SOME INEQUALITIES FOR REVERSIBLE MARKOV CHAINS

DAVID J. ALDOUS

1. *Introduction*

One of the most important results about finite ergodic Markov chains is the convergence of transition probabilities to the stationary distribution. The object of this paper is to investigate relations between the time taken to approach stationarity and certain properties of mean hitting times. Our main result, Theorem 5, shows that for reversible chains the following (informally stated) properties are equivalent.

(i) Convergence to stationarity is rapid.

(ii) Mean hitting times on single states are nearly uniform in the initial state.

(iii) Mean hitting times on a set A of states can be bounded in terms of the stationary measure of A.

Theorem 6 gives weaker results for general (that is, non-reversible) chains.

Let X_t be an ergodic Markov chain in continuous time, with finite state space $I = \{i, j, k, \ldots\}$. Let $Q = (q_{i,j})$ be the matrix of transition rates, let $p_{i,j}(t) \equiv p_t(i,j)$ be the transition probabilities, and let π be the stationary distribution. The total variation distance between distributions on I is

$$(1) \qquad \|\mu - \lambda\| = \frac{1}{2}\sum |\mu_j - \lambda_j| = \sup_A |\mu(A) - \lambda(A)| .$$

Because I is finite, the classical result on convergence to stationarity implies that

$$(2) \qquad \|p_t(i, \cdot) - \pi(\cdot)\| \to 0 \qquad \text{as } t \to \infty .$$

Recall that X_t is called *reversible* if

$$(3) \qquad \pi_i p_{i,j}(t) = \pi_j p_{j,i}(t), \qquad i,j \in I, \quad t \geqslant 0 .$$

Because I is finite, this is equivalent to

$$\pi_i q_{i,j} = \pi_j q_{j,i}, \qquad i,j \in I .$$

See [7, 8] for discussions of reversibility.

We now formalise properties (i)–(iii) by defining several parameters. Let

$$\tau_1 = \min\{t : \|p_t(i, \cdot) - \pi(\cdot)\| \leqslant (2e)^{-1} \text{ for all } i\} < \infty \qquad \text{by (2)} .$$

Received 13 June 1980; revised 10 August 1981.

This research was supported in part by N.S.F. Grant MCS 80-02698.

[J. London Math. Soc. (2), 25 (1982), 564–576]

Thus τ_1 measures the time until the transition probabilities are close in total variation to the stationary distribution: the constant $(2e)^{-1}$ has no special significance beyond algebraic convenience. Another way to measure the time taken to approach stationarity is to consider stopping times for which the stopped chain has exactly stationary distribution, and this suggests defining $\tau_2 = \max \alpha_i$, where

(4) $\alpha_i = \inf\{E_i T_i : T_i \text{ a stopping time such that } P_i(X_{T_i} = j) = \pi(j) \text{ for all } j\}$.

It is not quite obvious that such stopping times exist; a construction is given later.

The next parameters formalise the properties (ii) and (iii). We define

$$\tau_3 = \max_{i,k} \sum_j \pi_j |E_i H_j - E_k H_j|, \qquad \tau_4 = \max_{i,A} \pi(A) E_i H_A .$$

Here $H_A = \inf\{t : X_t \in A\}$ is the first hitting time on a subset A of I. We can now state our main result.

5 THEOREM. *There exist universal constants* $C_{r,s}$ *such that* $\tau_r \leqslant C_{r,s}\tau_s$, $1 \leqslant r, s \leqslant 4$, *for every reversible chain.*

The significance of Theorem 5 is qualitative—if one τ is small then so are the others, and so properties (i)–(iii) are equivalent. Universal inequalities seem rather novel in Markov chain theory, but have been the subject of intensive research in martingale theory [4].

Theorem 5 extends partially to non-reversible chains. Let $\pi_{\min} = \min \pi_i$.

6 THEOREM. *There exist universal constants* K_1, K_2, K_3 *such that for every chain*

$$\tau_2 \leqslant K_1 \tau_1 (1 + \log(1/\pi_{\min})), \qquad \tau_4 \leqslant \tau_3 \leqslant K_2 \tau_2 ,$$

$$\tau_2 \leqslant K_3 \tau_4 (1 + \log(1/\pi_{\min})) .$$

Examples 45 and 46 will show that the log terms cannot be omitted, and that there is no similar upper bound for τ_1 in terms of the other parameters.

