Mixing Time for a Markov Chain on Cladograms

David J. Aldous*
University of California
Department of Statistics
367 Evans Hall # 3860
Berkeley CA 94720-3860
aldous@stat.berkeley.edu
http://www.stat.berkeley.edu/users/aldous

October 14, 1998

Abstract

A cladogram is a tree with labeled leaves and unlabeled degree-3 branchpoints. A certain Markov chain on the set of n-leaf cladograms consists of removing a random leaf (and its incident edge) and reattaching it to a random edge. We show that the mixing time (time to approach the uniform stationary distribution) for this chain is at least $O(n^2)$ and at most $O(n^3)$.

AMS 1991 subject classification: 05C05, 60C05, 60J10.

Key words and phrases. Coupling, Markov chain, Markov chain Monte Carlo, mixing time, phylogenetic tree.

*Research supported by N.S.F. Grant DMS96-22859

1 Introduction

By *cladogram* we mean an unrooted tree with $n \geq 4$ *leaves* (degree-1 vertices) labeled $\{1, 2, \ldots, n\}$ and with $n-2$ unlabeled internal vertices (*branchpoints*) of degree 3. See figure 2 (ignore arrows for now). The name comes from biological systematics [13] where a cladogram represents the evolutionary relationship between n species. (A more common though less precise phrase is *phylogenetic tree*, but we use this phrase to denote a cladogram in which edge-lengths are positive real numbers: section 2.1. The term *semi-labeled tree* is also used.) Reconstructing phylogenetic trees from DNA data is of major biological interest: Holmes [15] gives a recent survey for statisticians. Some current work in the area [17, 18] uses Markov chain Monte Carlo. That is, one starts with a "base chain" on tree-space, and then seeks to sample from a complicated data-dependent distribution on tree-space by accepting or rejecting moves proposed by the base chain. As with other applications of Markov chain Monte Carlo [14], theoretical analysis of running times of such algorithms (the central issue being to estimate a mixing time) in the context of genuine data seems intractable. Nevertheless one can hope to gain insight by studying simple base chains on the set $\mathbf{T}_n$ of n-leaf cladograms. In this paper we study what is perhaps the conceptually simplest chain, informally "move a random leaf to a random position". In practice biologists typically use more complicated "branch-swapping" base chains, though Susan Holmes reports that the chain studied here is used in a program LVB [6]. A different chain is studied by Diaconis and Holmes [11] using an isomorphism between $\mathbf{T}_n$ and the set of matchings of $\{1, 2, \ldots, 2n\}$. An introductory presentation of tree space can be found in [8].

1.1 Model and results

Figure 1 illustrates one step of our chain. Start with a tree $\mathbf{t}$, partly shown on the left side of figure 1. Remove some leaf i. Precisely, i is attached by some edge e_1 to some branchpoint b where two other edges e_2, e_3 are incident; delete edge e_1 and branchpoint b, merging the two edges e_2, e_3 into a single edge e. The resulting tree has $2n-5$ edges. Choose some edge f, and re-attach leaf i to edge f. Precisely, create a branchpoint b' which splits edge f into two edges f_2, f_3 and attach leaf i to branchpoint b' via a new edge f_1. This creates a new tree $\mathbf{t}'$.

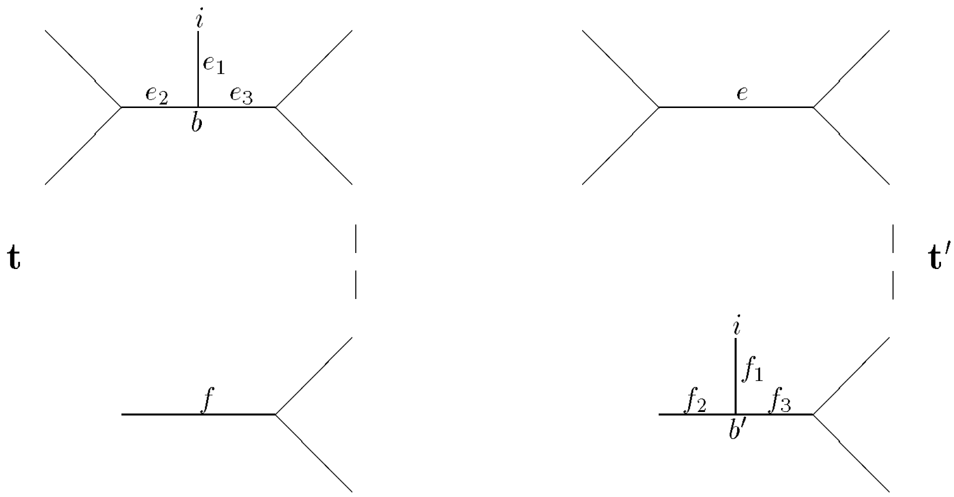


Figure 1

Write $\mathbf{t}' \sim \mathbf{t}$ if $\mathbf{t}'$ can be obtained in this way from $\mathbf{t}$, using some choice of (i, f), and if $\mathbf{t}' \neq \mathbf{t}$. It is easy to check that all the $n(2n-6)$ choices of (i, f) with $f \neq e$ give trees which are distinct from each other and from $\mathbf{t}$; of course the n choices with $f = e$ give back the original tree $\mathbf{t}$. Thus we can define our $\mathbf{T}_n$-valued chain $(X_s, s = 0, 1, 2, \ldots)$ verbally by

