Meeting Times for Independent Markov Chains

David J. Aldous

Technical Report No. 118
October 1987
(revised August 1990}

Research supported by NSF Grant DMS84-03239.

Department of Statistics
University of California
Berkeley, California

Abstract.

Start two independent copies of a reversible Markov chain from arbitrary initial states. Then the expected time until they meet is bounded by a constant times the maximum first hitting time for the single chain. This and a sharper result are proved, and several related conjectures are discussed.

1. Introduction.

Let (X_t) be an irreducible continuous-time pure jump Markov chain on finite state space $I = \{i, j, k, ...\}$ with stationary distribution π. Classical theory says $P(X_t = j) \to \pi_j$ as $t \to \infty$ for all j, regardless of the initial distribution. The modern "coupling" proof goes as follows. Let (Y_t) be an independent copy of the chain. Then (X_t, Y_t), considered as a chain on $I \times I$, is irreducible and hence the *meeting time*

$$T_M \equiv \min\{t : X_t = Y_t\}$$

is a.s. finite, regardless of the initial distributions. Now give Y_0 the stationary distribution and define

$$\begin{aligned} \hat{X}_t &= X_t,\ t < T_M \\ &= Y_t,\ t \geq T_M. \end{aligned}$$

Then $(\hat{X}_t)$ has the same distribution as (X_t). So

$$\begin{aligned} |P(X_t = j) - \pi_j| &= |P(\hat{X}_t = j) - P(Y_t = j)| \\ &\leq P(X_t \neq Y_t) \\ &= P(T_M > t) \to 0 \text{ as } t \to \infty. \end{aligned}$$

Asmussen (1987) gives a good account of this and other coupling arguments.

Given a simple proof of a fundamental result, it is natural to probe more deeply into the surrounding issues. The argument above can be quantified as follows. Let $d_i(t)$ be the total variation distance between π and the distribution of X_t given $X_0 = i$:

$$d_i(t) = \frac{1}{2}\sum_j |P_i(X_t = j) - \pi_j|.$$

Then one obtains

$$d_i(t) \leq P(T_M > t \mid X_0 = i, Y_0 \overset{D}{=} \pi). \tag{1}$$

This leads to the idea of *maximal coupling*: there is a dependent construction of X_t and Y_t such that equality holds in (1). See Thorisson (1986) for a recent account. This paper goes in a different direction, to study the meeting time of independent chains as a quantity in its own right, and to compare this quantity with other quantities associated with the Markov chain.

A natural object of study is the worst-case mean meeting time

$$\tau_M \equiv \max_{i,j} E(T_M \mid X_0 = i, Y_0 = j). \tag{2}$$

Inequality (1) can be used to relate this to a parameter τ_1 indicating the time taken for the distribution of the single chain to approach the stationary distribution. Define

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

(3) $$\tau_1 = \min\{t : d(t) \le 1/(2e)\}$$

(the constant $1/(2e)$ has no special significance beyond algebraic convenience). Then (1) and Markov's inequality give

(4) $$\tau_1 \le 2e\tau_M.$$

Aldous (1982) studied τ_1 and showed that for reversible chains it is "equivalent" to various other parameters τ, in the sense that

$$\tau_1 \le K\tau; \quad \tau \le K\tau_1$$

where here and throughout K denotes an absolute constant, not depending on the chain or the number of states (K varies from line to line).

In this paper we seek similar results for τ_M. It is easy to see that τ_M may be much larger than τ_1: consider the chain which holds at a state for an exponential (1) time and then jumps to a uniform random state. It seems natural to try to relate τ_M to hitting times

$$H_j = \min\{t : X_t = j\}$$

for the single chain. Let us consider two examples.

(5) **Example.** Consider continuous-time simple symmetric random walk on the integer lattice Z^d modulo q. Then the distance $X_t - Y_t$ between independent walks behaves precisely as X_{2t}, the single walk with transition rates doubled. Hence in this example $\tau_M = \frac{1}{2} \max_{i,j} E_i H_j$.

(6) **Example.** Consider the continuous-time analog of deterministic cycling. That is, take $I = \{0, 1, \ldots, N-1\}$ and transition rates $q_{i,i+1} = 1 = q_{N-1,0}$. Then $\max_{i,j} E_i H_j = N - 1$. Now if X_t, Y_t are independent walks then $X_t - Y_t$ modulo N is symmetric random walk. So by considering $X_0 - Y_0 = [\frac{1}{2} N]$, the central limit theorem shows that τ_M is of order N^2.

