Electronic Journal of Probability

Vol. 7 (2002) Paper no. 6, pages 1–17.

Journal URL
http://www.math.washington.edu/~ejpecp/
Paper URL
http://www.math.washington.edu/~ejpecp/EjpVol7/paper6.abs.html

RANDOM WALKS ON TREES AND MATCHINGS

Persi Diaconis and Susan P. Holmes

Statistics and Mathematics, Department
Stanford University,
CA 94305

Statistics Department
Stanford University
and INRA, Montpellier, France
susan@stat.stanford.edu

Abstract We give sharp rates of convergence for a natural Markov chain on the space of phylogenetic trees and dually for the natural random walk on the set of perfect matchings in the complete graph on $2n$ vertices. Roughly, the results show that $\frac{1}{2}n \log n$ steps are necessary and suffice to achieve randomness. The proof depends on the representation theory of the symmetric group and a bijection between trees and matchings.

Keywords Markov Chain, Matchings, Phylogenetic Tree, Fourier analysis, Zonal polynomials, Coagulation-Fragmentation.

AMS subject classification 60J10 60B15 (60J10 62F10 62F15 65C05 82C80).

Submitted December 21, 2001. Accepted for publication, January 2, 2002.

This work was partly supported by NSF grants DMS 0072360 and DMS 0072569.

1 Introduction:

A phylogenetic tree with l leaves is a rooted binary tree with l labeled leaves. For example, when $l = 3$, there are three such distinct trees:

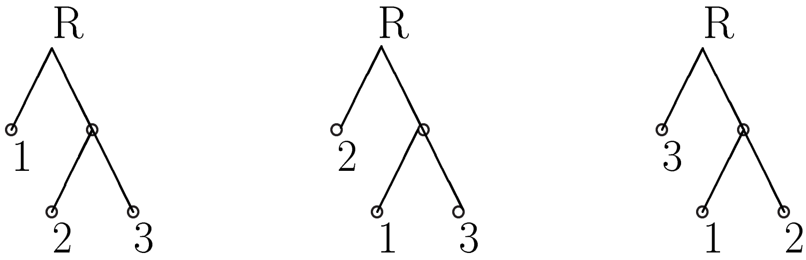


Figure 1: Different trees on 3 leaves

Set $n = \ell - 1$. Schröder (1870) showed there are $(2n)!/2^n n!$ phylogenetic trees.

A perfect matching on $2n$ points is a partition of $1, 2, \ldots, 2n$ into n two-element subsets. It is well known that there are $(2n)!/2^n n!$ distinct perfect matchings. When $n = 2$, the three perfect matchings are

$$\{1,2\}\{3,4\}; \{1,3\}\{2,4\}; \{1,4\}\{2,3\}$$

Diaconis & Holmes (1998) give an explicit bijection, explained below, between these two sets.

In this paper, we analyze a natural random walk on $\mathcal{M}_n$, the set of perfect matchings on $2n$ points, along with the isomorphic walk on trees. For matchings, a step in the walk is obtained by picking two matched pairs at random, a random entry of each pair, and transposing these entries. Thus, switching 2 and 3 moves $\{1, 2\}$ $\{3, 4\}$ to $\{1, 3\}$ $\{2, 4\}$.

For general $n \geq 2$ and $x, y \in \mathcal{M}_n$ define

$$(1) \qquad K(x,y) = \begin{cases} \frac{1}{n(n-1)} & \text{if } x \text{ and } y \text{ differ by a transposition;} \\ 0 & \text{otherwise.} \end{cases}$$

When $n = 2$, the transition matrix K is

	$\{1,2\}\{3,4\}$	$\{1,3\}\{2,4\}$	$\{1,4\}\{2,3\}$
$\{1,2\}\{3,4\}$	0	$\frac{1}{2}$	$\frac{1}{2}$
$\{1,3\}\{2,4\}$	$\frac{1}{2}$	0	$\frac{1}{2}$
$\{1,4\}\{2,3\}$	$\frac{1}{2}$	$\frac{1}{2}$	0

The Markov chain (1) has the uniform distribution $\pi(x) = 2^n n!/(2n)!$ as unique stationary distribution. Our main result determines sharp rates of convergence to stationarity.

Theorem 1 *For the Markov chain $K(x, y)$ of (1) on $\mathcal{M}_n$ the space of perfect matchings on $2n$ points, for any starting state x, if $m = \frac{1}{2}n(\log n + c)$, with $c > 0$, then*

$$(2) \qquad ||K_x^m - \pi|| \leq ae^{-c}$$

In (2) $K_x^m(y) = K^m(x,y) = \sum_z K^{m-1}(x,z)K(z,y)$ and a is a universal constant. The result is sharp; if $m = \frac{1}{2}n(\log n - c)$ for c positive, there is x^* and positive $\epsilon = \epsilon(c)$ such that

$$||K_{x^*}^m - \pi|| \geq \epsilon, \qquad \text{for all } n. \tag{3}$$

In (2) and (3) the norm is the total variation distance

$$||K_x^m - \pi|| = \frac{1}{2}\sum_y |K^m(x,y) - \pi(y)| = \max_A |K_x^m(A) - \pi(A)|. \tag{4}$$

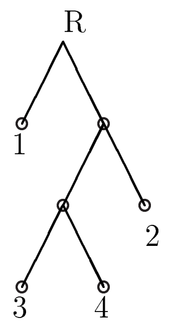
Section 2 contains background on the analytic theory of Markov chains and needed tools from representation theory. Theorem 1 is proved in Section 3 by bounding the total variation norm by the L^2 norm. This is expressed exactly in terms of symmetric group characters. Then standard calculus estimates finish the job. To conclude this introduction, we discuss background on phylogenetic trees, random matchings, diffusion problems, zonal polynomials, and the Metropolis algorithm. Each gives a different interpretation of Theorem 1.