> pick uniformly at random a leaf i to remove, then pick uniformly at random an edge f and re-attach i to f

and its transition matrix is

$$\begin{aligned} P_n(\mathbf{t}, \mathbf{t}') &= \tfrac{1}{n(2n-5)} \text{ if } \mathbf{t}' \sim \mathbf{t} \\ &= \tfrac{n}{n(2n-5)} \text{ if } \mathbf{t}' = \mathbf{t}. \end{aligned} \tag{1}$$

It is easy to see that the chain is symmetric, aperiodic and irreducible. Classical theory of finite Markov chains says

$$d_n(s) := \max_{\mathbf{t} \in \mathbf{T}_n} \max_{A \subset \mathbf{T}_n} |P(X_s \in A | X_0 = \mathbf{t}) - \pi_n(A)| \to 0 \text{ as } s \to \infty \tag{2}$$

where π_n is the uniform distribution on $\mathbf{T}_n$. Modern theory [5, 20] gives various formalizations of the notion of the time to approach stationarity. Define the (total variation) *mixing time*

$$\tau_n := \min\{s : d_n(s) \leq 1/4\}, \tag{3}$$

so that by general submultiplicativity arguments [5]

$$d_n(j\tau_n) \leq 2^{-j}, \ j = 1, 2, 3, \ldots.$$

Define the *relaxation time* as "1/spectral gap":

$$\tau_n^{\text{rel}} := 1/(1 - \lambda_{n,2})$$

where $1 = \lambda_{n,1} > \lambda_{n,2}$ are the largest eigenvalues of the transition matrix P_n. The purpose of this paper is to prove

Theorem 1 *There exist $0 < c_1$ and $c_2 < \infty$ such that for all large n*
(a) $\tau_n^{\text{rel}} \geq c_1 n^2$
(b) $\tau_n \leq c_2 n^3$.

In general ([1], [5] Chapter 4) we have $\tau_n^{\text{rel}} = O(\tau_n)$, so Theorem 1 sandwiches both parameters between order n^2 and order n^3. We conjecture that in fact both are order n^2: see section 4, where we point out an analogy with the voter model on the complete graph.

One can make a loose analogy between Markov chains on $\mathbf{T}_n$ and Markov chains on the permutation group of $\{1, \ldots, n\}$ (visualized as card-shuffling schemes) which have been studied in some detail [4, 7, 9, 10] and which suggest techniques for the tree-valued setting. But Theorem 1 may be surprising to those familiar with the card-shuffling literature. Our chain is loosely analogous to the shuffle "pick a card at random and move it to a random position" for which the mixing time can be shown to be order $n \log n$, the time until each card has been moved (cf. [9]). So one might expect our mixing times to be order $n \log n$ also, since "each leaf has been moved to a random position". However, although the leaf labels get randomized in order $n \log n$ steps, the "global shape" of the tree takes order n^2 steps to become randomized. See section 4 for further discussion.

Our proof of Theorem 1 involves several ingredients.

- The extremal characterization of spectral gap (section 2).
- Coupling and marking constructions, familiar from e.g. card-shuffling chains (section 3.1).
- Hitting times for birth-and-death chains (section 3.2).
- $n \to \infty$ distributional asymptotics for the uniform distribution on $\mathbf{T}_n$ (section 2.1).

2 The lower bound

Applying to our chain the well-known extremal characterization (e.g. [12]) of the eigenvalue $\lambda_{n,2}$ (and hence of the relaxation time) gives

$$\tau_n^{\mathrm{rel}} = \sup_{f_n : \mathbf{T}_n \to R} \frac{2 \ \mathrm{var}_{\pi_n} \ f_n}{\sum_{\mathbf{t}_1} \sum_{\mathbf{t}_2} \pi_n(\mathbf{t}_1) P_n(\mathbf{t}_1, \mathbf{t}_2)(f_n(\mathbf{t}_2) - f_n(\mathbf{t}_1))^2} \tag{4}$$

where P_n is the transition matrix and the numerator is the variance of $f_n(\mathcal{U}_n)$ when $\mathcal{U}_n$ has distribution π_n. We get a lower bound by considering a specific function of our choice. Given a branchpoint v of a tree $\mathbf{t} \in \mathbf{T}_n$, each of the three branches at v contains some number of leaves (these three numbers summing to n); write $m_{\mathbf{t}}(v)$ for the maximum of these three numbers. Consider

$$f_n(\mathbf{t}) := \min_{v \text{ branchpoint in } \mathbf{t}} m_{\mathbf{t}}(v). \tag{5}$$

