Inequalities for Rare Events in Time-Reversible Markov Chains I

David J Aldous *
Department of Statistics
University of California
Berkeley CA 94720

Mark Brown †
Department of Mathematics
City College, CUNY
New York NY 10031

Abbreviated title. Rare events in Markov chains.

AMS 1991 subject classifications. 60J27.
Key words and phrases. Markov chain, waiting time, hitting time, exponential approximation, completely monotone.

Stochastic Inequalities, edited by M. Shaked and Y.L. Tong, pages 1-16. IMS Lecture Notes in Statistics, volume 22, 1992.

*Research supported by N.S.F. Grant MCS90-01710
†Research supported by A.F.O.S.R. Grant 89-0083

Abstract

The distribution of waiting time until a rare event is often approximated by the exponential distribution. In the context of first hitting times for stationary reversible chains, the error has a simple explicit bound involving only the mean waiting time ET and the relaxation time τ of the chain. We recall general upper and lower bounds on ET and then discuss improvements available in the case $ET \gg \tau$ where the exponential approximation holds. In a sequel, Stein's method will be used to get explicit bounds on the Poisson approximation for the number of non-adjacent visits to a rare subset.

1 Introduction

The Poisson approximation for numbers of rare events which actually occur, and the exponential approximation for the waiting time until first occurrence of a rare event, are useful throughout many areas on probability – one view of this big picture is presented in Aldous (1989).Here we study explicit bounds in these approximations, in the special setting of hitting times of stationary reversible Markov chains. This paper deals with the exponential approximation and bounds on the mean waiting time; a sequel (Aldous and Brown (1991)) studies Poisson approximations using an implementation of the Chen-Stein method.

The following set-up and notation will be used throughout. $(X_t; t \geq 0)$ is an irreducible finite-state reversible Markov chain in continuous time. The state space is I and the transition rate matrix is $Q = (q(i,j); i,j \in I)$ where $q(i,i) = -\sum_{j \neq i} q(i,j)$. Let π be the stationary distribution. The symmetrizable matrix $-Q$ has real eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots$. Call $\tau = 1/\lambda_1$ the *relaxation time* of the chain. Let A be a fixed (proper, non-empty) subset of I, and let T_A be the first hitting time on A. So $0 < E_\pi T_A < \infty$.

Theorem 1

$$|P_\pi(T_A/E_\pi T_A > t) - e^{-t}| \leq \frac{\tau/E_\pi T_A}{1 + \tau/E_\pi T_A} \leq \tau/E_\pi T_A \quad \textit{for all } t > 0.$$

In the special case where A is a singleton this was proved by Brown (1990). That proof exploits the completely monotone property of the distribution of T_A. In section 3 we show how the general case can be reduced to the special case. In section 7 we give bounds on the density function $f(t)$ of T_A. Lemma 13 gives explicit bounds in the natural approximation

$$f(t) \approx \frac{1}{E_\pi T_A} \text{ on } \tau \ll t \ll E_\pi T_A.$$

Use of Theorem 1 and Lemma 13 requires an upper bound on τ, and bounds on $E_\pi T_A$. We have nothing new to say about the much-studied issue of upper bounding τ, and refer the reader to Diaconis and Stroock (1991) for interesting recent work. The problem of bounding $E_\pi T_A$ in terms of readily computable quantities has apparently not been studied much, aside from the simple and well-known general bounds stated in Lemma 2 below. The left two inequalities, and (1), follow easily from complete monotonicity (Keilson

(1979) Theorem 6.9C). The rightmost inequality is an easy consequence of extremal characterizations: see the proof of Lemma 10 below.

Lemma 2

$$\frac{1-\pi(A)}{q(A,A^c)} \leq \frac{E_\pi T_A}{1-\pi(A)} \leq E_\alpha T_A \leq \frac{\tau}{\pi(A)}$$

where

$$q(A,A^c) = \sum_{i\in A}\sum_{j\notin A} \pi(i)q(i,j)$$

and where α is the quasistationary distribution defined at (12).

The extreme bounds in Lemma 2 are often crude. In our setting where $\tau \ll E_\pi T_A$ the intermediate inequality, and its distributional version

$$P_\pi(T_A > t) \leq (1-\pi(A))\exp(-t/E_\alpha T_A) \qquad (1)$$

is often fairly sharp, but in practice is hard to use. The point is that, because of the extremal characterization (13) of $E_\alpha T_A$ as a *sup*, it is often easy to get a good *lower* bound on $E_\alpha T_A$, but much harder to get an upper bound. Our main new result is the following usable inequality.

