Chapter 4
Hitting and Convergence Time, and Flow Rate, Parameters for Reversible Markov Chains

David Aldous
Department of Statistics
University of California
Berkeley, CA 94720

James Allen Fill
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2692

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

$$\tau_1 = \inf\{t > 0 : \bar{d}(t) \le e^{-1}\}$$

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 "$\tau_2 = O(\mathrm{poly}(\log n))$" 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.

1 The maximal mean commute time τ^*

We start by repeating the definition

$$\tau^* \equiv \max_{ij}(E_iT_j + E_jT_i) \tag{1}$$

and recalling what we already know. Obviously

$$\max_{ij} E_iT_j \leq \tau^* \leq 2\max_{ij} E_iT_j$$

and by Chapter 3 Lemma yyy

$$\max_j E_\pi T_j \leq \tau^* \leq 4\max_j E_\pi T_j. \tag{2}$$

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 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

$$\tau^* \leq w \max_{i,j} \min_{\text{paths } \gamma_{ij}} r(\gamma_{ij}). \tag{3}$$

A more interesting idea is to combine the max-flow min-cut theorem (see e.g. [12] sec. 5.4) with Thompson's principle (Chapter 3 Corollary yyy). Given a weighted graph, define

$$c \equiv \min_A \sum_{i \in A} \sum_{j \in A^c} w_{ij} \tag{4}$$

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

$$\sum_j |f_{ij}| \leq 2 \text{ for all } i, \text{ and } = 1 \text{ for } i = a, b. \tag{5}$$

So

$$\begin{aligned} E_a T_b + E_b T_a &\leq w \sum_e \frac{f_e^2}{w_e} \text{ by Thompson's principle} \\ &\leq \frac{w}{c} \sum_e |f_e| \\ &\leq \frac{w}{c}\,(n-1) \text{ by (5).} \end{aligned}$$

and we have proved

Proposition 2 *For random walk on an n-vertex weighted graph,*

$$\tau^* \leq \frac{w(n-1)}{c}$$

for c defined at (4).

Lemma 1 and the Monotonicity Law also make clear a one-sided bound on the effect of changing edge-weights monotonically.

Corollary 3 *Let $\tilde{w}_e \geq w_e$ be edge-weights and let $\tilde{\tau}^*$ and τ^* be the corresponding parameters for the random walks. Then*

$$\frac{E_i\tilde{T}_j + E_j\tilde{T}_i}{E_iT_j + E_jT_i} \leq \frac{\tilde{w}}{w} \text{ for all } i, j$$

and so

$$\tilde{\tau}^*/\tau^* \leq \tilde{w}/w.$$

In the case of unweighted graphs the bound in Corollary 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) $\tilde{\tau}^*/\tau^* = O(1/n)$.

2 The average hitting time τ_0

As usual we start by repeating the definition

$$\tau_0 \equiv \sum_i \sum_j \pi_j \pi_i E_i T_j \tag{6}$$

and recalling what we already know. We know (a result not using reversibility: Chapter 2 Corollary yyy) the *random target lemma*

$$\sum_j \pi_j E_i T_j = \tau_0 \text{ for all } i \tag{7}$$

and we know the *eigentime identity* (Chapter 3 yyy)

$$\tau_0 = \sum_{m \geq 2} (1 - \lambda_m)^{-1} \text{ in discrete time} \tag{8}$$

$$\tau_0 = \sum_{m \geq 2} \lambda_m^{-1} \text{ in continuous time} \tag{9}$$

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

$$\tau^* \leq \frac{2\tau_0}{\min_j \pi_j} \tag{10}$$

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 (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 3 clearly holds, by averaging over i and j.

Corollary 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.

3 The variation threshold τ_1.

3.1 Definitions

Recall from Chapter 2 yyy that $|| \quad ||$ denotes variation distance and

$$d(t) \equiv \max_i ||P_i(X_t \in \cdot) - \pi(\cdot)||$$

$$\bar{d}(t) \equiv \max_{ij} ||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)||$$

$$d(t) \leq \bar{d}(t) \leq 2d(t)$$

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

We define the parameter

$$\tau_1 \equiv \min\{t : \bar{d}(t) \leq e^{-1}\}. \qquad (11)$$

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 $\tau_2 \leq \tau_1$ – see section 4). Submultiplicativity gives

Lemma 5 $d(t) \leq \bar{d}(t) \leq \exp(-\lfloor t/\tau_1 \rfloor) \leq \exp(1 - t/\tau_1),\ t \geq 0.$

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 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 (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

$$s(t) \equiv \min\{s : p_{ij}(t) \geq (1-s)\pi_j \text{ for all } i,j\}.$$

Then $s(\cdot)$ is submultiplicative, so we naturally define the separation threshold time to be

$$\tau_1^{(1)} \equiv \min\{t : s(t) \leq e^{-1}\}.$$

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

$$\tau_1^{(4)} \equiv \max_{i,k} \sum_j \pi_j |E_i T_j - E_k T_j| = \max_{i,k} \sum_j |Z_{ij} - Z_{kj}| \qquad (12)$$

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 (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 $\tau_1^{1,\mathrm{disc}}$. 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 [1].

Theorem 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 $\tau_1^{(2)} \leq \frac{e}{e-1}\tau_1^{1,disc}$.

This will be (partially) proved in section 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

$$P_i(X(V_i) \in \cdot\ |V_i = t) = \pi(\cdot) \text{ for all } t. \tag{13}$$

Under our standing assumption of reversibility there is a close connection between separation and variation distance, indicated by the next lemma.

Lemma 7 *(a) $\bar{d}(t) \leq s(t)$.*
(b) $s(2t) \leq 1 - (1 - \bar{d}(t))^2$.