Theorems 5 and 6 remain valid for positive-recurrent chains on a countable state space; though here the parameters may be infinite. The condition $\tau_1 < \infty$ is equivalent to

$$\sup_i \sum_j |p_{i,j}(t) - \pi_j| \to 0 \qquad \text{as } t \to \infty ,$$

which is often called "strong ergodicity" [6] or "uniform ergodicity" [5] in the literature. So Theorem 5 shows that for a reversible chain each of the conditions $\tau_r < \infty$ is necessary and sufficient for this property. A general necessary and sufficient condition is that [6]

$$\sup_i E_i H_j < \infty \qquad \text{for some (respectively all) } j .$$

But Example 48 will show there is no universal inequality relating τ_1 to $\sup_i E_i H_j$.

A classical way to analyse the transition probabilities of reversible chains is via the spectral representation [7]

$$(7) \qquad p_{i,j}(t) = \pi_j + \left(\frac{\pi_j}{\pi_i}\right)^{1/2} \sum_{r=1}^{N} u_i^{(r)} u_j^{(r)} e^{-t/\beta_r}, \qquad t \geqslant 0,$$

where $0 < \beta_1 \leqslant \beta_2 \leqslant \ldots \leqslant \beta_N$ and the $u_i^{(r)}$ are real numbers such that $\pi_j + \sum_r \{u_j^{(r)}\}^2 = 1$ for each j. So asymptotically $p_{i,j}(t) \approx \pi_j + \left(\frac{\pi_j}{\pi_i}\right)^{1/2} u_i u_j e^{-t/\beta}$, where $u_i = u_i^{(N)}$ and $\beta = \beta_N$. Thus β determines the asymptotic rate of convergence to stationarity, whereas the parameters τ describe features of the non-asymptotic behaviour.

8 Proposition. *For any reversible chain,*

(a) $\beta \leqslant \tau_1$;

(b) $\tau_1 \leqslant \beta(1 + \frac{1}{2}\log(1/\pi_{\min}))$.

Example 49 will show that the log term cannot be omitted.

Motivation for Theorem 5 came from the study of a particular chain, the random walk on the d-cube, discussed as Example 50. David Williams aroused my interest in this chain, and Jim Pitman observed that this chain had almost uniform mean hitting times.

A chain for which τ_1 is small might be called *rapidly mixing*. Such chains have other properties: for example, first hitting times are approximately exponentially distributed [1].

2. *Proofs*

We decompose Theorems 5 and 6 into a series of lemmas (12, 15, 16, 21, 22, 23). Observations (9) and (10) and Lemma 11 are preliminaries. Let

$$\Delta_t = \max_i \|p_t(i,\cdot) - \pi(\cdot)\|.$$

It is easy to verify that Δ_t is decreasing and that

$$(9) \qquad \Delta_{s+t} \leqslant 2\Delta_s\Delta_t.$$

By definition $\Delta_{\tau_1} = (2e)^{-1}$, and so (9) implies that $\Delta_{n\tau_1} \leqslant e^{-n}$. This gives an exponential bound

$$(10) \qquad \Delta_t \leqslant \exp(1 - t/\tau_1), \qquad t \geqslant 0.$$

Let $\operatorname{time}(t \leqslant T : X_t \in A)$ be the random variable describing the length of time X spends in set A before time T.

11 Lemma. *Suppose T is a stopping time such that $P_i(X_T = j) = \pi(j)$ for all j. Then $E_i \operatorname{time}(t \leqslant T : X_t = i) = \pi_i(E_iT + E_\pi H_i)$.*

Proof. Let $X_0 = i$. Let U_1 be the time of the first hit on i after T; inductively define U_n analogously for the process restarted at time U_{n-1}. In this way the process X_t is split into i.i.d. excursions of lengths $U_1, U_2, \ldots$. In each excursion, the length of time spent at i is distributed as $\text{time}(t \leqslant T : X_t = i)$. Using the strong law of large numbers, the asymptotic proportion of time spent at i is equal to $E_i \text{time}(t \leqslant T : X_t = i)/E_i U_1$. But this asymptotic proportion is π_i. Since $E_i U_1 = E_i T + E_\pi H_i$, the lemma follows.

12 Lemma. *$\tau_2 \leqslant C_{2,1}\tau_1$ for reversible chains.*

Proof. Suppose that we can find t_0 and δ such that

$$p_{i,k}(t_0) \geqslant \delta\pi_k, \qquad i, k \in I. \tag{13}$$

Then given i we can construct a stopping time T taking values in $\{t_0, 2t_0, 3t_0, \ldots\}$ such that

$$P_i(T = nt_0) = \delta(1-\delta)^{n-1},$$

$$P(X_T \in \cdot \mid T = nt_0, X_{(n-1)t_0} = j) = \pi;$$

and so

$$\tau_2 \leqslant E_i T = t_0/\delta. \tag{14}$$

We now construct t_0 and δ satisfying (13).

