129] considered our setting of a finite graph, and showed that Proposition 14.19 is a simple consequence of the duality relation.

Proposition 14.19 *The antivoter process has a unique stationary distribution (η_v), which satisfies*
(i) $E\eta_v = 0$
(ii) $c(v,w) \equiv E\eta_v\eta_w = P(N_{v,w} \text{ is even }) - P(N_{v,w} \text{ is odd })$.
If G is neither bipartite nor the n-cycle, then the set of all $2^n - 2$ non-unanimous configurations is irreducible, and the support of the stationary distribution is that set.

In particular, defining

$$S \equiv \sum_v \eta_v$$

so that S or $-S$ is the "margin of victory" in an election, we have $ES = 0$ and

$$\text{var } S = \sum_v \sum_w c(v,w). \tag{14.19}$$

On a bipartite graph with bipartition (A, A^c) the stationary distribution is

$$P(\eta_v = 1 \forall v \in A, \eta_v = -1 \forall v \in A^c) = P(\eta_v = -1 \forall v \in A, \eta_v = 1 \forall v \in A^c) = 1/2$$

and $c(v,w) = -1$ for each edge. Otherwise $c(v,w) > -1$ for every edge.

The antivoter process is in general a non-reversible Markov chain, because it can transition from a configuration in which v has the same opinion as all its neighbors to the configuration where v has the opposite opinion, but the reverse transition is impossible. Nevertheless we could use duality to discuss convergence time. But, following [129], the spatial structure of the stationary distribution is a more novel and hence more interesting question. Intuitively we expect neighboring vertices to be negatively correlated and the variance of S to be smaller than n (the variance if opinions were independent). In the case of the complete graph on n vertices, $N_{v,w}$ has (for $w \neq v$) the geometric distribution

$$P(N_{v,w} > m) = \left(1 - \frac{1}{n-1}\right)^m ; m \geq 0$$

from which we calculate $c(v,w) = -1/(2n-3)$ and var $S = \frac{n(n-2)}{2n-3} < n/2$. We next investigate var S in general.

14.4.1 Variances in the antivoter model

Write $\xi = (\xi_v)$ for a configuration of the antivoter process and write

$$S(\xi) = \sum_v \xi_v$$

$$a(\xi) = \text{ number of edges } (v,w) \text{ with } \xi_v = \xi_w = 1$$

$$b(\xi) = \text{ number of edges } (v,w) \text{ with } \xi_v = \xi_w = -1.$$

A simple counting argument gives

$$2(a(\xi) - b(\xi)) = rS(\xi). \tag{14.20}$$

Lemma 14.20 var $S = \frac{2}{r}E(a(\eta) + b(\eta))$, *where* η *is the stationary distribution.*

Proof. Writing (η_t) for the stationary process and $dS_t = S(\eta_{t+dt}) - S(\eta_t)$, we have

$$P(dS_t = +2|\eta_t) = b(\eta_t)dt$$
$$P(dS_t = -2|\eta_t) = a(\eta_t)dt$$

and so

$$\begin{aligned} 0 &= ES^2(\eta_{t+dt}) - ES^2(\eta_t) \text{ by stationarity} \\ &= 2ES(\eta_t)dS_t \ + E(dS_t)^2 \\ &= 4ES(\eta_t)(b(\eta_t) - a(\eta_t))dt \ + 4E(a(\eta_t) + b(\eta_t))dt \\ &= -2rES^2(\eta_t)dt + 4E(a(\eta_t) + b(\eta_t))dt \text{ by (14.20)} \end{aligned}$$

establishing the Lemma.

Since the total number of edges is $nr/2$, Lemma 14.20 gives the upper bound which follows, and the lower bound is also clear.

Corollary 14.21 *Let* $\kappa = \kappa(G)$ *be the largest integer such that, for any subset* A *of vertices, the number of edges with both ends in* A *or both ends in* A^c *is at least* κ*. Then*

$$\frac{2\kappa}{r} \leq \text{var } S \leq n.$$

Here κ is a natural measure of "non-bipartiteness" of G. We now show how to improve the upper bound by exploiting duality. One might expect some better upper bound for "almost-bipartite" graphs, but Examples 14.27 and 14.28 indicate this may be difficult.

Proposition 14.22 var $S < n/2$.

Proof. Take two independent stationary continuous-time random walks on the underlying graph G, and let $(X_t^{(1)}, X_t^{(2)}; t = \ldots, -1, 0, 1, 2, \ldots)$ be the jump chain, i.e. at each time we choose at random one component to make a step of the random walk on the graph. Say an "event' happens at t if $X_t^{(1)} = X_t^{(2)}$, and consider the inter-event time distribution L:

$$P(L = l) = P(\min\{t > 0 : X_t^{(1)} = X_t^{(2)}\} = l | X_0^{(1)} = X_0^{(2)}).$$

In the special case where G is vertex-transitive the events form a renewal process, but we use only stationarity properties (c.f. Chapter 2 yyy) which hold in the general case. Write

$$T = \min\{t \geq 0 : X_t^{(1)} = X_t^{(2)}\}$$

where the stationary chain is used. Then

$$P_{v,w}(T = t) \equiv P(T = t | X_0^{(1)} = v, X_0^{(2)} = w) = P(N_{v,w} = t)$$

and so by (14.19) and Proposition 14.19(ii),

$$\begin{aligned} \text{var } S &= \sum_v \sum_w (P_{v,w}(T \text{ is even}) - P_{v,w}(T \text{ is odd})) \\ &= n^2(P(T \text{ is even}) - P(T \text{ is odd})). \end{aligned}$$

If successive events occur at times t_0 and t_1, then

$$\begin{aligned} |\{s : t_0 < s \leq t_1 : t_1 - s \text{ is even }\}| - \{s : t_0 < s \leq t_1 : t_1 - s \text{ is odd}\}| &= 0 \text{ if } |t_1 - t_0| \text{ is even} \\ &= 1 \text{ if } |t_1 - t_0| \text{ is odd} \end{aligned}$$

and an ergodic argument gives

$$P(T \text{ is even}) - P(T \text{ is odd}) = P(L \text{ is odd})/EL.$$

But $EL = 1/P(\text{event}) = n$, so we have established

Lemma 14.23 $n^{-1}\text{var } S = P(L \textit{ is odd})$.

Now consider

$$T^- = \min\{t \geq 0 : X_{-t}^{(1)} = X_{-t}^{(2)}\}.$$

If successive events occur at t_0 and t_1, then there are $t_1 - t_0 - 1$ times s with $t_0 < s < t_1$, and another ergodic argument shows

$$P(T + T^- = l) = \frac{(l-1)P(L = l)}{EL}, \ l \geq 2.$$

So

$$\begin{aligned} n^{-1}(P(L \text{ is even}) - P(L \text{ is odd})) &= \frac{1}{EL} \sum_{l \geq 2} (-1)^l P(L = l) \text{ since } EL = n \\ &= \sum_{l \geq 2} \frac{(-1)^l}{l-1} P(T + T^- = l). \qquad (14.21) \end{aligned}$$

Now let $\phi(z)$ be the generating function of a distribution on $\{1, 2, 3, \ldots\}$ and let Z, Z^- be independent random variables with that distribution. Then

$$\sum_{l \geq 2} \frac{(-1)^l}{l-1} P(Z + Z^- = l) = \int_{-1}^{0} \frac{\phi^2(z)}{z^2}\, dz > 0. \tag{14.22}$$

Conditional on $(X_0^{(1)}, X_0^{(2)}) = (v, w)$ with $w \neq v$, we have that T and T^- are independent and identically distributed. So the sum in (14.21) is positive, implying $P(L \text{ is odd}) < 1/2$, so the Proposition follows from the Lemma.

Implicit in the proof are a corollary and an open problem. The open problem is to show that var S is in fact maximized on the complete graph. This might perhaps be provable by sharpening the inequality in (14.22).

Corollary 14.24 *On an edge-transitive graph, write* $c_{edge} = c(v, w) = E\eta_v\eta_w$ *for an arbitrary edge* (v, w). *Then*

$$\text{var } S = n(1 + c_{edge})/2$$

$$c_{edge} < 0.$$

Proof. In an edge-transitive graph, conditioning on the first jump from (v, v) gives

$$P(L \text{ is odd}) = P(N_{v,w} \text{ is even})$$

for an edge (v, w). But $P(N_{v,w} \text{ is even }) = (1 + c_{\text{edge}})/2$ by Proposition 14.19(ii), so the result follows from Lemma 14.23 and Proposition 14.22.

14.4.2 Examples and Open Problems

In the case of the complete graph, the number of $+1$ opinions evolves as the birth-and-death chain on states $\{1, 2, \ldots, n-1\}$ with transition rates

$$\begin{aligned} i \to i+1 \quad & \text{rate } \tfrac{(n-i)(n-1-i)}{n(n-1)} \\ \to i-1 \quad & \text{rate } \tfrac{i(i-1)}{n(n-1)} \end{aligned}$$

From the explicit form of the stationary distribution we can deduce that as $n \to \infty$ the asymptotic distribution of S is Normal. As an exercise in technique (see Notes) we ask

Open Problem 14.25 *Find sufficient conditions on a sequence of graphs which imply* S *has asymptotic Normal distribution.*

Example 14.26 *Distance-regular graphs.*

On a distance-regular graph of diameter Δ, define $1 = f(0), f(1), \ldots, f(\Delta)$ by

$$f(i) = c(v,w) = P(N_{v,w} \text{ is even }) - P(N_{v,w} \text{ is odd }), \text{ where } d(x,y) = i.$$

Conditioning on the first step of the random walks,

$$f(i) = -\left(p_{i,i+1} f(i+1) + p_{i,i} f(i) + p_{i,i-1} f(i-1)\right), \ 1 \le i \le \Delta \qquad (14.23)$$

where (c.f. Chapter 7 yyy) the $p_{i,j}$ are the transition probabilities for the birth-and-death chain associated with the discrete-time random walk. In principle we can solve these equations to determine $f(1) = c_{\text{edge}}$. Note that the bipartite case is the case where $p_{i,i} \equiv 0$, which is the case where $f(i) \equiv (-1)^i$ and $c_{\text{edge}} = -1$. A simple example of a non-bipartite distance-regular graph is the "2-subsets of a d-set" example (Chapter 7 yyy) for $d \ge 4$. Here $\Delta = 2$ and

$$p_{1,0} = \frac{1}{2(d-2)} \quad p_{1,1} = \tfrac{d-3}{2(d-2)} \quad p_{1,2} = \frac{d-2}{2(d-2)}$$
$$p_{2,1} = \tfrac{4}{2(d-2)} \quad p_{2,2} = \frac{2d-8}{2(d-2)}.$$

Solving equations (14.23) gives $c_{\text{edge}} = -1/(3d-7)$.

