SÉMINAIRE DE PROBABILITÉS (STRASBOURG)

DAVID J. ALDOUS

Random walks on finite groups and rapidly mixing Markov chains

Séminaire de probabilités (Strasbourg), tome 17 (1983), p. 243-297
<http://www.numdam.org/item?id=SPS_1983__17__243_0>

RANDOM WALKS ON FINITE GROUPS AND RAPIDLY MIXING MARKOV CHAINS

by

David Aldous*

Department of Statistics
University of California at Berkeley

1. Introduction

This paper is an expository account of some probabilistic techniques which are useful in studying certain finite Markov chains, and in particular random walks on finite groups. Although the type of problems we consider and the form of our results are perhaps slightly novel, the mathematical ideas are mostly easy and known: our purpose is to make them well-known! We study two types of problems.

(A) Elementary theory says that under mild conditions the distribution of a Markov chain converges to the stationary distribution. Consider the (imprecise) question: how long does it take until the distribution is close to the stationary distribution? One might try to answer this using classical asymptotic theory, but we shall argue in Section 3 that this answers the wrong question. Instead, we propose that the concept "time until the distribution is close to stationary" should be formalized by a parameter τ, defined at (3.3). Since it is seldom possible to express distributions of a chain at time t in tractable form, it is seldom possible to get τ exactly, but often τ can be estimated by the coupling technique. One situation where these problems arise naturally is in random card-shuffling, where τ can be interpreted as the number of random shuffles of a particular kind needed to make a new deck well-shuffled. In Section 4 we illustrate the coupling technique by analysing several card-shuffling schemes.

(B) Some chains have what we call the "rapid mixing" property: for a random walk on a group G, this is the property that τ is small compared to $\#G$, the size of the group. When this property holds, probabilistic techniques give simple yet widely-applicable estimates for hitting time

*Research supported by National Science Foundation Grant MCS80-02698.

distributions. These are discussed in Section 7. The fundamental result (7.1) (7.18) is that for a rapidly mixing random walk with uniform initial distribution, the first hitting time on a single state is approximately exponentially distributed with mean R#G. Here R, defined at (6.4), is a parameter which can be interpreted as the mean number of visits to the initial state in the short term. This result, and its analogue for rapidly mixing Markov chains, has partial extensions to more complicated problems involving hitting times on arbitrary sets of states, and hitting times from arbitrary initial distributions.

This paper is about approximations, which may puzzle the reader: since for finite Markov chains there are of course exact expressions for distributions at time t and hitting time distributions in terms of the transition matrix. However, we have in mind the case where the state space is large, e.g., 52! in the case of card-shuffling. Exact results in terms of $52! \times 52!$ matrices are seldom illuminating.

In principle, and sometimes in practice, random walks on groups can be studied using group representation theory, the analogue of the familiar Fourier theory in the real-valued case. Diaconis (1982) studies convergence to stationarity, and Letac (1981) studies hitting times, using this theory. Our arguments use only the Markov property; we are, so to speak, throwing away the special random walk structure. So naturally our results applied to a particular random walk give less precise information than can be obtained from the analytic study of that random walk, if such a study is feasible. Instead, our results reveal some general properties, such as exponential approximations for hitting times, which are not apparent from ad hoc analyses of particular cases.

Finally, we should point out two limitations of our techniques. To apply the Markov chain results it is usually necessary to know the stationary distribution, at least approximately: one reason for concentrating on random walk examples is that then the stationary distribution is uniform. Second, the rapid mixing property on which our hitting time results depend seems characteristic of complicated "high-dimensional" processes, rather than the elementary one-dimensional examples of Markov chains, for which our techniques give no useful information.

2. Notation

The general case we shall consider is that of a continuous-time irreducible Markov process (X_t) on a finite state space $G = \{i,j,k,\ldots\}$. Let $Q(i,j)$, $j \neq i$, be the transition rates, $q_i = \sum_{j\neq i} Q(i,j)$, and let $p_{i,j}(t) = P_i(X_t = j)$ be the transition probabilities. By classical theory there exists a unique stationary distribution π, and

(2.1) $$p_{i,j}(t) \rightarrow \pi(j) \quad \text{as } t \rightarrow \infty \ ;$$

(2.2) $$t^{-1}\text{time}(s \leq t\colon X_s = j) \rightarrow \pi(j) \text{ a.s. as } t \rightarrow \infty \ ,$$

where $\text{time}(s \leq t\colon X_s = j) = \int_0^t 1_{(X_s = j)}ds$ is the random variable measuring the amount of time before time t spent in state j.

The same results hold for a discrete-time chain (X_n), except that for the analogue of (2.1) we need aperiodicity:

(2.3) $$p_{i,j}(n) \rightarrow \pi(j) \quad \text{as } n \rightarrow \infty, \quad \text{provided } X \text{ is aperiodic.}$$

In Section 3 we study convergence to stationarity in the continuous-time setting; the results hold in the discrete-time aperiodic setting with no essential changes.

Given a discrete-time chain (X_n) with transition matrix $P(i,j)$ we can define a corresponding continuous-time process (X_t^*) with transition rates $Q(i,j) = P(i,j)$, $j \neq i$. In fact we can represent (X_t^*) explicitly as

(2.4) $$X_t^* = X_{N_t} \quad \text{where } N_t \text{ is a Poisson counting process of rate 1.}$$

Let T_A (resp. T_A^*) be the first hitting time of X (resp. X^*) on a set A from some initial distribution. Then $T_A = N_{T_A^*}$ by (2.4), and it is easy to see

(2.5) $$ET_A^* = ET_A \ ; \quad T_A/T_A^* \xrightarrow{p} 1 \quad \text{as} \quad T_A^* \xrightarrow{p} \infty \ .$$

In Section 7 we study hitting time distributions for continuous-time processes;

by (2.5) our results extend to discrete-time chains. It is important to realise that even though the results in Section 7 use rapid mixing, they may be used for periodic discrete-time chains by the observation (2.5) above, since it is only required that the corresponding continuous-time process be rapid mixing.

We shall illustrate our results by discussing the special case of random walks on finite groups. Suppose G has a group structure, under the operation $\oplus$. Let μ be a probability measure on G such that

(2.6) $$\text{support}(\mu) \text{ generates } G.$$

The discrete-time random walk on G associated with μ is the process

$X_{n+1} = X_n \oplus \xi_{n+1}$, where (ξ_n) are independent with distribution μ.

Equivalently, X_n is the Markov chain with transition matrix of the special form

$$P(i,j) = \mu(i^{-1} \oplus j) .$$

By (2.6) the chain is irreducible. The stationary distribution is the uniform distribution $\pi(i) = 1/\#G$. As at (2.4) there is a corresponding continuous-time random walk (X_t), and it is for this process that our general results are stated, although in the examples we usually remain with the more natural discrete-time random walks. The results in the general Markov case become simpler to state when specialized to the random walk case, because of the "symmetry" properties of the random walk. For example, $E_\pi T_i$, the mean first hitting time on i from the stationary distribution, is clearly not dependent on i in the random walk case.

When stating the specializations in the random walk case we shall assume

(2.7) $$q_i = 1 .$$

This is automatic if μ assigns probability zero to the identity; otherwise we need only change time scale by a constant factor to attain (2.7).

We shall avoid occasional uninteresting complications by assuming

(2.8) $$\max_i \pi(i) \le \frac{1}{2} ,$$

which in the random walk case is merely the assumption that G is not the trivial group.

We should make explicit our definition of hitting times:

$$T_i = \min\{t \ge 0: X_t = i\} ;$$
$$T_A = \min\{t \ge 0: X_t \in A\} ;$$

as distinct from the first return times

(2.9) $$T_i^+ = \min\{t > 0: X_t = i,\ X_{t-} \ne i\} .$$

Elementary theory gives

(2.10) $$E_i T_i^+ = 1/\pi(i)q_i ,$$

where we are using the convention $a/bc = a/(bc)$.

For sequences (a_n), (b_n) of reals,

$$a_n \sim b_n \text{ means } \lim a_n/b_n = 1 ;$$
$$a_n \lesssim b_n \text{ means } \limsup a_n/b_n \le 1 .$$

Finally, the total variation distance between two probability measures on G is

$$\|\mu-\nu\| = \max_{A\subset G} |\mu(A)-\nu(A)| = \tfrac{1}{2}\sum|\mu(i)-\nu(i)| \le 1 .$$

3. The time to approach stationarity

In the general Markov case write

(3.1) $$d_i(t) = \|P_i(X_t \in \cdot) - \pi\|$$

for the total variation distance between the stationary distribution and the distribution at time t for the process started at i. Let

$$d(t) = \max_i d_i(t) .$$

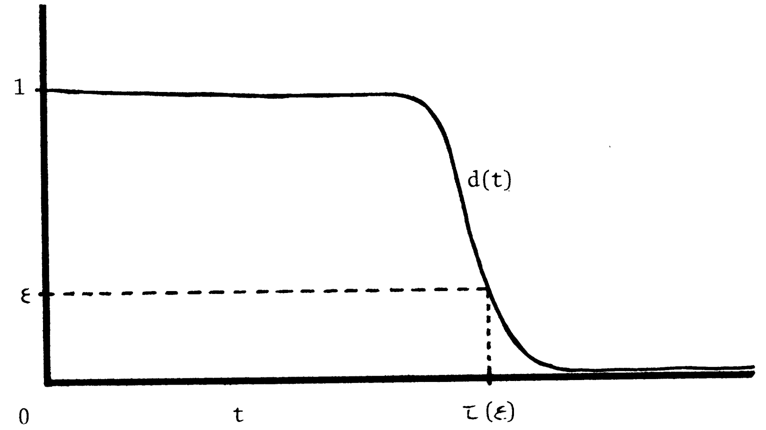
Note that in the random walk case $d_i(t)$ does not depend on i, by symmetry, so $d(t) = d_i(t)$. In general the elementary limit theorem (2.1) implies

$$d(t) \to 0 \quad \text{as} \quad t \to \infty .$$

Moreover, classical Perron-Frobenius theory gives

$$d(t) \sim C\lambda^t \quad \text{as } t \to \infty; \text{ some } C > 0,\ 0 < \lambda < 1 .$$

(In discrete time, λ is the largest absolute value, excepting 1, of the eigenvalues of the transition matrix.) Thus λ describes the asymptotic speed of convergence to stationarity. However, in our examples of rapidly mixing random walks the function $d(t)$ looks qualitatively like



That is, $d(t)$ makes a fairly abrupt switch from near 1 to near 0. It seems natural to use the time of this switch rather than the asymptotic behaviour of $d(t)$ to express the idea of "the time taken to approach uniformity". Informally, think of this switch occurring at a time τ. Formally, define

(3.2) $$\tau(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}$$

(3.3) $$\tau = \tau(1/2e)$$

where the constant $1/2e$ is used merely for algebraic convenience; replacing it by a different constant would merely alter other numerical constants in the sequel.

The idea that $d(t)$ makes an "abrupt switch" can be formalized by considering a sequence of processes. For example, in applying a particular shuffling scheme to an N-card deck we will get functions $d^N(t)$, $\tau^N(\varepsilon)$. In some examples we can prove (and we believe it holds rather generally) that there exist constants a_N such that

(3.4) $$\tau^N(\varepsilon) \sim a_N \text{ as } N \to \infty; \text{ for each } 0 < \varepsilon < 1 .$$

In other words, the scaled total variation distance function $d^N(t/a_N)$ converges to the step function $1_{(t<1)}$ for $t \neq 1$. An example is shown in Fig. 3.24.

The next lemma gives some elementary properties of $d(t)$, which can probably be traced back to Doeblin.

(3.5) LEMMA. *Define*

$$\rho_{i,j}(t) = \|P_i(X_t \in \cdot) - P_j(X_t \in \cdot)\| \;; \quad \rho(t) = \max_{i,j} \rho_{i,j}(t) .$$

Then (a) $\rho(t) \leq 2d(t)$.

(b) ρ *is submultiplicative:* $\rho(s+t) \leq \rho(s)\rho(t)$.

(c) $d(t)$ *is decreasing.*

Proof. Assertion (a) follows from the triangle inequality for the total variation distance. The other assertions can be proved algebraically, but the proof is more transparent when coupling ideas are used. The key idea is the following fact, whose proof is easy.

(3.6) LEMMA. *Let* Z_1, Z_2 *have distributions* ν_1, ν_2. *Then*

$$\|\nu_1 - \nu_2\| \leq P(Z_1 \neq Z_2) .$$

Conversely, given ν_1, ν_2, *we can construct* (Z_1, Z_2) *such that*

$$\|\nu_1 - \nu_2\| = P(Z_1 \neq Z_2) \;;$$

$$Z_n \text{ has distribution } \nu_n \quad (n = 1,2).$$

To prove (b), fix i_1, i_2, s, t. Construct (Z_s^1, Z_s^2) such that

$$Z_s^n \text{ has the distribution of } X_s \text{ given } X_0 = i_n \ ;$$
$$P(Z_s^1 \neq Z_s^2) = \rho_{i_1,i_2}(s) \ .$$

Then on the sets $A_j = \{Z_s^1 = j,\ Z_s^2 = j\}$ construct (Z_{s+t}^1, Z_{s+t}^2) such that

$$Z_{s+t}^1 = Z_{s+t}^2$$
$$P(Z_{s+t}^1 \in \cdot \mid A_j) = P_j(X_t \in \cdot) \ .$$

And on the sets $A_{j,k} = \{Z_s^1 = j,\ Z_s^2 = k\}$ $(j \neq k)$ construct (Z_{s+t}^1, Z_{s+t}^2) such that

$$P(Z_{s+t}^1 \in \cdot \mid A_{j,k}) = P_j(X_t \in \cdot) \ ; \quad P(Z_{s+t}^2 \in \cdot \mid A_{j,k}) = P_k(X_t \in \cdot) \ ;$$

(3.7)
$$P(Z_{s+t}^1 \neq Z_{s+t}^2 \mid A_{j,k}) = \rho_{j,k}(t) \le \rho(t) \ .$$

Now Z_{s+t}^1 (resp. Z_{s+t}^2) has the distribution of X_{s+t} given $X_0 = i_1$ (resp. i_2), and so

$$\begin{aligned} \rho_{i_1,i_2}(s+t) &\le P(Z_{s+t}^1 \neq Z_{s+t}^2) \\ &\le \rho(t) P(Z_s^1 \neq Z_s^2) \quad \text{by (3.7)} \\ &\le \rho(t)\rho(s) \ . \end{aligned}$$

To prove (c), use the same construction except for giving Z_s^2 the stationary distribution π, and having $P(Z_s^1 \neq Z_s^2) = d_{i_1}(s)$. Then

$$d_{i_1}(s+t) \le P(Z_{s+t}^1 \neq Z_{s+t}^2) \le P(Z_s^1 \neq Z_s^2) = d_{i_1}(s) \ .$$

This result is useful because it shows that an upper bound on ρ at a particular time t_0 gives an upper bound for later times:

$$\rho(t) \le (\rho(t_0))^n \ ; \quad nt_0 \le t \le (n+1)t_0 \ .$$

Translating this into an expression involving t explicitly,

$$\rho(t) \le (\rho(t_0))^{(t/t_0 - 1)} \ .$$

In particular the definition (3.3) of τ makes $\rho(\tau) \le e^{-1}$, and we obtain the following bound, which we shall use extensively.