$$p_{i,k}(2\tau_1) = \sum_j p_{i,j}(\tau_1)p_{j,k}(\tau_1) = \pi_k \sum_j p_{i,j}(\tau_1)p_{k,j}(\tau_1)/\pi_j$$

by reversibility. Now by the Cauchy–Schwarz inequality

$$\left(\sum_j p_{i,j}(\tau_1)p_{k,j}(\tau_1)/\pi_j\right)^{1/2} \geqslant \sum p_{i,j}^{1/2}(\tau_1)p_{k,j}^{1/2}(\tau_1) \geqslant \sum \min\left(p_{i,j}(\tau_1), p_{k,j}(\tau_1)\right)$$

$$= 1 - \|p_{\tau_1}(i,\cdot) - p_{\tau_1}(k,\cdot)\| \geqslant 1 - 2(2e)^{-1}$$

by definition of τ_1. Hence $p_{i,k}(2\tau_1) \geqslant \pi_k(1-e^{-1})^2$. Substituting into (13) and applying (14), the lemma is established for $C_{2,1} = 2(1-e^{-1})^{-2}$.

15 Lemma. *$\tau_2 \leqslant K_1\tau_1(1-\log\pi_{\min})$ for any chain.*

Proof. Let $t_0 = \tau_1(2-\log\pi_{\min})$. By (10) we have

$$\|p_{t_0}(i,\cdot) - \pi(\cdot)\| \leqslant e^{-1}\pi_{\min}.$$

So in particular $p_{i,k}(t_0) \geqslant (1-e^{-1})\pi_k$ for any i, k. Now substitute into (13) and apply (14) as before.

16 Lemma. *$\tau_3 \leqslant C_{3,2}\tau_2$ for any chain.*

Proof. If T is a stopping time such that $P_i(X_T \in \cdot) = \pi(\cdot)$ then

$E_iH_j \leqslant E_iT + E_\pi H_j$. Thus

(17) $$E_iH_j \leqslant \tau_2 + E_\pi H_j .$$

We shall prove that

(18) $$\sum_j \pi_j(E_\pi H_j - E_iH_j) \leqslant 3\tau_2 .$$

Straightforward arguments from (17) and (18) yield that

$$\sum \pi_j|E_\pi H_j - E_iH_j| \leqslant 5\tau_2 .$$

Thus $\tau_3 \leqslant 10\tau_2$.

To prove (18), fix $\varepsilon > 0$. By definition of τ_2, for each j there exists a stopping time T_j such that

(19) $$P_j(X_{T_j} = k) = \pi(k), \qquad k \in I,$$

and $E_jT_j \leqslant \tau_2 + \varepsilon$. Now fix i, let $X_0 = i$ and put $T = T_i$. For any j,

$$E_iH_j \geqslant E_i(H_j - T)1_{(H_j \geqslant T)} = \sum_k P_i(X_T = k,\, H_j \geqslant T)E_kH_j$$

$$= \sum (\pi_k - d_k)E_kH_j = E_\pi H_j - \sum d_kE_kH_j$$

where $d_k = P_i(X_T = k, H_j < T)$. Thus

$$E_\pi H_j - E_iH_j \leqslant \sum d_kE_kH_j \leqslant (\tau_2 + E_\pi H_j)\sum d_k \qquad \text{by (17)}$$

$$= (\tau_2 + E_\pi H_j)P_i(H_j < T).$$

Now, averaging over j,

(20) $$\sum_j \pi_j(E_\pi H_j - E_iH_j) \leqslant \tau_2 + \sum_j \pi_jE_\pi H_jP_i(H_j < T).$$

We must estimate the right hand term. For each j, on the set $\{X_T = j\}$ define S to be the stopping time T_j of (19) applied to the post-T process $(X_{T+t})_{t\geqslant 0}$. Then X_{T+S} has distribution π and is independent of $\sigma(X_s : s \leqslant T)$, and $E_i(T+S) \leqslant 2(\tau_2 + \varepsilon)$. Now fix j. Let X_t^* be the post-H_j process $(X_{H_j+t})_{t\geqslant 0}$, let $T^* = T - H_j$ and consider these processes only on the set $\{H_j < T\}$. On this set, $X_{T+S} = X^*_{T^*+S}$, and T^*+S is a stopping time for X^* such that $X^*_{T^*+S}$ has distribution π independently of $\sigma(X_t : t \leqslant H_j)$. Thus, still considering only the set $\{H_j < T\}$,

$$E_i(\text{time}\,(t \leqslant T+S : X_t = j) \mid X_u : u \leqslant H_j)$$

$$= E(\text{time}\,(t \leqslant T^*+S : X_t^* = j) \mid X_u : u \leqslant H_j)$$

$$= \pi_j\{E(T^*+S \mid X_u : u \leqslant H_j) + E_\pi H_i\} \qquad \text{by Lemma 11 applied to } X^*$$

$$\geqslant \pi_jE_\pi H_i .$$

Integrating this inequality over the set $\{H_j < T\}$,

$$\pi_j \cdot E_\pi H_j \cdot P_i(H_j < T) \leqslant E_i \operatorname{time}(t \leqslant T+S : X_t = j).$$

Summing over j, we see that the sum in the right side of (20) is at most $E_i(T+S)$, which was shown above to be at most $2(\tau_2+\varepsilon)$. This establishes (18).