It is straightforward to check that, because only one leaf is moved in a step of the chain,

$$\text{if } P_n(\mathbf{t}_1, \mathbf{t}_2) > 0 \text{ then } f_n(\mathbf{t}_2) \le f_n(\mathbf{t}_1) + 1.$$

By symmetry we also have $f_n(\mathbf{t}_1) \le f_n(\mathbf{t}_2) + 1$. So the denominator of (4) is at most 1. We shall show in section 2.1, as a corollary of known results, that

$$\pi_n\{\mathbf{t} : \tfrac{1}{n} f_n(\mathbf{t}) \le x\} \to P(A \le x) \tag{6}$$

where the limit distribution is non-degenerate on $[1/3, 1/2]$, and in fact has density

$$\frac{d}{dx} P(A \le x) = \frac{3x - 1}{\pi x^2 (1-x)^2 (1-2x)^{1/2}}, \ 1/3 < x < 1/2. \tag{7}$$

It then follows from (4) that

$$\tau_n^{\mathrm{rel}} \ge (2 \ \mathrm{var} \ A \ - o(1)) \ n^2$$

establishing part (a) of Theorem 1.

2.1 Asymptotics for random trees

Formula (7) appears in [3] equation (16) as the density of the mass A of the largest branch from the centroid of the *continuum random tree* (CRT).

We now describe briefly how (6) arises. Take a uniform random n-leaf cladogram, and assign random lengths (L_e) to the $2n-3$ edges e according to the density

$$P(L_e \in [l_e, l_e + dl_e] \ \forall e) = s e^{-s^2/2} \ dl_1 \ldots dl_{2n-3}, \quad \text{where } s = \sum_e l_e.$$

This gives a random phylogenetic tree $\mathcal{T}_n$. As described in [2], the family $(\mathcal{T}_n, 3 \leq n < \infty)$ is consistent in the sense that for $m < n$ the subtree of $\mathcal{T}_n$ spanned by leaves $\{1, 2, \ldots, m\}$ is distributed as $\mathcal{T}_m$. Now a realization of the CRT $\mathcal{T}_\infty$ supports a mass measure of total mass one, and the subtree of the CRT spanned by n vertices chosen i.i.d. from the mass measure is distributed as $\mathcal{T}_n$. So by constructing $(\mathcal{T}_n, 3 \leq n < \infty)$ simultaneously from a realization of $\mathcal{T}_\infty$ and i.i.d. sampled vertices, a Glivenko-Cantelli argument shows

$$A_n \xrightarrow{d} A \tag{8}$$

where A_n is the proportion of the n leaves of $\mathcal{T}_n$ in the largest branch from a centroid of $\mathcal{T}_n$, where a *centroid* is a branchpoint for which each of the three leaf-proportions is at most $1/2$. But A_n is distributed as $n^{-1} f_n(\mathbf{t})$ for a uniform random cladogram $\mathbf{t}$, so (8) implies (6).

3 The upper bound

We prove the upper bound in Theorem 1 via a coupling argument. That is, we construct two dependent versions $((X_s^1, X_s^2), s = 0, 1, 2, \ldots)$ of the chain with arbitrary initial states, show that $X_s^1 = X_s^2 \ \forall s \geq U$ for some random time U, and then upper bound U. The construction is given in section 3.1 and the upper bound is derived in section 3.3.

3.1 The coupling

Figure 2 illustrates a typical state $(\mathbf{t}^1, \mathbf{t}^2)$ of the coupled process. As part of the construction certain edges in each tree are *marked*, indicated by arrows on edges. The following two properties (a,b) will be maintained inductively.

- (a) In each tree, the unmarked edges form a subtree.

In each tree, the marked edges automatically form a forest consisting of tree-components.

- (b) There is a one-to-one correspondence (a *matching*) between marked edges of $\mathbf{t}^1$ and marked edges of $\mathbf{t}^2$, which maps tree-components to tree-components and preserves leaf labels.