(3.8) COROLLARY. $d(t) \leq \exp(1 - t/\tau)$, $t \geq 0$.

REMARKS. (a) We are here stating results for continuous time; the same results hold in the discrete time aperiodic case.

(b) Note that the exponential rate of convergence in finite state processes is a simple consequence of the basic limit theorem (2.1). The Perron-Frobenius theory is only needed if one wants an expression for the asymptotic exponent.

(c) Corollary 3.8 can be rephrased as

$$\tau(\varepsilon) \leq \tau(1 + \log(1/\varepsilon)) , \quad 0 < \varepsilon < 1 .$$

As mentioned in the Introduction, it is seldom possible to get useful exact expressions for $p_{i,j}(t)$, and hence for $d(t)$ or $\tau(\varepsilon)$. We shall instead discuss how to get bounds. The basic way to get *lower* bounds on $d(t)$ and $\tau(\varepsilon)$ is to use the obvious inequality

$$d(t) \geq |P(X_t \in A) - \pi(A)|$$

for some $A \subset G$ for which the right side may be conveniently estimated. We should however mention another general method which gives effortless, though usually rather weak, lower bounds. Recall that the *entropy* of a distribution μ is $\mathrm{ent}(\mu) = -\sum \mu(i) \log \mu(i)$. In particular, the uniform distribution π on G has $\mathrm{ent}(\pi) = \log \#G$. We quote two straightforward lemmas.

LEMMA. *Let* (X_n) *be the discrete-time random walk associated with* μ, *and let* μ_n *be the distribution of* X_n. *Then* $\mathrm{ent}(\mu_n) \leq n\, \mathrm{ent}(\mu)$.

LEMMA. *If* ν *is a distribution on* G *such that* $\|\nu - \pi\| \leq \varepsilon$ *then* $\mathrm{ent}(\nu) \geq (1-\varepsilon) \log \#G$.

From these lemmas we immediately obtain the lower bounds

$$(3.9) \qquad d(n) \geq 1 - \frac{n\, \mathrm{ent}(\mu)}{\log \#G} \ ; \qquad \tau(\varepsilon) \geq \frac{(1-\varepsilon) \log \#G}{\mathrm{ent}(\mu)}$$

for discrete-time random walks.

Our next topic is the coupling method, which is a widely-applicable method of getting *upper* bounds on τ. We remark that for the applications later to hitting time distributions we need only upper bounds on τ; and often rather crude upper bounds will suffice.

Let (X_t) be a Markov process. Fix states i, j. Suppose we can construct a pair of processes (Z_t^1, Z_t^2) such that

(3.10) Z^1 (resp. Z^2) is distributed as X given $X_0 = i$ (resp. j);

(3.11)
$$Z_t^1 = Z_t^2 \quad \text{on} \quad \{t \geq T\}\ , \quad \text{where}$$
$$T\ (= T^{i,j}) = \inf\{t: Z_t^1 = Z_t^2\}\ .$$

Call (Z^1, Z^2) a *coupling*, and T a *coupling time*. By Lemma 3.6

(3.12)
$$\rho_{i,j}(t) \leq P(Z_t^1 \neq Z_t^2) = P(T^{i,j} > t)\ .$$

Thus from estimates of the tails of the distributions of coupling times we can get estimates for $d(t)$. A crude way is to take expectations. Suppose we have constructed couplings for each pair i, j. Then

(3.13)
$$\tau \leq 2e\tau_c\ , \quad \text{where} \quad \tau_c = \max_{i,j} ET^{i,j}\ ,$$

because by (3.12) $\rho(t) \leq \tau_c/t$.

To summarize: to get good estimates of the time taken for the process to approach stationarity, we seek to construct couplings for which the coupling time is as small as possible.

We now outline the strategy we shall use in constructing couplings. It is conceptually simpler to discuss the discrete-time case first. Suppose we have a function $f: G \times G \to \{0,1,2,\ldots\}$ such that $f(i,j) = 0$ iff $i = j$: call f a *distance function*. Suppose that for each pair (i,j) there is a joint distribution

(3.14)
$$\theta_{i,j} = \mathcal{L}(V,W) \quad \text{such that}$$
$$\mathcal{L}(V) = P(i,\cdot);\quad \mathcal{L}(W) = P(j,\cdot);\quad V = W \text{ if } i = j.$$

Then we can construct the bivariate Markov process (Z_n^1, Z_n^2) such that

$$P((Z_{n+1}^1, Z_{n+1}^2) \in \cdot \mid (Z_n^1, Z_n^2) = (i,j)) = \theta_{i,j} .$$

This is plainly a coupling. Think of the process $D_n = f(Z_n^1, Z_n^2)$ as measuring the distance between the two processes; the coupling time is

$$T = \min(n: D_n = 0) .$$

All our couplings will be of this Markovian form. To specify the coupling, we need only specify the "one-step" distributions $\theta_{i,j}$. Of course there will be many possible choices for these joint distributions with prescribed marginals: since our aim is to make D_n decrease it is natural to choose the distribution (V,W) to minimize $Ef(V,W)$, and indeed it is often possible to arrange that $f(V,W) \le f(i,j)$ with some positive probability of a strict decrease. Once the coupling is specified, estimating the coupling time (and hence τ) is just estimating the time for the integer-valued process D_n to hit 0. Note, however, that D_n need not be Markov.

In the continuous-time setting, we merely replace the joint transition probabilities $\theta_{i,j}(k,\ell)$ by joint transition *rates* $\Lambda_{i,j}(k,\ell)$ such that

(3.15) $\sum_{\ell} \Lambda_{i,j}(k,\ell) = Q(i,k)$; $\sum_{k} \Lambda_{i,j}(k,\ell) = Q(j,\ell)$; $\Lambda_{i,i}(k,k) = Q(i,k)$.

We should mention the useful trick of time-reversal. Suppose (X_n) is the random walk associated with μ. Let $\mu^*(j) = \mu(j^{-1})$. Then the random walk (X_n^*) associated with μ^* is called the *time-reversed* process, because of the easily-established properties

(a) $P_j(X_n^* = k) = P_k(X_n = j)$;

(b) when X_0 and X_0^* are given the uniform distribution,

$$(X_0^*, X_1^*, \ldots, X_K^*) \overset{\mathcal{D}}{=} (X_K, X_{K-1}, \ldots, X_0) .$$

The next lemma shows that when estimating $d(n)$ we may replace the original random walk with its time-reversal, if this is more convenient to work with.

(3.16) LEMMA. *Let* $d(n)$ *(resp.* $d^*(n)$*) be the total variation function for a random walk* X_n *(resp. the time-reversed walk* X_n^**). Then* $d(n) = d^*(n)$.

Proof. Writing i for the identity of G,

$$\begin{aligned} d(n) &= \sum_j |P_i(X_n = j) - 1/\#G| \\ &= \sum_j |P_{j^{-1}}(X_n = i) - 1/\#G| \quad \text{by the random walk property} \\ &= \sum_j |P_j(X_n = i) - 1/\#G| \quad \text{re-ordering the sum} \\ &= \sum_j |P_i(X^*_n = j) - 1/\#G| \quad \text{by (a)} \\ &= d^*(n) . \end{aligned}$$

Of course it may happen that $\mu = \mu^*$, so the reversed process is the same as the original process: call such a random walk *reversible*. In the general continuous-time Markov setting, a process is reversible if it satisfies the equivalent conditions

$$\begin{aligned} \pi(i)Q(i,j) &= \pi(j)Q(j,i) \\ \pi(i)p_{i,j}(t) &= \pi(j)p_{j,i}(t) \end{aligned} \tag{3.17}$$

Although we lose the opportunity of taking advantage of our trick, reversible processes do have some regularity properties not necessarily possessed by non-reversible processes. For instance, another way to formalize the concept of "the time to approach stationarity" is to consider the random walk with $X_0 = i$ and consider stopping times S such that X_S is uniform; let $\hat{\tau}_i$ be the infimum of $E_i S$ over all such stopping times, and let $\hat{\tau} = \min_i \hat{\tau}_i$. It can be shown that $\hat{\tau}$ is equivalent to τ for reversible processes, in the following sense.

(3.18) PROPOSITION. *There exist constants* C_1, C_2 *such that* $C_1\tau \le \hat{\tau} \le C_2\tau$ *for all reversible Markov processes.*

This and other results on reversible processes are given in Aldous (1982a).

The rest of this section is devoted to one example, in which there is an exact analytic expression for $d(t)$ which can be compared with coupling estimates.

(3.19) EXAMPLE. *Random walk on the* N-*dimensional cube*. The vertices of the unit cube in N dimensions can be labelled as N-tuples $i = (i_1,\ldots,i_N)$ of 0's and 1's, and form a group G under componentwise addition modulo 2. There is a natural distance function $f(i,j) = \sum|i_r - j_r|$. Write $0 = (0,\ldots,0)$, $u^r = (0,\ldots,0,1,0,\ldots,0)$ with 1 at coordinate r,

$$\mu(u^r) = 1/N \quad 1 \leq r \leq N ,$$
$$\mu(j) = 0 \quad \text{otherwise.}$$

The random walk associated with μ is the natural "simple random walk" on the cube, which jumps from a vertex to one of the neighboring vertices chosen uniformly at random. The discrete-time random walk is periodic: we shall consider the continuous-time process, though similar results would hold for the discrete-time random walk modified to become aperiodic by putting

$$\mu(u^r) = 1/(N+1) \quad 1 \leq r \leq N$$
$$\mu(0) = 1/(N+1)$$

We now describe a coupling, which will give an upper bound for τ. Fix i, j; let $L = f(i,j)$ and let $C = \{c_1,\ldots,c_L\}$ be the set of coordinates c for which $j_c \neq i_c$. Define $\Lambda_{i,j}(k,\ell)$ as follows.

$$\Lambda_{i,j}(i \oplus u^c,\ j \oplus u^c) = 1/N , \quad c \notin C .$$

(if $L > 1$) $$\Lambda_{i,j}(i \oplus u^{c_r},\ j \oplus u^{c_{r+1}}) = 1/N , \quad 1 \leq r \leq L$$

(interpret c_{L+1} as c_1).

(if $L = 1$) $$\Lambda_{i,j}(i \oplus u^c, j) = \Lambda_{i,j}(i, j \oplus u^c) = 1/N , \quad c \in C.$$

Let (Z_t^1, Z_t^2) be the associated coupling, i.e. the Markov process with transition rates $\Lambda_{i,j}(k,\ell)$. It is plain that the distance process $D_t = f(Z_t^1, Z_t^2)$ evolves as the Markov process on $\{0,1,\ldots,N\}$ with transition rates $Q(n,n-2) = n/N$ $(2 \leq n \leq N)$, $Q(1,0) = 2/N$. It is not hard to deduce that the coupling time T is stochastically dominated by the sum

$$T^* = T_1^* + T_3^* + T_5^* + \cdots + T_M^* ; \quad M = \begin{cases} N & (N \text{ odd}) , \\ N-1 & (N \text{ even}), \end{cases}$$

where the summands are independent exponential random variables, T_m^* having mean N/m. To estimate the tail of the distribution of T^* we calculate

$$ET^* = N(1 + 1/3 + 1/5 + \cdots + 1/M) \sim \tfrac{1}{2}N \log(N)$$
$$\mathrm{var}(T^*) = N^2(1 + (1/3)^2 + \cdots + (1/M)^2) \sim CN^2 .$$

So for $\alpha > \frac{1}{2}$,

$$d(\alpha N \log(N)) \leq P(T^* > \alpha N \log(N))$$
$$\rightarrow 0 \quad \text{as} \quad N \rightarrow \infty \quad \text{by Chebyshev's inequality.}$$

So we conclude

$$\tau(\varepsilon) \underset{\sim}{<} \tfrac{1}{2}N \log(N) \; ; \quad 0 < \varepsilon < 1 .$$

We shall now show how to get an exact analytic formula for $d(t)$. Write the continuous-time random walk X_t componentwise as $(X_t^1,\ldots,X_t^N)$. It is easy to verify that the component processes X_t^c are independent Markov processes on $\{0,1\}$ with transition rates $Q(0,1) = Q(1,0) = 1/N$. So the component processes have transition probabilities

$$P_0(X_t^c = 0) = \tfrac{1}{2}\{1 + \exp(-2t/N)\} , \quad P_0(X_t^c = 1) = \tfrac{1}{2}\{1 - \exp(-2t/N)\} .$$

So the transition probabilities for the random walk are

(3.20) $$P_{\underset{\sim}{0}}(X_t = \underset{\sim}{j}) = 2^{-N}\{1 + \exp(-2t/N)\}^{N-L}\{1 - \exp(-2t/N)\}^L \; ; \quad L = f(\underset{\sim}{j},\underset{\sim}{0}) .$$

Thus we obtain the formula

(3.21) $$d(t) = 2^{-N-1} \sum_{L=0}^{N} \binom{N}{L} \left|\{1 + \exp(-2t/N)\}^{N-L}\{1 - \exp(-2t/N)\}^L - 1\right| .$$

Elementary but tedious calculus shows

(3.22) $$\lim_N d(t / \tfrac{1}{4}N \log(N)) = 1 , \quad t < 1$$
$$= 0 , \quad t > 1 ,$$

and hence

(3.23) $$\tau(\varepsilon) \sim \tfrac{1}{4}N \log(N) , \quad 0 < \varepsilon < 1.$$

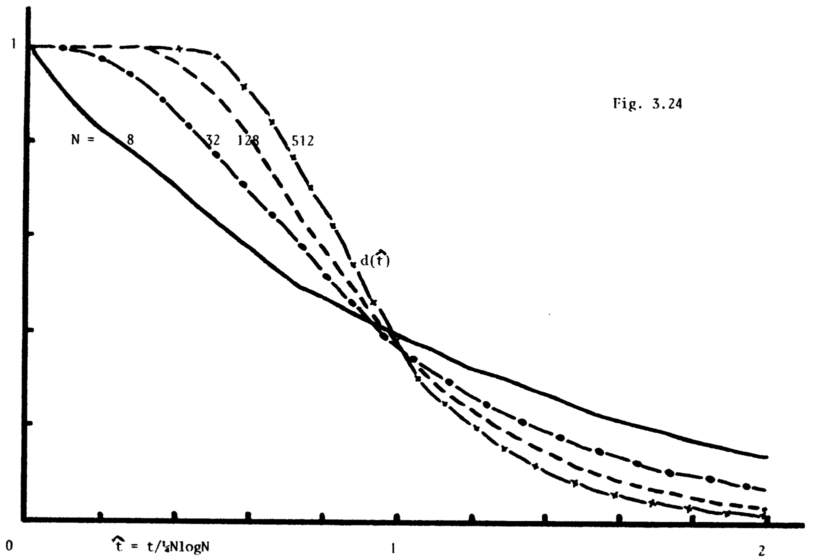


Fig. 3.24

Thus we see that the upper bound for τ derived by the coupling technique gives the correct order of magnitude, though not the correct constant, in this example.