21 LEMMA. $\tau_4 \leqslant \tau_3$ *for any chain.*

Proof. Fix a subset A of I and a state i outside A. For any j inside A, $E_iH_j = E_iH_A + E_\rho H_j$ where ρ is the P_i hitting distribution on A. Thus

$$\pi(A)E_iH_A = \sum_{j\in A} \pi_j E_i H_A = \sum_{j\in A} \pi_j(E_iH_j - E_\rho H_j) \leqslant \max_k \sum_{j\in A} \pi_j(E_iH_j - E_kH_j) \leqslant \tau_3 .$$

To complete the proof of Theorems 5 and 6, we must prove the two lemmas below.

22 LEMMA. $\tau_2 \leqslant K_3\tau_4(1-\log\pi_{\min})$ *for all chains.*

23 LEMMA. $\tau_1 \leqslant C_{1,4}\tau_4$ *for reversible chains.*

This is the most complicated part of our arguments, and we need some preliminaries. Observe that for any initial distribution ρ and any subset A of I,

$$P_\rho(H_A > e\tau_4/\pi(A)) \leqslant \pi(A)E_\rho H_A/e\tau_4 \leqslant e^{-1} .$$

Iterating this inequality gives an exponential bound

$$P_\rho(H_A > t) \leqslant \exp\left(1 - t\pi(A)/e\tau_4\right). \tag{24}$$

Next we describe a way of constructing stopping times such that the stopped process has a prescribed distribution. For the rest of the section, fix i and let $X_0 = i$. Let μ be a distribution on $I\setminus\{i\}$. Informally, think of $\mu(j)$ as a "quota" of probability to be allocated to state j by the following procedure: when the chain jumps to a new state, say j, we stop if the quota for j has not yet been filled, and continue otherwise. Thus we want a stopping time $T < \infty$ a.s. such that

$$T = \min\{H_j : \mu_{H_j}(j) < \mu(j)\}, \tag{25}$$

where

$$\mu_t(j) = P(X_T = j,\, T \leqslant t) \tag{26}$$

is the "quota" for j filled by time t. These imply that

$$P(X_T = j) = \mu(j) \qquad \text{for all } j . \tag{27}$$

That such a stopping time exists seems intuitively clear: the discrete-time analogue is discussed in [3], and we give a rigorous construction in continuous-time in Section 3.

In [3] it is shown that in discrete-time the construction minimises ET over all T satisfying (27); the same is true in continuous-time, though we shall not prove this since we do not need it. See also [9].

To attain the stationary distribution π, set $\mu(j) = \pi_j/(1-\pi_i), j \neq i$. Let T^* be the stopping time constructed above for μ, and put

$$T = \begin{cases} 0 & \text{on } F, \\ T^* & \text{on } F^c, \end{cases}$$

where $P(F) = \pi_i$ and the event F is independent of X. From (27),

(28) $$P(X_T = j) = \pi(j), \qquad j \in I.$$

Defining $\mu_t(j)$ as at (26), we have from (25) that

(29) $$T = \inf\{t : X_t \in I_t\} \qquad \text{provided } T > 0,$$

where $I_t = \{j : \mu_t(j) < \pi(j)\}$. We shall call T the *canonical* stopping time (for i). By construction,

(30) $$\mu_t(j) \leqslant \pi(j).$$

Here are some estimates for the distribution of T.

31 LEMMA. (a) $P(T > u) \leqslant \pi(I_u)$, $u \geqslant 0$.
(b) $P(T > v \mid T > u) \leqslant \exp\{1-\pi(I_v)\cdot(v-u)/e\tau_4\}$, $v \geqslant u \geqslant 0$.

Proof. (a) $P(T > u) = \sum \{\pi(j) - \mu_u(j)\} \leqslant \sum_{I_t} \pi(j) = \pi(I_t)$.

(b) By construction, $\{T > v\} \subset \{T > u : X \text{ does not hit } I_v \text{ during } (u, v)\}$. So if ρ denotes the distribution of X_u given $\{T > u\}$, then

$$P(T > v \mid T > u) \leqslant P_\rho(H_{I_v} > v-u).$$

Now apply (24) to complete the proof.