Proof. Part (a) is immediate from the definitions. For (b),

$$\begin{aligned}
\frac{p_{ik}(2t)}{\pi_k} &= \sum_j \frac{p_{ij}(t)p_{jk}(t)}{\pi_k} \\
&= \sum_j \pi_j \frac{p_{ij}(t)p_{kj}(t)}{\pi_j^2} \text{ by reversibility} \\
&\geq \left(\sum_j \pi_j \frac{p_{ij}^{1/2}(t)p_{kj}^{1/2}(t)}{\pi_j}\right)^2 \text{ by } EZ \geq (EZ^{1/2})^2 \\
&\geq \left(\sum_j \min(p_{ij}(t), p_{kj}(t))\right)^2 \\
&= (1 - ||P_i(X_t \in \cdot) - P_k(X_t \in \cdot)||)^2 \\
&\geq (1 - \bar{d}(t))^2. \quad \Box
\end{aligned}$$

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

$$\hat{d}(t) \equiv \max_{i,j} \frac{p_{ij}(t)}{\pi_j} - 1 = \max_i \frac{p_{ii}(t)}{\pi_i} - 1 \geq 0 \qquad (14)$$

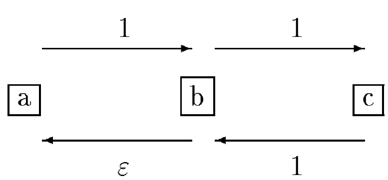
where the equality (Chapter 3 Lemma yyy) requires in discrete time that t be even. This yields the following one-sided inequalities, but Example 9 shows there can be no such reverse inequality.

Lemma 8 *(a)* $4||P_i(X_t \in \cdot) - \pi(\cdot)||^2 \leq \frac{p_{ii}(2t)}{\pi_i} - 1, \ t \geq 0.$
(b) $d(t) \leq \frac{1}{2}\sqrt{\hat{d}(2t)}, \ t \geq 0$

Proof. Part (b) follows from part (a) and the definitions. Part (a) is essentially just the "$|| \ ||_1 \leq || \ ||_2$" inequality, but let's write it out bare-hands.

$$\begin{aligned} 4||P_i(X_t \in \cdot) - \pi(\cdot)||^2 &= \left(\sum_j |p_{ij}(t) - \pi_j| \right)^2 \\ &= \left(\sum_j \pi_j^{1/2} \left| \frac{p_{ij}(t) - \pi_j}{\pi_j^{1/2}} \right| \right)^2 \\ &\leq \sum_j \frac{(p_{ij}(t) - \pi_j)^2}{\pi_j} \text{ by Cauchy-Schwarz} \\ &= -1 + \sum_j \frac{p_{ij}^2(t)}{\pi_j} \\ &= -1 + \frac{p_{ii}(2t)}{\pi_i} \text{ by Chapter 3 Lemma yyy.} \end{aligned}$$

Example 9 Consider a continuous-time 3-state chain with transition rates



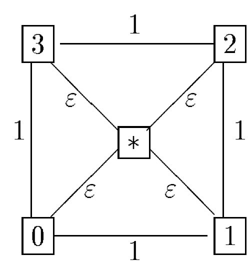
Here $\pi_a = \frac{\varepsilon}{2+\varepsilon}$, $\pi_b = \pi_c = \frac{1}{2+\varepsilon}$. It is easy to check that τ_1 is bounded as $\varepsilon \to 0$. But $p_{aa}(t) \to e^{-t}$ as $\varepsilon \to 0$, and so by considering state a we have $\hat{d}(t) \to \infty$ as $\varepsilon \to 0$ for any 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 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 $(i, *)$. $\square$

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 $\tau_1^{(2)} \to 3/2$.

3.2 Proof of Theorem 6

We will prove

Lemma 11 $\tau_1 \le \tau_1^{(1)} \le 4\tau_1$

Lemma 12 $\tau_1^{(3)} \le \tau_1^{(2)} \le \frac{e}{e-1}\tau_1^{(1)}$.

Lemma 13 $\tau_1^{(4)} \le 4\tau_1^{(3)}$

Lemma 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 12, 13 and 14 do not depend on reversibility. To complete the proof of Theorem 6 in continuous time we would need to show

$$\tau_1 \leq K\tau_1^{(5)} \text{ in continuous time} \tag{15}$$

for some absolute constant K. The proof I know is too lengthy to repeat here – see [1]. Note that (from its definition) $\tau_1^{(2)} \leq \tau_0$, so that (15) and the lemmas above imply $\tau_1 \leq 2K\tau_0$ in continuous time. We shall instead give a direct proof of a result weaker than (15):

Lemma 15 $\tau_1 \leq 66\tau_0$.

Turning to the assertions of Theorem 6 is discrete time, (b) is given by the discrete-time versions of Lemmas 11 and 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 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 11. The left inequality is immediate from Lemma 7(a), and the right inequality holds because

$$\begin{aligned} s(4\tau_1) &\leq 1 - (1 - \bar{d}(2\tau_1))^2 \text{ by Lemma 7}(b) \\ &\leq 1 - (1 - e^{-2})^2 \text{ by Lemma 5} \\ &\leq e^{-1}. \end{aligned}$$

Proof of Lemma 12. The left inequality is immediate from the definitions. For the right inequality, fix i. Write $u = \tau_1^{(1)}$, so that

$$p_{jk}(u) \geq (1 - e^{-1})\pi_k \text{ for all } j, k.$$

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

$$P_i(X_{U_i} \in \cdot, U_i = mu) = e^{-(m-1)}(1 - e^{-1})\pi(\cdot),\ m \geq 1.$$

Then $P_i(X_{U_i} \in \cdot) = \pi(\cdot)$ and $E_i U_i = u(1-e^{-1})^{-1}$. So $\tau_1^{(2)} \leq (1-e^{-1})^{-1}\tau_1^{(1)}$.