Figure 3.24 shows computer-calculated graphs of $d(t/\frac{1}{4}N \log N)$ for $N = 8, 32, 128, 512,$ to illustrate the convergence in (3.22).

REMARK. Our use of total variation distance to measure how close a distribution is to the stationary distribution may seem an arbitrary choice. What if we used another indicator, say entropy? In this example the entropy $\mathcal{E}_N(t)$ of the distribution of X_t has the form

$$\mathcal{E}_N(t) = N\phi(t/N)$$

for a certain $\phi(\cdot)$. Thus $\phi_N(\cdot)$ does not exhibit the "abrupt switch" of $d_N(\cdot)$ for large N. So it is hard to see how to define a parameter analogous to τ in terms of entropy; and it is not clear that the hitting time approximations of Section 7 would be valid under some definition of "rapid mixing" which used entropy rather than total variation distance.

4. Card-shuffling models

Imagine a deck of N cards, labelled 1 to N. The state of the deck may be described by a permutation π of $\{1,\dots,N\}$, the card labelled i being in position $\pi(i)$, where position 1 is the top of the deck and position N the bottom. So the card in position j is labelled $\pi^{-1}(j)$. A *shuffle* of the deck may also be described by a permutation σ, indicating that the card at position i has moved to position $\sigma(i)$. A probability distribution μ on the group G_N of permutations describes the *random shuffle* in which σ is picked according to the distribution μ. Write $X_n(i)$ for the position of the card labelled i after n independent such random shuffles. Then $X_n = (X_n(i))$ is the random walk on the group G_N associated with μ.

Let Π be the uniform distribution on G_N. Imagine starting with a *new* deck (i.e. with the card labelled i in position i). As in section 3 let $d(n)$ be the total variation distance between the distribution of X_n and Π. Think of the parameter τ at (3.3) as measuring the number of shuffles needed to get the deck well-shuffled. Our purpose in this section is to estimate τ for some specific shuffles μ. More precisely, we shall try to find the asymptotic behaviour of τ_N as the number of cards N tends to infinity. We shall get upper bounds by coupling. To describe couplings, we imagine two decks, in states π, σ, say, and then specify dependent random shuffles (Σ_1,Σ_2) of the two decks, each Σ_r having distribution μ. The joint distribution $\theta_{\pi,\sigma}$ of $(\pi\Sigma_1,\sigma\Sigma_2)$ is the transition matrix for the coupled processes. One way of getting lower bounds is to consider the motion of a particular card: this motion $Y_n = X_n(i)$, i fixed, forms the Markov chain on $\{1,\dots,N\}$ with transition matrix

$$P(j,k) = \mu\{\pi\colon \pi(j) = k\} \quad ; \tag{4.1}$$

the stationary distribution is uniform. Writing d_Y for the total variation distance function for Y_n, we have the obvious inequality

$$d(n) \geq d_Y(n) \ . \tag{4.2}$$

We shall need three famous results from elementary probability theory, which we now describe.

Given two decks, say a *match* occurs whenever one position is occupied by the same labelled card in both decks. Let $M(\pi,\sigma) = \{i: \pi(i) = \sigma(i)\}$ be the number of matches between decks in states π and σ. Then (Feller (1968) p. 107)

(4.3) CARD-MATCHING LEMMA. *For* X *uniform on* G_N, $M(X,\sigma) \xrightarrow{\mathcal{D}}$ *Poisson* (1) *as* $N \to \infty$.

Note that $f(\pi,\sigma) = N - M(\pi,\sigma)$, the number of *un*matched cards, is a natural distance function on G_N.

Second, let R_n be the number of distinct cards obtained in n uniform random draws with replacement from the deck. That is, $R_n = \#\{C_1,\ldots,C_n\}$, where C_i are i.i.d. uniform on $\{1,\ldots,N\}$. Let

$$L_j = \min\{n: R_n = N-j\}$$

be the number of draws needed to get all but some j cards. Then from Feller (1968) pp. 225 and 239

(4.4) COUPON-COLLECTORS LEMMA. *If* $0 \leq \alpha \leq 1$ *and if* $j = j(N)$ *satisfies* $0 < \lim j/N^{\alpha} < \infty$, *then* $\frac{L_j}{N\log(N)} \to 1-\alpha$ *in probability. In particular, for fixed* j *we have* $\frac{L_j}{N\log(N)} \to 1$ *in probability.*

Third, consider again random draws with replacement, and let U be the number of the first draw on which we obtain some previously-drawn card:

$$U = \min\{n: C_n = C_i \text{ for some } i < n\} .$$

(4.5) BIRTHDAY LEMMA. $U/N^{1/2} \xrightarrow{\mathcal{D}} V$, *where* $0 < V < \infty$.

We now describe and analyse some examples. Several of these are in Diaconis (1982).

(4.6) EXAMPLE. "*Top to random*". Here we shuffle by removing the top card and replacing it in a random position in the deck. For a formal description, for $1 \leq j \leq N$ define the permutation π^j by

$$\pi^j(1) = j$$
$$\pi^j(i) = i-1\ , \quad i \le j$$
$$\quad = i\ , \quad i > j.$$

Then the random shuffle is π^J, for J uniform on $\{1,\ldots,N\}$. We shall prove

$$(4.7) \qquad \tau(\varepsilon) \sim N \log(N)\ ; \quad 0 < \varepsilon < 1\ .$$

To analyse this example it is convenient to use the time-reversed process, as discussed in section 3. Here, the time-reversed process is "random to top". That is, a card is chosen uniformly at random, and moved to the top of the deck. To construct a coupling, consider two decks. Choose a label C uniformly from $\{1,\ldots,N\}$ and in each deck move the card labelled C to the top. Plainly this *is* a coupling. The coupling time T is the time at which the decks are completely matched. Now matches, once created, are not destroyed, so at the time L_0 at which each label has been chosen at least once, the decks are completely matched. So

$$d(n) \le P(T>n) \le P(L_0>n)\ .$$

By the Coupon-Collectors Lemma,

$$(4.8) \qquad d(\alpha N \log(N)) \to 0 \text{ as } N \to \infty; \quad \alpha > 1.$$

To get the lower bound, consider the set A_j of states π for which the bottom j cards have increasing labels: that is, $\pi^{-1}(N-j+1) < \pi^{-1}(N-j+2) < \cdots < \pi^{-1}(N)$. Suppose we start with a new deck. Let L_j be the number of shuffles until all but some j labels have been chosen. If $L_j > n$ then the bottom j cards after n "random to top" shuffles have never been chosen to be moved, so remain in their original relative order with increasing labels. So $P(X_n \in A_j) \ge P(L_j > n)$. Since $\Pi(A_j) = 1/j!$,

$$d(n) \ge P(L_j > n) - 1/j!\ .$$

Using the Coupon-Collectors Lemma, we find

$$d(\alpha N \log(N)) \to 1 \quad \text{as} \quad N \to \infty \ ; \quad \alpha < 1.$$

This and (4.8) establish (4.7).

In the example above the coupling is very simple. And in fact the upper bound could be obtained without using coupling, by observing that the order of the already-chosen cards in "random to top" shuffling is uniform. But here is a minor modification for which the coupling argument is equally trivial but where a direct argument seems hard. Diaconis (1982) records that Borel proposed this shuffle.

(4.9) EXAMPLE. "*Top to random, bottom to random*". Here we alternate between picking the top and the bottom card to be removed and replaced at random. Again we get

$$\tau(\varepsilon) \sim N \log(N) \ ; \quad 0 < \varepsilon < 1,$$

using the obvious modifications of the arguments above (for the lower bound, consider the set A_j^* of states for which some j successive cards have increasing labels).

(4.10) EXAMPLE. "*Transposing neighbours*". Here we pick at random a pair of adjacent cards, and transpose them. To eliminate periodicity, we also allow the possibility of doing nothing. Formally, let π^0 be the identity permutation, and π^j the permutation transposing j and $j+1$. Then the random shuffle is π^J, for J uniform on $\{0,\dots,N-1\}$. We shall prove

$$C_1 N^3 \lesssim \tau \lesssim C_2 N^3 \log(N) \tag{4.11}$$

for constants C_1, C_2.

We need first some results about the motion (Y_n) of a single card under this shuffle. This motion is the Markov chain on $\{1,\dots,N\}$ with transitions

$$\begin{aligned} P(j,j+1) &= P(j+1,j) = 1/N \quad && (1 \le j \le N-1) \\ P(j,j) &= 1 - 2/N && (2 \le j \le N-1) \\ P(1,1) &= P(N,N) = 1 - 1/N \ . \end{aligned}$$

This is a symmetric random walk with reflecting boundaries. It is a straightforward exercise in weak convergence theory to show that, suitably normalised, this converges weakly to Brownian motion B_t on $[0,1]$ with reflecting boundaries:

$$N^{-1} Y_{[2t/N^3]} \Rightarrow B_t .$$

The first assertion of the lemma below is now immediate, and the second is not hard.

(4.12) LEMMA. *Let* S_1 *be the number of shuffles until the card initially at the top reaches the* $[N/2]^{th}$ *position (i.e. the middle). Then*

$$S_1/N^3 \xrightarrow{\mathcal{D}} V , \quad \textit{where} \quad V > 0.$$

Let S_2 *be the number of shuffles until the card in an arbitrary initial position reaches the bottom. Then there exist constants* $K,\ \beta > 0,$ *such that*

$$P(S_2/N^3 > s) \leq Ke^{-\beta s} \ ; \quad s \geq 0,\ N \geq 1.$$

Suppose $Y_0 = 1$, and write $d_Y(n)$ for the total variation distance between the distribution of Y_n and uniform. Then

$$\begin{aligned} d_Y(n) &\geq |P(Y_n \leq [N/2]) - [N/2]/N| \\ &\geq P(S_1 > n) - \tfrac{1}{2} \end{aligned}$$

and so

$$d(\alpha N^3) \geq d_Y(\alpha N^3) \gtrsim P(V > \alpha) - \tfrac{1}{2}$$

by the first assertion of Lemma 4.12. For small α the right is greater than $1/2e$, and so we get the lower bound $\tau \gtrsim \alpha N^3$.

To get the upper bound, suppose we can produce a coupling (X_n^1, X_n^2) with the following two properties.

(a) Matches are not destroyed. That is, if $X_n^1(i) = X_n^2(i)$ then $X_m^1(i) = X_m^2(i)$ for $m > n$.

(b) A card in one deck cannot jump over the same card in the other deck. That is, if $X_0^1(i) \geq$ (resp. $\leq$) $X_0^2(i)$ then $X_n^1(i) \geq$ (resp. $\leq$) $X_n^2(i)$

for $n > 0$.

Given such a coupling, the coupling time is $T = \max T_i$, where T_i is the time until the cards labelled i are matched. But by (b) we have $T_i \le S_2^i$, the number of shuffles for the card labelled i to reach the bottom of the deck (in the deck where this card is initially higher). So

$$\begin{aligned} P(T > CN^3\log(N)) &\le \sum_i P(S_2^i/N^3 > C\log(N)) \\ &\le NKe^{-\beta C\log(N)} \qquad \text{by Lemma 3.10} \\ &\to 0 \quad \text{provided} \quad C > 1/\beta, \end{aligned}$$

and then $\tau \underset{\sim}{<} CN^3\log(N)$.

To exhibit a coupling satisfying (a) and (b), consider two decks in states π, σ, Let S be the set of j such that neither the cards in position j are matched nor the cards in position $j+1$ are matched. List S as $\{j_1,\dots,j_L\}$ and add $j_0 = 0$ to S. Let J be uniform on $\{0,\dots,N-1\}$ and define J^* by

$$\begin{aligned} J^* &= J \qquad \text{if } J \notin S \\ &= j_{k+1} \quad \text{if } J = j_k \in S \quad (\text{interpreting } j_{L+1} \text{ as } j_0). \end{aligned}$$

The coupling is produced by applying shuffle π^J to the first deck and π^{J^*} to the second deck. This is a coupling, because J^* is uniform. Property (a) is immediate. And the only way in which (b) could fail is if the same transposition π^j were applied to both decks when the card at position j in one deck had the same label as the card at position $j+1$ in the other deck: and the coupling is designed so this cannot happen.

REMARKS. (a) This shuffle generates a reversible random walk.

(b) The lower bound obtained by considering entropy (3.9) gives $\tau \underset{\sim}{>} CN$ in this example, which is rather crude.

(4.13) EXAMPLE. "*Random transpositions*". Here we shuffle by transposing a randomly chosen pair of cards. To avoid periodicity, we again allow the pair to be identical. For the formal description, let π_{j_1,j_2} be the permutation transposing j_1 and j_2. Then the shuffle is π_{J_1,J_2}, where

J_1 and J_2 are independent, uniform on $\{1,\dots,N\}$. We shall prove

(4.14) $$\tfrac{1}{2}N\log(N) \lesssim \tau \lesssim CN^2 \ ; \quad \text{for some constant } C.$$

Diaconis and Shahshahani (1981) use group representation techniques to analyse this shuffle. From their results one can obtain the precise result

(4.15) $$\tau(\varepsilon) \sim \tfrac{1}{2}N\log(N) \ ; \quad 0 < \varepsilon < 1 \ .$$

To describe the coupling, note that the random shuffle may be described as: pick a label C and a position J at random (independent, uniform), and then transpose the card labelled C with the card at position J. Given two decks in states π, σ, pick C and J and shuffle each deck as described above. Plainly this is a coupling: let (Y_1,Y_2) be the states of the decks after this shuffle. Then $Y_1(C) = Y_2(C) = J$. Thus we see

(a) if neither the cards labelled C were matched, nor the cards at position J were matched, in the decks π, σ, then at least one new match has been created, so $M(Y_1,Y_2) \geq M(\pi,\sigma)+1$;

(b) otherwise the number of matches remains the same, $M(Y_1,Y_2) = M(\pi,\sigma)$.