Proof of Lemma 22. Fix $i \in I$, let $X_0 = i$ and let T be the canonical stopping time. We must prove that

(32) $$ET \leqslant K_3\tau_4(1+\log(1/\pi_{\min})).$$

Define constants (t_n') by $P(T > t_n') = e^{-n}(1-\pi_i)$, $n \geqslant 1$. Define constants (t_n'') by

$$t_n'' = \begin{cases} 0, & n = 0, \\ \inf\{t : \pi(I_t) \leqslant e^{-1}\pi(I_{t_{n-1}''})\}, & \text{otherwise}. \end{cases}$$

Merge the sequences t_n', t_n'' into a single increasing sequence (t_m). By definition of (t_n')

we have $P(T > t'_{n+1} \mid T > t'_n) = e^{-1}$, and it follows that

$$P(T > t_{m+1} \mid T > t_m) \geq e^{-1}. \tag{33}$$

By definition of (t''_n), $\pi(I_t) \geq e^{-1}\pi(I_{t''_{n+1}})$, $t < t''_n$, and it follows that

$$\pi(I_t) \geq e^{-1}\pi(I_{t_{m-1}}), \qquad t < t_m. \tag{34}$$

To obtain (33) and (34) we are appealing to the fact that, if (a_n) is positive decreasing and $a_n/a_{n+1} \geq b$ for all n, then $a^*_n/a^*_{n+1} \geq b$ for any decreasing sequence (a^*_n) containing (a_n) as a subsequence.

Now fix m and let $\delta > 0$. Then

$$\begin{aligned} e^{-1} &\leq P(T > t_{m+1}-\delta \mid T > t_m) \qquad \text{by (33)} \\ &\leq \exp\big(1-(e\tau_4)^{-1}(t_{m+1}-\delta-t_m)\pi(I_{t_{m+1}}-\delta)\big) \quad \text{by Lemma 31(b)}. \end{aligned}$$

Letting $\delta \to 0$ and using (34), we have

$$e^{-1} \leq \exp\big(1-e^{-2}\tau_4^{-1}(t_{m+1}-t_m)\pi(I_{t_m})\big).$$

Putting $c = 2e^2\tau_4$ and rearranging,

$$t_{m+1}-t_m \leq c/\pi(I_{t_m}). \tag{35}$$

But $E\min(T, t_{m+1})-E\min(T, t_m) \leq (t_{m+1}-t_m)\cdot P(T > t_m)$, and so, summing over m, we obtain

$$ET \leq c\sum_{m\geq 0} P(T > t_m)/\pi(I_{t_m}). \tag{36}$$

To estimate this sum, fix $k \geq 1$ and consider the sum taken over $M_k = \{t_m : e^{-k} < \pi(I_{t_m}) \leq e^{1-k}\}$ only. The summands are at most 1, by Lemma 31(a). There is at most one element of the t'' sequence in M_k, by definition. And if $t'_r, t'_{r+1}, \ldots, t'_s$ are in M_k then

$$\begin{aligned} \sum_{n=r}^{s} P(T > t'_n)/\pi(I_{t'_n}) &\leq e/\pi(I_{t'_r})\cdot\sum P(T > t'_n) \\ &= e/\pi(I_{t'_r})\cdot(1+e^{-1}+e^{-2}+\ldots)P(T > t'_r) \qquad \text{by definition of } t' \\ &\leq e/(1-e^{-1}) \qquad \text{using Lemma 31(a)}. \end{aligned}$$

Hence the sum in (36) over M_k is at most $1+e/(1-e^{-1})$. But M_k is non-empty only for $k \leq 1-\log\pi_{\min}$. Evaluating (36) now establishes Lemma 22.

Proof of Lemma 23. First we need a variant of the above argument to show that

$$P(T > 2e^{2n+2}\tau_4) \leq e^{-n}. \tag{37}$$

Since $t_m \leqslant t'_m$ we have $\pi(I_{t_m}) \geqslant \pi(I_{t'_m}) = e^{-m}$. Also, the definitions of t'_n, t''_n and Lemma 31(a) imply that $t'_n \leqslant t''_n$, and so after the sequences are merged we have $t'_n \leqslant t_{2n}$. So from (35)

$$t'_n \leqslant c \sum_{m<2n} e^m \leqslant 2e^{2n+2}\tau_4 .$$

The definition of t'_n now gives (37).

The second ingredient is the next lemma, whose proof is deferred.