Corollary 14.24 said that in an edge-transitive graph, $c_{\text{edge}} < 0$. On a vertex-transitive graph this need not be true for every edge, as the next example shows.

Example 14.27 *An almost bipartite vertex-transitive graph.*

Consider the $m+2$-regular graph on $2m$ vertices, made by taking m-cycles $(v_1, \ldots, v_m)$ and $(w_1, \ldots, w_m)$ and adding all edges (v_i, w_j) between the two "classes". One might guess that, under the stationary distribution, almost all individuals in a class would have the same opinion, different for the two classes. But in fact the tendency for agreement between individuals in the same class is bounded: as $m \to \infty$

$$\begin{aligned} c(v_i, w_j) &\to -\frac{1}{9} \\ c(v_i, v_j) &\to \frac{1}{9}, \ j \ne i. \end{aligned} \qquad (14.24)$$

To prove this, consider two independent continuous-time random walks, started from opposite classes. Let N be the number of jumps before meeting and let $M \geq 1$ be the number of jumps before they are again in opposite classes. Then

$$P(M \text{ is odd }) = \frac{4}{m} + O(m^{-2}); \quad P(N < M) = \frac{1}{m} + O(m^{-2}).$$

So writing $M_1 = M, M_2, M_3, \ldots$ for the cumulative numbers of jumps each time the two walks are in opposite components, and writing

$$Q \equiv \min\{j : M_j \text{ is odd}\},$$

we have

$$P(\text{ walks meet before } Q) = \frac{1}{5} + O(m^{-1}).$$

Writing $Q_1 = Q, Q_2, Q_3, \ldots$ for the sucessive j's at which M_j changes parity, and

$$L \equiv \max\{k : M_{Q_k} < \text{ meeting time}\}$$

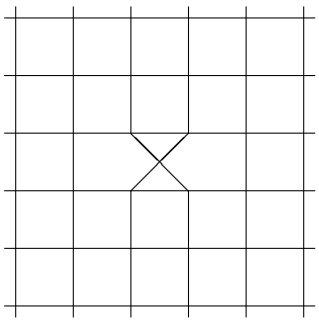
for the number of parity changes before meeting,

$$P(L = l) = \frac{1}{5}\left(\frac{4}{5}\right)^l + O(m^{-1}), \quad l \geq 0$$

So $P(\eta_{v_i,w_j} \text{ is odd}) = P(L \text{ is even}) \to \frac{5}{9}$ and (14.24) follows easily.

Example 14.28 *Another almost-bipartite graph.*

Consider the torus Z_m^d with $d \geq 2$ and with even $m \geq 4$, and make the graph non-bipartite by moving two edges as shown.



Let $m \to \infty$ and consider the covariance $c(v_m, w_m)$ across edges (v_m, w_m) whose distance from the modified edges tends to infinity. One might suspect

that the modification had only "local" effect, in that $c(v_m, w_m) \to -1$. In fact,

$$\begin{aligned} c(v_m, w_m) &\to -1, \quad d = 2 \\ &\to \beta(d) > -1, \ d \geq 3. \end{aligned}$$

We don't give details, but the key observation is that in $d \geq 3$ there is a bounded-below chance that independent random walks started from v_m and w_m will traverse one of the modified edges before meeting.

14.5 The interchange process

xxx notation: $\tilde{X}$ for process or underlying RW?

Fix a graph on n vertices. Given n distinguishable particles, there are $n!$ "configurations" with one particle at each vertex. The *interchange process* is the following continuous-time reversible Markov chain on configurations.

> On each edge there is a Poisson, rate 1, process of "switch times", at which times the particles at the two ends of the edge are interchanged.

The stationary distribution is uniform on the $n!$ configurations. We want to study the time taken to approach the uniform distribution, as measured by the parameters τ_2 and τ_1.

As with the voter model, there is an induced process obtained by declaring some subset of particles to be "visible", regarding the visible particles as indistinguishable, and ignoring the invisible particles. Interchanging two visible particles has no effect, so the dynamics of the induced process are as follows.

> On each edge there is a Poisson, rate 1, process of "switch times". At a switch time, if one endpoint is unoccupied and the other endpoint is occupied by a (visible) particle, then the particle moves to the other endpoint.

This is the finite analog of the *exclusion process* studied in the interacting particle systems literature. But in the finite setting, the interchange process seems more fundamental.

If we follow an individual particle, we see a certain continuous-time Markov chain $\tilde{X}_t$, say, with transition rate 1 along each edge. In the terminology of Chapter 3 yyy this is the *fluid model* random walk, rather than the usual continuized random walk. Write $\tilde{\tau}_2$ for the relaxation time of $\tilde{X}$. The contraction principle (Chapter 4 yyy) implies $\tau_2 \geq \tilde{\tau}_2$.

Open Problem 14.29 *Does $\tau_2 = \tilde{\tau}_2$ in general?*

If the answer is "yes", then the general bound of Chapter 4 yyy will give

$$\tau_1 \le \tilde{\tau}_2(1 + \frac{1}{2}\log n!) = O(\tilde{\tau}_2 n \log n)$$

but the following bound is typically better.

Proposition 14.30 $\tau_1 \le (2 + \log n)e \ \max_{v,w} \tilde{E}_v \tilde{T}_w$.

Proof. We use a coupling argument. Start two versions of the interchange process in arbitrary initial configurations. Set up independent Poisson processes $\mathcal{N}_e$ and $\mathcal{N}_e^*$ for each edge e. Say edge e is *special* at time t if the particles at the end-vertices in process 1 are the same two particles as in process 2, but in transposed position. The evolution rule for the coupled processes is

> Use the same Poisson process $\mathcal{N}_e$ to define simultaneous switch times for both interchange processes, except for special edges where we use $\mathcal{N}_e$ for process 1 and $\mathcal{N}_e^*$ for process 2.

Clearly, once an individual particle is matched (i.e. at the same vertex in both processes), it remains matched thereafter. And if we watch the process (X_t, Y_t) recording the positions of particle i in each process, it is easy to check this process is the same as watching two independent copies of the continuous-time random walk, run until they meet, at time U_i, say. Thus $\max_i U_i$ is a coupling time and the coupling inequality (14.5) implies

$$\bar{d}(t) \le P(\max_i U_i > t).$$

Now U_i is distributed as $M_{v(i),w(i)}$, where $v(i)$ and $w(i)$ are the initial positions of particle i in the two versions and where $M_{v,w}$ denotes meeting time for independent copies of the underlying random walk $\tilde{X}_t$. Writing $m^* = \max_{v,w} EM_{v,w}$, we have by subexponentiality (as at (14.16))

$$P(M_{v,w} > t) \le \exp\left(1 - \frac{t}{em^*}\right)$$

and so

$$\bar{d}(t) \le n \exp\left(1 - \frac{t}{em^*}\right).$$

This leads to $\tau_1 \le (2 + \log n)em^*$ and the result follows from Proposition 14.5.

14.5.1 Card-shuffling interpretation

Taking the underlying graph to be the complete graph on n vertices, the discrete-time jump chain of the interchange process is just the "card shuffling by random transpositions" model from Chapter 7 yyy. On any graph G, the jump chain can be viewed as a card-shuffling model, but note that parameters τ are multiplied by $|\mathcal{E}|$ (the number of edges in G) when passing from the interchange process to the card-shuffling model. On the complete graph we have $\max_{v,w} \tilde{E}_v \tilde{T}_w = \Theta(1)$ and $|\mathcal{E}| = \Theta(n^2)$, and so Proposition 14.30 gives the bound $\tau_1 = O(n^2 \log n)$ for card shuffling by random transpositions, which is crude in view of the exact result $\tau_1 = \Theta(n \log n)$. In contrast, consider the n-cycle, where $\max_{v,w} \tilde{E}_v \tilde{T}_w = \Theta(n^2)$ and $|\mathcal{E}| = n$. Here the jump process is the "card shuffling by random adjacent transpositions" model from Chapter 7 yyy. In this model, Proposition 14.30 gives the bound $\tau_1 = O(n^3 \log n)$ which as mentioned in Chapter 7 yyy is the correct order of magnitude.

Diaconis and Saloff-Coste [117] studied the card-shuffling model as an application of more sophisticated techniques of comparison of Dirichlet forms.

xxx talk about their results.

14.6 Other interacting particle models

As mentioned at the start of the chapter, the models discussed in sections 14.3 - 14.5 are special in that their behavior relates to the behavior of processes built up from independent random walks on the underlying graph. In other models this is not necessarily true, and the results in this book have little application.

xxx mention Ising model and contact process.

14.6.1 Product-form stationary distributions

Consider a continuous-time particle process whose state space is the collection of subsets of vertices of a finite graph (representing the subset of vertices occupied by particles), and where only one state can change occupancy at a time. The simplest stationary distribution would be of the form

$$\text{each vertex } v \text{ is occupied independently with probability } \theta/(1+\theta) \tag{14.25}$$

where $0 < \theta < \infty$ is a parameter. By considering the detailed balance equations (Chapter 3 yyy), such a process will be reversible with stationary distribution (14.25) iff its transition rates satisfy

For configurations $\mathbf{x}^0, \mathbf{x}^1$ which coincide except that vertex v is unoccupied in $\mathbf{x}^0$ and occupied in $\mathbf{x}^1$, we have $\frac{q(\mathbf{x}^0,\mathbf{x}^1)}{q(\mathbf{x}^1,\mathbf{x}^0)} = \theta$.

There are many ways to set up such transition rates. Here is one way, observed by Neuhauser and Sudbury [269]. For each edge (w, v) at time t with w occupied,

if v is occupied at time t, then with chance dt it becomes unoccupied by time $t + dt$

if v is unoccupied at time t, then with chance θdt it becomes occupied by time $t + dt$.

