COMPOSITION MARKOV CHAINS OF MULTINOMIAL TYPE

HUA ZHOU,* *Stanford University*

KENNETH LANGE,** *University of California, Los Angeles*

Abstract

Suppose that n identical particles evolve according to the same marginal Markov chain. In this setting we study chains such as the Ehrenfest chain that move a prescribed number of randomly chosen particles at each epoch. The product chain constructed by this device inherits its eigenstructure from the marginal chain. There is a further chain derived from the product chain called the composition chain that ignores particle labels and tracks the numbers of particles in the various states. The composition chain in turn inherits its eigenstructure and various properties such as reversibility from the product chain. The equilibrium distribution of the composition chain is multinomial. The current paper proves these facts in the well-known framework of state lumping and identifies the column eigenvectors of the composition chain with the multivariate Krawtchouk polynomials of Griffiths. The advantages of knowing the full spectral decomposition of the composition chain include (a) detailed estimates of the rate of convergence to equilibrium, (b) construction of martingales that allow calculation of the moments of the particle counts, and (c) explicit expressions for mean coalescence times in multi-person random walks. These possibilities are illustrated by applications to Ehrenfest chains, the Hoare and Rahman chain, Kimura's continuous-time chain for DNA evolution, a light bulb chain, and random walks on some specific graphs.

Keywords: Orthogonal polynomial; Krawtchouk polynomial; Ehrenfest chain; inhomogeneous Markov chain; coalescent time; random walk; convergence rate

2000 Mathematics Subject Classification: Primary 60J10
Secondary 60J22; 60J25; 60J27

1. Introduction

Many classical Markov chains are special cases of what we prefer to call composition chains. A composition chain follows the fate of n particles. If each particle can occupy one of d states then the composition chain records the number of particles $n_1, \ldots, n_d$ in each of the d states. In number theory a composition of a positive integer n into d parts is a representation $n = n_1 + \cdots + n_d$, where the order of the nonnegative summands is important. The composition chain has $\binom{n+d-1}{d-1}$ states. In this paper we relate the upper product chain with n distinguishable particles and d^n states to the lower composition chain. Our primary goal is to transfer the eigenstructure of the upper product chain to the composition chain. The natural assumption in many applications is that all particles evolve independently according to a common transition probability matrix (kernel) M, which, inspired by applications from genetics, we term the

mutation kernel. In some applications it is more convenient to move a random subset of the particles at each step while leaving the remaining particles fixed. If s is an integer between 1 and n, the simplest version of this strategy moves s randomly selected particles at each step. The case in which $s = n$ corresponds to moving all particles. Regardless of the choice of s, the equilibrium distribution of the lower chain is multinomial.

The count map from the upper chain to the lower chain implements a special case of state lumping previously described in the literature [1], [19]. With adequate symmetry, the lumped states function as new states of the lower chain. We briefly review lumping and demonstrate its relevance in transferring eigenvalues and eigenvectors, calculating powers of the kernel of the lumped chain, and generating martingales. In pursuing these goals, we clarify the mysterious nature of the multivariate Krawtchouk polynomials of Griffiths [9]. Our applications include the Ehrenfest chain and its generalizations, Gibbs sampling [20], the recent Hoare and Rahman chain [12], the light bulb model of Rao *et al.* [24], and the coalescence problem of Tian and Liu [25]. Further exploitation of the lumping paradigm may be possible; we mention some relevant open problems in the discussion.

The reader should keep in mind three key threads in our exposition. First, the equilibrium distribution of the composition chain is multinomial with cell probabilities determined by the equilibrium distribution of the mutation kernel M. Second, much of the theory carries over to continuous-time chains. As we go, we will indicate the necessary amendments. Third, Kronecker (tensor) products and matrix commutativity play important roles in simplifying calculations.