1.1 Phylogenetic Trees

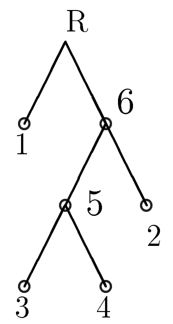
Leaf-labeled trees are a mainstay of modern genomics, depicting family trees or parental relations for l populations, species or genes. An overview, history and statistical developments is in Holmes (1999). See Page & Holmes (2000) for a good book-length treatment of trees and Aldous (1996, 2001) for a probabilistic view. Monte Carlo Markov chains for computing with trees and a variety of scenarios for natural probability distributions on trees lean on random walks such as (1) on the space of all trees. Aldous (2000) and Schweinsberg (2001) study a different walk for which they have given coupling and eigenvalue bounds quite different from (1).

In hope of duplicating the achievements of comparison theory in the analysis of random walk on groups (Diaconis & Saloff-Coste (1993a), Diaconis & Saloff-Coste (1993b), Aldous & Fill (2002)) we searched for a Markov chain on trees which permits a complete analysis. The present matching chain, carried over to trees, offers a candidate which we hope will prove useful.

We briefly describe the correspondence between matchings and trees. Begin with a tree with ℓ labeled leaves. Label the internal vertices sequentially with $\ell+1, \ell+2, \ldots, 2(\ell-1)$ choosing at each stage the ancestor which has both children labeled and who has the descendant lowest possible available label (youngest child). Thus the tree



is labeled



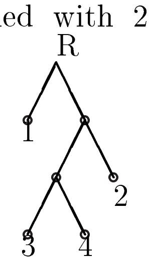
.

When all nodes are labeled, create a matching on $2n = 2(\ell - 1)$ vertices by grouping siblings. In the example above, this yields

$$\{3,4\}\{2,5\}\{1,6\}.$$

To go backward, given a perfect matching of $2n$ points, note that at least one matched pair has both entries from $\{1, 2, 3 \ldots, n+1\}$. All such labels are leaves; if there are several leaf-labeled pairs, choose the pair with the smallest label. Give the next available label $(n + 2 = \ell + 1)$ to their parent node. There are then a new set of available labeled pairs. Choose again the pair with the smallest label to take the next available label for its parent, and so on.

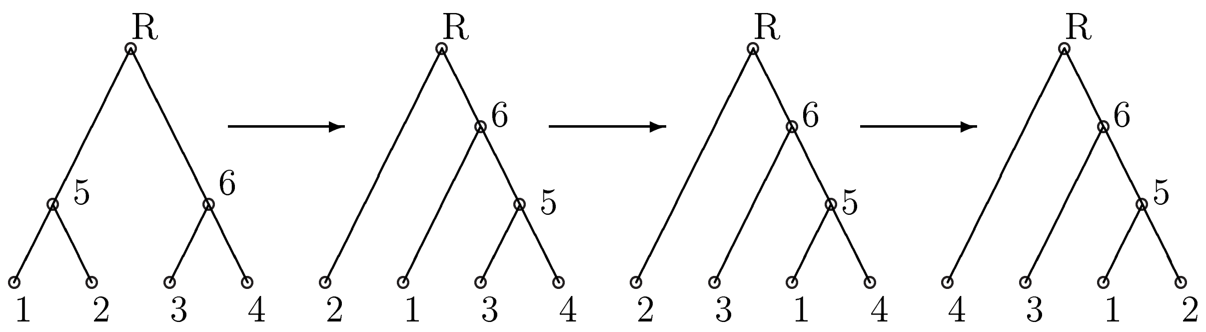
For example, $\{3, 4\}\{2, 5\}\{1, 6\}$ has $2n = 6$ and $\{3, 4\}$ has both entries from $\{1, 2, 3, 4\}$. The parent of these is labeled 5 and thus matched with 2 and then the parent of $\{2, 5\}$ is matched with 1, yielding



Under this correspondence, the Markov chain (1) on matchings leads to a Markov chain on trees. For example, if the first few steps of the walk on matchings are

$$\{1,2\}\{3,4\}\{5,6\} \longrightarrow \{1,5\}\{3,4\}\{2,6\} \longrightarrow \{1,4\}\{3,5\}\{2,6\} \longrightarrow \{4,6\}\{3,5\}\{1,2\},$$

the associated walk on trees is



Theorem 1 holds as stated for trees: after $\frac{1}{2}n(\log n + c)$ steps this walk is close to the uniform distribution.

1.2 Random Matchings

Let G be a graph with vertex set V and edge set E. A *perfect matching* is a set of disjoint edges containing all vertices. Matchings have evolved as an important tool in graph algorithms. See Lovasz & Plummer (1985) or (Pulleyblank, 1995, chapter 4).

The nearest neighbor process of Theorem 1.1 is a procedure for generating a random matching. A similar procedure can be run on more general graphs. Such procedures form the basis of an interesting new stochastic algorithm for approximating the number of perfect matchings in the graph. This last is a $\#{-}P$ complete problem and these stochastic algorithms offer the only currently feasible approach. Following work by Jerrum & Sinclair (1989), Jerrum et al. (2000) showed that the analogous random walk on bipartite graphs is rapidly mixing. The actual bound they found, while polynomial in n (number of vertices), is probably far from the truth. Theorem 1 gives the only example where the sharp bounds are known. We note further that our analysis determines all the eigenvalues of the matching graph.