If we exclude the empty configuration (which cannot be reached from other configurations) the state space is irreducible and the stationary distribution is given by (14.25) conditioned on being non-empty.

Convergence times for this model have not been studied, so we ask

Open Problem 14.31 *Give bounds on the relaxation time τ_2 in this model.*

14.6.2 Gaussian families of occupation measures

We mentioned in Chapter 3 yyy that, in the setting of a finite irreducible reversible chain (X_t), the fundamental matrix $\mathbf{Z}$ has the property

$$\pi_i Z_{ij} \text{ is symmetric and positive-definite .}$$

So by a classical result (e.g. [145] Theorem 3.6.4) there exists a mean-zero Gaussian family (γ_i) such that

$$E\gamma_i\gamma_j = \pi_i Z_{ij} \text{ for all } i, j. \tag{14.26}$$

What do such Gaussian random variables represent? It turns out there is a simple interpretation involving occupation measures of "charged particles". Take two independent copies $(X_t^+ : -\infty < t < \infty)$ and $(X_t^- : -\infty < t < \infty)$ of the stationary chain, in continuous time for simplicity. For fixed $u > 0$ consider the random variables

$$\gamma_i^{(u)} \equiv \frac{1}{2}\int_{-u}^{0} \left(1_{(X_t^+=i)} - 1_{(X_t^-=i)}\right) dt.$$

Picture one particle with charge $+1/2$ and the other particle with charge $-1/2$, and then $\gamma_i^{(u)}$ has units "charge $\times$ time". Clearly $E\gamma_i^{(u)} = 0$ and it is

easy to calculate

$$\begin{aligned} E\gamma_i^{(u)}\gamma_j^{(u)} &= \frac{1}{2}E\int_{-u}^{0}\int_{-u}^{0}\left(1_{(X_s=i,X_t=j)} - \pi_i\pi_j\right)\, ds\, dt \\ &= \frac{1}{2}\pi_i\int_{-u}^{0}\int_{-u}^{0}(P(X_t=j|X_s=i) - \pi_j)\, ds\, dt \\ &= \pi_i\int_0^u\left(1-\frac{r}{u}\right)(p_{ij}(r)-\pi_j)\, dr \end{aligned}$$

and hence

$$u^{-1}E\gamma_i^{(u)}\gamma_j^{(u)} \to \pi_i Z_{ij} \text{ as } u\to\infty. \tag{14.27}$$

The central limit theorem for Markov chains (Chapter 2 yyy) implies that the $u\to\infty$ distributional limit of $(u^{-1/2}\gamma_i^{(u)})$ is some mean-zero Gaussian family (γ_i), and so (14.27) identifies the limit as the family with covariances (14.26).

As presented here the construction may seem an isolated curiousity, but in fact it relates to deep ideas developed in the context of continuous-time-and-space reversible Markov processes. In that context, the *Dynkin isomorphism theorem* relates continuity of local times to continuity of sample paths of a certain Gaussian process. See [253] for a detailed account. And various interesting Gaussian processes can be constructed via "charged particle" models – see [2] for a readable account of such constructions. Whether these sophisticated ideas can be brought to bear upon the kinds of finite-state problems in this book is a fascinating open problem.

14.7 Other coupling examples

Example 14.32 *An m-particle process on the circle.*

Fix $m < K$. Consider m indistinguishable balls distributed amongst K boxes, at most one ball to a box, and picture the boxes arranged in a circle. At each step, pick uniformly at random a box, say box i. If box i is occupied, do nothing. Otherwise, pick uniformly at random a direction (clockwise or counterclockwise) search from i in that direction until encountering a ball, and move that ball to box i. This specifies a Markov chain on the $\begin{pmatrix} K \\ m \end{pmatrix}$ possible configurations of balls. The chain is reversible and the stationary distribution is uniform. Can we estimate the "mixing time" parameters τ_1 and τ_2? Note that as $K\to\infty$ there is a limit process involving m particles on the continuous circle, so we seek bounds which do not depend on K.

There is a simple-to-describe coupling, where for each of the two versions we pick at each time the same box and the same direction. The coupling has the usual property (c.f. the proof of Proposition 14.30) that the number of "matched" balls (i.e. balls in the same box in both processes) can only increase. But analyzing the coupling time seems very difficult. Cuellar-Montoya [105] carries through a lengthy analysis to show that $\tau_1 = O(m^{10})$. In the other direction, the bound

$$\tau_2 \geq \frac{m^3}{8\pi^2}$$

is easily established, by applying the extremal characterization (Chapter 3 yyy) to the function

$$g(\mathbf{x}) = \sum_{i=1}^{m} \sin(2\pi x_i/m)$$

where $\mathbf{x} = (x_1, \ldots, x_m)$ denotes the configuration with occupied boxes $\{x_1, \ldots, x_m\}$. It is natural to conjecture $\tau_2 = \Theta(m^3)$ and $\tau_1 = O(m^3 \log m)$.

The next example, from Jerrum [196] (xxx cite final version), uses a coupling whose construction is not quite obvious.

Example 14.33 *Permutations and words.*

Fix a finite alphabet A of size $|A|$. Fix m, and consider the set A^m of "words" $\mathbf{x} = (x_1, \ldots, x_m)$ with each $x_i \in A$. Consider the Markov chain on A^m in which a step $\mathbf{x} \to \mathbf{y}$ is specified by the following two-stage procedure.

Stage 1. Pick a permutation σ of $\{1, 2, \ldots, m\}$ uniformly at random from the set of permutations σ satisfying $x_{\sigma(i)} = x_i \forall i$.

Stage 2. Let $(c_j(\sigma); j \geq 1)$ be the cycles of σ. For each j, and independently as j varies, pick uniformly an element α_j of A, and define $y_i = \alpha_j$ for every $i \in c_j(\sigma)$.

Here is an alternative description. Write Π for the set of permutations of $\{1, \ldots, m\}$. Consider the bipartite graph on vertices $A^m \cup \Pi$ with edge-set $\{(\mathbf{x}, \sigma) : x_{\sigma(i)} = x_i \forall i\}$. Then the chain is random walk on this bipartite graph, watched every second step when it is in A^m.

From the second description, it is clear that the stationary probabilities $\pi(\mathbf{x})$ are proportional to the degree of $\mathbf{x}$ in the bipartite graph, giving

$$\pi(\mathbf{x}) \propto \prod_a n_a(\mathbf{x})!$$

where $n_a(\mathbf{x}) = |\{i : x_i = a\}|$. We shall use a coupling argument to establish the following bound on variation distance:

$$\bar{d}(t) \le m\left(1 - \frac{1}{|A|}\right)^t \tag{14.28}$$

implying that the variation threshold satisfies

$$\tau_1 \le 1 + \frac{1 + \log m}{-\log(1 - \frac{1}{|A|})} \le 1 + (1 + \log m)|A|.$$

The construction of the coupling depends on the following lemma, whose proof is deferred.

Lemma 14.34 *Givcen finite sets F^1, F^2 we can construct (for $u = 1, 2$) a uniform random permutation σ^u of F^u with cycles $(C_j^u; j \ge 1)$, where the cycles are labeled such that*

$$C_j^1 \cap F^1 \cap F^2 = C_j^2 \cap F^1 \cap F^2 \text{ for all } j.$$

We construct a step $(\mathbf{x}^1, \mathbf{x}^2) \to (\mathbf{Y}^1, \mathbf{Y}^2)$ of the coupled processes as follows. For each $a \in A$, set $F^{1,a} = \{i : x_i^1 = a\}$, $F^{2,a} = \{i : x_i^2 = a\}$. Take random permutations $\sigma^{1,a}$, $\sigma^{2,a}$ as in the lemma, with cycles $C_j^{1,a}, C_j^{2,a}$. Then $(\sigma^{1,a}, a \in A)$ define a uniform random permutation σ^1 of $\{1, \ldots, m\}$, and similarly for σ^2. This completes stage 1. For stage 2, for each pair (a, j) pick a uniform random element α_j^a of A and set

$$Y_i^1 = \alpha_j^a \text{ for every } i \in C_j^{1,a}$$

$$Y_i^2 = \alpha_j^a \text{ for every } i \in C_j^{2,a}.$$

This specifies a Markov coupling. By construction

$$\begin{array}{ll} \text{if } x_i^1 = x_i^2 & \text{then } Y_i^1 = Y_i^2 \\ \text{if } x_i^1 \ne x_i^2 & \text{then } P(Y_i^1 = Y_i^2) = 1/|A|. \end{array}$$

So the coupled processes $(\mathbf{X}^1(t), \mathbf{X}^2(t))$ satisfy

$$P(X_i^1(t) \ne X_i^2(t)) = \left(1 - \frac{1}{|A|}\right)^t 1_{(X_i^1(0) \ne X_i^2(0))}.$$

In particular $P(\mathbf{X}^1(t) \ne \mathbf{X}^2(t)) \le m(1 - 1/|A|)^t$ and the coupling inequality (14.5) gives (14.28).

xxx proof of Lemma – tie up with earlier discussion.

14.7.1 Markov coupling may be inadequate

Recall the discussion of the coupling inequality in section 14.1.1. Given a Markov chain and states i, j, theory (e.g. [233] section 3.3) says there exists a *maximal coupling* $X_t^{(i)}, X_t^{(j)}$ with a coupling time T for which the coupling inequality (14.5) holds with *equality*. But this need not be a *Markov* coupling, i.e. of form (14.6), as the next result implies. The point is that there exist fixed-degree expander graphs with $\tau_2 = O(1)$ and so $\tau_1 = O(\log n)$, but whose girth (minimal cycle length) is $\Omega(\log n)$. On such a graph, the upper bound on τ_1 obtained by a Markov coupling argument would be $\Theta(ET)$, which the Proposition shows is $n^{\Omega(1)}$.

Proposition 14.35 *Fix vertices i, j in a r-regular graph ($r \geq 3$) with girth g. Let $(X_t^{(i)}, X_t^{(j)})$ be any Markov coupling of discrete-time random walks started at i and j. Then the coupling time T satisfies*

$$ET \geq \frac{1 - (r-1)^{-d(i,j)/2}}{r-2} \; (r-1)^{\frac{g}{4} - \frac{1}{2}}.$$

Proof. We quote a simple lemma, whose proof is left to the reader.