38 LEMMA. *Suppose that X_t is reversible, and let L, U be positive constants. Let T be the canonical stopping time for i. Let $f_j(u) = P_i(X_u = j,\ T \leqslant L)$. Then there exists $u \leqslant L + U/2$ such that $\sum \pi_j^{-1} f_j^2(u) \leqslant 1 + 2L/U$.*

From the definition,

$$\sum f_j(u) = P_i(T \leqslant L). \tag{39}$$

Now we estimate

$$\begin{aligned}
\|p_u(i,\cdot)-\pi(\cdot)\| &= \frac{1}{2}\sum_j |p_{i,j}(u)-\pi_j| \\
&\leqslant \frac{1}{2}\left(\sum_j |p_{i,j}(u)-f_{i,j}(u)| + \sum_j |f_j(u)-\pi_j|\right) \\
&= \frac{1}{2}\left(P_i(T > L) + \sum_j |f_j(u)-\pi_j|\right)
\end{aligned}$$

by (39), since $f_j(u) \leqslant p_{i,j}(u)$. Next,

$$\begin{aligned}
\left(\sum |f_j(u)-\pi_j|\right)^2 &= \left(\sum \pi_j^{1/2}\left|\frac{f_j(u)-\pi_j}{\pi_j^{1/2}}\right|\right)^2 \\
&\leqslant \sum \pi_j^{-1}(f_j(u)-\pi_j)^2 \qquad \text{by the Cauchy–Schwarz inequality} \\
&= \sum \pi_j^{-1} f_j^2(u) - 2\sum f_j(u) + 1 \\
&\leqslant 2L/U + 2P(T > L) \qquad \text{by Lemma 38 and (39).}
\end{aligned}$$

Since $\|p_t(i,\cdot)-\pi(\cdot)\|$ is decreasing in t, these estimates give

$$\|P_{L+U/2}(i,\ \ )-\pi\| \leqslant \tfrac{1}{2}\{P_i(T > L) + (2L/U + 2P_i(T > L))^{1/2}\}.$$

Choosing $L = C\tau_4$ and $U = CL$ for a suitably large constant C, and using (37), we can make the right-hand side less than $(2e)^{-1}$, so that $\tau_1 \leqslant L + U/2$ and Lemma 23 is established.

Proof of Lemma 38. Define a measure ν on $I \times [0, L]$ by

$$\nu(\cdot,\cdot) = P_i(T \leqslant L, (X_T, T) \in (\cdot,\cdot)).$$

Also define $g_j(u) = P_i(X_{L+u/2} = j, T \leqslant L)$, $0 \leqslant u \leqslant U$. By conditioning on (X_T, T),

$$g_j(u) = \int p_{k,j}(L+u/2-s)d\nu(k,s)$$

where the integral is over (k, s) in $I \times [0, L]$. Thus

$$\sum \pi_j^{-1} g_j^2(u) = \iint \sum_j \pi_j^{-1}\, p_{k_1,j}(L+u/2-s_1)p_{k_2,j}(L+u/2-s_2)d\nu(k_1,s_1)d\nu(k_2,s_2).$$

Now by reversibility the integrand reduces to $\pi_{k_2}^{-1} p_{k_1,k_2}(2L+u-s_1-s_2)$. So, averaging over $[0, U]$,

$$U^{-1}\int_0^U \sum_j \pi_j^{-1} g_j^2(u)du \leqslant U^{-1}\iint \pi_{k_2}^{-1}\left(\int_0^{2L+U} p_{k_1,k_2}(u)du\right)d\nu(k_1,s_1)d\nu(k_2,s_2). \tag{40}$$

The integrand does not involve s_1 and s_2, and so since $\nu(k, [0, L])$ is less than or equal to π_k the right-hand expression is at most

$$U^{-1}\sum_{k_1,k_2} \pi_{k_2}^{-1}\left(\int_0^{2L+U} p_{k_1,k_2}(u)du\right)\pi_{k_1}\pi_{k_2}.$$

Since π is the stationary distribution, this reduces to $U^{-1}(2L+U) = 1+2L/U$. This is an upper bound for the average at (40), and hence dominates $\sum \pi_j^{-1} g_j^2(u)$ for some particular u: this proves Lemma 38.

Proof of Proposition 8. (a) Recall the definition $\Delta_t = \max_i \|p_t(i,\cdot)-\pi(\cdot)\|$. From the spectral representation (7) it is not hard to see that $\lim_{t\to\infty} \Delta_t e^{t/\beta} > 0$. But from (10) we have $\lim_{t\to\infty} \Delta_t e^{t/\tau_1} \leqslant e$. Thus $1/\beta \geqslant 1/\tau_1$, giving (a).