Remark. What the argument shows is that we can construct a strong stationary time V_i (in the sense of (13)) such that

$$E_i V_i = (1 - e^{-1})^{-1}\tau_1^{(1)}. \tag{16}$$

Proof of Lemma 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 $P_i(X(U_i) \in \cdot) = \mu(\cdot)$,

$$E_i T_j \leq E_i U_i + E_\mu T_j \leq \tau_1^{(3)} + E_\mu T_j.$$

The random target lemma (7) says $\sum_j E_i T_j \pi_j = \sum_j E_\mu T_j \pi_j$ and so

$$\sum_j \pi_j |E_i T_j - E_\mu T_j| = 2\sum_j \pi_j (E_i T_j - E_\mu T_j)^+ \leq 2\tau_1^{(3)}.$$

Writing $b(i)$ for the left sum, the definition of $\tau_1^{(4)}$ and the triangle inequality give $\tau_1^{(4)} \leq \max_{i,k}(b(i) + b(k))$, and the Lemma follows.

Proof of Lemma 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

$$\pi(A) E_i T_A = \sum_{j \in A} \pi_j E_i T_A = \sum_{j \in A} \pi_j (E_i T_j - E_\rho T_j)$$

$$\leq \max_k \sum_{j \in A} \pi_j (E_i T_j - E_k T_j) \leq \tau_1^{(4)}.$$

Proof of Lemma 15. For small $\delta > 0$ to be specified later, define

$$A = \{j : E_\pi T_j \leq \tau_0/\delta\}.$$

Note that Markov's inequality and the definition of τ_0 give

$$\pi(A^c) = \pi\{j : E_\pi T_j > \tau_0/\delta\} \leq \frac{\sum_j \pi_j E_\pi T_j}{\tau_0/\delta} = \frac{\tau_0}{\tau_0/\delta} = \delta. \tag{17}$$

Next, for any j

$$\begin{aligned} E_\pi T_j &= \int_0^\infty \left(\frac{p_{jj}(s)}{\pi_j} - 1 \right) ds \text{ by Chapter 2 Lemma yyy} \\ &\geq t \left(\frac{p_{jj}(t)}{\pi_j} - 1 \right) \text{ for any } t \end{aligned}$$

by monotonicity of $p_{jj}(t)$. Thus for $j \in A$ we have

$$\frac{p_{jj}(t)}{\pi_j} - 1 \leq \frac{E_\pi T_j}{t} \leq \frac{\tau_0}{\delta t}$$

and applying Chapter 3 Lemma yyy (b)

$$\frac{p_{jk}(t)}{\pi_k} \geq 1 - \frac{\tau_0}{\delta t}, \ j, k \in A.$$

Now let i be arbitrary and let $k \in A$. For any $0 \leq s \leq u$,

$$\frac{P_i(X_{u+t} = k | T_A = s)}{\pi_k} \geq \min_{j \in A} \frac{P_j(X_{u+t-s} = k)}{\pi_k} \geq 1 - \frac{\tau_0}{\delta(u+t-s)} \geq 1 - \frac{\tau_0}{\delta t}$$

and so

$$\frac{p_{ik}(u+t)}{\pi_k} \geq \left(1 - \frac{\tau_0}{\delta t} \right)^+ P_i(T_A \leq u). \tag{18}$$

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 14 and the second being an easy consequence of the definitions. Combining (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

$$\bar{d}(u+t) \leq 1 - \eta \pi(A) \leq 1 - \left(1 - \frac{\tau_0}{\delta t} \right)^+ \left(\pi(A) - \frac{2\tau_0}{u} \right)^+$$

$$\leq 1 - \left(1 - \frac{\tau_0}{\delta t}\right)^+ \left(1 - \delta - \frac{2\tau_0}{u}\right)^+ \text{ by (17).}$$

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 (15) in [1] uses these ingredients, plus the minimal hitting time construction in the recurrent balayage theorem (Chapter 2 yyy).

Outline proof of Lemma 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 $N(t) - t$. □

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 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 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,*

$$\max_i \left\| \frac{P_i(X_t \in \cdot) + P_i(X_{t+1} \in \cdot)}{2} - \pi(\cdot) \right\| \leq \psi(d(\phi(t))), \quad t = 0,1,2,\ldots$$

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

$$\beta \equiv \max(|\lambda_m| : 2 \leq m \leq n) = \max(\lambda_2, -\lambda_n). \tag{19}$$

The spectral representation then implies

$$|P_i(X_t = i) - \pi_i| \leq \beta^t, \ t = 0,1,2,\ldots. \tag{20}$$

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 18 *For an n-state chain, in discrete or continuous time,*

$$\tau_0 \leq n\tau_1^{(2)}$$

$$\tau^* \leq \frac{2\tau_1^{(2)}}{\min_j \pi_j}.$$

Lemmas 24 and 25 later are essentially stronger, giving corresponding upper bounds in terms of τ_2 instead of τ_1. But a proof of Lemma 18 is interesting for comparison with the cat-and-mouse game below.

Proof of Lemma 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

$$E(U_{s+1} - U_s | X_u, u \leq U_s) \leq \tau_1^{(2)}$$

$$(X(U_s); s \geq 1) \text{ are independent with distribution } \pi.$$

So $S_j \equiv \min\{s : X(U_s) = j\}$ has $E_{i_0} S_j = 1/\pi_j$, and so

$$E_{i_0} T_j \leq E_{i_0} U_{S_j} \leq \frac{\tau_1^{(2)}}{\pi_j}$$

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 19 *Let $0 = Y_0 \leq Y_1 \leq Y_2 \ldots$ be such that*

$$E(Y_{i+1} - Y_i | Y_j, j \leq i) \leq c, \; i \geq 0$$

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

$$EY_T \leq cET.$$

If in the hypothesis we replace "$\leq c$" 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 $m^* \geq \max_{i,j} E_iT_j$, since the mouse can just stand still.

Open Problem 20 *Does $m^* = \max_{i,j} E_iT_j$? 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

$$m^* \leq \frac{en\tau_1^{(1)}}{e-1}. \tag{21}$$

Because as remarked at (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 $0 = V_0 < V_1 < V_2 \ldots$ such that

$$E(V_{i+1} - V_i | V_j, j \leq i) \leq c, \ i \geq 0$$

$$(X(V_i), i \geq 1) \text{ are independent with the uniform distribution } \pi$$

$$(X(V_i), i \geq 1) \text{ are independent of } (V_i, i \geq 1).$$

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 $M \leq V_T$ where T is a stopping time with mean n, and (21) follows from Lemma 19. This topic will be pursued in Chapter 6 yyy.