Lemma 14.36 *Let ξ_1, ξ_2 be (dependent) random variables with $P(\xi_u = 1) = \frac{r-1}{r}$, $P(\xi_u = -1) = \frac{1}{r}$. Then*

$$E\theta^{\xi_1 + \xi_2} \leq \frac{r-1}{r}\theta^2 + \frac{1}{r}\theta^{-2}, \; 0 < \theta < 1.$$

In particular, setting $\theta = (r-1)^{-1/2}$, we have

$$E\theta^{\xi_1 + \xi_2} \leq 1.$$

Now consider the distance $D_t \equiv d(X_t^{(i)}, X_t^{(j)})$ between the two particles. The key idea is

$$\begin{aligned} E(\theta^{D_{t+1}} - \theta^{D_t} | X_t^{(i)}, X_t^{(j)}) &\leq 0 \text{ if } D_t \leq \lfloor g/2 \rfloor - 1 \\ &\leq (\theta^{-2} - 1)\theta^{\lfloor g/2 \rfloor} \text{ else.} \end{aligned} \tag{14.29}$$

The second inequality follows from the fact $D_{t+1} - D_t \geq -2$. For the first inequality, if $D_t \leq \lfloor g/2 \rfloor - 1$ then the incremental distance $D_{t+1} - D_t$ is distributed as $\xi_1 + \xi_2$ in the lemma, so the conditional expectation of $\theta^{D_{t+1} - D_t}$ is ≤ 1. Now define a martingale (M_t) via $M_0 = 0$ and

$$M_{t+1} - M_t = \theta^{D_{t+1}} - \theta^{D_t} - E(\theta^{D_{t+1}} - \theta^{D_t} | X_t^{(i)}, X_t^{(j)}).$$

Rearranging,

$$\begin{aligned} \theta^{D_t} - \theta^{D_0} &= M_t + \sum_{s=0}^{t-1} E(\theta^{D_{s+1}} - \theta^{D_s} | X_s^{(i)}, X_s^{(j)}) \\ &\leq M_t + (\theta^{-2} - 1)\theta^{\lfloor g/2 \rfloor} t \text{ by (14.29).} \end{aligned}$$

Apply this inequality at the coupling time T and take expectations: we have $EM_T = 0$ by the optional sampling theorem (Chapter 2 yyy) and $D_T = 0$, so

$$1 - \theta^{d(i,j)} \leq (\theta^{-2} - 1)\theta^{\lfloor g/2 \rfloor} ET$$

and the Proposition follows.

14.8 Notes on Chapter 14

Section 14.1. Coupling has become a standard tool in probability theory. Lindvall [233] contains an extensive treatment, emphasizing its use to prove limit theorems. Stoyan [315] emphasizes comparison results in the context of queueing systems.

Birth-and-death chains have more monotonicity properties than stated in Proposition 14.1 – see van Doorn [329] for an extensive treatment. The coupling (14.2) of a birth-and-death process is better viewed as a specialization of couplings of stochastically monotone processes, c.f. [233] Chapter 4.3.

Section 14.1.1. Using the coupling inequality to prove convergence to stationarity (i.e. the convergence theorem, Chapter 2 yyy) and the analogs for continuous-space processes is called the *coupling method*. See [233] p. 233 for some history. Systematic use to bound variation distance in finite-state chains goes back to Aldous [9]. repeated here. The coupling inequality is often presented as involving the chain started from an arbitrary point and the stationary chain, leading to a bound on $d(t)$ instead of $\bar{d}(t)$.

Section 14.3. The voter model on Z^d, and its duality with coalescing random walk, has been extensively studied – see [132, 231] for textbook treatments. The general notion of duality is discussed in [231] section 2.3. The voter model on general finite graphs has apparently been studied only once, by Donnelly and Welsh [128]. They studied the two-party model, and obtained the analog of Proposition 14.9(a) and some variations.

In the context of Open Problem 14.13 one can seek to use the randomization idea in Matthews' method, and the problem reduces to proving that,

in the coalescing of k randomly-started particles, the chance that the final join is between a $(k-1)$-cluster and a 1-cluster is small.

Section 14.3.5. On the *infinite* two-dimensional lattice, the meeting time M of independent random walks is such that $\log M$ has approximately an exponential distribution. Rather surprisingly, with a logarthmic time transformation one can get a analog of Proposition 14.17 on the infinite lattice – see Cox and Griffeath [104].

Section 14.4. Donnelly and Welsh [129] obtained Proposition 14.19 and a few other results, e.g. that, over edge-transitive graphs, c_{edge} is uniquely maximized on the complete graph.

In the context of Open Problem 14.25, there are many known Normal limits in the context of interacting particle systems on the infinite lattice, but it is not clear how well those techniques extend to general finite graphs. It would be interesting to know whether Stein's method could be used here (see Baldi and Rinott [38] for different uses of Stein's method on graphs).

Section 14.5. The name "interchange process" is my coinage: the process, in the card-shuffling interpretation, was introduced by Diaconis and Saloff-Coste [117].

The interchange process can of course be constructed from a Poisson process of directed edges, as was the voter model in section 14.3. On the n-path, this "graphical representation" has an interpretation as a method to create a pseudo-random permutation with paper and pencil – see Lange and Miller [220] for an entertaining elementary exposition.

Miscellaneous. One can define a wide variety of "growth and coverage" models on a finite graph, where there is some prescribed rule for growing a random subset $\mathcal{S}_t$ of vertices, starting with a single vertex, and the quantity of interest is the time T until the subset has grown to be the complete graph. Such processes have been studied as models for rumors, broadcast information and percolation – see e.g. Weber [335] and Fill and Pemantle [148].

Bibliography

[1] S.R. Adke and S.M. Manjunath. *An Introduction to Finite Markov Processes*. Wiley, 1984.

[2] R.J. Adler and R. Epstein. Some central limit theorems for Markov paths and some properties of Gaussian random fields. *Stochastic Process. Appl.*, 24:157–202, 1987.

[3] M. Ajtai, J. Komlós, and E. Szemerédi. Sorting in $c \log n$ parallel steps. *Combinatorica*, 3:1–19, 1983.

[4] M. Ajtai, J. Komlós, and E. Szemerédi. Deterministic simulation in logspace. In *Proc. 19th ACM Symp. Theory of Computing*, pages 132–140, 1987.

[5] D.J. Aldous. Markov chains with almost exponential hitting times. *Stochastic Process. Appl.*, 13:305–310, 1982.

[6] D.J. Aldous. Some inequalities for reversible Markov chains. *J. London Math. Soc. (2)*, 25:564–576, 1982.

[7] D.J. Aldous. Minimization algorithms and random walk on the d-cube. *Ann. Probab.*, 11:403–413, 1983.

[8] D.J. Aldous. On the time taken by random walks on finite groups to visit every state. *Z. Wahrsch. Verw. Gebiete*, 62:361–374, 1983.

[9] D.J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminaire de Probabilites XVII*, pages 243–297. Springer-Verlag, 1983. Lecture Notes in Math. 986.

[10] D.J. Aldous. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probab. Engineering Inform. Sci.*, 1:33–46, 1987.

[11] D.J. Aldous. Finite-time implications of relaxation times for stochastically monotone processes. *Probab. Th. Rel. Fields*, 77:137–145, 1988.

[12] D.J. Aldous. Hitting times for random walks on vertex-transitive graphs. *Math. Proc. Cambridge Philos. Soc.*, 106:179–191, 1989.

[13] D.J. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, 3:450–465, 1990.

[14] D.J. Aldous. The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.

[15] D.J. Aldous. Meeting times for independent Markov chains. *Stochastic Process. Appl.*, 38:185–193, 1991.

[16] D.J. Aldous. Random walk covering of some special trees. *J. Math. Analysis Appl.*, 157:271–283, 1991.

[17] D.J. Aldous. Threshold limits for cover times. *J. Theoretical Probab.*, 4:197–211, 1991.

[18] D.J. Aldous. On simulating a Markov chain stationary distribution when transition probabilities are unknown. In D.J. Aldous, P. Diaconis, J. Spencer, and J. M. Steele, editors, *Discrete Probability and Algorithms*, volume 72 of *IMA Volumes in Mathematics and its Applications*, pages 1–9. Springer-Verlag, 1995.

[19] D.J. Aldous and A. Bandyopadhyay. How to combine fast heuristic Markov chain Monte Carlo with slow exact sampling. Unpublished, 2001.

[20] D.J. Aldous and M. Brown. Inequalities for rare events in time-reversible Markov chains I. In M. Shaked and Y.L. Tong, editors, *Stochastic Inequalities*, volume 22 of *Lecture Notes*, pages 1–16. Institute of Mathematical Statistics, 1992.

[21] D.J. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93:333–348, 1986.

[22] D.J. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8:69–97, 1987.

[23] D.J. Aldous and B. Larget. A tree-based scaling exponent for random cluster models. *J. Phys. A: Math. Gen.*, 25:L1065–L1069, 1992.

[24] D.J. Aldous, L. Lovász, and P. Winkler. Mixing times for uniformly ergodic Markov chains. *Stochastic Process. Appl.*, 71:165–185, 1997.

[25] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász, and C. Rackoff. Random walks, universal traversal sequences, and the complexity of maze traversal. In *Proc. 20th IEEE Symp. Found. Comp. Sci.*, pages 218–233, 1979.

[26] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.

[27] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Comput. Complexity*, 5:60–75, 1995.

[28] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Struct. Alg.*, 5:271–284, 1994.

[29] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley, 1992.

[30] V. Anantharam and P. Tsoucas. A proof of the Markov chain tree theorem. *Stat. Probab. Letters*, 8:189–192, 1989.

[31] W.J. Anderson. *Continuous-Time Markov Chains*. Springer–Verlag, 1991.

[32] D. Applegate and R. Kannan. Sampling and integration of log-concave functions. In *Proc. 23rd ACM Symp. Theory of Computing*, pages 156–163, 1991.

[33] D. Applegate, R. Kannan, and N. Polson. Random polynomial time algorithms for sampling from joint distributions. Technical report, Carnegie-Mellon, 1990.

[34] S. Asmussen. *Applied Probability and Queues*. Wiley, 1987.

[35] S. Asmussen, P.W. Glynn, and H. Thorisson. Stationarity detection in the initial transient problem. *ACM Trans. Modeling and Computer Sim.*, 2:130–157, 1992.