The behavior in example 6 is in a sense a pathology caused by cyclicity: we can eliminate this by restricting attention to reversible chains. The exact equality in example 5 arises from spatial homogeneity and cannot be expected elsewhere, but it turns out there is a bound.

(7) **Proposition.** $\tau_M \le K \max_{i,j} E_i H_j$ *for all reversible chains.*

Our argument for (7) is indirect and yields a large K, but it is conceivable that $K = \frac{1}{2}$ suffices. Example 6 shows no such bound can hold for irreversible chains, but suggests

(8) **Conjecture.** $\tau_M \le KN \max_{i,j} E_i H_j$ *for all chains, where* $N =$ *number of states.*

This conjecture seems curiously difficult: the author can do no better than a N^3 bound.

Returning to the reversible case, adding a very rarely-visited state j may make $E_i H_j$ large without affecting τ_M, so the bound in (7) may not be the correct order of magnitude. There is a better bound, in which the $E_i H_j$ are averaged using the stationary distribution.

(9) **Proposition.** *For all reversible chains*

$$\tau_M \le K \{\sum_i \frac{\pi_i}{\tau_1 \vee E_\pi H_i}\}^{-1}.$$

Here $a \vee b \equiv \max(a, b)$ and E_π denotes the stationary initial distribution, so $E_\pi H_i = \sum_k \pi_k E_k H_i$. In words, the bound is the π-weighted harmonic mean of the $\tau_1 \vee E_\pi H_i$. Though complicated, the bound does involve only quantities associated with the single chain. We conjecture that this is the correct bound, in that the opposite inequality holds:

(10) **Conjecture.** *For all reversible chains*

$$\{\sum_i \frac{\pi_i}{\tau_1 \vee E_\pi H_i}\}^{-1} \le K \tau_M.$$

The author can obtain only the weaker result

$$\min_i E_\pi H_i \le K \tau_M.$$

The mathematical content of this paper is the proof of Proposition 9: we shall see that Proposition 7 is a consequence. The proof is an interesting use of the "harmonic mean formula" idea for estimating probabilities of rare events: see Aldous (1989a,b) for different applications. The form of the bound in (9) may look like an artifact of the proof, but example 12 below is rather convincing that (9) is the correct bound. Calculations with 2-state chains show that the τ_1 term in Proposition 9 cannot be omitted.

Although these meeting time questions have (apparently) not been studied before in this generality, a more complicated related question has been studied. Start a copy of the Markov chain from every state, and let the chains run independently except that chains coalesce when they meet. At some random time T_C all the chains have coalesced into one chain. This process, where the underlying chain is simple random walk on an infinite integer lattice, arises as a dual process to voter models - see Liggett (1985) - and in finite settings has been studied by Donnelly and Welsh (1983), Cox (1989). Write $\tau_C = ET_C$. Clearly $\tau_C \geq \tau_M$. It is easy to see that $\tau_C \leq K\tau_M \log N$, where N is the number of states. In natural examples, such as random walk on the d-dimensional torus, it turns out that $\tau_c \leq K\tau_M$. Ted Cox (private communication) has observed this is false in general (consider random walk on a ''star'' graph), but the following (partly vague) conjecture is open.

(11) **Conjecture.** *(a) For all reversible chains,* $\tau_C \leq K \max_{i,j} E_i H_j$.

(b) Under suitable symmetry conditions, $\tau_C \leq K\tau_M$.

We end this introduction with an instructive example.

(12) **Example.** Take state space $\{0, 1, \ldots, N-1; \Delta\}$ with transition rates

$q_{i,j} = 1$ if $j = i \pm 1$ modulo N

$q_{i,\Delta} = N^{-b}$, $q_{\Delta,i} = N^{-a-1}$, $0 \leq i \leq N-1$.

Here $0 < a < b < 2$ are fixed, and it is easy to see the order of magnitude (as $N \to \infty$) of the various quantities:

$\pi(\Delta) \approx N^{a-b}$; $\pi(i) \approx N^{-1}$ for $i \neq \Delta$

$E_\pi H_\Delta \approx N_b$; $E_\pi H_i \approx N^{1+\frac{1}{2}b}$ for $i \neq \Delta$

$\tau_1 \approx N^b$.