Theorem 3

$$P_\pi(T_A > t) \geq (1-\frac{\tau}{E_\alpha T_A})\exp(\frac{-t}{E_\alpha T_A}),\ t > 0.$$

Integrating over t gives

Corollary 4 *$E_\pi T_A \geq E_\alpha T_A - \tau$.*

The point is that these results may be applied without any prior estimate of $E_\pi T_A$, merely a lower bound on $E_\alpha T_A$. Note also that we may use Corollary 4 to rewrite Theorem 1 as

Corollary 5

$$|P_\pi(T_A/E_\pi T_A > t) - e^{-t}| \leq \frac{\tau}{E_\alpha T_A},\ t \geq 0.$$

Note also that Theorem 3 improves on the simple complete monotonicity result underlying the left inequality in Lemma 2:

$$P_\pi(T_A > t) \geq (1-\pi(A))\exp(-t\frac{q(A,A^c)}{1-\pi(A)}).$$

In section 4 we record results bounding $E_{\pi_0} T_A$ in terms of $E_\pi T_A$ for non-stationary initial distributions π_0. Let us emphasize that our results are "absolute" inequalities, i.e. do not involve any unspecified constants depending on the Markov chain under consideration. Without going into details, it follows that our results extend unchanged to *continuous*-space stationary reversible processes (under weak regularity conditions – say strong Markov and cadlag paths). For such processes we may have $\tau = \infty$, but the results are only interesting when $\tau < \infty$. One explanation is that the proofs extend, unchanged except for terminology. Another explanation is that, given a continuous-space process, one may express it as a weak limit of finite-state processes in such a way that relaxation times, first hitting times and the other parameters of interest converge.

Finally, we repeat that the existence of some exponential approximation has nothing to do with reversibility or even Markovianness: it is merely that in the reversible Markov context one can hope to get sharper general bounds. The non-reversible case, and applications to queueing networks, has recently been discussed by Iscoe and McDonald (1991), who give general explicit exponential approximations in terms of the spectral gap. Charles Stein, lecturing at Stanford in June 1991, outlined how to use the Chen-Stein method in the non-reversible case, giving bounds in terms of non-explicit "coupling times" which need ad hoc arguments to estimate. In older work, Flannery (1986) gave complicated bounds in terms of the maximal correlation function. Aldous (1982) gave bounds in terms of a uniform mixing coefficient. Exponential limits *without* explicit bounds can be proved in great generality. For instance, Korolyuk and Silvestrov (1984) and Cogburn (1985) prove exponential limits for hitting times to receding subsets of a fixed Harris-recurrent chain.

2 Background

It has recently become fashionable to treat reversible chains in terms of the associated *Dirichlet form* $\mathcal{E}$. That is, for functions $g : I \to R$ define

$$\begin{aligned}
\mathcal{E}(g,g) &= \frac{1}{2}\sum_i \sum_{j \neq i} \pi(i) q(i,j)(g(j)-g(i))^2 \\
&= \frac{1}{2}\lim_{t\to 0} t^{-1} E(g(X_t)-g(X_0))^2 \\
&= \lim_{t \to 0} t^{-1} E g(X_0)(g(X_0)-g(X_t)).
\end{aligned}$$

(X denotes the *stationary* chain). Write also

$$[g, g] = \sum_i \pi(i) g^2(i) = E g^2(X_0).$$

Various parameters of the chain have extremal characterizations, for example (Diaconis and Stroock (1991))

$$\tau = \sup\{[g, g]/\mathcal{E}(g, g) : \sum_i \pi(i) g(i) = 0\} \tag{2}$$

A more probabilistic interpretation is via the following *maximal correlation* property. For the stationary chain,

$$\begin{aligned}
\rho(t) &= \max\{\mathrm{cor}(Z_1, Z_2) : Z_1 \in \mathcal{F}(X_s, s \leq 0), Z_2 \in \mathcal{F}(X_s, s \geq t)\} \\
&= \max_{h,g} \mathrm{cor}(h(X_0), g(X_t)) \\
&= \exp(-t/\tau). \qquad (3)
\end{aligned}$$

As a standard consequence, if $Eg(X_0) = 0$ and $||g|| \equiv \sqrt{Eg^2(X_0)} < \infty$ then

$$||\mathbf{P_t} g|| \leq e^{-t/\tau} ||g||, \quad \text{where } (\mathbf{P}_t g)(i) = E(g(X_t)|X_0 = i). \tag{4}$$

Finally, we often use the fact that the tail distribution function $P_\pi(T_A > t)$ and the corresponding density function f are *completely monotone* (CM) functions, i.e. are of the form $\sum_i c_i \exp(-\gamma_i t)$ for nonnegative c_i, γ_i. Elementary qualitative properties of CM functions will be used without comment.