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 [32] noticed that τ_1 could be related to "multicommodity flow" issues, and we give a streamlined version of his result (essentially Corollary 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 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

$$\psi(\mathbf{f}^{a\to\pi}) = \max_{\mathrm{edge}(i,j)} \sum_a |Z_{ia} - Z_{ja}|.$$

The result now follows because by (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 11 - 13 to relate $\tau_1^{(4)}$ to τ_1, we can deduce a lower bound on τ_1 in terms of flows.

Corollary 22 $\tau_1 \ge \frac{e-1}{16e}\ \inf_{\mathbf{f}} \psi(\mathbf{f})$.

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

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 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

$$\tau_2 = \sup\{||g||_2^2/\mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0\}. \quad (22)$$

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 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 24 $\tau_2 \leq \tau_0 \leq (n-1)\tau_2$.

Lemma 25 $\tau^* \leq \frac{2\tau_2}{\pi_*}$.

Proof of Lemma 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 5 we have $|p_{ii}(t) - \pi_i| = O(\exp(-t/\tau_1))$. This shows $\tau_2 \leq \tau_1$. For the right inequality, the spectral representation gives

$$p_{ii}(t) - \pi_i \leq e^{-t/\tau_2}. \tag{23}$$

Recalling the definition (14) of $\hat{d}$,

$$\begin{aligned}
\bar{d}(t) &\leq 2d(t) \\
&\leq \sqrt{\hat{d}(2t)} \text{ by Lemma 8}(b) \\
&\leq \sqrt{\max_i \frac{e^{-2t/\tau_2}}{\pi_i}} \text{ by (14) and (23)} \\
&= \pi_*^{-1/2} e^{-t/\tau_2}
\end{aligned}$$

and the result follows. The upper bound on $\tau_1^{(2)}$ holds in continuous time by Lemmas 11 and 12, and so holds in discrete time because $\tau_1^{(2)}$ and τ_2 are unaffected by continuization.

Proof of Lemma 24. $\tau_2 \leq \tau_0$ 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

$$\tau_0 = \sum_j \pi_j E_\pi T_j \leq \sum_j (1 - \pi_j)\tau_2 = (n-1)\tau_2.$$

Proof of Lemma 25. Fix states a, b such that $E_a T_b + E_b T_a = \tau^*$ and fix a function $0 \leq g \leq 1$ 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 $g - c$,

$$\tau_2 \geq \frac{||g-c, g-c||_2^2}{\mathcal{E}(g-c, g-c)} = \frac{\text{var }_\pi g(X_0)}{\mathcal{E}(g,g)} = \tau^* \text{ var }_\pi g(X_0).$$

But

$$\text{var }_\pi g(X_0) \quad \geq \quad \pi_a c^2 + \pi_b (1-c)^2$$

$$\begin{aligned} &\geq \inf_{0\leq y\leq 1}(\pi_a y^2 + \pi_b(1-y)^2) \\ &= \frac{\pi_a\pi_b}{\pi_a+\pi_b} \\ &\geq \frac{1}{2}\min(\pi_a,\pi_b) \\ &\geq \pi_*/2 \end{aligned}$$

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 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 23 gives

Lemma 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 $\tau_1^{\mathrm{disc}} \geq \lceil \Delta/2 \rceil$. Combining with the upper bound in Lemma 26 leads to a relationship between the diameter and the eigenvalues of a weighted graph.

Corollary 27

$$\log \frac{1}{\beta} \leq \frac{2+\log(1/\pi_*)}{\Delta}.$$

This topic will be discussed further in Chapter yyy.

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 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

$$\begin{array}{lll} S_t \equiv & \int_0^t g(X_s)ds & \text{in continuous time} \\ S_t \equiv & \sum_{s=0}^{t-1} g(X_s) & \text{in discrete time.} \end{array}$$

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 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

$$\text{var }_\pi S_t \le t||g||_2^2(2\tau_2+\frac{1}{t}). \quad (24)$$

Proof. Consider first the continuous time case. A brief calculation using the spectral representation (Chapter 3 yyy) gives

$$E_\pi g(X_0)g(X_t) = \sum_{m\ge 2} g_m^2 e^{-\lambda_m t} \quad (25)$$

where $g_m = \sum_i \pi_i^{1/2} u_{im} g(i)$. So

$$\begin{aligned} t^{-1}\text{var }_\pi S_t &= t^{-1}\int_0^t\int_0^t E_\pi g(X_u)g(X_s)\ dsdu \\ &= 2t^{-1}\int_0^t (t-s)E_\pi g(X_0)g(X_s)ds \\ &= 2\int_0^t\left(1-\frac{s}{t}\right)\sum_{m\ge 2} g_m^2 e^{-\lambda_m s}\ ds \quad (26) \\ &= 2\sum_{m\ge 2}\frac{g_m^2}{\lambda_m}A(\lambda_m t) \quad (27) \end{aligned}$$

by change of variables in the integral defining $A(u)$. The right side increases with t to

$$\sigma^2 \equiv 2\sum_{m\ge 2} g_m^2/\lambda_m, \quad (28)$$

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 (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 (27), one needs an elementary calculation to get

$$t^{-1}\text{var }_\pi S_t = 2\sum_{m\geq 2} \frac{g_m^2}{1-\lambda_m}\ B(\lambda_m, t) \tag{29}$$

$$\text{where } B(\lambda, t) = \frac{1+\lambda}{2} - \frac{\lambda(1-\lambda^t)}{t(1-\lambda)}.$$

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 (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 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 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 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,

$$\sup_x \left| P_\pi \left(\frac{S_t}{\sigma t^{1/2}} \leq x \right) - P(Z \leq x) \right| \leq \psi(||g||_\infty, t/\tau_2)$$

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.