Now the first meeting time T_M for two independent chains X_t, Y_t can be regarded as $\min(T_1, T_2)$, where

$T_1 = \min\{t : X_t = Y_t = \Delta\}$

$T_2 = \min\{t : X_t = Y_t \neq \Delta\}$.

One can show

$$ET_1 \approx N^{2b-a};\ ET_2 \approx N^{1+\frac{1}{2}b}$$

and hence $\tau_M \approx N^{(2b-a) \wedge (1+\frac{1}{2}b)}$. Now looking at the bound in Proposition 9,

$$\pi(\Delta)/E_\pi H_\Delta \approx N^{a-2b};\ \sum_{i\neq\Delta} \pi(i)/E_\pi H_i \approx N^{-(1+\frac{1}{2}b)}$$

and the bound works out as $\approx N^{(2b-a)\wedge(1+\frac{1}{2}b)}$. Thus although the qualitative behavior of T_M changes according to whether $2b-a$ or $1+\frac{1}{2}b$ is larger, our bound tracks this change correctly.

(13) **Remark.** Though stated for finite-state chains, the fact that the constants K do not depend on the number of states implies the results extend to general state space. In most cases such extensions are uninteresting since the bounds will be infinite. An exception is that one can construct "Brownian motion" on certain compact fractal sets in R^d as a limit of random walks on graphs: see e.g. Lindstrom (1990), Barlow and Perkins (1988). If such a process hits single points a.s., then our results suggest that two independent processes will meet a.s., and this is indeed true (Krebs (1990)).

2. Ingredients of the Proof.

The proof to be given in section 3 is a concoction of three rather diverse ingredients, which will be set out in this section.

The first is the recurrent-potential formula for mean hitting times. In any finite state Markov chain,

(14) $$E_\pi H_i = R_i/\pi_i, \text{ where}$$

(15) $$R_i = \int_0^\infty (p_{ii}(s) - \pi_i)\, ds.$$

This can be deduced from matrix expressions for $E_j H_i$ in Kemeny and Snell (1960) in discrete time, and then extended to continuous time: a simpler argument based on renewal theory is in Aldous (1983). Though (15) does not assume reversibility, its use for bounding mean hitting times is helped by the fact

(16) in a reversible chain, $p_{ii}(t)$ decreases to π_i as $t \to \infty$.

This follows from the spectral representation: Keilson (1979) Section 3.3.

The second set of ingredients are bounds from Aldous (1982) which relate the parameter τ_1 of (3) to other quantities. As in section 1, K denotes an absolute constant, different from line to line.

(17) **Proposition.** *For reversible chains,*

(a) $\tau_1 \le K \max_{i,k} \sum_j \pi_j |E_i H_j - E_k H_j|$.

(b) *There exist stopping times* U_i *such that* $E_i U_i \le K \tau_1$ *and* $\text{dist}(X_{U_i} | X_0 = i) = \pi$.

Note that (a) implies the much weaker result

(18) $$\tau_1 \le K \max_{i,j} E_i H_j.$$

This enables us to deduce Proposition 7 from Proposition 9. For Proposition 9 certainly implies

$$\tau_M \le K \max_i (\tau_1 \vee E_\pi H_i)$$

and then (18) gives Proposition (7). Next, consider independent copies of the chain (X_t, Y_t) as a chain on state space $I \times I$, and let τ_1^* be defined as at (3) for this product chain. It is easy to show, using the submultiplicative property of 2d(t) (see Aldous (1982)), that $\tau_1^* \le K \tau_1$. So (17b) gives

(19) **Corollary.** *For independent copies* (X_t, Y_t) *of a reversible chain, and for any* i, j, *there exists a stopping time* U *such that*