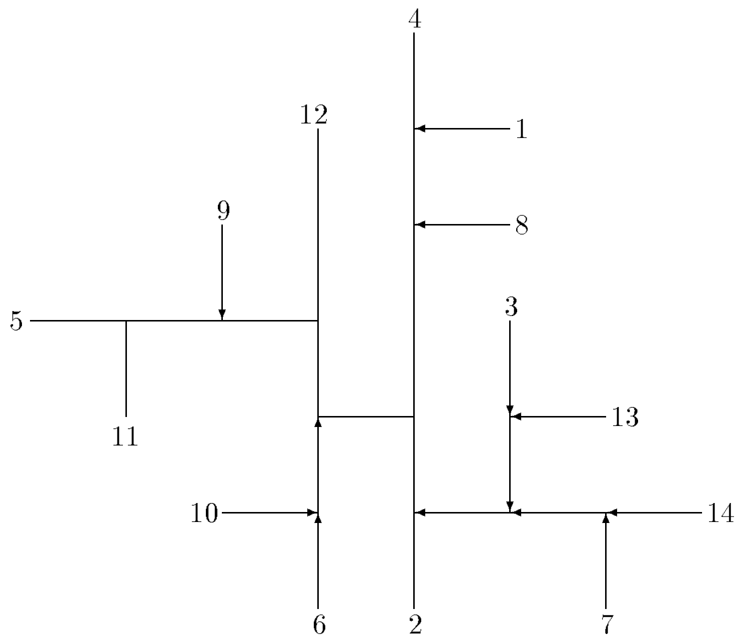
Thus in figure 2 the tree-components of marked edges contain leaves $\{3, 7, 13, 14\}$, $\{6, 10\}$, $\{1\}$, $\{8\}$, $\{9\}$.

Initially no edges are marked, so (a,b) hold trivially. We now specify a step of the coupled chain, starting from $(\mathbf{t}^1, \mathbf{t}^2)$, together with marking and unmarking rules. Pick uniformly at random $i \in \{1, 2, \ldots, n\}$ and remove the edge e_1^u at leaf i from each tree $\mathbf{t}^u$ $(u = 1, 2)$, leaving the branchpoint b^u as a degree-2 vertex (we are using the labeling of figure 1). By (b), the edges at i in $\mathbf{t}^1$ and $\mathbf{t}^2$ were either both marked or both unmarked, so the matching of marked edges extends to the resulting trees. Next, the edges (e_2^u, e_3^u) at b^u are made into one edge e^u, deleting the vertex b^u, and e^u is marked or not marked according to the rule:

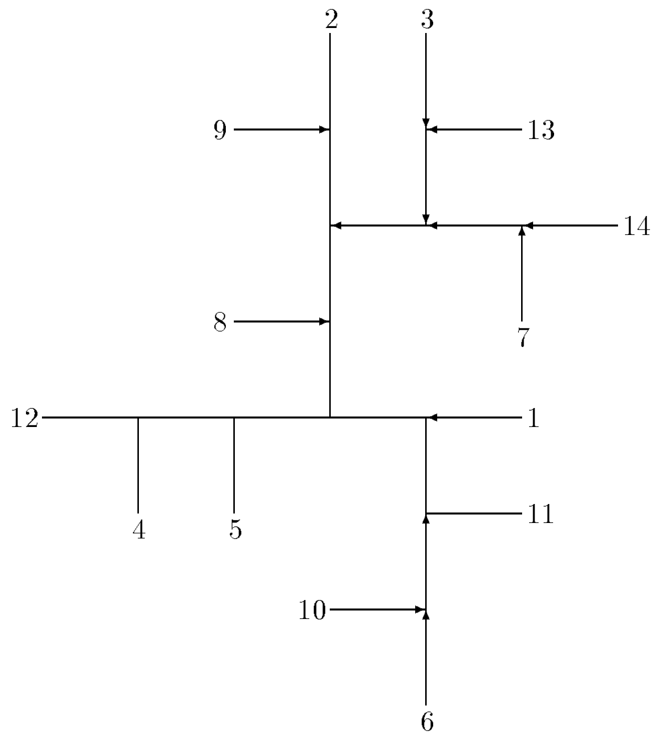
(α) e^u is marked if and only if both e_2^u and e_3^u were marked.

Note that this rule by itself would not preserve the matching. In figure 2 suppose leaf 11 were removed; then in the second tree the marked-edge component containing leaves $\{6, 10\}$ would be reduced to 2 edges, whereas in the first tree it would still have 3 edges. Thus we need to add the rule:

(β) if exactly one of the edges (e_2^u, e_3^u) were marked, then the matched edge in the other tree becomes unmarked.



$\mathbf{t}^1$



$\mathbf{t}^2$

Figure 2

With this rule, the two resulting trees $\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^2$ have a matching of marked edges preserving properties (a,b). We now re-attach leaf i to $\hat{\mathbf{t}}^1$ by choosing uniformly at random an edge f^1 of $\hat{\mathbf{t}}^1$, making a branchpoint $\hat{b}^1$ in that edge, thereby splitting it into two edges f_2^1, f_3^1, and attaching an new edge f_1^1 from $\hat{b}^1$ to leaf i. The corresponding edge f^2 of the second tree is chosen by the rule:

if f^1 is marked, then let f^2 be the matching edge of $\hat{\mathbf{t}}^2$; if f^1 is unmarked then let f^2 be a uniform random choice from the unmarked edges of $\hat{\mathbf{t}}^2$.
Then in the second tree, leaf i is re-attached to a new branchpoint in the middle of edge f^2.