[36] L. Babai. Local expansion of vertex-transitive graphs and random generation in finite groups. In *Proc. 23rd ACM Symp. Theory of Computing*, pages 164–174, 1991.

[37] L. Babai. Probably true theorems, cry wolf? *Notices Amer. Math. Soc.*, 41:453–454, 1994.

[38] P. Baldi and Y. Rinott. Asymptotic normality of some graph-related statistics. *J. Appl. Probab.*, 26:171–175, 1989.

[39] C. Bandle. *Isoperimetric Inequalities and Applications.* Pitman, Boston MA, 1980.

[40] M.T. Barlow. Random walks and diffusions on fractals. In *Proc. ICM Kyoto 1990*, pages 1025–1035. Springer–Verlag, 1991.

[41] M.T. Barlow. Diffusions on fractals. In P. Bernard, editor, *Lectures on Probability and Statistics*, volume 1690 of *Lecture Notes in Math.* Springer–Verlag, 1998.

[42] G. Barnes and U. Feige. Short random walks on graphs. *SIAM J. Discrete Math.*, 9:19–28, 1996.

[43] M.F. Barnsley and J.H. Elton. A new class of Markov processes for image encoding. *Adv. in Appl. Probab.*, 20:14–32, 1988.

[44] J.R. Baxter and R.V. Chacon. Stopping times for recurrent Markov processes. *Illinois J. Math.*, 20:467–475, 1976.

[45] K.A. Berman and M. Konsowa. Random paths, electrical networks and reversible Markov chains. *SIAM J. Discrete Math.*, 3:311–319, 1990.

[46] J. Besag and P.J. Greene. Spatial statistics and Bayesian computation. *J. Royal Statist. Soc. (B)*, 55:25–37, 1993. Followed by discussion.

[47] S. Bhatt and J-Y Cai. Taking random walks to grow trees in hypercubes. *J. Assoc. Comput. Mach.*, 40:741–764, 1993.

[48] N. L. Biggs. *Algebraic Graph Theory.* Cambridge University Press, 1974.

[49] N. L. Biggs. Potential theory on distance-regular graphs. *Combin. Probab. Comput.*, 2:243–255, 1993.

[50] L.J. Billera and P. Diaconis. A geometric interpretation of the Metropolis algorithm. Unpublished, 2000.

[51] N. H. Bingham. Fluctuation theory for the Ehrenfest urn. *Adv. in Appl. Probab.*, 23:598–611, 1991.

[52] D. Boivin. Weak convergence for reversible random walks in a random environment. *Ann. Probab.*, 21:1427–1440, 1993.

[53] B. Bollobás. *Random Graphs*. Academic Press, London, 1985.

[54] B. Bollobás. The isoperimetric number of random regular graphs. *European J. Combin.*, 9:241–244, 1988.

[55] E. Bolthausen. The Berry-Esseen theorem for functionals of discrete Markov chains. *Z. Wahrsch. Verw. Gebiete*, 54:59–73, 1980.

[56] A. Borodin, W.L. Ruzzo, and M. Tompa. Lower bounds on the length of universal traversal sequences. *J. Computer Systems Sci.*, 45:180–203, 1992.

[57] K. Borre and P. Meissl. *Strength Analysis of Leveling-type Networks*. Geodaetisk Institut, Copenhagen, 1974. Volume 50.

[58] O. Bottema. Uber die Irrfahrt in einem Strassennetz. *Math. Z.*, 39:137–145, 1935.

[59] R.C. Bradley, W. Bryc, and S. Janson. On dominations between measures of dependence. *J. Multivariate Anal.*, 23:312–329, 1987.

[60] L.A. Breyer and G.O. Roberts. From Metropolis to diffusions: Gibbs states and optimal scaling. Technical report, Statistical Lab., Cambridge U.K., 1998.

[61] G. Brightwell and P. Winkler. Extremal cover times for random walks on trees. *J. Graph Theory*, 14:547–554, 1990.

[62] G. Brightwell and P. Winkler. Maximum hitting time for random walks on graphs. *Random Struct. Alg.*, 1:263–276, 1990.

[63] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer–Verlag, 1987.

[64] A. Broder. How hard is it to marry at random? (on the approximation of the permanent). In *Proc. 18th ACM Symp. Theory of Computing*, pages 50–58, 1986.

[65] A. Broder. Generating random spanning trees. In *Proc. 30'th IEEE Symp. Found. Comp. Sci.*, pages 442–447, 1989.

[66] A. Broder and A.R. Karlin. Bounds on the cover time. *J. Theoretical Probab.*, 2:101–120, 1989.

[67] A. Broder, A.R. Karlin, P. Raghavan, and E. Upfal. Trading space for time in undirected $s-t$ connectivity. In *Proc. 21st ACM Symp. Theory of Computing*, pages 543–549, 1989.

[68] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In *Proc. 28th Symp. Foundations of Computer Sci.*, pages 286–294, 1987.

[69] A. Z. Broder, A. M. Frieze, and E. Upfal. Existence and construction of edge disjoint paths on expander graphs. *SIAM J. Comput.*, 23:976–989, 1994.

[70] A. Z. Broder, A. M. Frieze, and E. Upfal. Static and dynamic path selection on expander graphs: a random walk approach. *Random Struct. Alg.*, 14:87–109, 1999.

[71] A.E. Brouwer, A.M. Cohen, and A. Neumaier. *Distance Regular Graphs*. Springer–Verlag, 1989.

[72] M. Brown. Approximating IMRL distributions by exponential distributions, with applications to first passage times. *Ann. Probab.*, 11:419–427, 1983.

[73] M. Brown. Consequences of monotonicity for Markov transition functions. Technical report, City College, CUNY, 1990.

[74] M. Brown. Interlacing eigenvalues in time reversible Markov chains. *Math. Oper. Res.*, 24:847–864, 1999.

[75] M.J.A.M. Brummelhuis and H.J. Hilhorst. Covering of a finite lattice by a random walk. *Physica A*, 176:387–408, 1991.

[76] D. Brydges, S.N. Evans, and J.Z. Imbrie. Self-avoiding walk on a hierarchical lattice in four dimensions. *Ann. Probab.*, 20:82–124, 1992.

[77] R. Bubley and M. Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proc. 38'th IEEE Symp. Found. Comp. Sci.*, pages 223–231, 1997.

[78] R. Bubley and M. Dyer. Faster random generation of linear extensions. *Discrete Math.*, 201:81–88, 1999.

[79] R. Bubley, M. Dyer, and C. Greenhill. Beating the 2Δ bound for approximately counting colorings: A computer-assisted proof of rapid mixing. In *Proc. 9'th ACM-SIAM Symp. Discrete Algorithms*, pages 355–363, New York, 1998. ACM.

[80] R. Bubley, M. Dyer, and M. Jerrum. An elementary analysis of a procedure for sampling points in a convex body. *Random Struct. Alg.*, 12:213–235, 1998.

[81] F. Buckley and F. Harary. *Distance in Graphs.* Addison-Wesley, 1990.

[82] K. Burdzy and W.S. Kendall. Efficient Markovian couplings: Examples and counterexamples. *Ann. Appl. Probab.*, 10:362–409, 2000.

[83] R. Burton and R. Pemantle. Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances. *Ann. Probab.*, 21:1329–1371, 1993.

[84] E.A. Carlen, S. Kusuoka, and D.W. Stroock. Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincaré Probab. Statist.*, Suppl. 2:245–287, 1987.

[85] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. *Comput. Complexity*, 6:312–340, 1996/7. Extended abstract originally published in *Proc. 21st ACM Symp. Theory of Computing* (1989) 574-586.

[86] G. Chartrand and L. Lesniak. *Graphs and Digraphs.* Wadsworth, 1986.

[87] J. Cheeger. A lower bound for the lowest eigenvalue for the Laplacian. In R. C. Gunning, editor, *A Symposium in Honor of S. Bochner*, pages 195–199. Princeton Univ. Press, 1970.

[88] M. F. Chen. *From Markov Chains to Non-Equilibrium Particle Systems.* World Scientific, Singapore, 1992.

[89] M. F. Chen. Trilogy of couplings and general formulas for lower bound of spectral gap. In L. Accardi and C. Heyde, editors, *Probability Towards 2000*, number 128 in Lecture Notes in Statistics, pages 123–136. Springer–Verlag, 1996.

[90] M. F. Chen. Eigenvalues, inequalities and ergodic theory II. *Adv. Math. (China)*, 28:481–505, 1999.

[91] M.-H. Chen, Q.-M. Shao, and J.G. Ibrahim. *Monte Carlo Methods in Bayesian Computation*. Springer–Verlag, 2000.

[92] F.R.K. Chung. On concentrators, superconcentrators, and nonblocking networks. *Bell System Tech. J.*, 58:1765–1777, 1979.

[93] F.R.K. Chung. *Spectral Graph Theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Amer. Math. Soc., 1997.

[94] F.R.K. Chung, P. Diaconis, and R. L. Graham. A random walk problem arising in random number generation. *Ann. Probab.*, 15:1148–1165, 1987.

[95] F.R.K. Chung and S.-T. Yau. Eigenvalues of graphs and Sobolev inequalities. *Combin. Probab. Comput.*, 4:11–25, 1995.

[96] K.L. Chung. *Markov Chains with Stationary Transition Probabilities*. Springer–Verlag, second edition, 1967.

[97] A. Cohen and A. Widgerson. Dispensers, deterministic amplification, and weak random sources. In *Proc. 30'th IEEE Symp. Found. Comp. Sci.*, pages 14–19, 1989.

[98] A. Condon and D. Hernek. Random walks on colored graphs. *Random Struct. Alg.*, 5:285–303, 1994.

[99] D. Coppersmith, P. Doyle, P. Raghavan, and M. Snir. Random walks on weighted graphs and applications to on-line algorithms. *J. Assoc. Comput. Mach.*, 40:421–453, 1993.

[100] D. Coppersmith, U. Feige, and J. Shearer. Random walks on regular and irregular graphs. *SIAM J. Discrete Math.*, 9:301–308, 1996.

[101] D. Coppersmith, P. Tetali, and P. Winkler. Collisions among random walks on a graph. *SIAM J. Discrete Math.*, 6:363–374, 1993.