3 Proof of Theorem 1

Write

$$\rho_A = \frac{E_\pi T_A^2}{2(E_\pi T_A)^2} - 1.$$

Brown (1990) section 7 shows that for singletons $A = \{a\}$

$$|P_\pi(T_a/E_\pi T_a > t) - e^{-t}| \leq \frac{\rho_a}{\rho_a + 1}. \tag{5}$$

$$\rho_a \leq \tau / E_\pi T_a. \tag{6}$$

To prove Theorem 1 it suffices to show these remain true for general subsets A:

$$|P_\pi(T_A/E_\pi T_A > t) - e^{-t}| \leq \frac{\rho_A}{\rho_A + 1}. \tag{7}$$

$$\rho_A \leq \tau / E_\pi T_A. \tag{8}$$

Write $A^c = I \backslash A$. Consider the chain $\hat{X}$ in which the set A has been collapsed to a singleton a. Precisely, the state space is $\hat{I} = A^c \cup \{a\}$ and the transition rates are

$$\begin{array}{rcl} \hat{q}(i,j) = & q(i,j), & i,j \in I \backslash A \\ \hat{q}(i,a) = & \sum_{j \in A} q(i,j), & i \in I \backslash A \\ \hat{q}(a,i) = & \sum_{k \in A} \pi_k q(k,i)/\pi(A), & i \in I \backslash A \\ \hat{q}(a,a) = & -\sum_{i \in A^c} \hat{q}(a,i) & \end{array}$$

It is straightforward to verify (e.g. Keilson (1979) p.41) that $\hat{X}$ is reversible and has stationary distribution

$$\begin{array}{rcl} \hat{\pi}_i = & \pi_i, & i \in A^c \\ \hat{\pi}_a = & \pi(A) & \end{array}$$

By construction the P_π-distribution of T_A is the same as the $\hat{P}_{\hat{\pi}}$-distribution of $\hat{T}_a$. So (7) is immediate from (5). To get (8) the key fact, proved below, is

$$\hat{\tau} \leq \tau. \tag{9}$$

Then

$$\rho_A = \hat{\rho}_A \leq \frac{\hat{\tau}}{E_{\hat{\pi}} \hat{T}_a} \leq \frac{\tau}{E_\pi T_a}$$

which is (8).

Inequality (9) is one of a group of fairly well-known consequences of the extremal characterization (2) of the relaxation time. Given a function $\hat{g}$ on $\hat{I}$, consider its natural extension to a function g on I:

$$\begin{array}{rll} g(i) & = \hat{g}(i), & i \in A^c \\ & = \hat{g}(a), & i \in A. \end{array}$$

It is straightforward to verify

$$\sum_i \pi_i g(i) = \sum_i \hat{\pi}_i \hat{g}(i)$$

$$[g,g] = [\hat{g}, \hat{g}]$$

$$\mathcal{E}(g,g) = \hat{\mathcal{E}}(\hat{g}, \hat{g}).$$

Then (9) follows from (2).

4 Means from nonstationary starts.

If $E_\pi T_A \gg \tau$ then the approximation

$$E_{\pi_0} T_A \approx E_\pi T_A$$

holds for many non-stationary initial distributions π_0. In this section we study bounds for $E_{\pi_0} T_A$ obtainable by standard methods.

Lemma 6 *Let $h(i) = E_i T_A$. Then* $\mathrm{var}(h(X_0)) \leq \tau E_\pi T_A$.

It is easier to interpret this if we consider the standardized mean hitting time

$$\bar{h}(i) = E_i T_A / E_\pi T_A$$

for which Lemma 6 implies

$$\mathrm{var}\ (\bar{h}(X_0)) \leq \tau / E_\pi T_A.$$

So Chebyshev's inequality says that for "most" i the mean hitting time started at i is about the same as starting with the stationary distribution.