1.3 A Diffusion Problem

The original motivation for considering random matchings came from its association with the Bernoulli–Laplace diffusion model. There, one considers two urns, the left containing n red balls, the right containing n black balls. At each time, a ball is chosen at random in each urn and the two balls are switched. Diaconis & Shahshahani (1987) show that it takes $\frac{1}{4}\log n + cn$ switches to mix up the urns. An analysis for three urns is developed in Scarabotti (1997).

It is natural to consider problems with more urns. The present paper can be considered as involving n urns, each containing 2 balls. As explained below, the approach we use here works for any number of urns, provided all contain the same number of balls. However with three urns with a different number of balls in each, the problem is open.

1.4 Zonal Polynomials

Our original proof of Theorem 1 used the machinery of Gelfand pairs and the correspondence between the zonal spherical functions associated to the space of matchings with the zonal polynomial of Alan James. The symmetric group $\mathcal{S}_{2n}$ acts transitively on matchings by permuting coordinates. The subgroup fixing the matching $\{1,2\}\{3,4\}\ldots\{2n-1,2n\}$ is isomorphic to $\mathcal{B}_n$, the group of symmetries of an n-dimensional cube ($|\mathcal{B}_n| = 2^n n!$). The permutation representation of $\mathcal{S}_{2n}$ is multiplicity free and the eigenvalues of the random walk are the values of the spherical functions of the representation (Diaconis (1988)).

Using a form of the Schur-Weyl duality, these spherical functions may be identified with the coefficients of zonal polynomials expanded in the power sum symmetric functions.

The random walk led to new formulae for zonal polynomials and substantial new mathematics in joint work with Eric Lander. This has been brilliantly exposited by (Macdonald, 1995, Chapter 7). We recently realized that all of this machinery could be avoided. The simple proof which follows leans on early work (Diaconis & Shahshahani, 1981). It only needs basic representation theoretic tools. The new approach can be used backward to give different proofs of spherical function formulae. Using an extension of the random transposition results due to Roussel (2000) leads to new formulae for spherical functions and zonal polynomials.

1.5 A Metropolis walk on partitions; coagulation and fragmentation

Let $\mathcal{P}_n$ be the partitions of n, e.g. $\mathcal{P}_3 = \{3;\ 2,1;\ 1,1,1\}$. Partitions are written as $\lambda \vdash n$ with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r), \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$, or as $1^{a_1} 2^{a_2} \ldots n^{a_n}$, if λ has a_i parts labeled i, with $\sum i a_i = n$. The length is $\ell(\lambda) = r = \sum a_i$. If $z(\lambda) = \prod_{i=1}^r i^{a_i} a_i!$, then $\{1/z(\lambda)\}_{\lambda \vdash n}$ defines a probability distribution on $\mathcal{P}_n$. To see this, recall that the partitions index the conjugacy classes of the representation of the symmetric group $\mathcal{S}_n$, and the size of the conjugacy class corresponding to λ is $n!/z(\lambda)$. Since the conjugacy classes partition $\mathcal{S}_n$, $\sum_\lambda \frac{n!}{z(\lambda)} = n!$.

Repeated random transpositions induce a random walk on $\mathcal{S}_n$ with a uniform stationary distribution. If the walk is lumped to conjugacy classes, we get a Markov chain on partitions with $\frac{1}{z(\lambda)}$ as its stationary distribution. When $n = 3$, the stationary distribution and transition matrix are:

$$(\frac{1}{6}, \frac{1}{2}, \frac{1}{3}), \qquad \begin{array}{c|ccc} & 1^3 & 2,1 & 3 \\ \hline 1^3 & 0 & 1 & 0 \\ 2,1 & \frac{1}{3} & 0 & \frac{2}{3} \\ 3 & 0 & 1 & 0 \end{array}$$

This walk on partitions is a special case of a large class of models studied by chemists and physicists as a process of coagulation and fragmentation. See the reviews by Aldous (1999, 1998), Durrett et al. (1999) or Mayer-Wolfe et al. (2001). It does not seem to have been noticed that the eigenvalues of this walk were given in Diaconis & Shahshahani (1981) as the number in (13) below and the eigenvectors are the characters of the symmetric group. We will call this walk on partitions the conjugacy walk.

The walk (1) on matchings also induces a walk on partitions $\mathcal{P}_n$. To see this, let the identity matching be $\{1,2\}\{3,4\}\ldots\{2n-1,2n\}$. Given two matchings $x, y \in \mathcal{M}_n$, form a graph on $2n$ points, with blue edges for the matched pairs in x and red edges for the matched pairs in y. This graph decomposes into disjoint cycles, each of even length with alternating red/blue edges. These cycle lengths divided by 2 form a partition of n, which we call the distance between x and y. See Macdonald (1995) for more details.