(b) From the spectral representation (7),

$$p_{i,i}(t) \leqslant \pi_i + e^{-t/\beta}, \qquad t \geqslant 0. \tag{41}$$

Now $p_{i,i}(2t) = \sum_j p_{i,j}(t)p_{j,i}(t) = \sum_j \pi_i\pi_j^{-1}p_{i,j}^2(t)$ by reversibility. After applying (41) to $2t$ we obtain

$$\sum_j \pi_j^{-1}p_{i,j}^2(t) \leqslant 1+\pi_i^{-1}e^{-2t/\beta}. \tag{42}$$

Now $\sum_j |p_{i,j}(t)-\pi_j| = \sum_j \pi_j^{1/2}|\pi_j^{-1/2}p_{i,j}(t)-\pi_j^{1/2}| \leqslant 1\cdot D$ by the Cauchy–Schwarz inequality, where

$$D^2 = \sum \{\pi_j^{-1/2}p_{i,j}(t)-\pi_j^{1/2}\}^2 = \sum \pi_j^{-1}p_{i,j}^2(t)-2\sum p_{i,j}(t)+\sum \pi_j$$

$$\leqslant \pi_i^{-1}e^{-2t/\beta} \qquad \text{by (42)}.$$

Thus $\Delta_t \leqslant \frac{1}{2}e^{-t/\beta}\pi_{\min}^{-1/2}$, and (b) follows from the definition of τ_1.

3. *Construction of the canonical stopping time*

Fix $i \in I$, let $X_0 = i$, and let μ be a distribution of $I\setminus\{i\}$. We shall construct a stopping time T satisfying (25).

Set $i_1 = i$, $t_1 = 0$, $J_1 = I\setminus\{i_1\}$, $T_1 = H_{J_1}$. Suppose inductively that we have defined

states $i_1, \ldots, i_n$,

times $0 \leqslant t_1 \leqslant \ldots \leqslant t_n \leqslant \infty$,

stopping times $T_1, \ldots, T_n$

such that, writing $J_n = I\setminus\{i_1,\ldots,i_n\}$ and $\mu_t^n(j) = P(X_{T_n} = j,\ T_n \leqslant t)$, we have

$$\mu_{t_n}^n(j) \begin{cases} = \mu(j), & j = i_1,\ldots,i_n, \\ \leqslant \mu(j), & \text{otherwise}, \end{cases} \tag{43}$$

$$T_n \leqslant H_{J_n}. \tag{44}$$

Then, provided that $t_n < \infty$ and J_n is non-empty, we can define

$$T_{n+1} = \begin{cases} T_n & \text{if } T_n \leqslant t_n, \\ H_{J_n}\ (> t_n \text{ by (44)}) & \text{otherwise}, \end{cases}$$

$$t_{n+1} = \inf\{t \geqslant t_n : \mu_t^{n+1}(j) = \mu(j) \text{ for some } j \in J_n\}$$

and we can choose i_{n+1} to be an element of J_n for which $\mu_{t_{n+1}}^{n+1}(i_{n+1}) = \mu(i_{n+1})$. It is clear that (43) and (44) extend from n to $n+1$. The induction ends when for some N either $t_N = \infty$ or J_N is empty. In either case let $T = T_N$ and $\mu_t = \mu_t^N$. Note that $T < \infty$ a.s. because each $H_{J_n} < \infty$ a.s. For each j we have $P(X_T = j) = \mu_{t_N}^N(j) \leqslant \mu(j)$ by (43), and hence

$$P(X_T = j) = \mu(j), \qquad j \in I.$$

It remains to check that T satisfies (25). Define

$$T^* = \min\{H_j : \mu_{H_j}(j) < \mu(j)\}.$$

It is easy to check the following from the definitions: if $T > t_n$ then either $T = H_{J_n} \leqslant t_{n+1}$ or $H_{J_n} > t_{n+1}$ and $T > t_{n+1}$; if $T^* > t_n$ then either $T^* = H_{J_n} \leqslant t_{n+1}$ or $H_{J_n} > t_{n+1}$ and $T^* > t_{n+1}$. By induction on n we see that $T = T^*$, and thus T satisfies (25).

4. *Examples*

The first two examples show how Theorem 5 may break down without reversibility.

45 *Example* (motion round a circle). Let $I = \{0, 1, ..., N-1\}$; let $q_{0,1} = q_{1,2} = ... = q_{N-1,0} = N$; and let $q_{i,j} = 0$ for other $i \neq j$. Here π is uniform on I. By considering $T = H_J$, where J is distributed as π, we see that $\tau_2 \leqslant \frac{1}{2}$. So by Theorem 6, τ_3 and τ_4 are bounded (as N varies). But an argument using the central limit theorem shows that $\tau_1 \sim \delta N$ as $N \to \infty$ for some $\delta > 0$. Hence there can be no general upper bound for τ_1 in terms of the other parameters.