2. Background on lumped Markov chains

Consider two Markov chains with state spaces C and C^*. The second (lower) chain is embedded in the first (upper) chain by a map $f : C \to C^*$. This map partitions the states in C into equivalence classes under the equivalence relation $i \sim j$ when $f(i) = f(j)$. If $T = (t_{ij})$ denotes the kernel of the upper chain then it is natural to define the kernel of the lower chain to have entries

$$t^*_{f(i)f(j)} = \sum_{k \sim j} t_{ik}.$$

For the lumping to be probabilistically consistent, it is necessary and sufficient that

$$\sum_{k \sim j} t_{ik} = \sum_{k \sim j} t_{lk} \quad \text{for all } l \sim i.$$

This condition is satisfied only by chains with a special structure which include the product chains studied in Section 3. It is often helpful to restate this criterion as

$$\sum_{k \sim j} t_{ik} = \frac{1}{q_{f(i)}} \sum_{l \sim i} \sum_{k \sim j} t_{lk}, \tag{1}$$

where $q_{f(i)}$ is the number of states equivalent to i. The set of kernels that satisfy the consistency condition is clearly a convex compact set. A distribution u on the upper chain induces a distribution u^* on the lumped chain according to

$$u^*_{f(i)} = \sum_{j \sim i} u_j. \tag{2}$$

Mindful of these conventions, we state several facts. The reader can easily supply their proofs or consult the references [3], [19], or [23] for proofs.

Fact 1. *Suppose that λ is an eigenvalue of the upper chain with row eigenvector u. Then λ is an eigenvalue of the lower chain with row eigenvector u^* whose components are displayed in (2).*

Fact 2. *The lower chain is irreducible if the upper chain is irreducible. It is aperiodic if the upper chain is aperiodic.*

Fact 3. *If the upper chain is reversible with stationary distribution π then the lower chain is reversible with induced stationary distribution π^*.*

Fact 4. *The map $u \mapsto u^*$ on probability distributions satisfies the contraction inequality $\|u^* - v^*\|_{\mathrm{TV}} \le \|u - v\|_{\mathrm{TV}}$ for the total variation norms on the upper and lower state spaces.*

Some of these concepts can be rephrased using simple ideas from linear algebra [3], [19]. Let $P = (p_{ij})$ be the projection matrix taking a row vector indexed by the entries of C to a row vector indexed by the entries of C^*. Here $p_{ij} = 1$ if $f(i) = j$ and 0 otherwise. The averaging matrix $L = (l_{ij})$ reverses projection and lifts a row vector indexed by the entries of C^* to a row vector indexed by the entries of C; thus, $l_{ij} = q_i^{-1}$ if $f(j) = i$ and 0 otherwise. The pair of matrices P and L satisfies the equation $LP = I$, where I is the identity matrix. We can show that $L = (P^\top P)^{-1} P^\top$ by noting that P has orthogonal columns and hence full column rank. An easy calculation then shows that $(P^\top P)$ is diagonal with diagonal entry q_i corresponding to state $i \in C^*$. The remaining steps proving that $L = (P^\top P)^{-1} P^\top$ are straightforward. The matrix product PL equals the orthogonal projection $P(P^\top P)^{-1} P^\top$.

Fact 5. *The kernel of the lower chain is $T^* = LTP$. From left to right, this product reads lift, transition, and project. The corresponding r-step kernel is $T^{*r} = LT^r P$.*

Fact 6. *The consistency condition (1) is a disguised form of the identity $PLTP = TP$.*

Fact 7. *All eigenvalues of the lower chain are realized by projecting row eigenvectors from the upper chain to the lower chain.*

Fact 8. *We can lift column eigenvectors from the lower chain to the upper chain. If $T^* u^* = \lambda u^*$ then setting $u = P u^*$ yields $Tu = \lambda u$.*

Many lumpings are generated by group actions [2], [3]. Let the group G act on the state space C of a Markov chain with kernel T. The permutation group on $\{1, \ldots, m\}$ is often the relevant group.