Start the random walk of (1) at the identity matching, and report only the distance from the identity at each step. Thus when $n = 3$, a sample walk of matchings and partitions is

$$\begin{array}{cc} \mathcal{M}_n & \mathcal{P}_n \\ \{1,2\}\{3,4\}\{5,6\} & 1^3 \\ \{1,3\}\{2,4\}\{5,6\} & 1,2 \\ \{1,5\}\{2,4\}\{3,6\} & 3 \\ \{1,4\}\{2,5\}\{3,6\} & 3 \\ \{1,6\}\{2,5\}\{3,4\} & 1,2 \\ \vdots & \vdots \end{array}$$

This distance walk on partitions has stationary distribution

$$\pi(\lambda) = \frac{c^{-1}}{2^{\ell(\lambda)} z(\lambda)}, \text{ with } c = \frac{\binom{2n}{n}}{2^{2n}} \tag{5}$$

To see this, note that the distance of a matching is in one-to-one correspondence with the double coset in $\mathcal{B}_n \backslash \mathcal{S}_n / \mathcal{B}_n$. The double coset containing x of distance λ has size $2^{\ell(\lambda)} z(\lambda)$ by Macdonald (1995) . The stationary distribution is the special case of Ewens' sampling formula (with $\theta = \frac{1}{2}$) used in population genetics (see Ewens (1972)). When $n = 3$ the distance walk has stationary distribution and transition matrix:

$$(\frac{1}{15}, \frac{2}{5}, \frac{8}{15}) \qquad \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

We find the following relation between the conjugacy chain and the distance chain curious: if the conjugacy chain is changed to have stationary distribution $\pi(\lambda)$ by the Metropolis algorithm (Hammersley & Handscomb, 1964), the resulting metropolized chain is the distance chain. To explain, let the conjugacy chain have transition matrix $P(\lambda, \mu)$. The Metropolis procedure constructs an auxiliary matrix

$$A(\lambda, \mu) = \frac{\pi(\mu) P(\mu, \lambda)}{\pi(\lambda) P(\lambda, \mu)} \tag{6}$$

Then, define the Metropolis chain by

$$M(\lambda, \mu) = \begin{cases} P(\lambda, \mu) & \text{if } A(\lambda, \mu) \geq 1, \quad \lambda \neq \mu \\ P(\lambda, \mu) A(\lambda, \mu) & \text{if } A(\lambda, \mu) < 1, \quad \lambda \neq \mu \\ P(\lambda, \lambda) + \sum_{\mu: A(\lambda,\mu)<1} P(\lambda, \mu)(1 - A(\lambda, \mu)) & \text{if } \lambda = \mu \end{cases} \tag{7}$$

Our observation is that for all n, M is the distance chain. For example, when $n = 3$ the matrix $A(\lambda, \mu)$ is

(8)

	1^3	$2,1$	3
1^3	0	2	0
$2,1$	$\frac{1}{2}$	0	2
3	0	$\frac{1}{2}$	0

and using the recipe (7) with (6) yields (8). This is closely related to other examples. Diaconis & Ram (2000) and Diaconis & Hanlon (1992) show how the Metropolis algorithm sometimes results in group theoretically natural deformations. We thus can see the present paper as giving a sharp analysis of an instance of the Metropolis algorithm.

2 Background and Needed Tools

In broad outline, the proof of Theorem 1 is similar to the analysis of random transpositions in $\mathcal{S}_{2n}$. The expository account presented in Diaconis (1988) develops needed background from first principles and certainly provides sufficient background. Splendid treatments of the more recent developments in Markov chain theory appear in Aldous & Fill (2002) and Saloff-Coste (1997).

2.1 Markov Chain Background

Let $\mathcal{X}$ be a finite set and K be a matrix indexed by $\mathcal{X} \times \mathcal{X}$. Throughout, π is a stationary distribution for which K is ergodic and reversible: $\pi(x)K(x,y) = \pi(y)K(y,x)$. Because of reversibility, K has an orthonormal basis of eigenvectors $f_i(x)$ with

$$Kf_i(x) = \sum K(x,y)f_i(y) = \beta_i f_i(x).$$

Here β_i is the associated eigenvalue and both f_i and β_i are real. Arguing as in Diaconis & Saloff-Coste (1993a) we may express the chi-square distance as

$$(9) \qquad ||K_x^m - \pi||_2^2 = \sum_y \frac{|K^m(x,y) - \pi(y)|^2}{\pi(y)} = \sum_i^* f_i^2(x)\beta_i^{2m},$$

where the second sum is over all i with $\beta_i \neq 1$.
If a finite group G acts transitively on $\mathcal{X}$ and preserves K then the distance (9) does not depend on the initial state x. Multiply both sides of (9) by $\pi(x)$ and sum over x. The orthogonality of f_i yields

$$(10) \qquad ||K_x^m - \pi||_2^2 = \sum_i^* \beta_i^{2m}.$$

For theorem 1, $\mathcal{X}$ is the space of matchings, K is defined by (1.1) and π is uniform. The chain is symmetric and so, reversible. The group G is $\mathcal{S}_{2n}$, the symmetric group on $2n$ letters. Here σ in $\mathcal{S}_{2n}$ acts on matchings coordinate wise:

$$\sigma((i_1,i_2),(i_3,i_4),\ldots,(i_{2n-1},i_{2n})) = (\sigma(i_1),\sigma(i_2)),\ldots,(\sigma(i_{2n-1}),\sigma(i_{2n})).$$

This is a transitive action so (10) is in force.

Using the Cauchy-Schwartz inequality shows that the chi-square distance is an upper bound for the total variation norm of (4)

$$(11) \qquad ||K_x^m - \pi||^2 = (\frac{1}{2}\sum |K^m(x,y) - \pi(y)|)^2 \leq \frac{1}{4}\sum_y \frac{|K^m(x,y) - \pi(y)|^2}{\pi(y)}$$

The bounds (10), (11) and an explicit determination of the eigenvalues constitute the backbone of the proof.

2.2 Group Theory Background

For background in representation theory, we recommend Serre (1977) or Diaconis (1988). For characters of the symmetric group see Sagan (2001) or (Macdonald, 1995, Chapter 1).