Proof of Lemma 6. Applying (2) to $g(i) = h(i) - E_\pi T_A$ gives

$$\mathrm{var}\ (h(X_0)) \leq \tau \mathcal{E}(h, h).$$

But the third formulation of $\mathcal{E}$, along with a routine dominated convergence argument, gives

$$\mathcal{E}(h, h) = Eh(X_0) \times 1 = E_\pi T_A.$$

Proposition 7 gives an upper bound on ET from an arbitrary initial distribution π_0. Define

$$\chi^2(\pi_0) = \sum_i \frac{(\pi_0(i) - \pi(i))^2}{\pi(i)} = ||\frac{\pi_0 - \pi}{\pi}||^2.$$

If X_0 has distribution π_0 then X_t has distribution π_t which satisfies (after a brief calculation)

$$\frac{\pi_t - \pi}{\pi} = \mathbf{P}_t \left(\frac{\pi_0 - \pi}{\pi} \right)$$

and then by (4)

$$\chi(\pi_t) \leq \chi(\pi_0) e^{-t/\tau}. \qquad (10)$$

Proposition 7 *For any initial distribution π_0*

$$E_{\pi_0} T_A \leq E_\pi T_A + \tau + \frac{\tau}{2} \log^+ \left(\chi^2(\pi_0) \frac{E_\pi T_A}{\tau} \right).$$

Taking $\pi_0 = \delta_i$ gives

Corollary 8

$$E_i T_A \leq E_\pi T_A + \tau + \frac{\tau}{2} \log^+ \left(\frac{1 - \pi(i)}{\pi(i)} \; \frac{E_\pi T_A}{\tau} \right).$$

Another Corollary is given below.

Proof of Proposition 7. For any function g

$$\begin{aligned} E_{\pi_0} g(X_t) - E_\pi g(X_t) &= \sum_i \pi(i) \frac{\pi_t(i) - \pi(i)}{\pi(i)} (g(i) - E_\pi g(X_t)) \\ &\leq \sqrt{\chi^2(\pi_t) \mathrm{var}_\pi g(X_0)} \text{ by Cauchy-Schwarz} \\ &\leq e^{-t/\tau} \chi(\pi_0) \sqrt{\mathrm{var}_\pi g(X_0)} \text{ by (10).} \end{aligned}$$

Now put $g(i) = E_i T_A$, observe that

$$E_\pi g(X_t) = E_\pi T_A, \quad E_{\pi_0} T_A \leq t + E_{\pi_0} g(X_t)$$

and apply Lemma 6 to get

$$E_{\pi_0} T_A - E_\pi T_A \leq t + \chi(\pi_0) \sqrt{\tau E_\pi T_A} \; e^{-t/\tau}.$$

Minimizing the right side by putting

$$t = \frac{\tau}{2} \log^+ \left(\chi^2(\pi_0) \frac{E_\pi T_A}{\tau} \right)$$

gives the result.

The next Corollary shows that in our setting i's with large $E_i T_A$ are exponentially rare with respect to π. It is easy to see that no similar result can hold for *small* $E_i T_A$.

Corollary 9 *For all $b \geq 0$*

$$\pi\{i : E_i T_A \geq E_\pi T_A + \tau + b\} \leq e^{-2b/\tau} \frac{E_\pi T_A}{\tau}.$$

Proof. Fix a subset B and set $\pi_B(i) = \pi(i|B)$. Then

$$\chi^2(\pi_B) = \frac{1-\pi(B)}{\pi(B)} \leq \frac{1}{\pi(B)}$$

and so Proposition 7 gives

$$E_{\pi_B} T_A \leq E_\pi T_A + \tau + \frac{\tau}{2} \log^+ \left(\frac{E_\pi T_A}{\tau \pi(B)} \right). \tag{11}$$

Now fix $b > 0$ and consider

$$B = \{i : E_i T_A \geq E_\pi T_A + \tau + b\}$$

which we may assume to be non-empty. Then

$$E_\pi T_A + \tau + b \leq E_{\pi_B} T_A.$$

Combining with (11) gives the inequality below:

$$\frac{2b}{\tau} \leq \log^+ \left(\frac{E_\pi T_A}{\tau \pi(B)} \right) = \log(\frac{E_\pi T_A}{\tau \pi(B)})$$

the equality because the left side is positive. Rearrange.

Remark. One can use Proposition 7 to give an upper bound on $E_\alpha T_A$, but the bound is weaker than the bound $E_\alpha T_A \leq E_\pi T_A + \tau$ given by Corollary 4. So the (novel ?) argument in section 6 seems more powerful than the standard arguments above.

5 Properties of the quasistationary distribution

Let Q_A be the transition rate matrix Q restricted to A^c. Let λ_A be the smallest eigenvalue of $-Q_A$ and let α be the corresponding eigenvector

$$\alpha Q_A = -\lambda_A \alpha \tag{12}$$

normalized to be a probability distribution on A^c. Then α is the *quasistationary distribution*, and

$$P_\alpha(T_A > t) = \exp(-\lambda_A t)$$

$$E_\alpha T_A = 1/\lambda_A.$$

And there is a variational interpretation:

$$E_\alpha T_A = \sup\{[g,g]/\mathcal{E}(g,g) : g \geq 0, g = 0 \text{ on } A\} \tag{13}$$

where the sup is attained by

$$g(i) = \alpha(i)/\pi(i). \tag{14}$$

It is often easy to find *lower* bounds on $E_\alpha T_A$ by evaluating (13) for convenient g.