This procedure specifies a coupling of the chains. That is, each chain evolves according to the transition matrix (1). The marking rule for the re-attachment stage is:

(γ) if f^1 (and hence f^2) were marked then $f_2^1, f_3^1, f_2^2, f_3^2$ are marked; if f^1 (and hence f^2) were unmarked then $f_2^1, f_3^1, f_2^2, f_3^2$ are unmarked; in either case the newly-added edges f_1^1 and f_1^2 are marked.
It is easy to check that properties (a,b) are preserved. This completes the specification of one step of the coupled chain starting from $(\mathbf{t}^1, \mathbf{t}^2)$.

By property (b), when every edge is marked the two processes are identical. We want to upper bound how long this takes. Unfortunately it is difficult to analyze directly the process "number of marked edges after s steps", since this process is neither Markov nor monotone. We take an indirect approach which will involve comparisons with a certain birth-and-death chain, which we now discuss.

3.2 A birth-and-death chain

For each large n we consider the birth-and-death chain on states $\{1, 2, \ldots, n-1\}$ with transition probabilities

$$p_{1,2} = 1 \tag{9}$$

$$p_{x,x-1} = \tfrac{x}{n}\left(1 - \tfrac{2x-4}{2n-5}\right), \ 2 \le x \le n-1 \tag{10}$$

$$p_{x,x+1} = \left(1 - \tfrac{x}{n}\right) \tfrac{2x-2}{2n-5}, \ 2 \le x \le n-2 \tag{11}$$

with $p_{x,x} = 1 - p_{x,x+1} - p_{x,x-1}$. Write $E_x T_y$ for the mean hitting times for this chain (this notation suppresses the dependence on n).

Lemma 2 *As $n \to \infty$*
(a) $E_2 T_{n-1} = O(n^3)$
(b) E_2(number of visits to 1 before $T_{n-1}) = O(n^2)$.

Proof. Calculations like this fit into the classical elementary theory of Markov chains [16], but are more simple using the modern analogy with electrical networks. The point is that there is a simple formula for the end-to-end mean commute time:

$$E_1 T_{n-1} + E_{n-1} T_1 = \sum_{x=1}^{n-2} \tfrac{1}{\theta_x p_{x,x+1}} \tag{12}$$

where θ is the stationary distribution. See e.g. [19]. Since of course $\theta_x = \rho_x / \sum_y \rho_y$ where $\rho_1 > 0$ is arbitrary and

$$\rho_x = \rho_1 \prod_{y=1}^{x-1} \tfrac{p_{y,y+1}}{p_{y+1,y}}, \ 2 \leq x \leq n-1 \tag{13}$$

we may rewrite (12) as

$$E_1 T_{n-1} + E_{n-1} T_1 = \left(\sum_{x=1}^{n-1} \rho_x \right) \left(\sum_{x=1}^{n-2} \tfrac{1}{\rho_x p_{x,x+1}} \right). \tag{14}$$

It is convenient to choose $\rho_1 = 1/n$, and then

$$\rho_2 = \rho_1 p_{1,2} / p_{2,1} = (1/n)/(2/n) = 1/2.$$

For $2 \leq x \leq n-2$ we calculate

$$\frac{p_{x,x+1}}{p_{x+1,x}} = \frac{(x-1)(n-x)}{(x+1)(n-x-\frac{3}{2})}$$

and then (13) gives

$$\rho_x = \frac{1}{x(x-1)} \prod_{y=2}^{x-1} \frac{n-y}{n-y-\frac{3}{2}}, \ 4 \leq x \leq n-1.$$

Recall that $a_n = \Omega(b_n)$ means $b_n = O(a_n)$. From the formula above it is not hard to check that, as $n \to \infty$,

$$\rho_x = O\left(x^{-2} \left(\tfrac{n}{n-x} \right)^{3/2} \right) = \Omega\left(x^{-2} \left(\tfrac{n}{n-x} \right)^{3/2} \right) \tag{15}$$

uniformly in $2 \leq x \leq n-1$. It follows that

$$\sum_{2 \leq x \leq n/2} \rho_x = O\left(\sum_{2 \leq x \leq n/2} \tfrac{1}{x^2} \right) = O(1)$$

$$\sum_{n/2 \le x \le n-1} \rho_x = O\left(\frac{1}{n^2} \sum_{2 \le y \le n/2} (n/y)^{3/2}\right) = O(n^{-1/2})$$

and so

$$\sum_{x=1}^{n-1} \rho_x = O(1). \tag{16}$$

For $2 \le x \le n/2$ we have $p_{x,x+1} = \Omega(x/n)$ and so by (15)

$$\rho_x p_{x,x+1} = \Omega\left(\frac{1}{x^2}\frac{x}{n}\right) = \Omega(1/n^2).$$

For $n/2 \le x \le n-1$ we have $p_{x,x+1} = \Omega(\frac{n-x}{n})$ and so by (15)

$$\rho_x p_{x,x+1} = \Omega\left(\frac{1}{n^2}(\frac{n}{n-x})^{3/2}\frac{n-x}{n}\right) = \Omega(1/n^2).$$

Then

$$\sum_{x=1}^{n-2} \frac{1}{\rho_x p_{x,x+1}} = O(n^3).$$

Combining with (16) and (14) shows

$$E_1 T_{n-1} + E_{n-1} T_1 = O(n^3). \tag{17}$$

Since $E_1 T_{n-1} = 1 + E_2 T_{n-1}$, this establishes assertion (a).