Now the chance that the event in (a) happens is $(f(\pi,\sigma)/N)^2$, where $f(\pi,\sigma) = N - M(\pi,\sigma)$ is the number of unmatched cards. Let (Z_n^1,Z_n^2) be the coupled process, and $D_n = f(Z_n^1,Z_n^2)$ the number of unmatched cards in the coupled process. By (a) and (b), the process D_n is stochastically dominated by the Markov process D_n^* on $\{0,1,\dots,N\}$ with transition matrix

$$P(i,i-1) = (i/N)^2 \ ; \quad P(i,i) = 1-(i/N)^2 \ .$$

So the coupling time T is at most the first passage time T^* of D_n^* from N to 0. So

$$ET \leq ET^* = \sum_{i=1}^{N} (N/i)^2 \leq CN^2 \ ,$$

and (3.13) gives the upper bound in (4.14).

To get the lower bound, suppose we start with a new deck (state π^0, say). Let L_j be the number of shuffles needed until the j^{th} last card has been picked. By the Coupon-Collectors Lemma, recalling that two cards are picked on each shuffle,

(4.16) $$P(L_j > \alpha N \log(N)) \to 1 \; ; \quad \alpha < \frac{1}{2} .$$

Let A_j be the set of states π for which $\#\{i: \pi(i) = i\} \geq j$. Then $X_n \in A_j$ if $L_j > n$. So

$$\begin{aligned} d(n) &\geq P(X_n \in A_j) - P(X \in A_j) \; ; \text{ where X is uniform on } G_N \\ &\geq P(L_j > n) - P(M(X,\pi^0) \geq j) \end{aligned}$$

and (4.16) and the Card-Matching Lemma give

$$d(\alpha N \log(N)) \to 1 \; ; \quad \alpha < \frac{1}{2} .$$

This establishes the lower bound in (4.14).

REMARKS. (a) This shuffle also is reversible.

(b) For this example the lower bound (3.9) obtained from entropy considerations is $\tau \gtrsim CN$.

(4.17) EXAMPLE. "*Uniform riffle*". We now want to model the riffle shuffle, which is the way card-players actually shuffle cards: by cutting the deck into two roughly equal piles, taking one pile in each hand, and merging the two piles into one. If the top pile has L cards, this gives a permutation π such that

(4.18) $$\pi(1) < \pi(2) < \cdots < \pi(L) \quad \text{and} \quad \pi(L+1) < \pi(L+2) < \cdots < \pi(N) .$$

Call a shuffle satisfying (4.18) for some L a *riffle* shuffle. Such a shuffle can alternatively be described by a 0-1 valued sequence $(b(1),\ldots,b(N))$, where $b(j) = 0$ (resp. 1) indicates that the card at position j after the shuffle came from the top (resp. bottom) pile: formally,

$$\begin{aligned} \pi(1) &= \min\{j: b(j) = 0\} \\ \pi(i) &= \min\{j > \pi(i-1): b(j) = 0\} \; , \quad i \leq L = \#\{j: b(j) = 0\} \\ \pi(L+1) &= \min\{j: b(j) = 1\} \\ \pi(i) &= \min\{j > \pi(i-1): b(j) = 1\} \; , \quad L+1 < i \leq N . \end{aligned}$$

To model a random riffle shuffle we specify some probability measure μ on the set R of riffles. The easiest way is to take μ uniform on R. In terms of the second description, this means we take $(B(1),\dots,B(N))$ to be independent, $P(B(i)=1) = P(B(i)=0) = \frac{1}{2}$. Call this the *uniform* riffle. This process has been investigated in detail by Reeds (1982) (see also Diaconis (1982)), whose technique we shall use to prove

$$(4.19) \qquad \tau(\varepsilon) \sim \frac{3}{2}\log_2 N \ , \quad 0 < \varepsilon < 1 \ .$$

In actual riffle shuffles, successive cards tend to come from alternate piles: see Diaconis (1982), Epstein (1977) for discussion. A more realistic model would be to take $(B(i),\ 1 \le i \le N)$ to be Markov, with transition matrix $P(0,1) = P(1,0) = \theta$, say (Epstein suggests $\theta = 8/9$). The only result known for this model is the lower bound given by entropy (3.9): for fixed θ,

$$\tau(\varepsilon) \gtrsim \frac{(1-\varepsilon)}{\mathcal{E}(\theta)}\log_2 N \quad \text{as} \quad N \to \infty \ ,$$

where $\mathcal{E}(\theta) = -\theta\log_2\theta - (1-\theta)\log_2(1-\theta)$. It is natural to conjecture

$$\tau(\varepsilon) \sim C_\theta \log_2 N \quad \text{as} \quad N \to \infty \quad (\theta,\ \varepsilon \text{ fixed}) \ .$$

But the argument we shall use for the uniform riffle $(\theta = \frac{1}{2})$ does not extend to general θ, for which no reasonable upper bound is known.

The uniform riffle is another example for which it is easier to analyse the time-reversed process. This reversed shuffle can be described as follows. For each c write on the card labelled c the number $B_1(c)$, where $(B_1(c):\ 1 \le c \le N)$ are independent as before; form one pile consisting of the cards with 0 written on them, in their original order, thereby leaving another pile of cards with 1 written on them; and place the first pile on top of the second pile. Imagine now doing this reverse shuffle again with independent numbers $B_2(c)$; this will produce a deck with a sequence of cards on top which have $(B_1,B_2) = (0,0)$, followed by a sequence with $(1,0)$, followed by $(0,1)$, followed by $(1,1)$. Continuing, after n reverse shuffles let $D_n(c) = \sum_{m=1}^{n} 2^{m-1}B_m(c)$, and then

(4.20a) the random variables $(D_n(c): 1 \le c \le N)$ are independent, uniform on $\{0,\dots,2^n-1\}$

(4.20b) the order of the deck is such that D_n is increasing, and cards with identical values of D_n are in their original relative order.

We shall now use this description to get bounds on the total variation distance $d(n)$. We first present a coupling argument for a crude upper bound. Consider two decks, and apply the reverse shuffle to each using the same $(B_m(c))$. Let F_n be the event that the numbers $(D_n(c): 1 \le c \le N)$ are distinct. Then the coupling time T satisfies $T \le n$ on F_n, by (b). So $d(n) \le 1 - P(F_n)$. But the Birthday Lemma shows that $P(F_n) \to 1$ when $n \to \infty$, $N \to \infty$ in such a way that $N/(2^n)^{1/2} \to 0$. Hence $d(\alpha \log_2 N) \to 0$ for $\alpha > 2$, which gives the crude upper bound $\tau(\varepsilon) \lesssim 2 \log_2 N$.

We turn now to the lower bound. For a deck in state π let $\theta(\pi)$ be the number of adjacent pairs of cards with increasing labels:

$$\theta(\pi) = \#\{j: \pi^{-1}(j) < \pi^{-1}(j+1)\} = \sum a_j(\pi) ,$$

where a_j is the indicator function of $\{\pi: \pi^{-1}(j) < \pi^{-1}(j+1)\}$. Consider first X uniform on G_N. Then the random variables $\{a_j(X)$, j even (resp. odd)$\}$ are independent, and we easily get

(4.21) $$E\theta(X) = (N-1)/2 \ ; \quad \text{var } \theta(X) \le N/2 .$$

Now imagine starting with a new deck, and performing n reverse shuffles, leaving the deck in state X_n. Since D_n has at most 2^n distinct values, (b) implies $\theta(X_n) \ge N - 2^n$. From this and (4.21) we can immediately get $\tau(\varepsilon) \gtrsim \log_2 N$. However, a slightly more delicate analysis will improve this bound. We first quote a straightforward variation of the Birthday Lemma.

(4.22) LEMMA. *Let* (C_i) *be independent, uniform on* $\{1,\dots,M\}$. *Let* $U_N = \#\{n \le N: C_n = C_i$ *for some* $i < n\}$. *If* $N \to \infty$ *and* $M \sim N^\alpha$ *for some* $\alpha > 1$ *then*

$$EU_M \sim \tfrac{1}{2}N^{2-\alpha} \ ; \quad \text{var}(U_M) \sim \tfrac{1}{2}N^{2-\alpha} .$$

Let $V_N = \#\{n \leq N: C_n = C_i = C_j$ *for some* $i < j < n\}$. *If* $N \to \infty$ *and* $M \sim N^\alpha$ *for some* $\alpha > 3/2$ *then* $EV_N \to 0$.

Recall X_n is the state of the deck after n reverse shuffles. Let J_n be the (random) set of positions j for which the cards at positions j and $j+1$ after the shuffles have the same value of D_n:

$$J_n = \{j: D_n(X_n^{-1}(j)) = D_n(X_n^{-1}(j+1))\} .$$

Then, conditional on J_n,

(i) $a_j(X_n) = 1$; $j \in J_n$

(ii) the random variables $\{a_j(X_n),\ j \notin J_n,\ j$ even (resp. odd)$\}$ are independent.

From this we can calculate

$$(4.23) \qquad E(\theta(X_n)|J_n) = (N-1)/2 + \tfrac{1}{2}\#J_n \ ; \qquad \mathrm{var}(\theta(X_n)|J_n) \leq N/2 .$$

Now by (a) the distribution of $\#J_n$ is the same as the distribution of U_N in Lemma 4.22, for $M = 2^n$. So, putting

$$n = \alpha \log_2 N \ , \quad \text{for some } 1 < \alpha < \tfrac{3}{2} .$$

Lemma 4.22 gives

$$E\#J_N \sim \tfrac{1}{2}N^\beta \ ; \qquad \mathrm{var}\ \#J_N \sim \tfrac{1}{2}N^\beta \ ; \qquad \tfrac{1}{2} < \beta = 2-\alpha < 1 .$$

So using (4.23)

$$(4.24) \qquad E\theta(X_n) = (N-1)/2 + v_N N^{1/2} \ ; \quad \text{where } v_N \to \infty \ , \quad \mathrm{var}\ \theta(X_n) \lesssim N/2 .$$

Chebyshev's inequality applied to (4.21) and (4.24) gives

$$P(\theta(X) \geq (N-1)/2 + \tfrac{1}{2}v_N N^{1/2}) \to 0$$
$$P(\theta(X_n) \geq (N-1)/2 + \tfrac{1}{2}v_N N^{1/2}) \to 1$$

and so $d(n) \to 1$, giving the lower bound in (4.19).

We shall now return to the upper bound. Fix $\alpha > \frac{3}{2}$, $n = 1 + [\alpha \log_2 N]$, so $2^n \geq N^\alpha$. Let X_n be the state of the deck, described at (4.20), after

n reverse shuffles starting with a new deck. The random variables $(D_n(c) : 1 \le c \le N)$ define a random partition $\mathcal{A}$ of the shuffled deck into sets consisting of the positions of cards with common values of D_n. Thus if the numbers $D_n(c)$ are $(15,2,8,15,15,2)$, then when put in increasing order, they become $(2,2,8,15,15,15)$, and this defines the partition $\{1,2\}$, $\{3\}$, $\{4,5,6\}$. Denote a partition by $A = \{A_1, A_2, \ldots\}$, and let $|A|_r$ be the number of sets with exactly r elements in the partition A. Let P be the set of partitions consisting only of singletons and consecutive pairs. Using Lemma 4.22

(4.25) $$E|\mathcal{A}|_2 \lesssim N^{2-\alpha} \quad \text{as} \quad N \to \infty .$$

(4.26) $$P(\mathcal{A} \in P) \to 1 \quad \text{as} \quad N \to \infty .$$

And by conditioning on the set of distinct values taken by $(D_n(c) : 1 \le c \le N)$, we obtain

(4.27) for $A \in P$ the probability $P(\mathcal{A} = A)$ depends only on $|A|_2$.

Now for $m \ge 0$ let $W_1, \ldots, W_m$ be i.i.d. uniform on $\{1, \ldots, N-1\}$, and let $\mathcal{A}_m^*$ be the collection of sets $\{W_j, W_j+1\}$, $1 \le j \le m$. If these sets are disjoint, extend $\mathcal{A}_m^*$ to a partition by including the remaining elements of $\{1, \ldots, N\}$ as singletons. Given that $\mathcal{A}_m^*$ is such a partition, it is plainly distributed uniformly over the partitions $A \in P$ with $|A|_2 = m$. So by (4.27)

(4.28) $$\begin{aligned} P(\mathcal{A} = A \,\big|\, |\mathcal{A}|_2 = m, \mathcal{A} \in P) &= P(\mathcal{A}_m^* = A \mid \mathcal{A}_m^* \text{ is a partition}) \\ &\ge P(\mathcal{A}_m^* = A) , \quad A \in P . \end{aligned}$$

Now let π be a state of the deck, and as before let $\theta(\pi)$ be the number of successive pairs with increasing labels. Say a partition $A = \{A_1, A_2, \ldots\}$ is *consistent* with π if π^{-1} is increasing on each A_j. Fix γ, β such that $\gamma > \frac{1}{2} > \beta > 2-\alpha$, $\gamma + \beta < 1$.

(4.29) LEMMA. $P(\mathcal{A}_m^*$ *is some partition consistent with* $\pi) \ge$

$$\left(\tfrac{1}{2}\right)^m \{1 - \psi(\tfrac{\theta(\pi) - N/2}{N^\gamma}, \tfrac{m}{N^\beta}, N)\}$$

where $\psi(x,y,N) \to 0$ *as* $x \to 0$, $y \to 0$, $N \to \infty$.

PROOF. Given that the pairs $\{W_1, W_1+1\}, \ldots, \{W_{i-1}, W_{i-1}+1\}$ are disjoint and that π^{-1} is increasing on each, there are at least $\theta(\pi) - 3(i-1)$ choices for W_i which have π^{-1} increasing on $\{W_i, W_i+1\}$ and $\{W_i, W_i+1\}$ disjoint from the previous pairs. So

$$P(A_m^* \text{ is some partition consistent with } \pi) \geq \prod_{i=1}^{m} \left\{\frac{\theta(\pi)-3(i-1)}{N-1}\right\}$$

$$\geq \left(\frac{1}{2}\right)^m \prod_{i=1}^{m} \{1 + 2xN^{\gamma-1} - 6iN^{-1}\}, \quad \text{where } x = \frac{\theta(\pi) - N/2}{N^\gamma}.$$

Calculus shows the product tends to 1 as $N \to \infty$, $x \to 0$, $m/N^\beta \to 0$.