As explained above, $\mathcal{S}_{2n}$ acts transitively on the space of matchings, giving a permutation representation of $\mathcal{S}_{2n}$ on $\mathcal{L}(\mathcal{M}_n) = \{f : \mathcal{M}_n \longrightarrow \mathbb{R}\}$. Matchings may be thought of as a product of n disjoint transpositions and so as fixed-point free idempotent mappings of $\{1, 2, \ldots, 2n\}$ to itself, or as the elements of the conjugacy class of $\mathcal{S}_{2n}$ with all cycles of length two. The conjugacy action of the group is just as defined in Section 2.1. If B_n is the subgroup of $\mathcal{S}_{2n}$ fixing the matching $(1, 2)(3, 4) \ldots (2n-1, 2n)$, then B_n is isomorphic to the hyperoctahedral group of order $2^n n!$. Matchings may be identified with elements of the quotient $\mathcal{S}_{2n}/B_n$.

The irreducible representations of $\mathcal{S}_{2n}$ are indexed by partitions μ of $2n$. They will be denoted S^μ in what follows. A crucial fact is that the decomposition of $\mathcal{L}(\mathcal{M}_n)$ is known:

Theorem 2 *Let $\mathcal{M}_n = \mathcal{S}_{2n}/B_n$. Let $\mathcal{L}(\mathcal{M}_n)$ be all real functions on $\mathcal{M}_n$, considered as a representation of $\mathcal{S}_{2n}$. Then*

$$\mathcal{L}(M_n) = \bigoplus\nolimits_{\lambda \vdash n} \mathcal{S}^{2\lambda},$$

where the direct sum is over all partitions λ of n, $2\lambda = (2\lambda_1, 2\lambda_2, \ldots, 2\lambda_k)$ and $\mathcal{S}^{2\lambda}$ is the associated irreducible representation of the symmetric group $\mathcal{S}_{2n}$.

This result was used extensively by James (1982, 1968) who credits it to Littlewood or Thrall. A proof appears in (James & Kerber, 1981, page 224), see also Saxl (1981) or Inglis et al. (1990).

The final step of preparation relates the matching chain K of (1.1) to the random transposition chain on all of $\mathcal{S}_{2n}$. Consider the formal sum of all transpositions in $\mathcal{S}_{2n}$ weighted by $\frac{1}{\binom{2n}{n}}$:

$$T = \frac{1}{\binom{2n}{n}} \sum_{1 \leq i < j \leq 2n} (i, j).$$

This operates on $\mathcal{L}(\mathcal{M}_n)$ by left multiplication. Let T_n be the matrix of this as a linear map on $\mathcal{L}(\mathcal{M}_n)$ with the space having basis the delta functions of the matchings.

Proposition 1 *The transition matrix K of (1.1) and T_n satisfy*

$$K = \frac{2n-1}{2n-2}\left(T_n - \frac{1}{2n-1}I\right).$$

Proof This is best seen combinatorially. Operating by T_n corresponds to picking a random pair of indices $1 \le i < j \le 2n$ and transposing to get a new matching. This fixes a matching if i and j are matched to each other. The chance of this is easily seen to be $\frac{1}{2n-1}$. Deleting these diagonal elements from T_n and re-normalizing gives the result.□

Corollary 1 *The transition matrix K of (1.1) has an eigenvalue β_λ for each partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ of n, given by*

$$\beta_\lambda = \frac{1}{n(n-1)} \sum_{j=1}^{k} \lambda_j^2 - j\lambda_j.$$

The multiplicity of β_λ is determined by $\mu = 2\lambda$:

$$mult(\lambda) = \frac{(2n)!}{\prod_{(i,j)\in\mu} h(i,j)}, \tag{12}$$

with the product being over the cells of the shape μ, and $h(i,j)$ hook length $\mu_i + \mu'_j - i - j + 1$, where μ' is the transposed diagram.

Example Suppose $n = 3$, so $|\mathcal{M}_n| = 15$. The three partitions $3, 21, 1^3$ give $\beta_3 = 1, \beta_{21} = +\frac{1}{6}, \beta_{1^3} = -\frac{1}{2}$ with multiplicities $1, 9, 5$.

Proof of Corollary 2.1 If ρ_μ denotes the irreducible matrix representation of $\mathcal{S}_{2n}$ corresponding to μ, the Fourier transform of the element T is

$$\hat{T}(\mu) = \frac{1}{\binom{2n}{n}} \sum_{1 \le i < j < 2n} \rho_\mu(i,j).$$

Using Schur's lemma, this matrix is a constant times the identity:

$$\hat{T}(\mu) = c(\mu)I.$$

Taking the trace of both sides shows $c(\mu) = \chi_\mu(1,2)/\chi_\mu(id)$ with χ_μ the character for this representation. This character ratio was determined by Froebenius. See Ingram (1950) for a modern treatment. This gives

$$\frac{\chi_\mu(1,2)}{\chi_\mu(id)} = \frac{1}{2n(2n-1)} \sum_j \mu_j^2 - (2j-1)\mu_j \tag{13}$$

For the present application $\mu = 2\lambda$. Since $\hat{T}(\mu)$ is diagonal, the multiplicity of this eigenvalue is the dimension of ρ_μ, namely $\chi_\mu(id)$. This multiplicity is given by the hook length formula (12) (see Sagan (2001)).
Finally, using Proposition 2.1 and simple algebra completes the proof. □

Remark Corollary 1 determines the eigenvalues of the matching graph. This has vertices $\mathcal{M}_n$ and an edge from x to y if they differ by a single switch. There has been a healthy development on the combinatorial side, determining the eigenvalues for a variety of matching 'complexes'; see Wachs (2001) for a survey of closely related results.