To establish (b), let ξ be the time of the first visit to state 1 after the first visit to state $n-1$. By the ergodic theorem for Markov chains and the reward renewal theorem,

$$E_1(\text{ number of visits to 1 before time } \xi)/E\xi = \pi_1$$

because both sides equal the asymptotic proportion of time the chain spends in state 1. But, since $\rho_1 = 1/n$ and $\rho_2 = 1/2$, we have $\pi_1 = O(1/n)$. And (17) says $E\xi = O(n^3)$. So

$$E_1(\text{ number of visits to 1 before time } \xi) = O(n^2)$$

and the left side is an upper bound for the quantity in (b).

3.3 Bounding the coupling time

In the chain $(X_s, s = 0, 1, 2, \ldots)$ started from state $\mathbf{t}$, let $S(\mathbf{t})$ be the first time s such that the leaf (say i) moved on the s'th step is attached to some edge e incident to some leaf (say j), where j was moved at some time $s' < s$ and where the edge e has remained present during time $[s', s-1]$. The following bound is a variation of the classical *birthday problem* – we leave the details to the reader.

Lemma 3 $\max_{\mathbf{t} \in \mathbf{T}_n} ES(\mathbf{t}) = O(n^{1/2})$.

Let b be the branchpoint created at time $S = S(\mathbf{t})$, and let e_1 and e_2 be the edges from b to the two adjacent leaves. See figure 3.

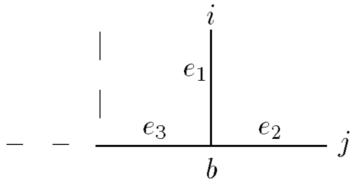


Figure 3

Define a process $(Z_u, u = 0, 1, 2, \ldots, \zeta)$ with $Z_0 = 2$ as follows. For as long as the branchpoint b exists, define Z_u to be the number of leaves of the tree X_{S+u} such that the path from b to the leaf goes via e_1 or e_2. When the branchpoint disappears, at time $S + \Delta$ say, we either stop or continue the process Z, depending on which of three cases holds. One case is that one of the edges $\{e_1, e_2\}$, say e_1, led to a leaf k at time $S + \Delta - 1$, and this leaf is moved at time $S + \Delta$. If the other edge e_2 also leads to a leaf, we stop the process, that is we set $\zeta = \Delta$, and declare $Z_\zeta = 1$.

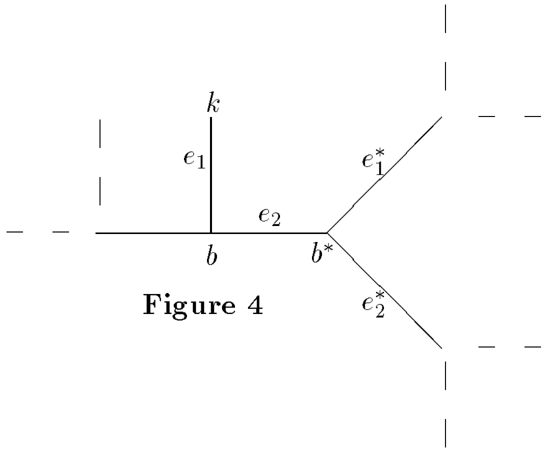


Figure 4

If instead the other edge e_2 led to more than one leaf at time $S + \Delta - 1$, as illustrated in figure 4, then when leaf k is moved and branchpoint b disappears, there was a branchpoint b^* and edges e_1^* and e_2^* at the other end of edge e_2. For $u \geq \Delta$ we then redefine Z_u to be the number of leaves of the tree X_{S+u} such that the path from b^* to the leaf goes via e_1^* or e_2^*. Continue with this definition until branchpoint b^* disappears, and then continue inductively. The final possibility is that b disappears by virtue of the third edge e_3 disappearing. But this can only happen if e_3 led to a leaf, that is if $Z = n - 1$, and we pre-empt this possibility by declaring that the process (Z_u) stops (at time $S + \zeta$) if it reaches the value $n - 1$.