4.2 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 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 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 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 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

$$P\left(|A(g, t_1, m_2)| > \varepsilon ||g||_2\right) \leq s(t_1) + P_\pi\left(\left|\frac{1}{m_2} \sum_{t=0}^{m_2-1} g(X_t)\right| > \varepsilon ||g||_2\right)$$

$$\leq s(t_1) + \frac{1}{m_2^2 \varepsilon^2 ||g||_2^2} \operatorname{var}_\pi \left(\sum_{t=0}^{m_2-1} g(X_t)\right).$$

Apply (24).

4.3 τ_2 and distinguished paths

The extremal characterization (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 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 32, was given by Diaconis and Stroock [16], and Sinclair [32] 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 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 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,*

$$\tau_2 \leq w \max_e \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) 1_{(e \in \gamma_{xy})}$$

$$\tau_2 \leq w \max_e \frac{1}{w_e} \sum_x \sum_y \pi_x \pi_y |\gamma_{xy}| 1_{(e \in \gamma_{xy})}.$$

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 $2 \text{ var } (Y_1) = E(Y_1 - Y_2)^2$ for i.i.d. Y's, and the first inequality is Cauchy-Schwarz.

$$\begin{aligned} 2||g||_2^2 &= \sum_x \sum_y \pi_x \pi_y (g(y) - g(x))^2 \\ &= \sum_x \sum_y \pi_x \pi_y \left(\sum_{e \in \gamma_{xy}} \Delta g(e) \right)^2 \\ &= \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) \left(\sum_{e \in \gamma_{xy}} \frac{1}{\sqrt{w_e r(\gamma_{xy})}} \sqrt{w_e} \Delta g(e) \right)^2 \qquad (30) \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) \sum_{e\in\gamma_{xy}} w_e(\Delta g(e))^2 && (31)\\
&= \sum_x \sum_y \pi_x \pi_y r(\gamma_{xy}) \sum_e w_e(\Delta g(e))^2 1_{(e\in\gamma_{xy})}\\
&\leq \kappa \sum_e w_e(\Delta g(e))^2 = \kappa\ 2w\mathcal{E}(g,g) && (32)
\end{aligned}
$$

where κ is the *max* in the first inequality in the statement of the Theorem. The first inequality now follows from the extremal characterization (22). The second inequality makes a simpler use of the Cauchy-Schwarz inequality, in which we replace (30,31,32) by

$$
\begin{aligned}
&= \sum_x \sum_y \pi_x \pi_y \left(\sum_{e\in\gamma_{xy}} 1 \cdot \Delta g(e) \right)^2\\
&\leq \sum_x \sum_y \pi_x \pi_y |\gamma_{xy}| \sum_{e\in\gamma_{xy}} (\Delta g(e))^2 && (33)\\
&\leq \kappa' \sum_e w_e(\Delta g(e))^2 = \kappa'\ 2w\mathcal{E}(g,g)
\end{aligned}
$$

where κ' is the *max* in the second inequality in the statement of the Theorem.

Remarks. (a) Theorem 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 (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 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 32 are often called *Poincaré inequalities* because, to quote [16], 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 [9])". 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 33 as $\tau_2 \leq w \max_e \frac{1}{w_e} F(e)$. 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 33,

$$\tau_2 |\vec{\mathcal{E}}| \leq w \sum_e \sum_x \sum_y \pi_x \pi_y E|\Gamma_{xy}| 1_{(e \in \Gamma_{xy})} = w \sum_x \sum_y \pi_x \pi_y E|\Gamma_{xy}|^2$$

where $|\vec{\mathcal{E}}|$ is the number of directed edges. Now $w = |\vec{\mathcal{E}}|$, so we may reinterpret this inequality as follows.

Corollary 34 *For random walk on an arc-transitive graph, $\tau_2 \leq ED^2$, 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 34 should hold for more general graphs, but the following example illustrates a difficulty.

Example 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 32 gives an upper bound of $2m$, while in fact $\tau_2 = m/2$. On the other hand the conclusion of Corollary 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 34 can hold.

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.

5.1 Definition and easy inequalities

Define

$$\tau_c = \sup_A \frac{\pi(A)\pi(A^c)}{Q(A,A^c)} \tag{34}$$

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 36 *The* sup *in (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 $\frac{Q(A,A^c)}{\pi(A)\pi(A^c)} > \gamma$. $\square$

To see how τ_c arises, note that the extremal characterization of τ_2 (22) applied to $g = 1_A$ implies

$$\frac{\pi(A)\pi(A^c)}{Q(A, A^c)} \leq \tau_2$$

for any subset A. But much more is true: Chapter 3 yyy may be rephrased as follows. For any subset A,

$$\frac{\pi(A)\pi(A^c)}{Q(A, A^c)} \leq \frac{\pi(A) E_\pi T_A}{\pi(A^c)} \leq \pi(A) E_{\alpha_A} T_A \leq \tau_2$$

where α_A is the quasistationary distribution on A^c defined at Chapter 3 yyy. So taking *sups* gives

Corollary 37

$$\tau_c \leq \sup_A \frac{\pi(A) E_\pi T_A}{\pi(A^c)} \leq \sup_A \pi(A) E_{\alpha_A} T_A \leq \tau_2.$$

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

$$\sup_{A: \pi(A) \leq 1/2} \frac{\pi(A)}{Q(A, A^c)}$$

which has been used in the literature but which would introduce a spurious factor of 2 into the inequality $\tau_c \leq \tau_2$.