Fact 9. *Suppose that $t_{ij} = t_{gi,gj}$ for all $g \in G$ and that $i \sim l$ if and only if $gi = l$ for some g. Then the orbit space C^* under G forms a lumped chain from the chain C with kernel T.*

In practice, many chains are reversible. If a chain has kernel T and equilibrium distribution π, then detailed balance is equivalent to the symmetry of the matrix DTD^{-1}, where D is the diagonal matrix $\mathrm{diag}(\sqrt{\pi})$. Taking the spectral decomposition $DTD^{-1} = U^\top \Gamma U$ yields a matrix $V = UD$ whose rows are the row eigenvectors of T. These rows v^i satisfy the orthogonality relations

$$\langle v^i, v^j \rangle_{1/\pi} = v^i D^{-2} (v^j)^\top = u^i (u^j)^\top = 1_{\{i=j\}}$$

and constitute a basis of the Hilbert space $\ell^2_{1/\pi}$. Here $1_{\{\cdot\}}$ is the indicator function. The column eigenvectors are the columns of the matrix

$$W = D^{-1}U^\top = D^{-1}D^{-1}V^\top = \mathrm{diag}(\pi^{-1})V^\top.$$

Thus, a trivial rescaling allows us to pass from row eigenvectors to column eigenvectors, or vice versa. The column eigenvectors satisfy the orthogonality condition

$$\langle w^i, w^j \rangle_\pi = (w^i)^\top D^2 w^j = v^i D^{-2}D^2 D^{-2}(v^j)^\top = v^i D^{-2}(v^j)^\top = 1_{\{i=j\}}.$$

It is a short step from these observations to show that the row and column eigenvectors are biorthogonal and that the kernel can be written as the sum $T = \sum_i \gamma_i w^i v^i$, where γ_i is the ith diagonal entry of Γ. In general, $T^k = \sum_i \gamma_i^k w^i v^i$.

The eigenvalues and eigenvectors of a reversible ergodic chain determine how fast it converges to equilibrium starting from an initial distribution μ [4].

Fact 10. *The distance from equilibrium in $\ell^2_{1/\pi}$ bounds the total variation distance from equilibrium in the sense that*

$$\|\mu - \pi\|_{\mathrm{TV}} \leq \tfrac{1}{2}\|\mu - \pi\|_{1/\pi}.$$

Fact 11. *If the eigenvalues are arranged so that the sequence $|\gamma_i|$ is decreasing then we have the chi-square bound*

$$\begin{aligned}
\|\mu T^k - \pi\|^2_{1/\pi} &= \sum_{j \geq 2} \gamma_j^{2k} \langle \mu - \pi, v^j \rangle^2_{1/\pi} \\
&= \sum_{j \geq 2} \gamma_j^{2k} [(\mu - \pi)w^j]^2 \\
&\leq \gamma_2^{2k} \|\mu - \pi\|^2_{1/\pi}.
\end{aligned}$$

For obvious reasons, $\|\mu T^k - \pi\|^2_{1/\pi}$ is called the χ^2 distance from equilibrium.

Our final fact gives a sufficient condition for the preservation of orthogonality. The condition suggests that we should look for eigenvectors that respect the equivalence relation induced by lumping.

Fact 12. *A row vector v on the upper chain is invariant on equivalence classes if and only if $vPL = v$. If the equilibrium distribution π and two row vectors v and w are constant on equivalence classes, then*

$$\langle v, w \rangle_{1/\pi} = \langle v^*, w^* \rangle_{1/\pi^*}.$$

For the most part, the same facts hold for continuous-time Markov chains if we substitute infinitesimal generators for transition probability kernels. The validity of the consistency condition (1) for the off-diagonal entries of the infinitesimal generator is equivalent to its validity for the full infinitesimal generator, since all row sums of the infinitesimal generator are 0. The part of Fact 2 dealing with aperiodicity is no longer pertinent. Fact 5 reads $e^{t\Omega^*} = Le^{t\Omega}P$, where Ω and Ω^* are the upper and lower infinitesimal generators. Fact 11 reads