The upshot is that we have defined a process $(Z_u, u = 0, 1, 2, \ldots, \zeta)$ with $Z_0 = 2$ and $\zeta = \min\{u : Z_u = 1 \text{ or } n - 1\}$. We shall now show that, until ζ, this process is the birth-and-death process of section 3.2. Given $Z = x$ there are exactly $2x - 2$ edges in the subtree consisting of the branches at b through e_1 and e_2. With chance $1 - \frac{x}{n}$, a leaf outside this subtree is picked to be moved. Since the resulting tree has $2n - 5$ edges, when the leaf is re-attached there is chance $\frac{2x-2}{2n-5}$ that it is re-attached to the subtree, causing a transition $x \to x + 1$. On the other hand, with chance x/n a leaf inside this subtree is picked to be moved, leaving $2x - 4$ edges in the subtree: note this is true even in the case (figure 4) where the removed edge is incident at b and the process (Z_u) is redefined in terms of b^*. Since the resulting tree has $2n - 5$ edges, when the leaf is re-attached there is chance $1 - \frac{2x-4}{2n-5}$ that it is re-attached outside the subtree, causing a transition $x \to x - 1$. This establishes formulas (10,11). Note that in the case where there are two leaves adjacent to b, the chance that one is removed (making $Z = 1$ by convention above) equals $2/n$, in accordance with formula (10) for the transition $2 \to 1$.

Now suppose $(X_s, s = 0, 1, 2, \ldots)$ is one of a pair (X_s, X'_s) of coupled processes, as in section 3.1. We claim

$$\text{if } Z_\zeta = n - 1 \text{ then } X_{S+\zeta} = X'_{S+\zeta}. \tag{18}$$

Recall from section 3.1 the marking rule for edges of X_s. At time S the two edges e_1 and e_2 are marked. At times $S + u$ the process Z_u counts leaves in some subset A_u of the tree X_{S+u}, the subset being determined by the current branchpoint. One can check inductively on u that all the edges in A_u remain marked. So if $Z_\zeta = n - 1$ then there is a matching of $(n-1)$-leaf subtrees, implying the trees are identical, which is assertion (18).

We may think of the path segment $(X_s, 0 \le s \le S + \zeta)$ as a *trial* which is *successful* if $Z_\zeta = n - 1$. If unsuccessful, we wish to repeat trials until achieving success. We formalize this idea as follows. The definitions of S and (Z_u) are as functionals of the process (X_s), so we can write

$$S = \bar{S}(X_s, 0 \le s < \infty)$$

$$(Z_u, 0 \le u \le \zeta) = \bar{Z}(X_s, 0 \le s < \infty).$$

In terms of the process $(X_s, s = 0, 1, 2, \ldots)$ define $0 < S_1 < S_1 + \zeta_1 < S_2 < S_2 + \zeta_2 < S_3 \ldots$ where $S_1 = S$, $\zeta_1 = \zeta$,

$$S_m = S_{m-1} + \zeta_{m-1} + \bar{S}(X_{S_{m-1}+\zeta_{m-1}+u}, 0 \le u < \infty)$$

$$(Z^m_u, 0 \le u \le \zeta_m) = \bar{Z}(X_{S_m+u}, 0 \le u < \infty).$$

So the time period $S_{m-1} + \zeta_{m-1} < s \le S_m + \zeta_m$ is the period of the m'th trial. So the number of trials until success is

$$M := \min\{m : Z^m_{\zeta_m} = n - 1\}$$

and the time of success is

$$U := S_M + \zeta_M.$$

If (X_s) is one component of coupled processes (X_s, X'_s), then by applying (18) to Z^m we have $Z_U = Z'_U$. By taking (X'_s) to be the stationary process we have, by the usual *coupling inequality*,

$$\max_{A \subset \mathbf{T}_n} |P(X_s \in A | X_0 = \mathbf{t}) - \pi_n(A)| \le P(U > s | X_0 = \mathbf{t}).$$

It follows that
$$\tau_n \leq 4 \max_{\mathbf{t}\in\mathbf{T}_n} E(U|X_0 = \mathbf{t}). \tag{19}$$
Recall we showed that the chain $(Z_u, o \leq u \leq \zeta)$ was the birth-and-death chain of section 3.2, run until hitting 1 or $n-1$. Now let $(Z^*_u, 0 \leq u \leq Q)$ be the concatenation of the processes $(Z^m_u, 0 \leq u \leq \zeta_m)$ for $1 \leq m \leq M$. That is,
$$Z^*_{\zeta_1+\ldots\zeta_{m-1}+(m-1)+u} = Z^m_u,\ 0 \leq u \leq \zeta_m, 1 \leq m \leq M.$$
Then the process $(Z^*_u, 0 \leq u \leq Q)$ is exactly the chain of section 3.2 run until time T_{n-1}. (By construction, $Z^m - 1_{\zeta_{m-1}} = 1$ and $Z^m_0 = 2$, giving the deterministic step $1 \rightarrow 2$ corresponding to (9)). So Lemma 2(a) implies
$$E\sum_{m=1}^{M} \zeta_m = O(n^3). \tag{20}$$
We can write
$$U = \sum_{m=1}^{M}(D_m + \zeta_m) \tag{21}$$
where
$$D_m = \bar{S}(X_{S_{m-1}+\zeta_{m-1}+u}, 0 \leq u < \infty).$$
Now
$$E\left(D_m|X_u, u \leq S_{m-1} + \zeta_{m-1}, X_{S_{m-1}+\zeta_{m-1}} = \mathbf{t}\right) = ES(\mathbf{t})$$
and the usual martingale analog of Wald's equation gives
$$E\left(\sum_{m=1}^{M} D_m|X_0 = \mathbf{t}\right) \leq EM \times \max_{\mathbf{t}\in\mathbf{T}_n} ES(\mathbf{t}). \tag{22}$$
But M is distributed as the number of visits of the birth-and-death chain to state 1 before time T_{n-1}, and so by Lemma 2(b) and Lemma 3 the right side of (22) is $O(n^{5/2})$. Combining with (20), we see from (21) that $\max_{\mathbf{t}\in\mathbf{T}_n} EU = O(n^3)$. Then (19) implies Theorem 1(b).