[102] M.K. Cowles and B.P. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.*, 91:883–904, 1996.

[103] J.T. Cox. Coalescing random walks and voter model consensus times on the torus in Z^d. *Ann. Probab.*, 17:1333–1366, 1989.

[104] J.T. Cox and D. Griffeath. Mean field asymptotics for the planar stepping stone model. *Proc. London Math. Soc.*, 61:189–208, 1990.

[105] S. L. Cuéllar-Montoya. A rapidly mixing stochastic system of finite interacting particles on the circle. *Stochastic Process. Appl.*, 67:69–99, 1997.

[106] D.M. Cvetkovic, M. Doob, I. Gutman, and A. Torgasev. *Recent Results in the Theory of Graph Spectra*. North-Holland, 1988. Annals of Discrete Math. 36.

[107] D.M. Cvetkovic, M. Doob, and H. Sachs. *Spectra of Graphs*. Academic Press, 1980.

[108] C. Dellacherie and P.-A. Meyer. *Probabilités et Potentiel: Théorie Discrète du Potentiel*. Hermann, Paris, 1983.

[109] L. Devroye. *Nonuniform Random Number Generation*. Springer–Verlag, 1986.

[110] L. Devroye and A Sbihi. Random walks on highly-symmetric graphs. *J. Theoretical Probab.*, 3:497–514, 1990.

[111] L. Devroye and A Sbihi. Inequalities for random walks on trees. In A. Frieze and T. Luczak, editors, *Random Graphs*, volume 2, pages 35–45. Wiley, 1992.

[112] P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward CA, 1988.

[113] P. Diaconis. Notes on the hit-and-run algorithm. Unpublished, 1996.

[114] P. Diaconis and J.A. Fill. Examples for the theory of strong stationary duality with countable state spaces. *Prob. Engineering Inform. Sci.*, 4:157–180, 1990.

[115] P. Diaconis and J.A. Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18:1483–1522, 1990.

[116] P. Diaconis, R.L. Graham, and J.A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Struct. Alg.*, 1:51–72, 1990.

[117] P. Diaconis and L. Saloff-Coste. Comparison theorems for random walk on finite groups. *Ann. Probab.*, 21:2131–2156, 1993.

[118] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3:696–730, 1993.

[119] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6:695–750, 1996.

[120] P. Diaconis and L. Saloff-Coste. Nash inequalities for finite Markov chains. *J. Theoretical Probab.*, 9:459–510, 1996.

[121] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, 57:20–36, 1998.

[122] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57:159–179, 1981.

[123] P. Diaconis and M. Shahshahani. Time to reach stationarity in the Bernouilli-Laplace diffusion model. *SIAM J. Math. Anal.*, 18:208–218, 1986.

[124] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1:36–61, 1991.

[125] I.H. Dinwoodie. A probability inequality for the occupation measure of a reversible Markov chain. *Ann. Appl. Probab.*, 5:37–43, 1995.

[126] W. Doeblin. Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkanique*, 2:77–105, 1938.

[127] P. Donnelly, P. Lloyd, and A. Sudbury. Approach to stationarity of the Bernouilli-Laplace diffusion model. *Adv. in Appl. Probab.*, 26:715–727, 1994.

[128] P. Donnelly and D. Welsh. Finite particle systems and infection models. *Math. Proc. Cambridge Philos. Soc.*, 94:167–182, 1983.

[129] P. Donnelly and D. Welsh. The antivoter problem: Random 2-colourings of graphs. In B. Bollobás, editor, *Graph Theory and Combinatorics*, pages 133–144. Academic Press, 1984.

[130] C. Dou and M. Hildebrand. Enumeration and random random walks on finite groups. *Ann. Probab.*, 24:987–1000, 1996.

[131] P.G. Doyle and J.L. Snell. *Random Walks and Electrical Networks.* Mathematical Association of America, Washington DC, 1984.

[132] R. Durrett. *Lecture Notes on Particle Systems and Percolation.* Wadsworth, Pacific Grove CA, 1988.

[133] R. Durrett. *Probability: Theory and Examples.* Wadsworth, Pacific Grove CA, 1991.

[134] M. Dyer and A. Frieze. Computing the volume of convex bodies: A case where randomness provably helps. In B. Bollobás, editor, *Probabilistic Combinatorics And Its Applications*, volume 44 of *Proc. Symp. Applied Math.*, pages 123–170. American Math. Soc., 1991.

[135] M. Dyer, A. Frieze, and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. In *Proc. 21st ACM Symp. Theory of Computing*, pages 375–381, 1989.

[136] M. Dyer, A. Frieze, and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.*, 38:1–17, 1991.

[137] M. Dyer and C. Greenhill. A genuinely polynomial-time algorithm for sampling two-rowed contingency tables. Technical report, University of Leeds, U.K., 1998. Unpublished.

[138] M.L. Eaton. Admissability in quadratically regular problems and recurrence of symmetric Markov chains: Why the connection? *J. Statist. Plann. Inference*, 64:231–247, 1997.

[139] R.G. Edwards and A.D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. D*, 38:2009–2012, 1988.

[140] B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Statist.*, 10:586–596, 1981.

[141] S.N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley, New York, 1986.

[142] U. Feige. A tight lower bound on the cover time for random walks on graphs. *Random Struct. Alg.*, 6:433–438, 1995.

[143] U. Feige. A tight upper bound on the cover time for random walks on graphs. *Random Struct. Alg.*, 6, 1995.

[144] U. Feige. Collecting coupons on trees, and the cover time of random walks. *Comput. Complexity*, 6:341–356, 1996/7.

[145] W. Feller. *An Introduction to Probability Theory and its Applications*, volume II. Wiley, 2nd edition, 1971.

[146] M. Fiedler. Bounds for eigenvalues of doubly stochastic matrices. *Linear Algebra Appl.*, 5:299–310, 1972.

[147] M. Fiedler, C.R. Johnson, T.L. Markham, and M. Neumann. A trace inequality for M-matrices and the symmetrizability of a real matrix by a positive diagonal matrix. *Linear Alg. Appl.*, 71:81–94, 1985.

[148] J. A. Fill and R. Pemantle. Percolation, first-passage percolation and covering times for Richardson's model on the n-cube. *Ann. Appl. Probab.*, 3:593–629, 1993.

[149] J.A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1:62–87, 1991.

[150] J.A. Fill. Time to stationarity for a continuous-time Markov chain. *Prob. Engineering Inform. Sci.*, 5:45–70, 1991.

[151] J.A. Fill. Strong stationary duality for continuous-time Markov chains. part I: Theory. *J. Theoretical Probab.*, 5:45–70, 1992.

[152] L. Flatto, A.M. Odlyzko, and D.B. Wales. Random shuffles and group representations. *Ann. Probab.*, 13:154–178, 1985.

[153] R.M. Foster. The average impedance of an electrical network. In *Contributions to Applied Mechanics*, pages 333–340, Ann Arbor, MI, 1949. Edwards Brothers, Inc.

[154] D. Freedman. *Markov Chains*. Springer–Verlag, 1983. Reprint of 1971 Holden-Day edition.

[155] J. Friedman. On the second eigenvalue and random walk in random regular graphs. *Combinatorica*, 11:331–362, 1991.

[156] J. Friedman, editor. *Expanding Graphs*. Amer. Math. Soc., 1993. DIMACS volume 10.

[157] J. Friedman, J. Kahn, and E. Szemeredi. On the second eigenvalue in random regular graphs. In *Proc. 21st ACM Symp. Theory of Computing*, pages 587–598, 1989.

[158] A. Frieze, R. Kannan, and N. Polson. Sampling from log-concave distributions. *Ann. Appl. Probab.*, 4:812–837, 1994.

[159] M. Fukushima. *Dirichlet Forms and Markov Processes.* North-Holland, 1980.

[160] A. Gelman, G.O. Roberts, and W.R. Gilks. Efficient Metropolis jumping rules. In *Bayesian Statistics*, volume 5, pages 599–608. Oxford University Press, 1996.

[161] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472, 1992. With discussion.

[162] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721–741, 1984.

[163] C.J. Geyer. Markov chain Monte Carlo maximum likelihood. In E. Keramigas, editor, *Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface*, pages 156–163. Interface Foundation, 1991.

[164] A. Giacometti. Exact closed form of the return probability on the Bethe lattice. *J. Phys. A: Math. Gen.*, 28:L13–L17, 1995.

[165] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*, London, 1996. Chapman and Hall.

[166] W.R. Gilks, G.O. Roberts, and E.I. George. Adaptive direction sampling. *The Statistician*, 43:179–189, 1994.

[167] D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27:1203–1220, 1998.

[168] F. Gobel and A.A. Jagers. Random walks on graphs. *Stochastic Process. Appl.*, 2:311–336, 1974.

[169] S. Goldstein. Maximal coupling. *Z. Wahrsch. Verw. Gebiete*, 46:193–204, 1979.

[170] J. Goodman and A. Sokal. Multigrid Monte Carlo method: Conceptual foundations. *Phys. Rev. D*, 40:2037–2071, 1989.

[171] D. Griffeath. A maximal coupling for Markov chains. *Z. Wahrsch. Verw. Gebiete*, pages 95–106, 1974-75.

[172] D. Griffeath. *Additive and Cancellative Interacting Particle Systems*, volume 724 of *Lecture Notes in Math.* Springer–Verlag, 1979.

[173] D. Griffeath and T. M. Liggett. Critical phenomena for Spitzer's reversible nearest particle systems. *Ann. Probab.*, 10:881–895, 1982.

[174] P. Griffin. Accelerating beyond the third dimension: Returning to the origin in a simple random walk. *Mathematical Scientist*, 15:24–35, 1990.

[175] G. Grimmett. Random graphical networks. In *Networks and Chaos — Statistical and Probabilistic Aspects*, pages 288–301. Chapman and Hall, London, 1993. Monogr. Statist. Appl. Prob. 50.

[176] G. Grimmett and H. Kesten. Random electrical networks on complete graphs. *J. London Math. Soc. (2)*, 30:171–192, 1984.

[177] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 1982.

[178] H. Haken. *Synergetics*. Springer-Verlag, 1978.

[179] K.J. Harrison and M.W. Short. The last vertex visited in a random walk on a graph. Technical report, Murdoch University, Australia, 1992.