$$\begin{aligned}
\|\mu e^{t\Omega} - \pi\|^2_{1/\pi} &= \sum_{j \geq 2} \exp\{2\gamma_j t\} \langle \mu - \pi, v^j \rangle^2_{1/\pi} \\
&= \sum_{j \geq 2} \exp\{2\gamma_j t\} [(\mu - \pi)w^j]^2 \\
&\leq \exp\{2\gamma_2 t\} \|\mu - \pi\|^2_{1/\pi},
\end{aligned}$$

where $\gamma_2 < 0$ is the second largest eigenvalue of a reversible chain with equilibrium distribution π. The biorthogonal representation $\Omega = \sum_i \gamma_i w^i v^i$ leads to the compact formula

$$e^{t\Omega} = \sum_i \exp\{\gamma_i t\} w^i v^i = \sum_i \exp\{\gamma_i t\} w^i (w^i)^\top \operatorname{diag}(\pi) \tag{3}$$

for the finite-time transition probability matrix.

3. Lumping of product Markov chains

In dealing with the upper Markov chain generating a composition Markov chain, it is convenient to list the Cartesian product states in the upper chain in dictionary order. When the particles move independently and simultaneously ($s = n$), this choice allows us to write the product kernel as the tensor product $M \otimes \cdots \otimes M$. If π is the equilibrium distribution for a single particle then $\pi \otimes \cdots \otimes \pi$ is the equilibrium distribution for the upper chain. The tensor product multiplication rule

$$(\pi \otimes \cdots \otimes \pi)(M \otimes \cdots \otimes M) = (\pi M) \otimes \cdots \otimes (\pi M) = \pi \otimes \cdots \otimes \pi$$

shows how these conventions meld perfectly. If we take $s < n$ then the kernel M_s of the upper chain can be represented as a convex combination of Kronecker products. Let M_S be the tensor product matrix with M inserted in the slots corresponding to the elements of the set S and the identity matrix I inserted in the remaining slots. For example, with $n = 3$ and $S = \{1, 3\}$, $M_S = M \otimes I \otimes M$. In this notation we have

$$M_s = \binom{n}{s}^{-1} \sum_{|S|=s} M_S. \tag{4}$$

Lumping can be approached from two directions. A permutation σ on $\{1, \ldots, n\}$ takes a vector $u_1 \otimes \cdots \otimes u_n$ into the vector $u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(n)}$. This group action is compatible with lumping and satisfies the consistency condition (1) because M_s is defined symmetrically over all subsets S of size s. The same conclusion can be reached without invoking group actions by observing that consistency is obvious for $s = n$. If $s < n$, consider a subset S of size s. Since the counts contributed by the particles outside S do not change on application of M_S, for all intents the proof of lumpability of M_S reduces to the case in which $s = n$. Finally, since the set of lumpable kernels is convex, lumpability holds for M_s as well. We stress that more complicated kernels also lead to lumpability.

According to Fact 2, the composition chain is ergodic (irreducible and aperiodic) whenever M_s is ergodic. In general, a chain with kernel K is ergodic if and only if some power K^r has all entries positive. Suppose that the mutation matrix M is ergodic with positivity exponent r. By considering successive subsets $S = \{ms + 1, \ldots, (m + 1)s\}$, we can demonstrate that, for sufficiently large k, M_s^k exceeds a positive multiple of $M^{k_1} \otimes \cdots \otimes M^{k_n}$, where each $k_i \geq r$. Thus, M_s is ergodic whenever M is ergodic.