Lemma 39 below shows that the final inequality of Corollary 37 can be reversed. In contrast, on the n-cycle $\tau_c = \Theta(n)$ whereas the other quantities in Corollary 37 are $\Theta(n^2)$. This shows that the "square" in Theorem 40 below cannot be omitted in general. It also suggests the following question (cf. τ_1 and $\tau_1^{(5)}$)

Open Problem 38 *Does there exist a constant K such that*

$$\tau_2 \leq K \sup_A \frac{\pi(A) E_\pi T_A}{\pi(A^c)}$$

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 39

$$\tau_2 \leq \sup_{A:\pi(A)\geq 1/2} E_{\alpha_A} T_A$$

and so in particular

$$\tau_2 \leq 2 \sup_A \pi(A) E_{\alpha_A} T_A.$$

Proof. $\tau_2 = ||h||_2^2/\mathcal{E}(h,h)$ for the eigenvector h associated with λ_2. Put

$$A = \{x : h(x) \leq 0\}$$

and assume $\pi(A) \geq 1/2$, by replacing h by $-h$ if necessary. Write $h_+ = \max(h,0)$. We shall show

$$\tau_2 \leq ||h_+||_2^2/\mathcal{E}(h_+,h_+) \tag{35}$$

and then the extremal characterization Chapter 3 yyy

$$E_{\alpha_A} T_A = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \geq 0, g = 0 \text{ on } A\} \tag{36}$$

implies $\tau_2 \leq E_{\alpha_A} T_A$ for this specific A.

The proof of (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

$$\partial\langle h_+, \mathbf{P}_t h\rangle = -\mathcal{E}(h_+, h) \leq -\mathcal{E}(h_+, h_+)$$

where the inequality follows from the inequality $(a^+ - b^+)^2 \leq (a^+ - b^+)(a-b)$ for real a, b. On the other hand, $\langle h_+, h\rangle \leq \langle h_+, h_+\rangle = ||h_+||_2^2$, and the eigenvector h satisfies $\partial(\mathbf{P}_t h) = -\lambda_2 h$, so

$$\partial\langle h_+, \mathbf{P}_t h\rangle \geq -\lambda_2 ||h_+||_2^2.$$

Combining these inequalities leads to (35).

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 40 (Cheeger's inequality) *$\tau_2 \leq 8q^*\tau_c^2$, where $q^* \equiv \max_i q_i$.*

This result follows by combining Lemma 39 above with Lemma 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 [16] – see Notes for more history.

Lemma 41 *For any subset A,*

$$E_{\alpha_A} T_A \leq \frac{2q^*\tau_c^2}{\pi^2(A)}.$$

Proof. Fix A and g with $g \geq 0$ and $g = 0$ on A.

$$\begin{aligned}
&\left(\sum\sum_{x\neq y} |g^2(x) - g^2(y)|\pi_x q_{xy}\right)^2 \\
&\leq \sum\sum_{x\neq y}(g(x)-g(y))^2\pi_x q_{xy} \times \sum\sum_{x\neq y}(g(x)+g(y))^2\pi_x q_{xy} \\
&\qquad \text{by the Cauchy-Schwarz inequality} \\
&= 2\mathcal{E}(g,g)\sum\sum_{x\neq y}(g(x)+g(y))^2\pi_x q_{xy} \\
&\leq 4\mathcal{E}(g,g)\sum\sum_{x\neq y}(g^2(x)+g^2(y))\pi_x q_{xy} \\
&= 8\mathcal{E}(g,g)\sum_x \pi_x q_x g^2(x) \\
&\leq 8q^*\mathcal{E}(g,g)\ ||g||_2^2.
\end{aligned}$$

On the other hand

$$\begin{aligned}
&\sum\sum_{x\neq y} |g^2(x) - g^2(y)|\pi_x q_{xy} \\
&= 2\sum\sum_{g(x)>g(y)} (g^2(x) - g^2(y))\pi_x q_{xy}
\end{aligned}$$

$$
\begin{aligned}
&= 4\sum \sum_{g(x)>g(y)} \left(\int_{g(y)}^{g(x)} t\,dt \right) \pi_x q_{xy} \\
&= 4 \int_0^\infty t \left(\sum \sum_{g(y)\le t<g(x)} \pi_x q_{xy} \right) dt \\
&= 4 \int_0^\infty t\; Q(B_t, B_t^c)\; dt \text{ where } B_t \equiv \{x : g(x) > t\} \\
&\geq 4 \int_0^\infty t\; \frac{\pi(B_t)\;\pi(B_t^c)}{\tau_c}\; dt \text{ by definition of } \tau_c \\
&\geq 4 \int_0^\infty t\; \frac{\pi(B_t)\;\pi(A)}{\tau_c}\; dt \text{ because } g = 0 \text{ on } A \\
&= \frac{2\pi(A)||g||_2^2}{\tau_c}.
\end{aligned}
$$

Rearranging,

$$\frac{||g||_2^2}{\mathcal{E}(g,g)} \leq \frac{2q^*\tau_c^2}{\pi^2(A)}$$

and the first assertion of the Theorem follows from the extremal characterization (36) of $E_{\alpha_A} T_A$.

5.3 τ_c and hitting times

Lemma 25 and Theorem 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 41, does better.

Proposition 42

$$\tau^* \leq \frac{4(1+\log n)}{\min_j \pi_j}\; \tau_c.$$

Example 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 $0 \leq g \leq 1$ with $g(a) = 0, g(b) = 1$. Order the states as $a = 1, 2, 3, \ldots, n = b$ so that $g(\cdot)$ is increasing.

$$
\begin{aligned}
\mathcal{E}(g,g) &= \sum \sum_{i<k} \pi_i q_{ik} (g(k) - g(i))^2 \\
&\geq \sum \sum \sum_{i\leq j<k} \pi_i q_{ik} (g(j+1) - g(j))^2
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_j (g(j+1)-g(j))^2 Q(A_j, A_j^c), \text{ where } A_j \equiv [1,j] \\
&\geq \sum_j (g(j+1)-g(j))^2 \frac{\pi(A_j)\pi(A_j^c)}{\tau_c} \qquad (37)
\end{aligned}
$$

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 (37)

$$
1 \leq \tau_c \mathcal{E}(g,g) \sum_j \frac{1}{\pi(A_j)\pi(A_j^c)}. \qquad (38)
$$

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 (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 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 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 42 is sharp up to the numerical constant.