Lemma 10 gives easy relations between $E_\pi T_A$ and $E_\alpha T_A$. Part (b) implies that $\alpha \approx \pi$ when $\tau / E_\pi T_A$ is small, using the general identity

$$\sum \frac{(\alpha_i - \pi_i)^2}{\pi_i} = \sum \frac{\alpha - i^2}{\pi_i} \ - 1.$$

Lemma 10 *(a)* $\frac{E_\pi T_A}{1-\pi(A)} \leq E_\alpha T_A \leq \tau/\pi(A)$.
(b) Now suppose $\tau < E_\pi T_A$. *Then*

$$\sum_i \alpha^2(i)/\pi(i) \leq (1 - \tau/E_\pi T_A)^{-1}.$$

Proof. Assertion (a) repeats part of Lemma 2. We shall give the usual proof of the right inequality, because ingredients are needed for subsequent parts. As at (14) set $g(i) = \alpha(i)/\pi(i)$, so

$$E_\alpha T_A = [g,g]/\mathcal{E}(g,g). \tag{15}$$

Applying (2)

$$\tau \geq \frac{[g-1,g-1]}{\mathcal{E}(g-1,g-1)} = \frac{[g,g]-1}{\mathcal{E}(g,g)}.$$

Using the expression for $\mathcal{E}(g,g)$ given in (15), we get

$$\frac{1}{[g,g]} \geq 1 - \frac{\tau}{E_\alpha T_A}. \tag{16}$$

Since α is a probability distribution on A^c we have

$$1 = E 1_{A^c}(X_0) g(X_0)$$

and so by Cauchy-Schwarz

$$1^2 \leq (E 1_{A^c}(X_0)) \times [g,g] = (1-\pi(A))[g,g].$$

Using (16),

$$1 - \frac{\tau}{E_\alpha T_A} \leq 1 - \pi(A)$$

which gives the right inequality in (a).

Now suppose $\tau / E_\pi T_A < 1$. Since $E_\alpha T_A \geq E_\pi T_A$ we can invert (16) to get

$$\begin{aligned} [g,g] &\leq (1 - \tau / E_\alpha T_A)^{-1} \\ &\leq (1 - \tau / E_\pi T_A)^{-1} \end{aligned} \tag{17}$$

which is (b).

6 Proof of Theorem 3.

We now come to Theorem 3, the main result which we believe to be new. There is no loss of generality in supposing that A is a singleton $\{a\}$, since (as in the proof of Theorem 1) we could replace the original chain by the "collapsed" chain $\hat{X}$.

The proof involves study of a certain irreducible reversible Markov chain on states $\{0, 1, 2, \ldots, r\}$ where the only allowable transitions are $0 \leftrightarrow i$. In graph theory terminology, the graph of such allowable transitions is a "star", so we shall call the chain X^* and write $\pi^*, \tau^*, q^*(i,0)$, etc, for quantities associated with X^*. We shall construct X^* such that $\mathcal{L}_{\pi^*} T_0^* = \mathcal{L}_\pi T_a$, (i.e. the distribution of T_0^* under π^* coincides with the distribution of T_a under π) and $E_{\alpha^*} T_0^* = E_\alpha T_a$. From the explicit form of X^* it is easy to see that the analog of Theorem 3 holds for X^*, and to complete the proof we demonstrate in Lemma 12 that $\tau^* \leq \tau$.

We shall make use of standard facts from the spectral representation for finite state irreducible reversible Markov chains. Keilson (1979) Chapter 3 and section 6.9 covers the relevant material. We first recall that

$$P_\pi(T_a > t) = \sum_{i=1}^{m} p_i \exp(-\gamma_i t) \tag{18}$$

where $p_i \geq 0$ and $0 < \gamma_1 < \gamma_2 < \ldots < \gamma_m$ are the distinct eigenvalues of $-Q_a$, the restriction of $-Q$ to $\{a\}^c$. We first show that the term in (18) corresponding to failure rate γ_1 has strictly positive coefficient p_1, and this identifies γ_1 as $1/E_\alpha T_a$.