We can express the distribution of X_n by conditioning on the partition $\mathcal{A}$, using description (4.20), as

$$N!P(X_n = \pi) = \sum_A P(\mathcal{A} = A)(2!)^{|A|_2}(3!)^{|A|_3}\cdots 1_{(A \text{ consistent with } \pi)}$$

$$\geq \sum_A P(\mathcal{A} = A, \mathcal{A} \in \mathcal{P}) 2^{|A|_2} 1_{(A \text{ consistent with } \pi)}$$

$$= P(\mathcal{A} \in \mathcal{P}) \sum_m P(|\mathcal{A}|_2 = m \mid \mathcal{A} \in \mathcal{P}) 2^m \sum_A P(\mathcal{A} = A \mid |\mathcal{A}|_2 = m, \mathcal{A} \in \mathcal{P}) \cdot 1_{(A \text{ consistent with } \pi)}$$

$$\geq P(\mathcal{A} \in \mathcal{P}) \sum_m P(|\mathcal{A}|_2 = m \mid \mathcal{A} \in \mathcal{P}) \left(1 - \psi\left(\frac{\theta(\pi) - N/2}{N^\gamma}, \frac{m}{N^\beta}, N\right)\right), \tag{4.30}$$

by (4.28) and Lemma 4.29.

By (4.21) we can find $\varepsilon_N \to 0$ such that the set F_N of states π such that $\left|\frac{\theta(\pi) - N/2}{N^\gamma}\right| \leq \varepsilon_N$ satisfies $\#F_N/N! \to 1$.

By (4.25) we can find $\delta_N \to 0$ such that $P(|\mathcal{A}|_2 \leq \delta_N N^\beta) \to 0$. Applying these observations and (4.26) to (4.30) we obtain

$$N!P(X_n = \pi) \geq 1 - \lambda_N, \quad \pi \in F_N, \quad \text{where } \lambda_N \to 0.$$

So the total variation distance $d(n)$ between X_n and the uniform distribution satisfies $d(n) \leq \lambda_N + (1 - \frac{\#F_N}{N!}) \to 0$, establishing the upper bound in (4.19).

(4.31) EXAMPLE. "*Overhand shuffle*". Here is an example of a random shuffle for which no good upper bound for τ is known. Overhand shuffling is where the deck is divided into a number of blocks, and the order of the blocks is reversed. To make a model, let $2 \le K \le N/2$ be a parameter which will represent the mean length of the blocks. Let $(V_i: 1 \le i < N)$ be independent, $P(V_i = 1) = 1/K$, and let $V_0 = V_N = 1$. Let

$$J_1 = 0 \ ; \quad J_i = \min\{j > J_{i-1}: V_j = 1\} \ ; \quad B_i = \{j: J_i < j \le J_{i+1}\} \ .$$

Then B_i represents the i^{th} block, and the random shuffle is:

$$\pi(j) = (N-J_{i+1}) + (j-J_i) \ ; \qquad j \in B_i \ .$$

The only result known is the following lower bound, whose proof we shall merely indicate.

$$\tau \ge C \max(K,(N/K)^2) \ ; \quad \text{some constant } C \ .$$

Note that the right side is minimized by $K = N^{2/3}$, for which $\tau \ge CN^{2/3}$.

First, consider two cards which are initially adjacent. On each shuffle, the chance they are separated is at most $2/K$, and this leads to the inequality $\tau \ge CK$. Second, consider the motion Y_n of a particular card after n shuffles, where we measure its position from the top for even n and from the bottom for odd n. Then Y_n is a Markov process on $\{1,\dots,N\}$ which, away from 1 and N, is approximately a random walk whose increments have mean 0 and standard deviation $2^{1/2}K$. It can be shown that Y_n has standard deviation at most $CKn^{1/2}$, and this leads to the other inequality.

REMARK. One would like to conjecture that for any "reasonable" way of shuffling cards, τ is at most polynomial in N. But it is not clear what "reasonable" means. Note that for our applications to hitting times, we only need τ small compared to $N!$

5. Rapidly mixing Markov chains

In this section we mention a few Markov chain examples, and discuss informally the "rapid mixing" property.

(5.1) EXAMPLE. "*Ehrenfest urn model*". We discuss the continuous-time version, which is the Markov process Y_t on $\{0,1,\dots,N\}$ with transition rates

$$Q(i,i+1) = 1 - i/N \ , \quad Q(i,i-1) = i/N \ .$$

Think of N balls distributed among two boxes, with a Poisson (rate 1) process of selections of balls chosen uniformly at random and transferred to the other box; Y_t describes the number of balls in a particular box at time t. Now we can represent Y_t as $f(X_t)$, where X_t is the random walk on the N-dimensional cube (Example 3.19), and $f(i_1,\dots,i_N) = \sum i_r$. In fact, the random walk describes the process of balls in boxes where the balls are labelled $1,\dots,N$, and state $i = (i_1,\dots,i_N)$ indicates that ball r is in box i_r.

From this representation we see that the stationary distribution π for Y is the Binomial $(N,\frac{1}{2})$ distribution. And $d(t)$ is the same for Y as for X, so

$$\tau(\varepsilon) \sim \tfrac{1}{4}N \log(N) \tag{5.2}$$

by (3.23).

(5.3) EXAMPLE. "*Random subsets*". Let $1 \le M < N$, $N \ge 3$, and let $\mathcal{B}$ be the set of all subsets B of $\{1,2,\dots,N\}$ with $\#B = M$. Consider a random subset B evolving by elements being deleted and replaced by outside elements. Formally, consider the $\mathcal{B}$-valued process X_t with transition rates

$$\begin{aligned} Q(B,B') &= \frac{1}{M(N-M)} \ ; \quad \#(B \cap B') = M-1 \\ &= 0 \qquad\quad ; \quad \text{other } B' \neq B. \end{aligned}$$

The stationary distribution is uniform on $\mathcal{B}$. The reader may like to construct a coupling argument similar to that of Example 3.19 to show

$$\tau \lesssim CN \log(1+\min(M,N-M)) \quad \text{as } N \to \infty; \text{ for some constant } C. \tag{5.4}$$

(5.5) EXAMPLE. "*Sequences in coin-tossing*". Let (ξ_i) be independent, $P(\xi_i = H) = 1/2$, $P(\xi_i = T) = 1/2$, representing repeated tosses of a fair coin. For fixed $N \geq 1$ the process $X_n = (\xi_{n+1}, \ldots, \xi_{n+N})$ is a Markov chain on $\{H,T\}^N$. For this chain the stationary distribution is uniform and

(5.6) $$d(n) = 1 - (\tfrac{1}{2})^{N-n}, \quad 0 \leq n \leq N$$
$$= 0, \quad n \geq N.$$

(5.7) EXAMPLE. "*Random walk in a* d-*dimensional box*". We want to consider the random walk on the d-dimensional integers restricted to a box of side N by boundaries. Formally, let $G = \{i = (i_1, \ldots, i_d): 0 \leq i_r < N\}$ and consider the Markov chain with transition matrix

$$P(i,j) = 1/(2d+1) \quad \text{for } \sum |i_r - j_r| = 1;$$
$$= 0 \quad \text{for other } j \neq i;$$
$$P(i,i) = 1 - \sum_{j \neq i} P(i,j) .$$

(We use $1/(2d+1)$ instead of $1/2d$ to avoid periodicity problems.) The stationary distribution is uniform, and using the Central Limit Theorem we see

(5.8) $$\tau \sim C_d N^2 \text{ as } N \to \infty; \text{ for fixed } d.$$

(5.9) EXAMPLE. "*Rubic's cube*". One may consider the random walk on Rubic's cube obtained by choosing one of the 27 possible rotations at random at each step. It would be interesting to estimate τ for this random walk. Perhaps one of the algorithms to "solve" (i.e. reach a specific state of) the cube could be used to construct a coupling. But this seems difficult.

We now introduce informally the "rapid mixing" proeprty. For a discrete-time random walk, this is the property

(5.10) $$\tau \text{ is small compared to } \#G .$$

The intuitive idea here is that the distribution of the chain approaches stationarity while only a small proportion of states have been visited. For the general discrete-time chain, we measure "proportion of states" using the stationary distribution, and so formulate the rapid mixing property as

(5.11) τ is small compared to $\min(1/\pi(i))$.

For continuous-time processes we must take into account the rate at which transitions occur. Recall $q_i = \sum_{j \neq i} Q(i,j)$ is the rate of leaving state i. In the general Markov case the rapid mixing property becomes

(5.12) τ is small compared to $\min(1/\pi(i)q_i)$.

Recall (2.7) that in the random walk case we normalize to make $q_i \equiv 1$, so then (5.12) is the same as (5.10).

Almost all the examples mentioned have this rapid mixing property. It is particularly noticeable in the card-shuffling examples, where $\#G = N!$ but τ is at most polynomial in N. An exception is the random walk in the d-dimensional box for $d = 1$ or 2. Indeed, it is easy to see that the familiar examples of 1-dimensional Markov processes do *not* have the rapid mixing property. For instance, consider the single server queue process on $\{0,1,\dots,N\}$, with transition rates

$$Q(i,i-1) = 1 \; ; \quad Q(i,i+1) = \lambda < 1 \; ;$$

and stationary distribution $\pi(i) = \lambda^i(1-\lambda)/(1-\lambda^{N+1})$. Very roughly, τ must be of the same order as the passage time from N to 0, which is of order $N/(1-\lambda)$: to put it another way, the process starting at N must pass through most states before approaching the stationary distribution.

We thus have a curious paradox: the rapid mixing property, which we use in the sequel to get approximations for hitting times, seems characteristic of complicated high-dimensional processes rather than simple one-dimensional processes. A possible explanation is that rapid mixing is a kind of "local transience" property, and we recall that mean zero random walks are transient only in three or more dimensions. This analogy is pursued a little in the next section.

6. The mean occupation function

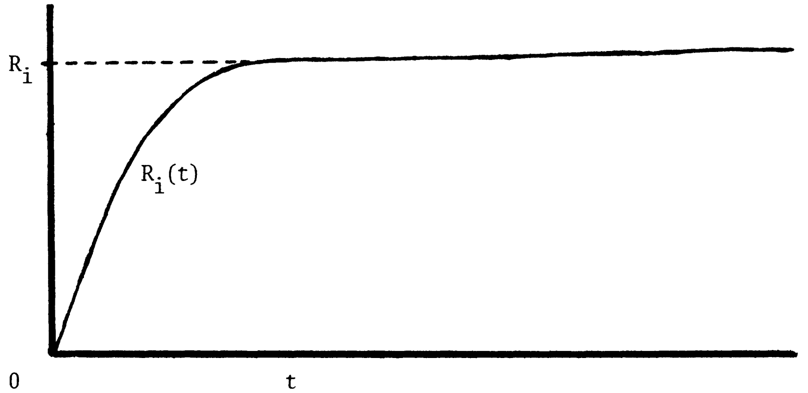
In this section we discuss the mean occupation function $R_i(t)$, which plays a major role in the behaviour of rapidly mixing Markov processes. For a Markov process (X_t) and a state i define

$$(6.1) \qquad R_i(t) = \int_0^t p_{i,i}(s)\,ds = E_i \text{time}(s \le t: X_s = i)$$

where $\text{time}(s \le t: X_s = i)$ is the random variable indicating the amount of time X spends at state i before time t.

In the next paragraph we describe informally the behaviour of $R_i(t)$ in a rapidly mixing process: the rest of the section contains lemmas formalizing these assertions.

The function $R_i(t)$ looks roughly like



That is, $R_i(t)$ initially tends to increase to a value which must be at least $1/q_i$, the mean length of the initial sojourn in i (6.2). It then starts to level off, and remains essentially constant over the interval of t large compared to τ but small compared to $1/\pi(i)q_i$ (6.5). So we can define a parameter R_i as the approximate value of $R_i(t)$ on this interval. Interpret R_i as the mean length of time (mean number of visits, in the random walk case) spent at i in the short term. For another interpretation, recall that in the infinite state space setting the condition $R_i(\infty) < \infty$ is equivalent to transience, and then $R_i(\infty) = 1/q_i(1-\rho_i)$, where ρ_i is the probability of return to i. Analogously, in the rapid mixing case we may think of R_i as approximately $1/q_i(1-\rho_i^*)$, where ρ_i^* is the probability of return to i in the short term (6.17). In particular, if ρ_i^* is close to 0 then R_i is close to $1/q_i$. Finally, note that in the random walk case R_i and $R_i(t)$ are quantities R and $R(t)$ not dependent on i.

We now start the formalities. First, $P_{i,i}(s) \geq e^{-q_i s}$, so

$$(6.2) \qquad R_i(t) \geq \int_0^t e^{-q_i s} ds = q_i^{-1}(1-e^{-q_i t}) \ .$$

Second, by integrating the inequality $|p_{i,i}(s)-\pi(i)| \leq d(s)$ we get, for $t_1 \leq t_2$,

$$(6.3) \qquad |R_i(t_2) - R_i(t_1) - (t_2-t_1)\pi(i)| \leq \int_{t_1}^{\infty} d(s)ds$$
$$\leq \tau \exp(1-t_1/\tau) \quad \text{by (3.8).}$$

So the limit

$$(6.4) \qquad R_i = \lim_{t\to\infty} R_i(t) - t\pi(i)$$

exists and is finite. This quantity occurs in the traditional analytic treatment of Markov process theory; one reason for its significance will become clear in the next section. To compute R_i directly would require knowing $p_{i,i}(t)$, which is rarely available explicitly in practice. But by (6.3) we see

$$(6.5) \qquad |R_i - R_i(t)| \leq t\pi(i) + \tau \exp(1 - t/\tau) \ .$$

If t is large compared to τ then the second term on the right is small; if t is small compared to $1/\pi(i)q_i$ then the first term on the right is small compared to $1/q_i$; in the rapidly mixing case we can find t satisfying both these conditions and then $R_i(t)$ approximates R_i. This is the informal description of R_i given earlier. Specifically, from (6.5) we get

$$(6.6) \qquad |R_i - R_i(\tau_i^*)| \leq 2\tau_i^*\pi(i) \ ; \quad \text{for} \quad \tau_i^* = \tau(1 - \log \pi(i))$$

in general; and in the random walk case

$$(6.7) \qquad |R - R(\tau^*)| \leq 2\tau^*/\#G \ ; \quad \text{for} \quad \tau^* = \tau(1 + \log \#G) \ .$$

In Section 8 we shall see examples where R is estimated in this way.

Our informal discussion earlier suggested that for rapidly mixing processes, R_i should not be much smaller than $1/q_i$. Lemma 6.8 formalizes this idea. To state such a result we introduce a notational device, to be used extensively in the next section. Call a function $\psi(x) \geq 0$, $x \geq 0$ *vanishing* if $\psi(x) \to 0$ as $x \to 0$, and adopt the convention that a function asserted to be vanishing is a "universal" function, that is to say the function does not depend on the particular process under consideration. The symbol ψ will denote different functions in different assertions.

(6.8) LEMMA. $R_i \geq q_i^{-1}\{1 - \psi(q_i\tau\pi(i)\log(1+q_i\tau))\}$, *for some vanishing* ψ.