6 Induced and product chains

Here we record the behavior of our parameters under two natural operations on chains.

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

$$\hat{p}_{\hat{\imath},\hat{\jmath}} = P_\pi(f(X_1) = \hat{\jmath} | f(X_0) = \hat{\imath}) = \frac{\sum\sum_{i,j:f(i)=\hat{\imath},f(j)=\hat{\jmath}} \pi_i p_{i,j}}{\sum_{i:f(i)=\hat{\imath}} \pi_i}. \quad (39)$$

More informatively, we are matching the stationary flow rates:

$$P_{\hat{\pi}}(\hat{X}_0 = \hat{\imath}, \hat{X}_1 = \hat{\jmath}) = P_\pi(f(X_0) = \hat{\imath}, f(X_1) = \hat{\jmath}). \quad (40)$$

The reader may check that (39) and (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.

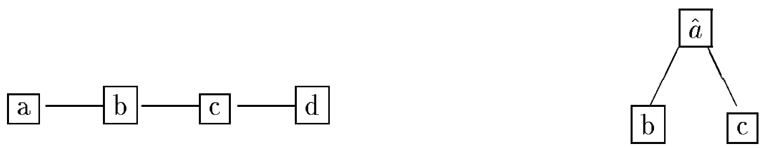
Proposition 44 (The contraction principle) *The values of τ^*, τ_0, τ_2 and τ_c in an induced chain are less than or equal to the corresponding values in the original chain.*

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. $\Box$

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 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 $K = 1$?*

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 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

$$P_{i_1, i_2}(\mathbf{X}_t = (j_1, j_2)) = P_{i_1}^{(1)}(X_t^{(1)} = j_1) P_{i_2}^{(2)}(X_t^{(2)} = j_2). \qquad (41)$$

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

$$\tau_2 = \max(\tau_2^{(1)}, \tau_2^{(2)}). \qquad (42)$$

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

$$\tau_2 = 2 \max(\tau_2^{(1)}, \tau_2^{(2)}). \tag{43}$$

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 (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 (41), the total variation distance $\bar{d}$ of section 3 satisfies

$$\bar{d}(t) = 1 - (1 - \bar{d}^{(1)}(t))(1 - \bar{d}^{(2)}(t))$$

and we deduce the crude bound

$$\tau_1 \leq 2 \max(\tau_1^{(1)}, \tau_1^{(2)})$$

where superscripts refer to the graphs $G^{(1)}, G^{(2)}$ and not to the parameters in section 3.1. For the discrete-time chain, there is an extra factor of 2 from "slowing down" (cf. (42,43)), leading to

$$\tau_1 \leq 4 \max(\tau_1^{(1)}, \tau_1^{(2)}).$$

Here our conventions are a bit confusing: this inequality refers to the discrete-time product chain, but as in section 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 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 $(\lambda_i; 1 \leq i \leq n)$ then the product chain has eigenvalues $(\lambda_i + \lambda_j; 1 \leq i, j \leq n)$ and so by the eigentime identity

$$\begin{aligned}
\tau_0^{\mathrm{product}} &= \sum_{i,j \geq 1; (i,j) \neq (1,1)} \frac{1}{\lambda_i + \lambda_j} \\
&= 2\tau_0 + \sum_{i,j \geq 2} \frac{1}{\lambda_i + \lambda_j} \\
&= 2\tau_0 + 2 \sum_{i=2}^{n} \sum_{j=i}^{n} \frac{1}{\lambda_i + \lambda_j} \\
&\leq 2\tau_0 + \sum_{i=2}^{n} (n - i + 1) \frac{2}{\lambda_i} \\
&\leq 2\tau_0 + (n-1) 2\tau_0 = 2n\tau_0.
\end{aligned}$$

Thus in discrete time

$$\tau_0^{\mathrm{product}} \leq 4n\tau_0. \qquad (44)$$

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 (43) with the extremal characterization (22) of the relaxation time for the product chain, a brief calculation gives the following result.

Corollary 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*

$$\frac{\mathrm{var}\ (Z)}{\frac{1}{2n}\sum_{i=1}^{n} E(Z - Z^{(i)})^2} \le n\tau_2.$$

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 46 becomes (45) below, because for each i the distribution of $Z - Z^{(i)}$ is unchanged by substituting X_0 for $Y^{(i)}$.

Corollary 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*

$$\mathrm{var}\ (Z) \le \frac{1}{2}\sum_{i=1}^{n} E(Z - Z^{(i)})^2 \tag{45}$$

If f is symmetric then

$$\mathrm{var}\ (Z) \le \sum_{i=0}^{n} E(Z^{(i)} - \bar{Z})^2 \tag{46}$$

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 (46) to be the celebrated *Efron-Stein* inequality in statistics, and in fact (45) is a known variant (see Notes).

Proof. As observed above, (45) is a special case of Corollary 46. So it is enough to show that for symmetric f the right sides of (45) and (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 (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 (46) equals

$$(n+1)(a - 2EZ^{(i)}\bar{Z} + E\bar{Z}^2) = na - nb.$$

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.

7 Notes on Chapter 4

Section 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 [17] sec. 5.6 for an introduction, and [8, 30] 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 [5], 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 [4, 5] and Matthews [28]. Development of theory, mostly for stochastically monotone chains on 1-dimensional state space, is in Diaconis and Fill [14, 15], Fill [20, 21] and Matthews [29].

The recurrent balayage theorem (Chapter 1 yyy) can be combined with the mean hitting time formula to get

$$\tau_1^{(2)} = \max_{ij} \frac{-Z_{ij}}{\pi_j}. \tag{47}$$

Curiously, this elegant result doesn't seem to help much with the inequalities in Theorem 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 [1]. Recently, Lovasz and Winkler [27] 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 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 7(b). But it is not clear whether there is an analog of Theorem 6 relating the $\hat{d}$-threshold to other quantities.