Lemma 11 *$p_1 > 0$.*

Proof. For $x \neq a, y \neq a$ write $q_t(x,y) = P_x(X_t = y, T_a > t)$. By the spectral representation for Q_a

$$q_t(x,x) = \sum_{j=1}^{m} c_j(x,x) \exp(-\gamma_j t)$$

where $c_1(x,x) \geq 0$ and where $\sum_{x \neq a} c_1(x,x) =$ multiplicity of $\gamma_1 \geq 1$. Thus $c_1(x_0,x_0) > 0$ for some $x_0 \neq a$. Again by the spectral representation, for $y \neq x$

$$q_t(x,y) = \sum_{j=1}^{m} c_j(x,y) \exp(-\gamma_j t)$$

for certain constants $c_j(x,y)$. It follows that

$$0 \leq \lim_{t \to \infty} (\exp(\gamma_1 t) q_t(x,y)) = c_1(x,y).$$

Since $P_\pi(T_a > t) = \sum_{x,y \neq a} \pi(x) q_t(x,y)$, it follows that

$$p_1 = \sum_{x,y \neq a} \pi(x) c_1(x,y) \geq \pi(x_0) c_1(x_0,x_0) > 0$$

establishing the Lemma.

Now eliminate the zero-coefficient terms from (18) and relabel, so (18) becomes

$$P_\pi(T_a > t) = \sum_{i=1}^{r} p_i \exp(-\gamma_i t) \tag{19}$$

where $p_i > 0$, $1 \leq i \leq r$, and $0 < 1/E_\alpha T_a = \gamma_1 < \gamma_2 < \ldots < \gamma_r$ and $\sum_{i=1}^{r} p_i = P_\pi(T_a > 0) = 1 - \pi(a)$. Each γ_i is an eigenvalue of $-Q_a$.

Let X^* be the Markov chain on states $\{0, 1, \ldots, r\}$ with transition rate matrix Q^* defined by

$$\begin{aligned}
q^*(i,0) &= \gamma_i, \ i \neq 0 \\
q^*(0,i) &= \frac{\gamma_i p_i}{\pi(a)}, \ i \neq 0 \\
q^*(i,i) &= -\gamma_i, \ i \neq 0 \\
q^*(0,0) &= -\sum_{i=1}^{r} \frac{\gamma_i p_i}{\pi(a)} \\
q^*(i,j) &= 0 \text{ elsewhere.}
\end{aligned}$$

Here the quantities γ_i, p_i are those appearing in (19). The chain X^* is irreducible and reversible, with stationary distribution $\pi^* = (p_0, p_1, \ldots, p_r)$, where $p_0 = \pi(a)$.

Clearly the eigenvalues of Q_0^* are the (γ_i), and so in particular

$$E_{\alpha^*} T_0^* = 1/\gamma_1 = E_\alpha T_a \tag{20}$$

using Lemma 11. We also have the distributional identity

$$\mathcal{L}_{\pi^*} T_0^* = \mathcal{L}_\pi T_a \tag{21}$$

because

$$\begin{aligned} P_{\pi^*}(T_0^* > t) &= \sum_{i=1}^{r} \pi^*(i) P_i^*(T_0 > t) \\ &= \sum_{i=1}^{r} p_i \exp(-\gamma_i t) \\ &= P_\pi(T_a > T) \text{ by (19).} \end{aligned}$$

Next, note that the star chain has Dirichlet form

$$\mathcal{E}(g,g) = \sum_{i=1}^{r} p_i \gamma_i (g(i) - g(0))^2. \tag{22}$$

The particular function $g(i) = 1_{(i=1)} - p_1$ satisfies $\sum_{i=0}^{r} \pi^*(i) g(i) = 0$, and so by (2) and (22)

$$\tau^* \geq \frac{[g,g]}{\mathcal{E}(g,g)} = \frac{\sum_{i=0}^{r} p_i g^2(i)}{\sum_{i=1}^{r} p_i \gamma_i (g(i) - g(0))^2} = \frac{p_1(1-p_1)}{p_1 \gamma_1} = \frac{1-p_1}{\gamma_1}.$$

In other words

$$p_1 \geq 1 - \gamma_1 \tau^*. \tag{23}$$

Granted Lemma 12 below, we have

$$\begin{aligned} P_\pi(T_a > t) &\geq p_1 \exp(-\gamma_1 t) \text{ by (19)} \\ &\geq (1 - \gamma_1 \tau^*) \exp(-\gamma_1 t) \text{ by (23)} \\ &\geq (1 - \gamma_1 \tau) \exp(-\gamma_1 t) \text{ by Lemma 12} \\ &= (1 - \frac{\tau}{E_\alpha T_a}) \exp(-\frac{t}{E_\alpha T_a}) \text{ by (20)} \end{aligned}$$

establishing Theorem 3.