Specializing to the random walk case,

$$R \geq 1 - \psi(\tfrac{\tau}{\#G}\log(1+\tau)) \ . \tag{6.9}$$

Results like this could equivalently be formulated as limit theorems for sequences of processes. For instance, Lemma 6.8 is equivalent to:

Let X^n *be processes on state spaces* G^n; *let* $i^n \in G^n$; *suppose*

$$q^n_{i_n}\tau^n\pi^n(i_n)\log(1+q^n_{i_n}\pi^n(i_n)) \to 0 \ ;$$

then $R^n_{i_n} \gtrsim 1/q^n_{i_n}$.

Both formulations have the same interpretation:

> If $\pi(i)$ is small compared to $1/q_i\tau\log(1+q_i\tau)$ then R_i is not much less than $1/q_i$.

The formulation involving vanishing functions seems to convey this idea more directly.

PROOF OF LEMMA 6.8. By (6.2) and (6.5)

$$R_i \geq q_i^{-1}(1-e^{-q_i t}) - t\pi(i) - \exp(1-t/\tau) \ ; \quad t \geq 0 \ . \tag{6.10}$$

We want to evaluate this at a time t_0 which is large compared to $\tau\log(1+q_i\tau)$

but small compared to $1/q_i\pi(i)$. To do so, define

$$\alpha = \pi(i)q_i\log(1+q_i\tau) \tag{6.11}$$
$$t_0 = \alpha^{-1/2}\tau\log(1+q_i\tau) = \alpha^{1/2}/q_i\pi(i)\ .$$

Note that $d(t) \geq p_{i,i}(t) - \pi(i) \geq e^{-q_it} - \frac{1}{2}$ by assumption (2.8), so the definition of τ gives

$$q_i\tau \geq c = -\log(\tfrac{1}{2}+\tfrac{1}{2e}) > 0\ . \tag{6.12}$$

Evaluating (6.10) at $t = t_0$,

$$q_iR_i \geq 1 - \exp(-\alpha^{1/2}c\log(1+c)) - \alpha^{1/2} - eq_i\tau(1+q_i\tau)^{-\alpha^{-1/2}}\ . \tag{6.13}$$

Now each of the functions

$$\begin{aligned}&\exp(-\alpha^{1/2}c\log(1+c))\\ &\alpha^{1/2}\\ &\sup_{y>0} y(1+y)^{-\alpha^{-1/2}}\end{aligned} \tag{6.14}$$

is a vanishing function of α, and the result follows.

We remark that for non-rapidly mixing processes there is the weaker lower bound

$$R_i \geq \frac{1}{2q_i} \tag{6.15}$$

which cannot be improved: see Section 7. Of course for non-rapidly mixing processes, R_i does not have the intuitive meaning described earlier. We also remark that for reversible processes (6.15) can be improved. In a reversible process the function $p_{i,i}(t)$ is decreasing (in fact, completely monotone: Keilson (1979)). So

$$\begin{aligned}R_i &= \lim \int_0^t \{p_{i,i}(s)-\pi(i)\}ds\\ &\geq \int_0^{t_0}\{e^{-q_is}-\pi(i)\}ds\ , \quad \text{where } e^{-q_it_0} = \pi(i)\\ &\geq q_i^{-1}\{1-\pi(i)(1-\log\pi(i))\} \quad (X_t \text{ reversible}).\end{aligned} \tag{6.16}$$

Let F_i be the distribution function of the first return to i:

$$F_i(t) = P_i(T_i^+ \le t) .$$

(6.17) LEMMA. $|R_i q_i(1 - F_i(\tau^*)) - 1| \le \psi(\frac{\tau^* q_i \pi(i)}{1-F_i(\tau^*)})$ *where* ψ *is vanishing and* $\tau^* = \tau(1 - \log \pi(i))$.

In other words, for rapidly mixing processes we can approximate R_i by $1/q_i(1-F_i(\tau^*))$, as discussed informally earlier.

As a preliminary, we need

(6.18) LEMMA. (a) $q_i R_i(t) \le \frac{1}{1-F_i(t)}$

(b) $q_i R_i(n(t+1/q_i)) \ge (1-e^{-n})\frac{1-\{F_i(t)\}^{n+1}}{1-F_i(t)}$, $n \ge 1$.

PROOF. Let $X_0 = i$. Let $[U_n, V_n)$ be the n^{th} sojourn interval at i. Then

$$\begin{aligned} R(t) &\le \sum_{n\ge1} E(V_n - U_n) 1_{(U_n \le t)} \\ &= \sum_{n\ge1} q_i^{-1} P(U_n \le t) \\ &\le q_i^{-1} \sum_{n\ge1} P(U_m - U_{m-1} \le t;\ 1 < m \le n) \\ &= q_i^{-1} \sum_{n\ge1} \{F_i(t)\}^{n-1} \\ &= q_i^{-1}\{1 - F_i(t)\}^{-1}, \text{ giving (a).} \end{aligned}$$

To prove (b),

$$\begin{aligned} q_i R_i(n(t+1/q_i)) &\ge q_i E \sum_{m=1}^{n+1} \{(V_m - U_m) \wedge n/q_i\} 1_{(U_m \le (m-1)t)} \\ &\ge E(Z \wedge n) \sum_{m=1}^{n+1} P(U_r - U_{r-1} \le t,\ 1 < r \le m) \end{aligned}$$

where Z has exponential (1) distribution,

$$= (1-e^{-n}) \sum_{m=1}^{n+1} \{F_i(t)\}^{m-1}, \text{ giving (b).}$$

PROOF OF LEMMA 6.17. By Lemma 6.18(a) and (6.6),

$$q_i R_i - (1-F_i(\tau^*))^{-1} \le 2 q_i \tau^* \pi(i) ,$$

giving one side of the inequality. For the other, write $\alpha = q_i\tau^*\pi(i)(1-F_i(\tau^*))^{-1}$ and let n be the integer part of

$$\alpha^{-1/2}(1-F_i(\tau^*))^{-1} = \alpha^{1/2}(q_i\tau^*\pi(i))^{-1} .$$

Note $n \geq \{\psi(\alpha)\}^{-1}$, for some vanishing ψ. Setting $t_1 = n(\tau^* + 1/q_i)$, Lemma 6.18(b) gives

$$(6.19)\qquad \begin{aligned}(1 - F_i(\tau^*))q_iR_i(t_1) &\geq (1-e^{-n})(1 - \{F_i(\tau^*)\}^{n+1}) \\ &\geq 1 - \psi(\alpha) ,\end{aligned}$$

using the fact that $n(1-F_i(\tau^*)) \geq \{\psi(\alpha)\}^{-1}$ for some vanishing ψ. Finally, by (6.5) $q_i(R_i(t_1)-R_i) \leq \theta_1 + \theta_2$, say, where

$$\begin{aligned}\theta_1 &= q_i\tau^*\pi(i) \leq \alpha \\ \theta_2 &= q_i\tau^*\exp(1 - t_1/\tau) \\ &\leq q_i\tau^*e\ \exp(-n\tau^*/\tau) \\ &\leq \frac{\alpha}{\pi(i)}e\{\pi(i)\}^n \leq \psi(\alpha)\end{aligned}$$

which with (6.19) establishes the lower bound in Lemma 6.17.

Lemma 6.17 implies that if the process started at i is unlikely to return in the short term, then R_i should be about $1/q_i$. Our final two lemmas in this section give upper bounds in this situation. The first is applicable if the transition rates into i from other states are all small.

(6.20) LEMMA. $q_iR_i \leq 1 + \psi(\alpha)$, *where* $\alpha = (q^* + q_i\pi(i))\,\tau\log(1+q_i\tau)$ *and* $q^* = \max_{j\neq i} q_{j,i}$.

PROOF. Set $t_2 = \alpha^{-1/2}\,\tau\log(1+q_i\tau)$, so $t_2 \leq \alpha^{1/2}/q^*$ and $t_2 \leq \alpha^{1/2}/q_i\pi(i)$. Since the rate of return to i is at most q^*, we have $F_i(t) \leq q^*t$. By Lemma 6.18(a),

$$q_iR_i(t_2) \leq (1-q^*t_2)^{-1} \leq (1-\alpha^{1/2})^{-1} \leq 1 + \psi(\alpha) .$$

And by (6.5)

$$q_iR_i - q_iR_i(t_2) \le q_i t_2 \pi(i) + q_i\tau \exp(1 - t_2/\tau)$$
$$\le \alpha^{1/2} + q_i\tau e(1+q_i\tau)^{-\alpha^{-1/2}} \le \psi(\alpha) .$$

The final lemma is applicable when there is a distance function f such that $f(X_t,i)$ tends to increase away from $X_0 = i$.

(6.21) LEMMA. *Let* f *be a distance function on* G. *Let* $0 < s < 1$. *Suppose* c *is a constant such that for each* $j \ne i$,

$$c \ge \sum_{k:\ k\ne j} \{s^{f(k,i)} - s^{f(j,i)}\} q_{j,k} .$$

Then $q_iR_i(t) \le \{1 - (s+ct)\}^{-1}$, $0 \le t < (1-s)/c$.

PROOF. Fix i. Consider the process

$$Y_t = s^{f(X(t\wedge T_i),i)} - ct .$$

The definition of c ensures that Y_t is a supermartingale. So for $j \ne i$, $E_jY_t \le E_jY_0 = s^{f(j,i)} \le s$. But $f(X(t\wedge T_i),i) = 0$ on $\{T_i \le t\}$, so $E_jY_t \ge P_j(T_i \le t) - ct$. This implies

$$P_j(T_i \le t) \le s + ct ; \quad j \ne i.$$

Hence $F_i(t) \le s + ct$, and the result follows from Lemma 6.18(a).

7. Hitting times

Mean hitting times E_iT_j, and more generally hitting distributions, have been studied for many years, but there is no single method which yields tractable results in all cases. Kemeny and Snell (1959) give elementary matrix results; Kemperman (1961) presents an array of classical analytic techniques. Our purpose is to give approximations which are applicable to rapidly mixing processes. Keilson (1979) gives a different style of approximation which seems applicable to different classes of processes.

We first give two well-known exact results, which concern the case of hitting a single state from the stationary initial distribution.

(7.1) PROPOSITION. $E_\pi T_i = R_i/\pi(i)$

In the random walk case, $E_\pi T_i = R\#G$.

(7.2) PROPOSITION. $P_\pi(T_i \in dy) = q_i\pi(i)(1-F_i(y))$

Proposition 7.1 is useful because it shows we can estimate $E_\pi T_i$ by estimating R_i. Proposition 7.2 is less useful, because estimating $F_i(y)$ in practice may be hard. We shall give "probabilistic" proofs, quoting renewal theory. First, a lemma about reward renewal processes. Informally, if you are paid random amounts of money after random time intervals, then your long-term average income per unit time should be
E(money paid per interval)/E(duration of interval).

(7.3) LEMMA. *Let* (V_n,W_n), $n \geq 1$, *be positive random variables. Let* $Z(t)$ *be an increasing process such that* $Z(\sum_1^n V_i) = \sum_1^n W_i$.

(a) *If* (V_n,W_n), $n \geq 1$, *are i.i.d. and* $EV_1 = v$, $EW_1 = w$, *then* $\lim t^{-1}Z(t) = w/v$ *a.s.*

(b) *Suppose* $\sup EW_n^2 < \infty$, $\sup EV_n^2 < \infty$, *and there exist constants* v, w *such that* $E(V_n|F_{n-1}) \leq v$, $E(W_n|F_{n-1}) \geq w$ *for all* n, *where* $F_n = \sigma(V_m,W_m;\ m \leq n)$. *Then* $\liminf t^{-1}Z(t) \geq w/v$ *a.s.*

PROOF. In case (a), the strong law of large numbers says that a.s.

$$\bar{V}_n = n^{-1}\sum_1^n V_i \to v\ ,\quad \bar{W}_n = n^{-1}\sum_1^n W_i \to w\ ,\quad \bar{V}_{n+1}-\bar{V}_n \to 0\ ,$$

and the result follows easily. In case (b) we can use the strong law for square-integrable martingales (Stout (1974) Theorem 3.3.1) to show that a.s.

$$\limsup \bar{V}_n \leq v\ ,\quad \liminf \bar{W}_n \geq w\ ,\quad \bar{V}_{n+1}-\bar{V}_n \to 0\ ,$$

and again the result follows easily.

PROOF OF PROPOSITION 7.1. Fix i, $t_1 > 0$, let $\rho(\cdot) = P_i(X_{t_1} \in \cdot)$ and let

$$U_1 = \min\{t:\ X_t = i\}$$
$$U_n = \min\{t \geq U_{n-1}+t_1:\ X_t = i\}\ .$$

Let Y^n be the block of X over the interval $[U_n, U_{n+1}]$; that is,

$$Y^n_s = X_{U_n+s}\ , \quad 0 \le s < U_{n+1} - U_n\ .$$

The blocks (Y^n), $n \ge 1$, are i.i.d. So we can apply Lemma 6.3(a) to

$$V_n = U_{n+1} - U_n$$
$$W_n = \text{time}(s\colon U_n \le s < U_{n+1},\ X_s = i)$$
$$Z(t) = \text{time}(s\colon U_1 \le s < t,\ X_s = i)$$

and the lemma shows

(7.4) $$\lim t^{-1} Z(t) = EV_1/EW_1 \quad \text{a.s.}$$

Now $EV_1 = t_1 + E_\rho T_i$, $EW_1 = R_i(t_1)$, and $\lim t^{-1}Z(t) = \pi(i)$. Substituting into (7.4) and rearranging,

(7.5) $$E_\rho T_i = \{R_i(t_1) - \pi(i)t_1\}/\pi(i)\ .$$

Letting $t_1 \to \infty$, we have $\|\rho - \pi\| \to 0$, so $E_\rho T_i \to E_\pi T_i$, and the result follows. ∎

PROOF OF PROPOSITION 7.2. Let $X_0 = i$. Let $S_0 = 0$,

$$S_n = \text{time of } n^{th} \text{ return to } i$$
$$Y(t) = \min\{S_n - t\colon S_n \ge t\}\ .$$

Then $Y(t)$ has distribution $P_{\rho_t}(T_i \in \cdot)$, where $\rho_t = P_i(X_t \in \cdot)$. So $Y(t) \xrightarrow{\mathcal{D}} P_\pi(T_i \in \cdot)$ as $t \to \infty$.

But (S_n) are the epochs of a renewal process with inter-renewal distribution $P_i(T_i^+ \in \cdot)$, and for such a process (Karlin and Taylor (1975)) we have

$$Y(t) \xrightarrow{\mathcal{D}} Y\ ,$$

where $P(Y \in dy) = P_i(T_i^+ \ge y)/E_i T_i^+$.

The result follows from (2.10).

We can deduce a useful lower bound.

(7.6) COROLLARY. $E_\pi T_i \geq (2q_i\pi(i))^{-1}$.