3 Proof of Theorem

We begin by giving a direct proof of the lower bound (3). The lower bound is proved by finding a set $A \subset \mathcal{M}_n$ and $x_* \in \mathcal{M}_n$ such that $\pi(A)$ is large and $K^m_{x_*}(A)$ is small. From (4), $||K^m_x - \pi|| \geq |K^m_x(A) - \pi(A)|$. To produce A, define a function T on matchings by

$$T(x) = |i, 1 \leq i \leq n, \{(2i-1, 2i) \in x\}|.$$

Thus $T(x) = \sum_{i=1}^n T_i(x)$ with

$$T_i(x) = \begin{cases} 1 & \text{if } (2i-1, 2i) \text{ are matched in } x \\ 0 & \text{otherwise} \end{cases}$$

Under the uniform distribution $E_u(T_i) = \frac{1}{2n-1}$ and $E_u(T) = \frac{n}{2n-1} \sim \frac{1}{2}$ this T is the analog of the number of fixed points of a permutation. Arguing as in Barbour et al. (1992) a straight-forward use of Stein's method shows that for n large, T has an approximate Poisson $(\frac{1}{2})$ distribution. We choose $A = \{x : T(x) = 0\}$, whence $\pi(A) \sim e^{-\frac{1}{2}}$.
To bound $K^m_{x_*}(A)$ where $x_* = (1,2)(3,4)\cdots(2n-1,2n)$, we show that there is a good chance that there is some pair $(2i-1, 2i)$ which has not been hit in the first m steps. Toward this end, let $(I_1, J_1), \cdots (I_m, J_m)$ be the transpositions chosen in the first m steps of the walk. For $1 \leq i \leq n$ let

$$Y_i = \begin{cases} 1 & \text{if } (2i-1, 2i) \in \bigcup_{k=1}^m \{(I_k, J_k)\} \\ 0 & \text{otherwise} \end{cases}$$

Thus

$$E(Y_i) = \left(\frac{\binom{2n-2}{2}}{\binom{2n}{2}} \right)^m = \left((1 - \frac{1}{n})(1 - \frac{2}{2n-1}) \right)^m \sim \left(1 - \frac{2}{n} \right)^m,$$

so the random variable $Y = \sum_{i=1}^n Y_i$ has mean $nE(Y_i) \sim n\left(1 - \frac{2}{n}\right)^m$. By the Coupon-collectors problem, if $m = \frac{n}{2}(\log n + c)$ with c fixed in $\mathbb{R}$, Y has an approximate Poisson(e^{-c}) distribution. Again, this can be quantified using Stein's method as in Barbour et al. (1992). Of course if $Y \geq 1$ the Markov chain started at x_* certainly has

$T(x) \neq 0$. Thus $K^m_{x^*}(A^c) \geq P\{Y \geq 1\} = 1 - e^{-e^{-c}} + o(1)$, so $K^m_{x_*}(A) \leq e^{-e^{-c}} + o(1)$. Combining bounds, for $m = \frac{1}{2}n(\log n + c)$, with $c < 0$,

$$||K^m_{x_*} - \pi|| \geq |\pi(A) - K^m_{x_*}(A)|| = e^{\frac{1}{2}} - e^{-e^{-c}} + o(1)$$

Remark A slight refinement of this argument shows that $||K^m_{x_*} - \pi|| \longrightarrow 1$ if $m = \frac{1}{2}n(\log n - c_n)$ with c_n tending slowly to infinity. This completes the proof for the lower bound. $\square$

To prove the upper bound, we use Corollary (1), the bound (11) and equality (10) to see that it suffices to bound

(14) $$\sum_{\lambda \vdash n, \lambda \neq n} mult(\lambda)\beta_\lambda^{2m}.$$

For this, we use the following bounds on $mult(\lambda)$ and β_λ:

(15) $$\sum_{\lambda \vdash n} mult(\lambda) = \frac{(2n)!}{2^n n!} \sim \sqrt{2}(\frac{2}{e}n)^n.$$

(16) $$\sum_{\lambda \vdash n, \lambda_1 = n-j} mult(\lambda) \leq \binom{2n}{2j}\frac{(2j)!}{2^j j!} \leq \frac{2^j n^{2j}}{j!}$$

(17) $$|\beta_\lambda| \leq \begin{cases} 1 - \frac{j(2n-j)}{n^2} & \text{if } \lambda_1 \geq \frac{n}{2} \\ \frac{1+o(1)}{2} & \text{if } \lambda_1 \leq \frac{n}{2} \end{cases}$$

The bound (15) follows from the decomposition of Theorem 2 and Stirling's formula. For (16), we bound the number of standard tableaux Q of shape 2λ with $\lambda_1 = n - j$ by noting that there are at most $\binom{2n}{2j}$ ways of picking the elements not in the first row and at most $\frac{(2j)!}{2^j j!}$ ways of arranging these in a standard tableau of shape $(2\lambda_2, 2\lambda_3, \ldots)$ and then there is at most one way to fill in the first row with the remaining elements (c.f. 15). The final inequality in (16) is elementary. The inequalities in (17) are a direct consequence of the monotonicity results of (Diaconis & Shahshahani (1987)):

(18) $$|\beta_\lambda| \leq \begin{cases} \beta_{\lambda_1, n-\lambda_1} & \text{if } \lambda_1 \geq \frac{n}{2} \\ \beta_{\frac{n}{2},\frac{n}{2}} & \text{if } \lambda_1 \leq \frac{n}{2} \end{cases}$$

together with a direct computation of the right-hand side of the formula in Corollary (1).
Completion of the proof of Theorem 1
The sum in (14) is bounded in three zones:

Zone I	$\lambda_1 \le \frac{n}{2}$,	
Zone II	$\frac{n}{2} < \lambda_1 \le \theta n$,	with $\theta = \frac{1}{\sqrt{e}}$.
Zone III	$\theta n < \lambda_1 \le n-1$,	

Begin with Zone I, bound

$$|\beta_\lambda|^{2m} \le \left(\frac{1+o(1)}{2}\right)^{2m}$$

using (17). Then (15) gives the sum over Zone I bounded above by

$$\sqrt{2}\left(\frac{1}{2}\right)^{2m}\left(\frac{2}{e}n\right)^n = \sqrt{2}\exp\{-2(\log 2)m + n\log(\frac{2}{e}n)\} \tag{19}$$

For $n \ge 10$, the term in the exponent is bounded above by

$$-2c - .2n(\log n + O(1)).$$

It follows that the sum over Zone I is bounded above by

$$Ae^{-2c}, \text{ for universal A, for } n \ge 10.$$

Consider Zone II, $\frac{n}{2} < \lambda_1 \le \frac{1}{\sqrt{e}}n$. From (16) and (17), for any λ in this Zone with $\bar{\theta} = 1 - \frac{1}{\sqrt{e}}$

$$|\beta_\lambda| \le 1 - \frac{\bar{\theta}n(2n - \bar{\theta}n)}{n^2} = 1 - \bar{\theta}(2-\bar{\theta}) = \frac{1}{e}.$$

Using this and (15), the sum over Zone II is bounded above by

$$\sqrt{2}\left(\frac{2}{e}\right)^n\left(\frac{1}{e}\right)^{2m} = \sqrt{2}\left(\frac{2}{e}\frac{1}{e^c}\right)^n \le Ae^{-2c} \text{ for } n \ge 2.$$

Finally, consider Zone III, $\frac{n}{\sqrt{e}} < \lambda_1 \le n-1$. Writing $j = n - \lambda_1$, $1 \le j \le n$, using $1 - x \le e^{-x}$ with (16-17), the sum to be bounded is thus:

$$\sum_{1 \le j \le \bar{\theta}n} \frac{2^j n^{2j}}{j!} e^{-j\frac{2n-j}{n}(\log n + c)}.$$

Next

$$j(\log n + c)\frac{(2n-j)}{n} \ge 2(\log n + c) + (2(j-1) - \frac{j^2}{n})(\log n + c)$$

From this, the general term in the sum over Zone III is bounded above by:

$$n^2 e^{-2(\log n + c)} 2^j n^{2(j-1) - (2(j-1) - \frac{j^2}{n})} = e^{-2c}\frac{2^j n^{\frac{j^2}{n}}}{j!}.$$

We show that the sum

$$\sum_{1 \le j \le \bar{\theta}n} \frac{2^j n^{\frac{j^2}{n}}}{j!} \tag{20}$$

is uniformly bounded, which will complete the proof. Call the general term t_j. For $2 \le j \le \frac{n}{2}$, the ratios

$$\frac{t_{j+1}}{t_j} = \frac{2}{j+1} n^{\frac{2j+1}{n}}.$$

are decreasing for $2 \le j < k$ and increasing for $j > k$, with $k = \frac{n}{2\log n} - 1$. We show that t_3/t_2 and $t_{\bar\theta n+1}/t_{\bar\theta n}$ are smaller than a constant $q < 1$. Then

$$t_2 + \cdots + t_{\bar\theta n} \le \frac{t_2 + t_k}{1-q}.$$

Since t_2, t_k (and t_1) are all uniformly bounded, this will suffice. First, $t_3/t_2 = 2/3n^{\frac{5}{n}} < 1$, for $n \ge 17$. Next,

$$\frac{t_{\bar\theta n+1}}{t_{\bar\theta n}} = \frac{2}{\bar\theta n + 1} e^{\frac{(2\bar\theta n+1))}{n}\log n} = \frac{2n^{2\bar\theta}}{\bar\theta n + 1} e^{\frac{\log n}{n}}$$

Since $\bar\theta = 1 - \frac{1}{\sqrt{e}} \doteq .3935 < .4$, this term tends to 0 and is smaller than 1 for n large enough. Finally routine use of Stirling's formula shows t_i is uniformly bounded. Thus the sum over Zone III is bounded above by

$$Ae^{-2c}$$

□

Remark: Of course, for moderate n, the bound (14) can be computed numerically.

For n smaller than 10^6 or so, intermediate bounds (19) and (20) would give explicit, accurate error bounds. The less explicit form ae^{-2c} is simply easier to look at and think about.

Acknowledgments

We are very thankful to David Aldous, Joe Blitzstein and Warren Ewens for carefully reading the manuscript.

References

Aldous, D. J. (1996). *Probability Distributions on Cladograms*, pages 1–18. in Random discrete structures (Aldous and Pemantle, eds.) Number 76 in IMA Series. Springer Verlag, NY.

Aldous, D. J. (1998). Stochastic coalescence. In *Proceedings of the International Congress of Mathematicians, Vol. III (Berlin, 1998)*, number Extra Vol. III, pages 205–211 (electronic).

ALDOUS, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli* **5**, 3–48.

ALDOUS, D. J. (2000). Mixing time for a Markov chain on Cladograms. *Combin. Probab. Computing* **9**, 191–204.

ALDOUS, D. J. (2001). Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statist. Sci.* **16**, 23–34.