Section 3.2. The parts of Theorem 6 involving $\tau_1^{(1)}$ and $\tau_1^{(3)}$ are implicit rather than explicit in [1]. That paper had an unnecessarily complicated proof of Lemma 13. The proof of (15) in [1] 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

$$\tau_1^{(1)} \leq 10 \min_j \max_i \; E_i T_j \tag{48}$$

which has some relevance to later examples (yyy).

Section 3.3. The analog of Open Problem 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 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 (15) into discrete time and using the CM proxy property.

Section 3.4. The cat-and-mouse game was treated briefly in Aleliunas et al [6], who gave a bare-hands proof of a result like (21). Variations in which the cat is also allowed an arbitrary strategy have been called "princess and monster" games – see Isaacs [23] for results in a different setting.

Section 3.5. Sinclair [32] points out that "hard" results of Leighton and Rao [25] on multicommodity flow imply

$$\inf_{\mathbf{f}} \psi(\mathbf{f}) \leq K\tau_2 \log n. \tag{49}$$

This follows from Corollary 22 and Lemma 23 when π is uniform, but Sinclair posed

Open Problem 49 (i) Is there a simple proof of (49) in general?
(ii) Does (49) hold with the diameter Δ in place of $\log n$?

Section 4. As an example of historical interest, before this topic became popular Fiedler [19] proved

Proposition 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).

This upper bound is sharp. On the other hand, Proposition 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 4.1. 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 [3] 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 ([17] Thm. 2.4.10) has an analog for chains. Write σ^2 for the asymptotic variance rate in Proposition 29 and write Z for a standard Normal r.v.

Proposition 51 *There is a constant K, depending on the chain, such that*

$$\sup_x |P_\pi(\frac{S_t}{\sigma t^{1/2}} \leq x) - P(Z \leq x)| \leq Kt^{-1/2}$$

for all $t \geq 1$ 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 [10]. 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 30.

The variance results in Proposition 29 are presumably classical, being straightforward consequences of the spectral representation. Their use in algorithmic settings such as Corollary 31 goes back at least to [2].

Section 4.3. Systematic study of the optimal choice of weights in the Cauchy-Schwarz argument for Theorem 32 may lead to improved bounds in examples. Alan Sokal has unpublished notes on this subject.

Section 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 38 we mention the following result. Suppose that in the definition (section 4.1) of the maximal correlation function $\rho(t)$ we considered only *events*, i.e. suppose we defined

$$\tilde{\rho}(t) \equiv \sup_{A,B} \operatorname{cor}(1_{(X_0 \in A)}, 1_{(X_t \in B)}).$$

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 [11].

Section 5.2. The history of Cheeger-type inequalities up to 1987 is discussed in [24] section 6. Briefly, Cheeger [13] 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 [7] to the relationship between eigenvalues and expansion properties of graphs. Lawler and Sokal [24], and independently Jerrum and Sinclair [31], 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 [16], whose proof we followed for Lemmas 39 and 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 [26, 22].

Section 6.2. See Efron-Stein [18] for the origin of their inequality. Inequality (45), or rather the variant mentioned above Corollary 47 involving the $2n$ i.i.d. variables

References

[1] D.J. Aldous. Some inequalities for reversible Markov chains. *J. London Math. Soc. (2)*, 25:564–576, 1982.

[2] D.J. Aldous. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probab. Engineering Inform. Sci.*, 1:33–46, 1987.

[3] D.J. Aldous. Finite-time implications of relaxation times for stochastically monotone processes. *Probab. Th. Rel. Fields*, 77:137–145, 1988.

[4] D.J. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93:333–348, 1986.

[5] D.J. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8:69–97, 1987.

[6] 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.

[7] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.

[8] S. Asmussen. *Applied Probability and Queues.* Wiley, 1987.

[9] C. Bandle. *Isoperimetric Inequalities and Applications.* Pitman, Boston MA, 1980.

[10] E. Bolthausen. The Berry-Esseen theorem for functionals of discrete Markov chains. *Z. Wahrsch. Verw. Gebiete*, 54:59–73, 1980.

[11] R.C. Bradley, W. Bryc, and S. Janson. On dominations between measures of dependence. *J. Multivariate Anal.*, 23:312–329, 1987.

[12] G. Chartrand and L. Lesniak. *Graphs and Digraphs.* Wadsworth, 1986.

[13] 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.

[14] 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.

[15] P. Diaconis and J.A. Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18:1483–1522, 1990.

[16] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1:36–61, 1991.

[17] R. Durrett. *Probability: Theory and Examples.* Wadsworth, Pacific Grove CA, 1991.

[18] B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Statist.*, 10:586–596, 1981.

[19] M. Fiedler. Bounds for eigenvalues of doubly stochastic matrices. *Linear Algebra Appl.*, 5:299–310, 1972.

[20] J.A. Fill. Time to stationarity for a continuous-time Markov chain. *Prob. Engineering Inform. Sci.*, 5:45–70, 1991.

[21] J.A. Fill. Strong stationary duality for continuous-time Markov chains. part I: Theory. *J. Theoretical Probab.*, 5:45–70, 1992.

[22] A. Frieze, R. Kannan, and N. Polson. Sampling from log-concave distributions. *Ann. Appl. Probab.*, 4:812–837, 1994.

[23] R. Isaacs. *Differential Games.* Wiley, 1965.

[24] 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.

[25] 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.

[26] 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.

[27] L. Lovász and P. Winkler. Efficient stopping rules for Markov chains. In *Proc. 27th ACM Symp. Theory of Computing*, page xxx, 1995.

[28] P.C. Matthews. A strong uniform time for random transpositions. *J. Theoretical Probab.*, 1:411–423, 1988.

[29] P.C. Matthews. Strong stationary times and eigenvalues. *J. Appl. Probab.*, 29:228–233, 1992.

[30] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, 1993.

[31] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82:93–133, 1989.

[32] A. J. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1:351–370, 1992.