4 Remarks

1. Though our proofs are ad hoc, there are tantalizing analogies between our chain and well-studied interacting particle processes. Consider first the following process. There are n balls. Initially each is a different color. At each step, two balls are picked at random, and the first ball is re-colored to have the same color as the second ball. Eventually all balls are the same color: this takes V_n steps, say. This process is a re-phrasing of the voter model on the complete graph, and well known results for the voter model imply that EV_n is order n^2. Now a fuzzy mental outline of the argument in section 3.3 is as follows.

> Take each edge of the initial tree to have a different color. When a leaf is moved, give it the color of the edge to which it is re-attached. When all edges have the same color, the chain has coupled with a companion chain.

So by analogy with the voter model, it should take only $O(n^2)$ steps to couple. The difficulty with formalizing this outline is that when branchpoints disappear, adjacent edges with different colors may merge, so we cannot identify an edge at time s as a "descendant" of a unique edge at time 0. The possibility of such mergers causes a slight negative drift in the birth-and-death chain of section 3.2, making the mean hitting time $O(n^3)$ instead of the $O(n^2)$ for the driftless chain.

2. Note that our section 3.3 argument is plainly inefficient in that (in the jargon of "colors" above) we track just one color at a time to see if it takes over or goes extinct; and in the latter case we repeat with another color. It would be better to track all colors simultaneously, but the usual analysis of the voter model (using duality with coalescing random walk, here Kingman's coalescent) seems hard to adapt to our setting.

3. A different approach to the analysis of our chain is via weak convergence. As $n \to \infty$ the uniform randon n-cladogram, with rescaled edge-lengths, converges weakly to the continuum random tree of [2]. It is plausible that our chain, with time rescaled by a factor n^2, will converge to some diffusion process on the space of continuum trees. We may pursue this approach elsewhere.

References

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

[2] D.J. Aldous. The continuum random tree III. *Ann. Probab.*, 21:248–289, 1993.

[3] D.J. Aldous. Recursive self-similarity for random trees, random triangulations and Brownian excursion. *Ann. Probab.*, 22:527–545, 1994.

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

[5] D.J. Aldous and J.A. Fill. Reversible Markov chains and random walks on graphs. Book in preparation, 1997.

[6] D. Barker. LVB 1.0: Reconstructing evolution with parsimony and simulated annealing. http://www.icmb.ed.ac.uk/sokal.html.

[7] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Prob.*, 2:294–313, 1992.

[8] M.A. Charleston. Landscape characteristics of tree space. http://taxonomy.zoology.gla.ac.uk/m̃ac/landscape/landscape.html.

[9] P. Diaconis. *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics, Hayward CA, 1988.

[10] P. Diaconis, J. Fill, and J. Pitman. Analysis of top in at random shuffles. *Combinatorics, Probability and Computing*, 1:135–155, 1992.

[11] P. Diaconis and Holmes S. Random walks on phylogenetic trees. In preparation, 1998.

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

[13] N. Eldredge and J. Cracraft. *Phylogenic Patterns and the Evolutionary Process.* Columbia University Press, New York, 1980.

[14] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*, London, 1996. Chapman and Hall.

[15] Susan P. Holmes. Phylogenies: An overview. Technical report, Biometrics, Cornell, 1997.

[16] J.G. Kemeny and J.L. Snell. *Finite Markov Chains.* Van Nostrand, 1960.

[17] S. Li, D.K. Pearl, and H. Doss. Phylogenetic tree construction using MCMC. Technical Report 583, Statistics, Ohio State, 1996.

[18] B. Mau, M.A. Newton, and B. Larget. Bayesian phylogenetic inference via Markov chain Monte Carlo. Technical Report 961, Statistics, U.W. Madison, 1996.

[19] J. L. Palacios and P. Tetali. A note on expected hitting times for birth and death chains. *Stat. Probab. Letters*, 30:119–125, 1996.

[20] J. S. Rosenthal. Convergence rates for Markov chains. *SIAM Review*, 37:387–405, 1995.