Because $(\pi \otimes \cdots \otimes \pi)M_s = (\pi \otimes \cdots \otimes \pi)$, the upper chain has the product equilibrium distribution regardless of the value of s. Equation (2) shows that the product equilibrium distribution translates into the multinomial equilibrium distribution on the composition chain. It turns out that reversibility of the upper chain is both a necessary and sufficient condition for reversibility of the lower chain. Necessity is proved by considering two states of the composition chain. In the first composition, all particles occupy state i; in the second composition, $n - s$

particles occupy state i and s particles occupy state j. If detailed balance holds in the composition chain then

$$\pi_i^n m_{ij}^s = \binom{n}{s} \pi_i^{n-s} \pi_j^s \binom{n}{s}^{-1} m_{ji}^s.$$

Simple algebra now yields the detailed balance condition $\pi_i m_{ij} = \pi_j m_{ji}$ relevant to the mutation matrix and, hence, to the product chain.

These conclusions are summarized in the next proposition.

Proposition 1. *Under the kernel M_s, the composition chain is ergodic whenever M is ergodic. If the mutation matrix M has equilibrium distribution π then the composition chain has as its equilibrium distribution the multinomial distribution with n trials and success probability π_i for state i. The composition chain is reversible if and only if the mutation kernel M is reversible.*

We now consider the more interesting problem of determining the eigenstructure of the composition chain. To ease the notational burden, we will use multi-index notation to indicate count vectors. Thus, a tilded letter, e.g. $\tilde{m}$, stands for a vector of counts $(m_1, \ldots, m_d)$ and its absolute value $|\tilde{m}|$ stands for the sum of the counts $m_1 + \cdots + m_d$. If p is a vector with d components then $p^{\tilde{m}} = \prod_{i=1}^d p_i^{m_i}$.

If $M = V^{-1}\Gamma V$ diagonalizes M then

$$M_s = (V^{-1} \otimes \cdots \otimes V^{-1})\left(\binom{n}{s}^{-1} \sum_{|S|=s} \Gamma_S\right)(V \otimes \cdots \otimes V) \tag{5}$$

diagonalizes M_s, where Γ_S is derived from Γ in the same manner as M_S is derived from M. Diagonalization (5) provides us with the row eigenvectors of M_s. To construct one, take a tensor product $v^{i_1} \otimes \cdots \otimes v^{i_n}$, where v^i is the ith row of V. The corresponding eigenvalue appears in the equation

$$v^{i_1} \otimes \cdots \otimes v^{i_n} M_s = \left(\binom{n}{s}^{-1} \sum_{|S|=s} \prod_{j \in S} \gamma_{i_j}\right) v^{i_1} \otimes \cdots \otimes v^{i_n}.$$

The eigenvalues for two vectors $v^{i_1} \otimes \cdots \otimes v^{i_n}$ and $v^{j_1} \otimes \cdots \otimes v^{j_n}$ coincide if the index vectors $(i_1, \ldots, i_n)$ and $(j_1, \ldots, j_n)$ generate the same count vector $\tilde{n}$. In this case the common eigenvalue is

$$\beta_{\tilde{n}} = \binom{n}{s}^{-1} \sum_{|S|=s} \prod_{j \in S} \gamma_{i_j} = \binom{n}{s}^{-1} \sum_{\tilde{k}} \prod_{l=1}^d \binom{n_l}{k_l} \gamma_l^{k_l}, \tag{6}$$

where $\tilde{k}$ ranges over all multi-indices with $|\tilde{k}| = s$.

Given lumpability, Fact 1 implies that a row eigenvector $v^{i_1} \otimes \cdots \otimes v^{i_n}$ of the product chain maps to a row eigenvector $(v^{i_1} \otimes \cdots \otimes v^{i_n})P$ of the composition chain with the eigenvalue $\beta_{\tilde{n}}$ displayed in (6). To calculate the projected eigenvector, let u^i be the standard unit basis vector of $\mathbb{R}^d$. Multilinearity implies that

$$[(v^{i_1} \otimes \cdots \otimes v^{i_n})P]_{\tilde{m}} = \sum_{j_1} \cdots \sum_{j_n} v_{j_1}^{i_1} \cdots v_{j_n}^{i_n}[(u^{j_1} \otimes \cdots \otimes u^{j_n})P]_{\tilde{m}}$$

$$= \sum_{(k_{ij})} \prod_{i=1}^d \binom{n_i}{k_{i1} \cdots k_{id}} \prod_{j=1}^d (v_j^i)^{k_{ij}},$$

where (k_{ij}) indicates a $d \times d$ table of nonnegative integers satisfying the row and column sum constraints $\sum_j k_{ij} = n_i$ and $\sum_i k_{ij} = m_j$.