[180] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[181] R. Holley and D. Stroock. Logarithmic Sobolev inequalitites and stochastic Ising models. *J. Statist. Phys.*, 46:1159–1194, 1987.

[182] D.F. Holt. A graph which is edge-transitive but not arc-transitive. *J. Graph Theory*, 5:201–204, 1981.

[183] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[184] B.D. Hughes. *Random Walks and Random Environments*. Oxford University Press, 1995.

[185] K. Hukushima and K. Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *J. Physics Soc. Japan*, 65:1604–1608, 1996.

[186] J.J. Hunter. *Mathematical Techniques of Applied Probability*. Academic Press, 1983.

[187] J. P. Imhof. On the range of Brownian motion and its inverse process. *Ann. Probab.*, 13:1011–1017, 1985.

[188] R. Impagliazzo and D. Zuckerman. How to recycle random bits. In *Proc. 30'th IEEE Symp. Found. Comp. Sci.*, pages 248–253, 1989.

[189] M. Iosifescu. *Finite Markov Processes and Their Applications*. Wiley, 1980.

[190] R. Isaacs. *Differential Games*. Wiley, 1965.

[191] D. Isaacson and R. Madsen. *Markov Chains: Theory and Applications*. Wiley, 1976.

[192] I. Iscoe and D. McDonald. Asymptotics of exit times for Markov jump processes I. *Ann. Probab.*, 22:372–397, 1994.

[193] I. Iscoe, D. McDonald, and K. Qian. Capacity of ATM switches. *Ann. Appl. Probab.*, 3:277–295, 1993.

[194] S. Janson. *Gaussian Hilbert Spaces*. Number 129 in Cambridge Tracts in Mathematics. Cambridge University Press, 1997.

[195] E. Janvresse. Spectral gap for Kac's model of the Boltzmann equation. *Ann. Probab.*, 29:288–304, 2001.

[196] M. Jerrum. Uniform sampling modulo a group of symmetries using Markov chain simulation. In J. Friedman, editor, *Expanding Graphs*, pages 37–48. A.M.S., 1993. DIMACS, volume 10.

[197] M. Jerrum. Mathematical foundations of the Markov chain Monte Carlo method. In *Probabilistic Methods for Algorithmic Discrete Mathematics*, number 16 in Algorithms and Combinatorics, pages 116–165. Springer–Verlag, 1998.

[198] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18:1149–1178, 1989.

[199] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In D. Hochbaum, editor, *Approximation Algorithms for NP-Hard Problems*, pages 482–520, Boston MA, 1996. PWS.

[200] M.R. Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Struct. Alg.*, 7:157–165, 1995.

[201] M.R. Jerrum, L.G. Valiant, and V.V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theor. Computer Sci.*, 43:169–188, 1986.

[202] C.D. Meyer Jr. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Review*, 17:443–464, 1975.

[203] N. Kahale. Eigenvalues and expansion of regular graphs. *J. ACM*, 42:1091–1106, 1995.

[204] N. Kahale. Large deviation bounds for Markov chains. *Combin. Probab. Comput.*, 6:465–474, 1997.

[205] J.D. Kahn, N. Linial, N. Nisan, and M.E. Saks. On the cover time of random walks on graphs. *J. Theoretical Probab.*, 2:121–128, 1989.

[206] R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Struct. Alg.*, 11:1–50, 1997.

[207] S. Karlin, B. Lindquist, and Y-C Yao. Markov chains on hypercubes: Spectral representations and several majorization relations. *Random Struct. Alg.*, 4:1–36, 1993.

[208] S. Karlin and H.M. Taylor. *A First Course in Stochastic Processes.* Academic Press, second edition, 1975.

[209] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Processes.* Academic Press, 1981.

[210] R.M. Karp, M. Luby, and N. Madras. Monte Carlo approximation algorithms for enumeration problems. *J. Algorithms*, 10:429–448, 1989.

[211] A. Karzanov and L. Khachiyan. On the conductance of order Markov chains. *Order*, 8:7–15, 1991.

[212] J. Keilson. *Markov Chain Models - Rarity and Exponentiality*. Springer–Verlag, 1979.

[213] F.P. Kelly. *Reversibility and Stochastic Networks*. Wiley, 1979.

[214] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Van Nostrand, 1960.

[215] J.G. Kemeny, J.L Snell, and A.W. Knapp. *Denumerable Markov Chains*. Springer–Verlag, 2nd edition, 1976.

[216] J.H.B. Kemperman. *The Passage Problem for a Stationary Markov Chain*. University of Chicago Press, 1961.

[217] C. Kipnis and S.R.S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104:1–19, 1986.

[218] W.B. Krebs. Brownian motion on the continuum tree. *Probab. Th. Rel. Fields*, 101:421–433, 1995.

[219] H.J. Landau and A.M. Odlyzko. Bounds for eigenvalues of certain stochastic matrices. *Linear Algebra Appl.*, 38:5–15, 1981.

[220] L. Lange and J.W. Miller. A random ladder game: Permutations, eigenvalues, and convergence of markov chains. *College Math. Journal*, 23:373–385, 1992.

[221] G. Lawler. On the covering time of a disc by simple random walk in two dimensions. In *Seminar in Stochastic Processes 1992*, pages 189–208. Birkhauser, 1993.

[222] G.F. Lawler and A.D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov proceses. *Trans. Amer. Math. Soc.*, 309:557–580, 1988.

[223] T. Leighton and S. Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *Proc. 29th IEEE Symp. Found. Comp. Sci.*, pages 422–431, 1988.

[224] G. Letac. Problemes classiques de probabilite sur un couple de Gelfand. In *Analytic Methods in Probability Theory*. Springer–Verlag, 1981. Lecture Notes in Math. 861.

[225] G. Letac. Les fonctions spheriques d'un couple de Gelfand symmetrique et les chaines de Markov. *Adv. in Appl. Probab.*, 14:272–294, 1982.

[226] G. Letac. A contraction principle for certain Markov chains and its applications. In *Random Matrices and Their Applications*, volume 50 of *Contemp. Math.*, pages 263–273. American Math. Soc., 1986.

[227] G. Letac and L. Takács. Random walks on a 600-cell. *SIAM J. Alg. Discrete Math.*, 1:114–120, 1980.

[228] G. Letac and L. Takács. Random walks on a dodecahedron. *J. Appl. Probab.*, 17:373–384, 1980.

[229] P. Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8:849–867, 1998.

[230] T.M. Liggett. Coupling the simple exclusion process. *Ann. Probab.*, 4:339–356, 1976.

[231] T.M. Liggett. *Interacting Particle Systems*. Springer–Verlag, 1985.

[232] T. Lindstrøm. Brownian motion on nested fractals. *Memoirs of the A.M.S.*, 420, 1989.

[233] T. Lindvall. *Lectures on the Coupling Method*. Wiley, 1992.

[234] J.S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6:113–119, 1996.

[235] J.S. Liu. *Monte Carlo Techniques in Scientific Computing*. Springer–Verlag, 2001.

[236] J.S. Liu, F. Liang, and W.H. Wong. The use of multiple-try method and local optimization in Metropolis sampling. *JASA*, xxx:xxx, xxx.

[237] L. Lovász. *Combinatorial Problems and Exercises*. North-Holland, 1993. Second Edition.

[238] L. Lovász and M. Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *Proc. 31st IEEE Symp. Found. Comp. Sci.*, pages 346–355, 1990.

[239] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Struct. Alg.*, 4:359–412, 1993.

[240] L. Lovász and P. Winkler. A note on the last new vertex visited by a random walk. *J. Graph Theory*, 17:593–596, 1993.

[241] L. Lovász and P. Winkler. Efficient stopping rules for Markov chains. In *Proc. 27th ACM Symp. Theory of Computing*, pages 76–82, 1995.

[242] L. Lovász and P. Winkler. Exact mixing in an unknown Markov chain. *Electronic J. Combinatorics*, 2:#R15, 1995.

[243] A. Lubotzky. *Discrete Groups, Expanding Graphs and Invariant Measures*. Birkhauser, 1994. Progress in Mathematics, vol. 125.

[244] M. Luby and E. Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets. Technical report, ICSI, Berkeley CA, 1999.

[245] R. Lyons. Random walks, capacity, and percolation on trees. *Ann. Probab.*, 20:2043–2088, 1992.

[246] R. Lyons and R. Pemantle. Random walk in a random environment and first-passage percolation on trees. *Ann. Probab.*, 20:125–136, 1992.

[247] R. Lyons, R. Pemantle, and Y. Peres. Ergodic theory on Galton–Watson trees: Speed of random walk and dimension of harmonic measure. *Ergodic Theory Dynam. Systems*, 15:593–619, 1995.

[248] R. Lyons, R. Pemantle, and Y. Peres. Biased random walks on Galton–Watson trees. *Probab. Th. Rel. Fields*, 106:249–264, 1996.

[249] R. Lyons, R. Pemantle, and Y. Peres. Unsolved problems concerning random walks on trees. In K. Athreya and P. Jagers, editors, *Classical and Modern Branching Processes*, pages 223–238. Springer–Verlag, 1996.

[250] R. Lyons and Y. Peres. *Probability on Trees and Networks*. Cambridge University Press, 2002. In preparation.

[251] T.J. Lyons. A simple criterion for transience of a reversible Markov chain. *Ann. Probab.*, 11:393–402, 1983.

[252] N. Madras and G. Slade. *The Self-Avoiding Walk*. Birkhauser, 1993.

[253] M. B. Marcus and J. Rosen. Sample path properties of the local times of strongly symmetric Markov processes via Gaussian processes. *Ann. Probab.*, 20:1603–1684, 1992.

[254] E. Marinari and G. Parisi. Simulated tempering: a new Monte Carlo scheme. *Europhysics Letters*, 19:451–458, 1992.

[255] P.C. Matthews. *Covering Problems for Random Walks on Spheres and Finite Groups*. PhD thesis, Statistics, Stanford, 1985.

[256] P.C. Matthews. Covering problems for Brownian motion on spheres. *Ann. Probab.*, 16:189–199, 1988.

[257] P.C. Matthews. Covering problems for Markov chains. *Ann. Probab.*, 16:1215–1228, 1988.

[258] P.C. Matthews. A strong uniform time for random transpositions. *J. Theoretical Probab.*, 1:411–423, 1988.