$\text{dist}(X_U, Y_U | X_0 = i, Y_0 = j) \; = \; \pi \times \pi;$

$E(U | X_0 = i, Y_0 = j) \; \le \; K \tau_1.$

The third ingredient is the starting idea of what the author calls "harmonic mean formulas" for estimating first hitting times. Let (Z_t) be a stationary process, and suppose A is such that the sojourns of Z in A and in A^c form successive non-trivial time intervals. Write L_t for the Lebesgue measure of $\{0 \le s \le t : Z_s \in A\}$. Then

$$1_{(L_t > 0)} \; = \; \int_0^t L_t^{-1} 1_{(Z_s \in A)} \, ds$$

(interpreting the integrand as 0 for $Z_s \in A^c$). So taking expectations,

$$P(Z_s \in A \text{ for some } 0 \le s \le t) \; = \; P(L_t > 0)$$

(20) $$= \int_0^t E(L_t^{-1}; Z_s \in A) \, ds.$$

3. Proof of Proposition 9.

We first give the proof under the extra assumption

(21) $$\max_i \pi_i \; \le \; 2 \min_i \pi_i$$

and will then show the general case can be reduced to this case by a "splitting states" technique.

Let X_t, Y_t be independent copies of the chain with the stationary initial distribution π. Applying (20) to the stationary process (X_t, Y_t) and to $A = \{(k,k) : k \in I\}$ gives

$$P(X_s = Y_s \text{ for some } 0 \le s \le t) = \int_0^t E(L_t^{-1}; X_s = Y_s)\, ds$$

where $L_t = \int_0^t 1_{(X_s = Y_s)}\, ds$;

$$= \int_0^t \sum_i E(L_t^{-1} \mid Z_s = Y_s = i)\, \pi_i^2\, ds$$

$$\ge (2N)^{-2} \sum_i \int_0^t E(L_t^{-1} \mid X_s = Y_s = i)\, ds$$

where N is the number of states, since (21) implies $\pi_i \ge (2N)^{-1}$;

$$\ge (2N)^{-2} \sum_i \int_0^t \{E(L_t \mid X_s = Y_s = i)\}^{-1}\, ds \tag{22}$$

by Jensen's inequality.

So we consider, for $0 \le s \le t$,

$$E(L_t \mid X_s = Y_s = i) \le 2 \int_0^t P(X_s = Y_s \mid X_0 = Y_0 = i)\, ds \quad \text{using reversibility}$$

$$= 2 \int_0^t \sum_j p_{ij}^2(s)\, ds \quad \text{by independence}$$

$$= 2 \int_0^t \sum_j p_{ij}(s)\, p_{ji}(s)\, \pi_j / \pi_i\, ds \quad \text{by reversibility}$$

$$\le 4 \int_0^t \sum_j p_{ij}(s)\, p_{ji}(s)\, ds \quad \text{by (21)}$$

$$= 4 \int_0^t p_{ii}(2s)\, ds$$

$$= 2 \int_0^{2t} p_{ii}(s)\, ds$$

$$= 2 \int_0^{2t} (p_{ii}(s) - \pi_i)\, ds + 4t\pi_i$$

$$\le \; 2R_i + 4t\pi_i \;\text{ by (16), for } R_i \text{ as at (15);}$$

$$\le \; 4\pi_i (E_\pi H_i + t) \;\text{ by (14)}$$

$$\le \; 16N^{-1} \max (E_\pi H_i, t) \;\text{ since } \pi_i \le 2/N \text{ by (21).}$$

Putting this together with (22), and putting $t = \tau_1$,

$$P(X_s = Y_s \text{ for some } 0 \le s \le \tau_1) \;\ge\; (64N)^{-1} \tau_1 \sum_i (\max (E_\pi H_i, \tau_1))^{-1}$$

(23) $$\ge\; (128)^{-1} \tau_1 / \tau_h \;=\; \alpha, \text{ say,}$$

where $\tau_h = \{\sum_i \pi_i / \max (E_\pi H_i, \tau_1)\}^{-1}$ is the desired bound for Proposition 9, and where we used $\pi_i \le 2/N$ again.

The inequality (23) applies to the case where X_0, Y_0 are independent with distribution π. Consider now the case where X_0 and Y_0 are arbitrary. Using Corollary 19 we can construct stopping times $S_n = \sum_{m=1}^{n} U_m$ such that

$U_m \;\ge\; \tau_1$;

(X_{S_n}, Y_{S_n}) has distribution $\pi \times \pi$ and is independent of $\mathbf{F}_{n-1} =$

$\sigma (X_t, Y_t : t \le S_{n-1} + \tau_1)$;

(24) $E(U_n | \mathbf{F}_{n-1}) \;\le\; K\tau_1$.