This discussion omits all mention of orthogonality and norms. In the presence of reversibility, these come into play. On a tensor product space with identical factors, the induced inner product satisfies the identity $\langle x^1 \otimes \cdots \otimes x^n, y^1 \otimes \cdots \otimes y^n \rangle = \langle x^1, y^1 \rangle \cdots \langle x^n, y^n \rangle$. This partial definition extends by bilinearity to the full tensor product space. In view of our discussion of reversible chains, the most pertinent inner product for each particle is $\langle \cdot, \cdot \rangle_{1/\pi}$. The inner product induced on the tensor product by this choice is the inner product $\langle \cdot, \cdot \rangle_{1/\nu}$ corresponding to the product equilibrium distribution $\pi \otimes \cdots \otimes \pi = \nu$. To construct an orthonormal basis of the tensor product space, we take the spectral decomposition $V^{-1} \Gamma V$ of the mutation matrix M and assemble all possible vectors $v^{i_1} \otimes \cdots \otimes v^{i_n}$ from the rows of V.

If ν^* is the multinomial distribution generated by ν then we need to relate the two inner products $\langle x, y \rangle_{1/\nu}$ and $\langle x^*, y^* \rangle_{1/\nu^*}$. The key is to symmetrize and invoke Fact 12. We symmetrize a basis element $v^{i_1} \otimes \cdots \otimes v^{i_n}$ of the product chain by taking

$$u = \frac{1}{n!} \sum_\sigma v^{i_{\sigma_1}} \otimes \cdots \otimes v^{i_{\sigma_n}}, \tag{7}$$

where the sum extends of all permutations σ of $\{1, \ldots, n\}$. Symmetry dictates that u is constant on equivalence classes and projects to the same vector as the original vector. If $(i_1, \ldots, i_n)$ generates the count vector $\tilde{n}$ then

$$\langle u, u \rangle_{1/\nu} = \frac{n!}{(n!)^2} n_1! \cdots n_d! = \binom{n}{\tilde{n}}^{-1}.$$

It follows that $x = \sqrt{\binom{n}{\tilde{n}}} u$ has unit norm in $\ell^2_{1/\nu}$. According to Fact 12, its projection x^* also has unit norm $\ell^2_{\nu^*}$. When we construct another symmetrized vector y that maps to $y^* \neq x^*$, then x and y are orthogonal and x^* and y^* are orthogonal as well.

In the presence of reversibility, it is straightforward to recover the column eigenvectors of the composition chain. To the row eigenvector v^i for a single particle there corresponds the column eigenvector $w^i = \mathrm{diag}(\pi^{-1})(v^i)^\top$. The column eigenvector corresponding to the symmetrized vector u in (7) has entries

$$z_{\tilde{m}} = \frac{1}{\pi^{\tilde{m}}} \binom{n}{\tilde{m}}^{-1} \sum_{(k_{ij})} \prod_{i=1}^d \binom{n_i}{k_{i1} \cdots k_{id}} \prod_{j=1}^d (v^i_j)^{k_{ij}}$$

$$= \frac{1}{\pi^{\tilde{m}}} \binom{n}{\tilde{n}}^{-1} \sum_{(k_{ij})} \prod_{j=1}^d \binom{m_j}{k_{1j} \cdots k_{dj}} \prod_{i=1}^d (v^i_j)^{k_{ij}}$$