PROOF. Fix $c > 0$. Consider the class C of distributions on $[0,\infty)$ which have a decreasing density $f(t)$ with $f(0) = c$. The distribution in C with minimal mean is plainly the distribution uniform on $[0,c^{-1}]$. So every distribution in C has mean at least $(2c)^{-1}$. The result now follows from Proposition 7.2.

In view of Proposition 7.1, the Corollary is equivalent to

(7.7) $$R_i \geq 1/2q_i \ .$$

Inequalities (7.6) and (7.7) cannot be improved, even for the random walk case: consider the cyclic motion $Q(0,1) = Q(1,2) = \cdots = Q(N-1,N) = Q(N,0) = 1$. Of course, in the rapidly mixing case R_i is essentially at least $1/q_i$ by Lemma 6.8.

We now start the approximation results. The first says that for rapidly mixing processes the exact value $R_i/\pi(i)$ of the mean hitting time on i from the stationary distribution is an approximate upper bound for the mean hitting time from an *arbitrary* initial distribution.

(7.8) PROPOSITION. *For any state* i *and any initial distribution* ν,

$$E_\nu T_i \leq \frac{R_i}{\pi(i)}\{1 + \psi(q_i\pi(i)\tau)\}$$

where ψ *is vanishing.*

In the random walk case, this says $E_\nu T_i \leq R\#G\{1 + \psi(\tau/\#G)\}$. In words, when τ is small compared to $\#G$ then the mean hitting time on a state from any other state cannot be much more than $R\#G$.

We need the following lemma.

(7.9) LEMMA. *Fix* t, *and let* $\rho_i = P_i(X_t \in \cdot)$. *Then*

$$\max_i E_i T_A \leq t + \max_i E_{\rho_i} T_A \leq (1-d(t))^{-1}(t + E_\pi T_A) \ .$$

PROOF. First recall

(7.10) $$|E_\rho T_A - E_\pi T_A| \le \|\rho-\pi\| \max_j E_j T_A .$$

So

(7.11) $$E_{\rho_i} T_A \le E_\pi T_A + d(t) \max_j E_j T_A .$$

But obviously $E_i T_A \le t + E_{\rho_i} T_A$ (giving the first inequality), so $\max_i E_i T_A \le t + E_\pi T_A + d(t) \max_j E_j T_A$ by (7.10). Rearranging,

$$\max_i E_i T_A \le (1-d(t))^{-1}(t+E_\pi T_A) .$$

Substituting into (7.11) gives the second inequality.

PROOF OF PROPOSITION 7.8. By Lemma 7.9,

$$\frac{E_\nu T_i}{E_\pi T_i} \le (1-d(t))^{-1}(1 + t/E_\pi T_i) , \quad t > 0 .$$

So by Proposition 7.1 and Corollary 7.6,

$$E_\nu T_i \cdot \pi(i)/R_i \le (1-d(t))^{-1}(1+2q_i t\pi(i)) , \quad t > 0 .$$

Evaluating the right side at t large compared to τ, small compared to $1/q_i\pi(i)$, we see from (3.8) that the right side is at most $1 + \psi(\tau q_i \pi(i))$.

Consider for fixed i how the mean hitting times $E_j T_i$ vary with j. Proposition 7.1 says that the π-average of these hitting times is $R_i/\pi(i)$; Proposition 7.8 says that each $E_j T_i$ is not much more than $R_i/\pi(i)$; these imply that $E_j T_i$ must be approximately equal to $R_i/\pi(i)$ for π-most j. It is straightforward to formalize and prove such a result: let us just state the random walk case.

(7.12) COROLLARY. *There is a vanishing function* ψ *such that for random walks*

$$\#\{j: |\frac{E_j T_i}{R\#G} - 1| > \varepsilon\} \le \varepsilon\#G , \quad \textit{for} \quad \varepsilon = \psi(\tau/\#G) .$$

So rapidly mixing processes have the property that $E_j T_i$ is almost constant, over most j. It can be shown that for reversible processes this property is actually equivalent to rapid mixing, see Aldous (1982a).

Of course one cannot expect to have E_jT_i approximately equal to $R_i/\pi(i)$ for *all* j, since there will often be states j such that the process started at j is likely to hit i quickly.

We now consider the time to hit subsets of states, rather than single states. Here even approximations are hard to find: let us give some lower bounds on the mean time to hit a subset from the stationary initial distribution.

(7.13) PROPOSITION. *Suppose* $q_i \equiv 1$. *Then*

(a) $E_\pi T_A \geq \frac{1}{2\pi(A)} - \frac{3}{2}$

(b) $E_\pi T_A \geq \min_{i\in A} R_i \cdot \frac{1}{\pi(A)}\{1 - \psi(\pi(A)\tau \log(1+\tau))\}$, *where* ψ *is vanishing.*

PROOF. (a) By (2.5) it suffices to prove this for a discrete-time chain. There, $P_\pi(T_A = n) \leq P_\pi(X_n \in A) = \pi(A)$, and so $P_\pi(T_A < n) \leq n\pi(A)$. Let $N = [\frac{1}{\pi(A)}]$. Then

$$\begin{aligned} E_\pi T_A &\geq \sum_{n=1}^{N} P_\pi(T_A \geq n) \\ &\geq \sum_{1}^{N}(1 - n\pi(A)) \\ &= N - \frac{1}{2}N(N+1)\pi(A) \\ &\geq \frac{1}{\pi(A)} - 1 - \frac{1}{2}\,\frac{1}{\pi(A)}\left(\frac{1}{\pi(A)} + 1\right)\pi(A) \end{aligned}$$

giving (a).

The proof of (b) is similar to the proofs of Lemma 6.8 and Proposition 7.1. Analogously to the latter proof, fix t_1 and set

$$\begin{aligned} U_1 &= \min\{t: X_t \in A\} \\ U_n &= \min\{t \geq U_{n-1}+t_1: X_t \in A\} \\ V_n &= U_{n+1} - U_n \\ W_n &= \text{time}\{s: U_n \leq s < U_{n+1},\ X_s \in A\} \\ Z(t) &= \text{time}\{s: U_1 \leq s < t,\ X_s \in A\} \\ F_n &= \sigma(X_s: s \leq U_{n+1}) \\ \rho_i(\cdot) &= P_i(X_{t_1} \in \cdot) \quad . \end{aligned}$$

Then

$$E(V_n|F_{n-1}) \le v(t_1) = t_1 + \max_i E_{\rho_i} T_A$$
$$E(W_n|F_{n-1}) \ge w(t_1) = \min_{i\in A} R_i(t_1) \ .$$

Also $W_n \le t_1$; and $\max_i E_i T_A^2 < \infty$ because the state space is finite, so $\max_n EV_n^2 < \infty$. So we can apply Lemma 7.3(b) to obtain

$$\pi(A) = \lim t^{-1} Z(t) \ge w(t_1)/v(t_1) \ .$$

Estimating $v(t_1)$ by Lemma 7.9 and rearranging,

$$E_\pi T_A \ge \frac{w(t_1)}{\pi(A)}(1 - d(t_1)) - t_1 \ . \tag{7.14}$$

We want to evaluate this at a time t_1 large compared to $\tau \log(1+\tau)$ but small compared to $1/\pi(A)$. So set

$$\alpha = \pi(A)\tau \log(1+\tau)$$
$$t_1 = \alpha^{-1/2}\tau \log(1+\tau) = \alpha^{1/2}/\pi(A) \ .$$

Then, setting $w = \min_{i\in A} R_i$,

$$|w - w(t_1)| \le t_1\pi(A) + \exp(1 - t_1/\tau) \quad \text{by (6.5)}$$
$$\le \psi(\alpha)$$

and since $w \ge \frac{1}{2}$ by (7.7),

$$\left|\frac{w(t_1)}{w} - 1\right| \le \psi(\alpha) \ . \tag{7.15}$$

Also $t_1\pi(A)/w = \alpha^{1/2}/w$

$$\le \psi(\alpha) \ . \tag{7.16}$$

And by (3.8), $d(t_1) \le \psi(\alpha)$. Putting this, (7.16) and (7.15) into (7.14) gives the result.

In the rapidly mixing random walk case, Proposition 7.13 gives an approximate lower bound of $R\#G/\#A$ for $E_\pi T_A$. If the subset A is "sparse" in the sense that, starting at one element of A, the random walk is unlikely to hit any different element of A in the short term, then this lower bound is approximately the correct value of $E_\pi T_A$. Such a result can be proved by the techniques of Proposition 7.13: but since the conditions are hard to check in practice, we shall merely state one form of this idea.

(7.17) PROPOSITION. *There is a vanishing function* ψ *such that for random walks*

$$\left|\frac{\#A}{R\#G}E_\pi T_A - 1\right| \leq \psi\left(\frac{\#A}{\#G}\tau \log \#G + b_A\right)$$

where $b_A = \max_{i \in A} P_i(T_{A \setminus \{i\}} \leq \tau(1 + 2\log \#G))$.

We shall now discuss the distribution of hitting times T_A. At first sight, the difficulty of estimating the mean $E_\pi T_A$ for general A suggests that one could say little about the distribution. But it turns out that, in the rapidly mixing case, the hitting time distribution on A from a stationary initial distribution is approximately exponential, provided $\pi(A)$ is sufficiently small.

(7.18) PROPOSITION. *There is a vanishing function* ψ *such that*

$$\sup_{t \geq 0} |P_\pi(T_A > t) - \exp(-t/E_\pi T_A)| \leq \psi(\tau/E_\pi T_A) .$$

In other words, the distribution is approximately exponential provided $E_\pi T_A$ is large compared to τ. In the random walk case, it is sufficient by Proposition 7.13 that $\#A$ be small compared to $\#G/\tau$. In particular, our informal definition of "rapid mixing" (5.10) ensures that in a rapidly mixing random walk the exponential approximation for the hitting time on a single state will be valid.

Proposition 7.18 is proved in Aldous (1982b), and we will not repeat the details. The main idea is that the conditional distributions $\nu_t = P_\pi(X_t \in \cdot \,|\, T_A > t)$ must stay close to π, because the tendency of ν_t to drift away from π (due to paths hitting A being eliminated) is counteracted by the rapid mixing. So $P_\pi(T_A > t+s \,|\, T_A > t) = P_{\nu_t}(T_A > s)$ is approximately $P_\pi(T_A > s)$, and this makes T_A be close to exponential.

We now discuss the distribution of T_A for rapidly mixing processes when the initial distribution ν is arbitrary: our remarks are formalized in Proposition 7.19 below. There is a certain probability p, say, that the process hits A in the short term (compared to $E_\pi T_A$). Given this does not happen, the distribution of T_A is approximately exponential, mean $E_\pi T_A$. In other words, the P_ν-distribution of $T_A/E_\pi T_A$ will be a mixture of a distribution concentrated near 0 (with weight p) and a distribution close to the exponential mean 1 distribution (with weight $1-p$). So $E_\nu T_A$ is approximately $(1-p)E_\pi T_A$. So assuming $E_\pi T_A$ is known, estimates of either p or $E_\nu T_A$ give estimates of the other.

Of course if p is close to 1, these arguments tell us only that $E_\nu T_A$ is small compared to $E_\pi T_A$.

(7.19) PROPOSITION. *For arbitrary* ν, A, *let* $\hat{m} = E_\nu T_A$, $m = E_\pi T_A$. *There is a vanishing function* ψ *such that*

$$\sup_{t \geq \varepsilon m} |P_\nu(T_A > t) - \frac{\hat{m}}{m}\exp(-t/m)| \leq \varepsilon$$

where $\varepsilon = \psi(\tau/m)$.

Analogously to (7.12), when A is a "small" subset in a rapidly mixing process, then $E_j T_A$ will be close to $E_\pi T_A$ for "most" j, and so for "most" j the P_j-distribution of T_A will be approximately exponential.

PROOF OF PROPOSITION 7.19. Set $\alpha = \tau/m$, and suppose $\alpha < 1$. Set

$$t_0 = \tau\alpha^{-1/3}$$
$$J = \{j: P_j(T_A \leq t_0) \leq \alpha^{1/3}\} .$$

We assert

$$E_\nu \min(T_A, T_J) \leq \alpha^{1/3} m . \tag{7.20}$$

Indeed, by definition of J we have, for $j \notin J$,

$$P_\nu(\min(T_A,T_J) > (n+1)t_0 | \min(T_A,T_J) > nt_0,\ X_{nt_0} = j) \leq 1 - \alpha^{1/3}$$

and so $P_\nu(\min(T_A,T_J) > nt_0) \le (1-\alpha^{1/3})^n$, giving

$$E_\nu \min(T_A,T_J) \le t_0\alpha^{-1/3} = \alpha^{1/3}m .$$

Next we assert

$$(7.21) \qquad |P_j(T_A > t) - \exp(-t/m)| \le \psi(\alpha) \ ; \quad j \in J, \ t \ge 0 .$$

For, setting $\rho_j = P_j(X_{t_0} \in \cdot)$,

$$|P_j(T_A > t) - P_{\rho_j}(T_A > t-t_0)| \le \alpha^{1/3} , \ j \in J, \text{ by definition of } J;$$
$$|P_{\rho_j}(T_A > t-t_0) - P_\pi(T_A > t-t_0)| \le \|\rho_j - \pi\| \le d(t_0) \le \psi(\alpha) \text{ by (3.8)};$$
$$|P_\pi(T_A > t-t_0) - \exp(-(t-t_0)/m)| \le \psi(\alpha) \text{ by Proposition 7.18};$$
$$|\exp(-(t-t_0)/m) - \exp(-t/m)| \le t_0/m = \alpha^{2/3} .$$

Next, set $t_2 = \alpha^{1/4}m$ and let B be the event $\{T_J \le \min(T_A,t_2)\}$. For $t \ge t_2$,

$$\begin{aligned} |P_\nu(T_A > t) - P_\nu(T_A > t, B)| &\le P_\nu(T_A > t, \ T_J \ge t_2) \\ &\le \alpha^{1/3}m/t_2 \quad \text{by (7.20)} \\ (7.22) \qquad &\le \psi(\alpha) . \end{aligned}$$

And for $t \ge t_2$,

$$\min_{j\in J} P_j(T_A > t) \le P_\nu(T_A > t|B) \le \max_{j\in J} P_j(T_A > t-t_2) ,$$

so from (7.21)

$$\begin{aligned} |P_\nu(T_A > t|B) - \exp(-t/m)| &\le \psi(\alpha) + t_2/m \\ &\le \psi(\alpha) \ ; \quad t \ge t_2 . \end{aligned}$$

Using (7.22),

$$(7.23) \qquad |P_\nu(T_A > t) - P(B)\cdot\exp(-t/m)| \le \psi(\alpha) \ ; \quad t \ge t_2.$$