Then the meeting time T_M satisfies

(25) $$T_M \;\le S_\xi + \tau_1,$$

where $\xi = \min \{n : X_{S_n+u} = Y_{S_n+u}$ for some $0 \le u \le \tau_1\}$. By (24) and the optional sampling theorem,

$$ES_\xi \;\le\; K\tau_1 \, E\xi.$$

But by (23) and the independence property of our construction,

$$P(\xi > m) \;\le\; (1-\alpha)^m, \;\; m \ge 1,$$

and so $E\xi \le \alpha^{-1}$. Thus from (25) and (23),

$$ET_M \;\le\; \tau_1 + K\tau_1/\alpha$$

$$\le\; \tau_1 + K\tau_h$$

(26) $\leq K\tau_h$ since $\tau_h \geq \tau_1$ by definition.

This completes the proof under assumption (21). Consider now a reversible chain (X_t) on I with arbitrary π: we shall show that (26) remains true with the same K. We can choose integers $M_i \geq 1$ such that (π_i / M_i) satisfies (21). Define a chain $Z_t = (\hat{X}_t, V_t)$ on $I^* = \{(i,m) : i \in I, 1 \leq m \leq M_i\}$ with transition rates

$(i,m) \to (j,m')$ rate q_{ij}/M_j $(i \neq j \in I;\ 1 \leq m \leq M_i,\ 1 \leq m' \leq M_j)$

$(i,m) \to (i,m')$ rate $\gamma (i \in I;\ 1 \leq m \neq m' \leq M_i)$,

where q_{ij} are the transition rates of X and γ is arbitrary. Then $(\hat{X}_t)$ is a copy of (X_t). And (Z_t) is reversible and has stationary distribution $\pi^*(i,m) = \pi_i / M_i$. So Z satisfies (21) and hence (26). This is true for any value of γ. As $\gamma \to \infty$ there is probability $\to 1$ that, during a visit of X_t to i, V_t will visit all states $1 \leq m \leq M_i$. It is easy to deduce that, writing H^γ and T_M^γ for hitting and meeting times for Z,

$$E_{\pi^*}(H_{i,m}^\gamma) \to E_\pi H_i \text{ as } \gamma \to \infty.$$

Also, for all γ,

$$E(T_M | X_0 = i, Y_0 = j) \leq E(T_M^\gamma | \hat{X}_0 = i, \hat{Y}_0 = j).$$

Thus we can pass to the limit in (26) and see that (26) holds for (X_t).

References

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

Aldous, D.J. (1983). Random walks on finite groups and rapidly mixing Markov chains. In: *Seminaire de Probabilites XVII* (Springer Lecture Notes 986) pp243-297.

Aldous, D.J. (1989a). *Probability Approximations via the Poisson Clumping Heuristic.* Springer, New York.

Aldous, D.J. (1989b). The harmonic mean fomula for probabilities of unions: applications to sparse random graphs. *Discrete Math.* **76** 167-176.

Asmussen, S. (1987). *Applied Probability and Queues.* Wiley, New York.

Barlow, M.T. and Perkins, E.A. (1988). Brownian motion on the Sierpinski gasket. *Probab. Th. Rel. Fields* **79** 543-624.

Cox, J.T. (1989). Coalescing random walks and voter model consensus times on the torus in Z^d. *Ann. Probab.* **17** 1333-1366.

Donnelly, P. and Welsh, D. (1983). Finite particle systems and infection models. *Math. Proc. Cambridge Phil. Soc.* **94** 167-182.

Keilson, J. (1979). *Markov Chain Models: Rarity and Exponentiality.* Springer, New York.

Kemeny, J.G. and Snell, J.L. (1960). *Finite Markov Chains.* Van Nostrand, New York.

Krebs, W.B. (1990). Hitting time bounds for Brownian motion on a fractal. Technical Report, Florida State Univ.

Liggett, T.M. (1985). *Interacting Particle Systems.* Springer, New York.

Lindstrom, T. (1990). Brownian motion on nested fractals. *Memoirs of the Amer. Math. Soc.* **420**.

Thorisson, H. (1986). On maximal and distributional coupling. *Ann. Probab.* **14** 873-876.