Lemma 12 *The eigenvalues of Q^* form a subset of those of Q. Consequently $\tau^* \leq \tau$.*

Proof. A classical Markov chain identity (Keilson (1979) p. 77) relates Laplace transforms of hitting times to Laplace transforms of transition densities. Writing

$$\tilde{p}(s) = \int_0^\infty P_a(X_t = a) e^{-st} dt$$

the identity asserts that for arbitrary initial distribution μ

$$\tilde{p}(s) \int_0^\infty e^{-st} dP_\mu(T_a < t) = \int_0^\infty P_\mu(X_t = a) e^{-st} dt.$$

Taking $\mu = \pi$ the right side becomes $\pi(a)/s$ and the identity shows that the distribution $\mathcal{L}_\pi T_a$ determines $\tilde{p}(s)$. So the distributional identity (21) implies

$$\tilde{p}(s) = \tilde{p}^*(s) \equiv \int_0^\infty P_0(X_t^* = 0) e^{-st} ds.$$

Now in terms of the matrix Q

$$\begin{aligned}
\tilde{p}(s) &= \int_0^\infty e^{-st} (e^{Qt})_{aa} \, dt \\
&= \left(\int_0^\infty e^{(Q-sI)t} \, dt \right)_{aa} \\
&= -(Q - sI)^{-1}_{aa} \\
&= -\frac{\det(Q_a - sI)}{\det(Q - sI)}.
\end{aligned}$$

Equating this with the corresponding expression for the star chain gives

$$\frac{\det(Q - sI)}{\det(Q^* - sI)} = \frac{\det(Q_a - sI)}{\det(Q_0^* - sI)}. \tag{24}$$

The eigenvalues $(\gamma_1, \ldots, \gamma_r)$ of Q_0^* are also, by (21), eigenvalues of Q_a. So the characteristic polynomial of Q_0^* is a divisor of the characteristic polynomial of Q_a. But then by (24) we see that the characteristic polynomial of Q^* is a divisor of the characteristic polynomial of Q. Thus the eigenvalues of Q^* are also eigenvalues of Q, establishing the Lemma.

Question. Does the relation between X and X^* fall into the general notion of "duality" described in Liggett (1985) page 84?

7 Bounds on the density function

For use in Aldous and Brown (1991) we need delicate estimates of the density function $f(t), t > 0$ of $T = T_A$ for the stationary chain, which imply that $f(t) \approx 1/E_\pi T_A$ on the range $\tau \ll t \ll E_\pi T_A$. By scaling we can reduce to the case $ET = 1$.

Lemma 13 *If $ET = 1$ then*
(a) $f(t) \leq 1 + \tau/(2t),\ t > 0$.
(b) $f(t) \geq 1 - 2\tau - t,\ t > 0$.
(c) $-f'(t) \leq 2$ for $t \geq \tau(5 + 2\log(1/\tau))$, provided $\tau \leq 1$.

Proof. f is CM and so is convex (Keilson (1979) p. 66). Then

$$\begin{aligned} 2tf(t) &\leq \int_0^{2t} f(s)ds \text{ by convexity} \\ &= P_\pi(0 < T < 2t) \\ &\leq P_\pi(0 \leq T < 2t) \\ &\leq 1 - e^{-2t} + \tau \text{ by Theorem 1} \\ &\leq 2t + \tau \end{aligned}$$

giving (a).

To prove (b), note that Corollary 4 and Lemma 2 imply

$$1 \leq E_\alpha T_A \leq 1 + \tau. \tag{25}$$

By CM, the "hazard rate" is decreasing (Keilson (1979) p. 75):

$$\frac{f(t)}{P_\pi(T_A > t)} \downarrow \frac{1}{E_\alpha T_A} \text{ as } t \to \infty.$$

So

$$\begin{aligned} f(t) &\geq \frac{1}{E_\alpha T_A} P_\pi(T_A > t) \\ &\geq \frac{1}{E_\alpha T_A}(\exp(-\frac{t}{E_\alpha T_A}) - \tau) \text{ by Theorem 3} \\ &\geq \frac{1}{E_\alpha T_A}(1 - \frac{t}{E_\alpha T_A} - \tau) \\ &\geq \frac{1}{E_\alpha T_A}(1 - t - \tau) \text{ by (25)} \\ &\geq \frac{1}{1+\tau}(1 - t - \tau) \text{ by (25)} \\ &\geq 1 - t - 2\tau \end{aligned}$$

giving (b). In the inequalities above we implicitly assumed $1 - t - \tau \geq 0$, since otherwise the result is trivial.