[259] P.C. Matthews. Mixing rates for Brownian motion in a convex polyhedron. *J. Appl. Probab.*, 27:259–268, 1990.

[260] P.C. Matthews. Strong stationary times and eigenvalues. *J. Appl. Probab.*, 29:228–233, 1992.

[261] J.E. Mazo. Some extremal Markov chains. *Bell System Tech. J.*, 61:2065–2080, 1982.

[262] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machine. *J. Chem. Phys.*, 21:1087–1091, 1953.

[263] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer–Verlag, 1993.

[264] J.W. Moon. Random walks on random trees. *J. Austral. Math. Soc.*, 15:42–53, 1973.

[265] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[266] C. St. J. A. Nash-Williams. Random walks and electric currents in networks. *Proc. Cambridge Philos. Soc.*, 55:181–194, 1959.

[267] R.M. Neal. *Bayesian Learning for Neural Networks*. Number 118 in Lecture Notes in Statistics. Springer–Verlag, 1996.

[268] R.M. Neal. Sampling from multimodal distributions using tempered transitions. *Statistics and Computing*, 6:353–366, 1996.

[269] C. Neuhauser and A. Sudbury. The biased annihilating branching process. *Adv. in Appl. Probab.*, 25:24–38, 1993.

[270] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[271] I. Pak. Random walk on finite groups with few random generators. *Electon. J. Probab.*, 4(1):1–11, 1999.

[272] J. L. Palacios. Bounds on expected hitting times for a random walk on a connected graph. *Linear Algebra Appl.*, 141:241–252, 1990.

[273] J. L. Palacios. A bound for the covering time of random walks on graphs. *Statistics and Probab. Lett.*, 14:9–11, 1992.

[274] J. L. Palacios. Expected cover times of random walks on symmetric graphs. *J. Theoretical Probab.*, 5:597–600, 1992.

[275] J. L. Palacios. Fluctuation theory for the Ehrenfest urn model via electrical networks. *Adv. in Appl. Probab.*, 25:472–476, 1993.

[276] J.L. Palacios. On a result of Aleliunas et al. concerning random walks on graphs. *Prob. Engineering Inform. Sci.*, 4:489–492, 1990.

[277] L. Pearce. Random walks on trees. *Discrete Math.*, 30:269–276, 1980.

[278] R. Pemantle. Choosing a spanning tree for the integer lattice uniformly. *Ann. Probab.*, 19:1559–1574, 1991.

[279] R. Pemantle. Uniform random spanning trees. In J. L. Snell, editor, *Topics in Contemporary Probability*, pages 1–54, Boca Raton, FL, 1995. CRC Press.

[280] P. Peskun. Optimal Monte Carlo sampling using Markov chains. *Biometrika*, 60:607–612, 1973.

[281] N. Pippenger. Superconcentrators. *SIAM J. Comput.*, 6:298–304, 1977.

[282] J.W. Pitman. On coupling of Markov chains. *Z. Wahrsch. Verw. Gebiete*, 35:313–322, 1976.

[283] J.W. Pitman. Occupation measures for Markov chains. *Adv. in Appl. Probab.*, 9:69–86, 1977.

[284] U. Porod. L_2 lower bounds for a special class of random walks. *Probab. Th. Rel. Fields*, 101:277–289, 1995.

[285] U. Porod. The cut-off phenomenon for random reflections. *Ann. Probab.*, 24:74–96, 1996.

[286] J. Propp and D. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Struct. Alg.*, 9:223–252, 1996.

[287] Y. Rabani, Y. Rabinovich, and A. Sinclair. A computational view of population genetics. In *Proc. 36'th IEEE Symp. Found. Comp. Sci.*, pages 83–92, 1995.

[288] P. Revesz. *Random Walk in Random and Non-Random Scenery.* World Scientific, Singapore, 1990.

[289] D. Revuz. *Markov Chains.* North-Holland, second edition, 1984.

[290] C.P. Robert, editor. *Discretization and MCMC Convergence Assesment.* Number 135 in Lecture Notes in Statistics. Springer–Verlag, 1998.

[291] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer–Verlag, 2000.

[292] G.O. Roberts. Optimal Metropolis algorithms for product measures on the vertices of a hypercube. *Stochastics Stochastic Rep.*, 62:275–283, 1998.

[293] G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7:110–120, 1997.

[294] G.O. Roberts and R.L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110, 1996.

[295] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales: Foundations*, volume 1. Wiley, second edition, 1994.

[296] Y. Roichman. On random random walks. *Ann. Probab.*, 24:1001–1011, 1996.

[297] V. I. Romanovsky. *Discrete Markov Chains.* Wolthers-Noordhoff, 1970. English translation of Russian original.

[298] J. S. Rosenthal. Random rotations: Characters and random walks on $SO(N)$. *Ann. Probab.*, 22:398–423, 1994.

[299] S. Ross. *Stochastic Processes.* Wiley, 1983.

[300] S. M. Ross. A random graph. *J. Appl. Probab.*, 18:309–315, 1981.

[301] H. Rost. The stopping distributions of a Markov process. *Inventiones Math.*, 14:1–16, 1971.

[302] O. S. Rothaus. Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 42:102–109, 1981.

[303] W. Rudin. *Real and Complex Analysis.* McGraw–Hill Book Co., New York, 3rd edition, 1987.

[304] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, pages 301–413. Springer, Berlin, 1997.

[305] P. Sarnak. *Some Applications of Modular Forms.* Cambridge University Press, 1990. Cambridge Tracts in Math. 99.

[306] A.M. Sbihi. *Covering Times for Random Walks on Graphs.* PhD thesis, McGill University, 1990.

[307] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82:93–133, 1989.

[308] A. J. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1:351–370, 1992.

[309] A. J. Sinclair. *Algorithms for Random Generation and Counting.* Birkhauser, 1993.

[310] R.L. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32:1296–1308, 1984.

[311] A.D. Sokal. Monte Carlo methods in statistical mechanics: Foundations and new algorithms. In *Cours de Troisieme Cycle de la Physique en Suisse Romande, Lausanne*, 1989.

[312] R. Solovay and V. Strassen. A fast Monte-Carlo test for primality. *SIAM J. Comput.*, 6:84–85, 1977.

[313] R. Stanley. *Enumerative Combinatorics, Vol. 2*. Cambridge University Press, 1999.

[314] W.J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1995.

[315] D. Stoyan. *Comparison Methods for Queues and Other Stochastic Models*. Wiley, 1983.

[316] W.G. Sullivan. L^2 spectral gap and jump processes. *Z. Wahrsch. Verw. Gebiete*, 67:387–398, 1984.

[317] D.E. Symer. Expanded ergodic Markov chains and cycling systems. Senior thesis, Dartmouth College, 1984.

[318] R. Syski. *Passage Times for Markov Chains*. IOS Press, Amsterdam, 1992.

[319] L. Takács. Random flights on regular polytopes. *SIAM J. Alg. Discrete Meth.*, 2:153–171, 1981.

[320] L. Takács. Random flights on regular graphs. *Adv. in Appl. Probab.*, 16:618–637, 1984.

[321] M. Tanushev and R. Arratia. A note on distributional equality in the cyclic tour property for Markov chains. *Combin. Probab. Comput.*, 6:493–496, 1997.

[322] A. Telcs. Spectra of graphs and fractal dimensions I. *Probab. Th. Rel. Fields*, 85:489–497, 1990.

[323] A. Telcs. Spectra of graphs and fractal dimensions II. *J. Theoretical Probab.*, 8:77–96, 1995.

[324] P. Tetali. Random walks and the effective resistance of networks. *J. Theoretical Probab.*, 4:101–109, 1991.

[325] P. Tetali. Design of on-line algorithms using hitting times. Bell Labs, 1994.

[326] P. Tetali. An extension of Foster's network theorem. *Combin. Probab. Comput.*, 3:421–427, 1994.

[327] P. Tetali and P. Winkler. Simultaneous reversible Markov chains. In *Combinatorics: Paul Erdos is Eighty*, volume 1, pages 433–451. Bolyai Society Mathematical Studies, 1993.

[328] H. Thorisson. *Coupling, Stationarity and Regeneration.* Springer–Verlag, 2000.

[329] Erik van Doorn. *Stochastic Monotonicity and Queueing Applications of Birth-Death Processes*, volume 4 of *Lecture Notes in Stat.* Springer–Verlag, 1981.

[330] A.R.D. van Slijpe. Random walks on regular polyhedra and other distance-regular graphs. *Statist. Neerlandica*, 38:273–292, 1984.

[331] A.R.D. van Slijpe. Random walks on the triangular prism and other vertex-transitive graphs. *J. Comput. Appl. Math.*, 15:383–394, 1986.

[332] N. Th. Varopoulos, L. Saloff-Coste, and T. Coulhon. *Analysis and Geometry on Groups.* Cambridge University Press, 1992.

[333] E. Vigoda. Improved bounds for sampling colorings. C.S. Dept., U.C. Berkeley, 1999.

[334] I.C. Walters. The ever expanding expander coefficients. *Bull. Inst. Combin. Appl.*, 17:79–86, 1996.

[335] K. Weber. Random spread of information, random graph processes and random walks. In M. Karonski, J. Jaworski, and A. Rucinski, editors, *Random Graphs '87*, pages 361–366. Wiley, 1990.

[336] H.S. Wilf. The editor's corner: The white screen problem. *Amer. Math. Monthly*, 96:704–707, 1989.

[337] H.S. Wilf. Computer-generated proofs of 54 binomial coefficient identites. U. Penn., 1990.

[338] D.B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.*, 14:274–325, 2004.

[339] W. Woess. *Random Walks on Infinite Graphs and Groups.* Number 138 in Cambridge Tracts Math. Cambridge Univ. Press, 2000.

[340] O. Yaron. Random walks on trees. Technical report, Hebrew University, Jerusalem, 1988.

[341] D. Zuckerman. Covering times of random walks on bounded degree trees and other graphs. *J. Theoretical Probab.*, 2:147–157, 1989.

[342] D. Zuckerman. On the time to traverse all edges in a graph. *Information Proc. Letters.*, 38:335–337, 1991.

[343] D. Zuckerman. A technique for lower bounding the cover time. *SIAM J. Discrete Math.*, 5:81–87, 1992.