46 *Example* (climbing a greasy ladder). Let $I = \{1, ..., N\}$; let $q_{i,i+1} = 1$ $(1 \leqslant i < N)$; let $q_{i,1} = 1$ $(1 < i < N)$; let $q_{N,1} = 2$; and let $q_{i,j} = 0$ for other $i \neq j$. Here $\pi_i = 2^{-i}/(1-2^{-N})$. Plainly $E_i H_1 \leqslant 1$, and it follows easily that τ_1, τ_3 and τ_4 are bounded as N varies. But a rather complicated analysis of the canonical stopping times of Section 2 shows that $\tau_2 \sim \delta N$ for some $\delta > 0$. Hence the log terms in Theorem 6 cannot be omitted.

47 *Conjecture.* There exists a K such that $\tau_3 \leqslant K\tau_1$ for every chain.

If so, then with Theorem 6 and the above examples we have a complete picture of the inequalities obtaining between the parameters.

48 *Example* (uniform jump). Let $I = \{1, ..., N\}$; and let $q_{i,j} = 1$ for $i \neq j$. Here π is uniform on I. In this example τ_1 is bounded as N varies, while $E_i H_j = N-1$, $i \neq j$. But in Example 45, $E_i H_j \leqslant 1$ whereas $\tau_1 \sim \delta N$. So there can be no universal inequalities relating τ_1 to $\max_i E_i H_j$.

It would be interesting to know whether there is any parameter involving mean hitting times which is equivalent (in the sense of Theorem 5) to τ_1 for general chains.

49 *Example* (random walk with drift). Let $I = \{0, 1, ..., N\}$; let $q_{i,i+1} = 1$ and $q_{i+1,i} = 2$ for $i = 0,1, ..., N-1$; and let $q_{i,j} = 0$ for other $i \neq j$. This is a birth-and-death process, and hence is reversible. It can be shown that as $N \to \infty$ we have $\tau_1 \to \infty$ but β remains bounded (for β as in Proposition 8). So the log term in Proposition 8(b) cannot be dropped.

50 *Example* (random walk on the d-cube). Here $I = \{0, 1\}^d$; $|i-j|$ denotes the number of coordinates in which i and j differ; $q_{i,j}$ is d^{-1} if $|i-j| = 1$, and is zero for other $i \neq j$; and π is uniform on I. Think of a particle resting at a vertex of a hypercube for an exponential (mean 1) time, and then jumping to a randomly-chosen neighbouring vertex. This chain is reversible, and other special structure (for example symmetry of the hypercube, independence of the coordinate processes)

makes it comparatively easy to analyse directly. We find that $\tau_1 \sim d\log(d)/4$ as $d \to \infty$. Mean hitting times are of the form $E_i H_j = f(|i-j|)$, where f can be found from the recursion

$$f(0) = 0\,, \qquad f(1) = 2^d - 1\,, \qquad f(r) = \left[d(f(r)-1) - r \cdot f(r-1)\right]/(d-r)\,.$$

As $d \to \infty$, $f(r)/2^d \to 1$ uniformly in $r \geq 1$. But even for such a nice chain, Theorem 5 provides information which would be hard to get directly—for example, bounds on $E_i H_A$ for arbitrary subsets A. For further properties of this random walk see [2].

References

1. D. J. Aldous, "Markov chains with almost exponential hitting times", *Stochastic Process Appl.*, to appear.
2. D. J. Aldous, "On the time taken by the random walk on the *d*-cube to visit all vertices", preprint, University of California at Berkeley, 1981.
3. J. R. Baxter and R. V. Chacon, "Stopping times for recurrent Markov processes", *Illinois J. Math.*, 20 (1976), 467–475.
4. A. M. Garsia, *Martingale inequalities* (Benjamin, Reading, Mass., 1973).
5. E. Hille and R. S. Phillips, *Functional analysis and semigroups*, American Mathematical Society Colloquium Publications 31 (American Mathematical Society, Providence, 1957).
6. C. Huang and D. Isaacson, "Ergodicity using mean visit times", *J. London Math. Soc.* (2), 14 (1976), 570–576.
7. J. Keilson, *Markov chain models—rarity and exponentiality* (Springer, New York, 1979).
8. F. P. Kelly, *Reversibility and stochastic networks* (Wiley, New York, 1979).
9. P. A. Meyer, "Travaux de H. Root en theorie du balayage", *Seminaire de probabilités* V, Lecture Notes in Mathematics 191 (Springer, Berlin, 1971) pp. 237–250.

Department of Statistics,
University of California,
Berkeley,
Ca. 94720,
U.S.A.