The proof of (c) has several ingredients. Excursions of the stationary chain inside A alternate with excursions outside A. Let $r(l)dl$ be the rate of excursions outside A of length $\in (l, l + dl)$. Then an easy renewal theory argument (Aldous and Brown (1991) Lemma 3) shows

$$f(l) = -f'(l). \tag{26}$$

Now consider the joint density $\theta(t_1, t_2)$ of (T_A^-, T_A), where $T_A^- = \min\{t \geq 0 : X_{-t} \in A\}$ and the stationary chain is extended to $-\infty < t < \infty$. By conditioning on X_0 we see

$$\theta(t_1, t_2) = E f_{t_1}(X_0) f_{t_2}(X_0).$$

But in the notation of (26), $\theta(t_1, t_2) = r(t_1 + t_2)$. So putting $t_1 = t_2 = t$,

$$E f_t^2(X_0) = \theta(t, t) = r(2t) = -f'(2t) \tag{27}$$

the final equality by (26).

For the second ingredient, let $f_t(i)$ be the density of T_A under P_i. We shall show

$$E_\pi f_{s+t}^2(X_0) \leq f^2(t) + e^{-2s/\tau} E_\pi f_t^2(X_0). \tag{28}$$

For $s, t > 0$ consider

$$g_{s,t}(i)dt = P_i(X_u \in A^c \text{ for } s \leq u \leq s + t, X_{s+t+dt} \in A).$$

Then $g_{s,t} = \mathbf{P}_s f_t$ in the notation of (4), and the assertion of (4) gives (for the stationary chain)

$$\text{var } g_{s,t}(X_0) \leq e^{-2s/\tau} \text{var } f_t(X_0).$$

Plainly $f_{s+t} \leq g_{s,t}$ and $E g_{s,t}(X_0) = E f_t(X_0) = f(t)$, so

$$\begin{aligned} E f_{s+t}^2(X_0) &\leq E g_{s,t}^2(X_0) \\ &= f^2(t) + \text{var } g_{s,t}(X_0) \\ &\leq f^2(t) + e^{-2s/\tau} \text{var } f_t(X_0) \end{aligned}$$

and (28) follows. Next we shall show

$$-f'(2t) \leq t^{-2}/2. \tag{29}$$

For the function $-f'(u)$ is decreasing, so

$$f(u) \geq (2t - u)(-f'(2t)) \text{ on } 0 < u < 2t$$

and hence

$$1 \geq \int_0^{2t} f(u)du \geq 2t^2(-f'(2t))$$

giving (29).

Combining (27-29) shows that, for any $s, t > 0$,

$$-f'(2s + 2t) \leq f^2(t) + e^{-2s/\tau}t^{-2}/2.$$

Putting $t = 2\tau$ and appealing to part (a)

$$-f'(2s + 4\tau) \leq (5/4)^2 + e^{-2s/\tau}/(8\tau^2).$$

Then putting $s = \tau(\log(1/\tau) + 1/2)$ establishes part (c).

References

[1] D.J. Aldous. Markov chains with almost exponential hitting times. *Stochastic Process. Appl.*, 13:305–310, 1982.

[2] D.J. Aldous. *Probability Approximations via the Poisson Clumping Heuristic.* Springer-Verlag, 1989.

[3] D.J. Aldous and M. Brown. Inequalities for rare events in time-reversible Markov chains II. *Stochastic Process. Appl.*, 44:15–25, 1993.

[4] M. Brown. Consequences of monotonicity for Markov transition functions. Technical report, City College, CUNY, 1990.

[5] R. Cogburn. On the distribution of first passage and return times for small sets. *Annals of Probability*, 13:1219–1223, 1985.

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

[7] B.R. Flannery. *Quasi-stationary Distributions for Markov Chains.* PhD thesis, U.C. Berkeley, 1986.

[8] I. Iscoe and D. McDonald. Asymptotics of exit times for Markov jump processes I. *Ann. Probab.*, 22:372–397, 1994.

[9] J. Keilson. *Markov Chain Models - Rarity and Exponentiality.* Springer-Verlag, 1979.

[10] D.V. Korolyuk and D.S. Sil'vestrov. Entry times into asymptotically receding domains for ergodic Markov chains. *Theory Prob. Appl.*, 19:432–442, 1984.

[11] T.M. Liggett. *Interacting Particle Systems.* Springer-Verlag, 1985.

[2, 3, 4, 6, 9, 8, 7, 1, 11] [5, 10]