It remains to estimate $P(B)$. First,

$$\begin{aligned} \max_i E_i T_A &\le \frac{t_0 + m}{1-d(t_0)} \quad \text{by Lemma 7.9} \\ (7.24) \qquad &\le m(1+\psi(\alpha)) . \end{aligned}$$

Second, note the elementary inequality

$$(7.25) \qquad \|P(Y\in\cdot) - P(Y\in\cdot|D)\| \le 1 - P(D) .$$

Now for $j \in J$ and $\rho = P_j(X_{t_0} \in \cdot | T_A > t_0)$,

$$\begin{aligned} \|\pi-\rho\| &\le \|P_j(X_{t_0} \in \cdot) - \pi\| + P_j(T_A \le t_0) \quad \text{by (7.25)} \\ &\le d(t_0) + \alpha^{1/3} \le \psi(\alpha) ; \\ E_j T_A \ge E_\rho T_A &\ge m - \|\rho-\pi\| \max_i E_i T_A \end{aligned}$$

and so by (7.24)

$$E_j T_A \ge m(1-\psi(\alpha)) ; \quad j \in J .$$

Using (7.24) again,

$$(7.26) \qquad |E_i T_A - m| \le m\psi(\alpha) ; \quad j \in J .$$

Now from the definition of B,

$$P(B) \min_{j\in J} E_j T_A \le ET_A 1_B \le P(B)\{\max_{j\in J} E_j T_A + t_2\} .$$

Combining this fact with (7.26),

$$(7.27) \qquad |E_\nu T_A 1_B - P(B)m| \le m\psi(\alpha) .$$

Now Ω is covered by $\{B, B_1, B_2\}$, where

$$\begin{aligned} B_1 &= \{t_2 \le \min(T_A, T_J)\} \\ B_2 &= \{T_A \le \min(T_J, t_2)\} . \end{aligned}$$

So we estimate the contribution to the expectation of T_A made over these sets.

$$\begin{aligned} E_\nu T_A 1_{B_1} &\le P(B_1)(t_2 + \max_i E_i T_A) \\ &\le \psi(\alpha)(t_2 + \max_i E_i T_A) \quad \text{by the argument for (7.22)} \\ &\le \psi(\alpha)m \quad \text{using (7.24)} \\ E_\nu T_A 1_{B_2} &\le t_2 = \alpha^{1/4} m . \end{aligned}$$

Combining these with (7.27),

$$|E_\nu T_A - P(B)m| \le m\psi(\alpha) .$$

This estimate for $P(B)$, substituted into (7.23), establishes the Proposition.

REMARK. By applying Proposition 7.19 to the distribution ν of the position after the first jump out of state i, we see that in a rapidly mixing process the distribution of T_i^+, the return time to i, is approximately a mixture of an exponential distribution and a distribution comparatively small. Precisely, we obtain

(7.28) COROLLARY. $\sup_{t \ge \varepsilon R_i/\pi(i)} |P_i(T_i^+ > t) - \frac{\exp\{-t\pi(i)/R_i\}}{q_i R_i}| \le \varepsilon$

for $\varepsilon = \psi(\tau\pi(i)/R_i)$.

Then from Proposition 7.2 one can obtain estimates of the density function of $P_\pi(T_i \in \cdot)$.

It seems reasonable to hope that the ideas here will be useful in studying properties of rapidly mixing processes other than first hitting time distributions. Let us merely mention one slightly different result. Let $V = \max_i T_i$ be the time taken for the process to visit every state. The following result, proved in Aldous (1983), says that in the random walk case V is approximately $R\#G \log \#G$ provided $\log \tau$ is small compared to $\log \#G$.

(7.29) PROPOSITION. *There is a vanishing function* ψ *such that for random walks*

$$E|\frac{V}{R\#G \log \#G} - 1| \le \psi(\frac{\log(1+\tau)}{\log \#G}) .$$

8. Hitting times - Examples

Here we apply the theory of Sections 6 and 7 to the examples described previously.

EXAMPLE 3.19. *Random walk on the* N*-dimensional cube.* In this example, the explicit formula (3.20) for $p_{i,i}(t)$ gives an explicit formula for $R^N(t)$:

$$R^N(t) = \int_0^t 2^{-N}(1+e^{-2s/N})^N ds .$$

Calculus gives

$$R^N(t_N) \to 1 \quad \text{for} \quad t_N \to \infty, \quad t_N/2^N \to 0 .$$

Recalling from (3.23) that $\tau \sim \frac{1}{4}N \log N$, we have from (6.7)

$$R^N \to 1 \quad \text{as} \quad N \to \infty .$$

In other words, for large N there is only a small chance of the process returning to its starting state in the short term.

We can now apply the results of Section 7. Proposition 7.1 says

$$E_\pi T_i \sim 2^N \quad \text{as} \quad N \to \infty .$$

Proposition 7.18 says that the P_π-distribution of $T_i/2^N$ converges to exponential as $N \to \infty$. In this example, one could obtain this result analytically. But Proposition 7.18 also says that for subsets A_N such that $\#A_N N \log(N)/2^N \to 0$, the P_π-distribution of $T_{A_N}/E_\pi T_{A_N}$ converges to exponential; even in such a simple example analytic techniques do not readily yield such results.

Donnelly (1982), in the context of a problem in genetics, compares the exponential approximation with the exact distribution in several particular cases: the approximation is rather good, even in low dimensions.

EXAMPLE 5.1. *Ehrenfest urn model.* Kemperman (1961) investigates this example in detail by analytic techniques. Let us indicate how some of the results are special cases of our general results.

Consider hitting times on i_N, where $i_N/N \to c < \frac{1}{2}$ as $N \to \infty$. We assert

(8.1) $$R_{i_N} \sim (1-2c)^{-1} \quad \text{as} \quad N \to \infty .$$

The idea is that, starting (X_t) at i_N, the process $(X_t - i_N)$ behaves initially like the simple random walk on $\mathbf{Z}$ with drift: $Q(j,j-1) = c$, $Q(j,j+1) = 1-c$. This transient process has $R(\infty) = (1-2c)^{-1}$, and it is not hard to justify (8.1).

Recall that π is Binomial $(N,\frac{1}{2})$ and $\tau \sim \frac{1}{4}N\log N$. We can now apply the results of Section 7. Proposition 7.1 says

$$(8.2)\qquad E_\pi T_{i_N} \sim (1-2c)^{-1}2^N/\binom{N}{i_N} = m_N \quad \text{say} ,$$

and $\log m_N \sim N(\log 2 + c\log c + (1-c)\log(1-c))$. Proposition 7.18 says

(8.3) the P_π-distribution of T_{i_N}/m_N converges to exponential (1).

Moreover Proposition 7.8 shows $\max_j E_j T_{i_N} \le m_N(1+\varepsilon_N)$, where $\varepsilon_N \to 0$. Since $E_j T_{i_N}$ is plainly monotone in $j > i_N$, it follows that (8.2) holds also for the process started at $j_N \ge N/2$. Then Proposition 7.19 shows that (8.3) also holds for the process started at $j_N \ge N/2$. Finally, consider the first return time $T^+_{i_N}$. Corollary 7.28 shows

$$T^+_{i_N}/m_N \xrightarrow{\mathcal{D}} Y ,$$

where $P(Y=0) = 2c$, $P(Y>t) = (1-2c)e^{-t}$, $t \ge 0$.

Let us now consider the card-shuffling models. As explained at (2.5), the continuous-time theory of Section 7 extends to discrete-time random walks. In card-shuffling models it is often true that

$$(8.4)\qquad R^N \to 1 \quad \text{as} \quad N\to\infty ;$$

in other words when starting with a new deck one is unlikely to get back to the new deck state in the short term. When (8.4) holds, Propositions 7.1 and 7.18 show that the P_π-distribution of T_i is asymptotically exponential with mean $N!$, as $N\to\infty$.

In the cases of the uniform riffle shuffle (4.17) and random transpositions (4.13), assertion (8.4) is an immediate consequence of Lemma 6.21, since

$$\text{(for uniform riffle)} \qquad q^* = 2^{-N} , \quad \tau \sim \tfrac{3}{2}\log_2 N$$
$$\text{(for random transpositions)} \qquad q^* = 2/N^2 , \quad \tau \sim \tfrac{1}{2}N\log N .$$

Let us now prove (8.4) for the "transposing neighbours" shuffle (4.10), using

Lemma 6.21. Let $f(\pi,\sigma) = \#\{i: \pi(i) \neq \sigma(i)\}$ be the number of unmatched cards in decks π, σ. Fix π, σ and let $m = f(\pi,\sigma)$. Let X_1 be the distribution of the deck initially in state π after one shuffle, and let $Y = f(X_1,\sigma)$. To apply Lemma 6.21 we need c, $0 < s < 1$ such that

$$(8.5) \qquad c \geq Es^Y - s^m \; ; \quad m \geq 2 \; .$$

(Note m cannot equal 1.) So we want to estimate the distribution of Y. Plainly $m-2 \leq Y \leq m+2$. And the number of successive pairs which are both matched is at least $N-1-2m$. If such a pair is transposed, then two new cards become unmatched. So $P(Y = m+2) \geq 1 - (2m+1)/N$. Hence we obtain

$$(8.6) \qquad \begin{aligned} Es^Y &\leq s^{m-2} \; ; & m \geq 2 \\ Es^Y &\leq \tfrac{2m+1}{N} s^{m-2} + \left(1 - \tfrac{2m+1}{N}\right) s^{m+2} \; ; & 2 \leq m < N/2 \; . \end{aligned}$$

Setting $s = N^{-1/3}$ and $m_0 = [\frac{1}{2}(N^{1/3}-2)]$ we have, for $m \leq m_0$,

$$\begin{aligned} Es^Y - s^m &\leq s^m \{\tfrac{2m+1}{N} s^{-2} + s^2 - 1\} \\ &\leq 0 \quad \text{after some algebra.} \end{aligned}$$

Thus (8.5) holds for $c = s^{m_0 - 2}$. Applying Lemma 6.21,

$$R(t) \leq \{1 - (s+ct)\}^{-1} \; .$$

Applying this to $\tau^* = \tau(1 + \log(N!)) \lesssim N^5$, we have $s + c\tau^* \to 0$ as $N \to \infty$, and so $R(\tau^*) \to 1$. And (6.7) gives

$$|R - R(\tau^*)| \leq 2\tau^*/N! \to 0$$

establishing (8.4) for this model.

EXAMPLE 5.5. *Sequences in coin-tossing.* For a prescribed sequence $i = (i_1,\ldots,i_N)$ of Heads and Tails, let $\hat{T}_i$ be the number of tosses of a fair coin until sequence i appears. Studying $\hat{T}_i$ is a classical problem in elementary probability: see Feller (1968). We shall derive some known results. As at (5.5) let X_n be the Markov chain of sequences of length N, with uniform initial distribution. Let $T_i = \min\{n \geq 0: X_n = i\}$, and note

$\hat{T}_i = T_i + N$. The discrete-time analogue of Proposition 7.1 is

$$E_\pi T_i = R_i/\pi(i) \; ; \quad R_i = \lim_n \sum_{m=0}^{n} (p_{i,i}(m) - \pi(i)) \; .$$

In this example we have

$$\pi(i) = 2^{-N}$$
$$p_{i,i}(m) = 2^{-N} \; , \quad m \geq N$$
$$= 1_{(i_q = i_{q+m} : \; 1 \leq q \leq N-m)} \; , \quad 0 \leq m < N.$$

Hence we find

$$E\hat{T}_i = 2^N \{1 + \sum_{m=1}^{N-1} 2^{-m} 1_{(i_q = i_{q+m} : \; 1 \leq q \leq N-m)}\} \; .$$

This is well-known: see Li (1980) for recent extensions and references. Proposition 7.18 says that as $N \to \infty$ the distribution of $\hat{T}_i / E\hat{T}_i$ converges to exponential: this fact is implicit in the generating function approach to this problem (Gerber and Li (1981)) but seems not to have been explicitly noted. Moreover, Li (1980) discusses the time $\hat{T}_A$ until some one of a set A of sequences of length N occurs: by Propositions 7.19 and 7.13 the distribution of $\hat{T}_{A_N} / E\hat{T}_{A_N}$ converges to exponential when $\#A_N / 2^N \to 0$.

EXAMPLE 5.7. *Random walk in a* d*-dimensional box.* Fix $d \geq 3$. Consider points $x = x^N$ in boxes of side N, which are away from the sides in the sense $\min_{1 \leq i \leq d} (x_i^N, N - x_i^N) \to \infty$ as $N \to \infty$. For such points it is not difficult to see that $R_x \to (1 - F_d)^{-1}$, where F_d is the return probability for the unrestricted d-dimensional simple random walk. Thus Proposition 7.1 implies $E_\pi T_x \sim N^d (1 - F_d)^{-1}$, and Proposition 7.18 implies that the distribution of T_x / ET_x converges to exponential.

References

ALDOUS, D. J. (1982a). Some inequalities for reversible Markov chains. *J. London Math. Soc.* **25** 564-576.

ALDOUS, D. J. (1982b). Markov chains with almost exponential hitting times. *Stochastic Processes Appl.* **13**, to appear.

ALDOUS, D. J. (1983). On the time taken by a random walk on a finite group to visit every state. *Zeitschrift fur Wahrscheinlichkeitstheorie*, to appear.

DIACONIS, P. (1982). Group theory in statistics. Preprint.

DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Zeitschrift fur Wahrscheinlichkeitstheorie* **57** 159-179.

DONNELLY, K. (1982). The probability that a relationship between two individuals is detectable given complete genetic information. *Theoretical Population Biology*, to appear.

EPSTEIN, R. A. (1977). *The Theory of Gambling and Statistical Logic* (Revised Edition). Academic Press.

FELLER, W. (1968). *An Introduction to Probability Theory* (3rd Edition). Wiley.

GERBER, H. U. and LI, S.-Y. R. (1981). The occurrence of sequence patterns in repeated experiments and hitting times in a Markov chain. *Stochastic Processes Appl.* **11** 101-108.

KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes.* Academic Press.

KEILSON, J. (1979). *Markov Chain Models--Rarity and Exponentiality.* Springer-Verlag.

KEMENY, J. G. and SNELL, J. L. (1959). *Finite Markov Chains.* Van Nostrand.

KEMPERMAN, J. (1961). *The First Passage Problem for a Stationary Markov Chain.* IMS Statistical Research Monograph 1.

LETAC, G. (1981). Problèmes classiques de probabilité sur un couple de Gelfand. *Analytical Methods in Probability Theory*, ed. D. Duglé et al. Springer Lecture Notes in Mathematics 861.

LI, S.-Y. R. (1980). A martingale approach to the study of occurrence of sequence patterns in repeated experiments. *Ann. Probability* **8** 1171-1176.

REEDS, J. (1982). Unpublished notes.

STOUT, W. F. (1974). *Almost Sure Convergence.* Academic Press.