ALDOUS, D. J. & FILL, J. (2002). Reversible Markov chains and random walks on graphs. URL= `http://www.mts.jhu.edu/~fill`.

BARBOUR, A. D., HOLST, L., & JANSON, S. (1992). *Poisson approximation.* The Clarendon Press Oxford University Press, New York. Oxford Science Publications.

DIACONIS, P. (1988). *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics, Hayward, Calif.

DIACONIS, P. & HANLON, P. (1992). Eigen-analysis for some examples of the Metropolis algorithm. In *Hypergeometric functions on domains of positivity, Jack polynomials, and applications (Tampa, FL, 1991)*, pages 99–117. Amer. Math. Soc., Providence, RI.

DIACONIS, P. W. & HOLMES, S. P. (1998). Matchings and phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **95**, 14600–14602 (electronic).

DIACONIS, P. & RAM, A. (2000). Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. *Michigan Math. Journ.* **48**, 157–190.

DIACONIS, P. & SALOFF-COSTE, L. (1993a). Comparison techniques for random walk on finite groups. *Ann. Probab.* **21**, 2131–2156.

DIACONIS, P. & SALOFF-COSTE, L. (1993b). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3**, 696–730.

DIACONIS, P. & SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. W.* **57**, 159–179.

DIACONIS, P. & SHAHSHAHANI, M. (1987). Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.* **18**, 208–218.

DURRETT, R., GRANOVSKY, B. L., & GUERON, S. (1999). The equilibrium behavior of reversible coagulation-fragmentation processes. *J. Theoret. Probab.* **12**, 447–474.

EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

HAMMERSLEY, J. M. & HANDSCOMB, D. C. (1964). *Monte Carlo Methods.* Chapman and Hall, London.

HOLMES, S. (1999). Phylogenies: An overview. In Halloran, E. & Geisser, S., editors, *Statistics and Genetics*, number 81 in IMA. Springer Verlag, NY.

INGLIS, N. F. J., RICHARDSON, R. W., & SAXL, J. (1990). An explicit model for the complex representations of S_n. *Arch. Math. (Basel)* **54**, 258–259.

INGRAM, R. E. (1950). Some characters of the symmetric group. *Proc. Amer. Math. Soc.* **1**, 358–369.

JAMES, G. & KERBER, A. (1981). *The representation theory of the symmetric group*. Addison-Wesley, Reading, MA.

JAMES, A. T. (1968). Calculation of zonal polynomial coefficients by use of the Laplace-Beltrami operator. *Ann. Math. Statist* **39**, 1711–1718.

JAMES, A. T. (1982). Analysis of variance determined by symmetry and combinatorial properties of zonal polynomials. In *Statistics and probability: essays in honor of C. R. Rao*, pages 329–341. North-Holland, Amsterdam.

JERRUM, M. & SINCLAIR, A. (1989). Approximating the permanent. *SIAM Journal on Computing* **18**, 1149–1178.

JERRUM, M., SINCLAIR, A., & VIGODA, E. (2000). A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Electronic Colloquium on Computational Complexity (ECCC)* .

LOVASZ, L. & PLUMMER, M. D. (1985). *Matching Theory.* North Holland, Amsterdam.

MACDONALD, I. G. (1995). *Symmetric functions and Hall polynomials.* The Clarendon Press Oxford University Press, New York, second edition. With contributions by A. Zelevinsky, Oxford Science Publications.

MAYER-WOLFE, E., ZEITOUNI, O., & ZERNER, M. (2001). Asymptotics of certain coagulation-fragmentation processes and invariant Poisson-Dirichlet measures. Technical report, Dept. of Mathematics, Stanford University, Stanford, CA94305.

PAGE, R. & HOLMES, E. (2000). *Molecular Evolution, A Phylogenetic Approach.* Blackwell Science.

PULLEYBLANK, W. (1995). Matchings and extensions. In Graham, R. L., Grötschel, M., & Lovász, L., editors, *Handbook of combinatorics. Vol. 1, 2*, pages Vol. I: cii+1018 pp.; Vol. II: pp. i–cii and 1019–2198, Amsterdam. Elsevier Science B.V.

ROUSSEL, S. (2000). Phénomène de cutoff pour certaines marches aléatoires sur le groupe symétrique. *Colloq. Math.* **86**, 111–135.

SAGAN, B. E. (2001). *The symmetric group.* Springer-Verlag, New York, second edition. Representations, combinatorial algorithms, and symmetric functions.

SALOFF-COSTE, L. (1997). Lectures on finite Markov chains. In *Lectures on probability theory and statistics*, pages 301–413, Berlin. Lecture Notes in Math., 1665, Springer-Verlag. Lectures from the 26th Summer School on Probability Theory held in Saint-Flour,Edited by P. Bernard.

Saxl, J. (1981). On multiplicity-free permutation representations. In *Finite geometries and designs (Proc. Conf., Chelwood Gate, 1980)*, pages 337–353. Cambridge Univ. Press, Cambridge.

Scarabotti, F. (1997). Time to reach stationarity in the Bernoulli-Laplace diffusion model with many urns. *Adv. Math.* **18**, 351–371.

Schröder, E. (1870). Veir combinatorische probleme. *Zeit. für. Math. Phys.* **15**, 361–376.

Schweinsberg, J. (2001). An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms. Technical Report 572, Dept Statistics, UC Berkeley.

Serre, J.-P. (1977). *Linear representations of finite groups.* Springer-Verlag, New York. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.

Wachs, M. (2001). Topology of matching, homology of matching, combinatorial Laplacian of the matching complex. URL `http://www.math.miami.edu/~wachs`.