$$= \binom{n}{\tilde{n}}^{-1} \sum_{(k_{ij})} \prod_{j=1}^d \binom{m_j}{k_{1j} \cdots k_{dj}} \prod_{i=1}^d \left(\frac{v^i_j}{\pi_j}\right)^{k_{ij}}$$

$$= \binom{n}{\tilde{n}}^{-1} \sum_{(k_{ij})} \prod_{j=1}^d \binom{m_j}{k_{1j} \cdots k_{dj}} \prod_{i=1}^d (w^i_j)^{k_{ij}}$$

$$= \binom{n}{\tilde{n}}^{-1} K_{\tilde{n}}(\tilde{m}).$$

The function $K_{\tilde{n}}(\tilde{m})$ is a multivariate Krawtchouk polynomial; we will say more about these in a moment. If instead of starting with u we start with the normalized vector x, then z is a unit vector. In other words, the scaled Krawtchouk polynomial

$$K_{\tilde{n}}^0(\tilde{m}) = \binom{n}{\tilde{n}}^{-1/2} K_{\tilde{n}}(\tilde{m})$$

has unit norm in $\ell_{\nu^*}^2$. Furthermore, these polynomials constitute an orthonormal basis of column vectors.

Our second proposition summarizes these findings.

Proposition 2. *The row eigenvectors of the product chain M_s project to row eigenvectors of the composition chain with the shared eigenvalue $\beta_{\tilde{n}}$ displayed in (6). If the mutation matrix is reversible then the symmetrized and scaled eigenvectors of the product chain project to an orthonormal basis of the composition chain. These in turn map to a dual orthonormal basis of scaled Krawtchouk polynomials $K_{\tilde{n}}^0(\tilde{m})$ of the composition chain with the same eigenvalues.*

4. The role of Krawtchouk polynomials

Griffiths [9] defined the multivariate Krawtchouk polynomials $K_{\tilde{n}}(\tilde{m})$ via their generating function

$$G(\tilde{m}, y) = \sum_{|\tilde{n}|=n} K_{\tilde{n}}(\tilde{m}) y^{\tilde{n}} = \prod_{i=1}^d \left(\sum_{j=1}^d y_j w_i^j \right)^{m_i},$$

under the restrictions $y_1 = 1$ and $w^1 = \mathbf{1}$. Equating coefficients of $y^{\tilde{n}}$ shows that

$$K_{\tilde{n}}(\tilde{m}) = \sum_{(k_{ij})} \prod_{j=1}^d \binom{m_j}{k_{1j} \cdots k_{dj}} \prod_{i=1}^d (w_j^i)^{k_{ij}}.$$

Of course, this is exactly how we defined $K_{\tilde{n}}(\tilde{m})$ earlier. For the case in which $d = 2$, the generating function reduces to

$$G(\tilde{m}, y) = (1 + y_2 w_1^2)^{m_1} (1 + y_2 w_2^2)^{m_2} = \left(1 - y_2 \sqrt{\frac{\pi_2}{\pi_1}} \right)^{m_1} \left(1 + y_2 \sqrt{\frac{\pi_1}{\pi_2}} \right)^{m_2}$$

with $m_1 + m_2 = n$. Extracting the coefficient of $y_2^{n_2}$ yields

$$K_{n_1, n_2}(m_1, m_2) = \sum_{l_1 + l_2 = n_2} \binom{m_1}{l_1} \binom{m_2}{l_2} \left(-\sqrt{\frac{\pi_2}{\pi_1}} \right)^{l_1} \left(\sqrt{\frac{\pi_1}{\pi_2}} \right)^{l_2}$$

$$= \binom{n}{n_2} \left(\sqrt{\frac{\pi_1}{\pi_2}} \right)^{n_2} \sum_{l_1 + l_2 = n_2} \binom{m_1}{l_1} \binom{m_2}{l_2} \binom{n}{n_2}^{-1} \left(-\frac{\pi_2}{\pi_1} \right)^{l_1},$$

$$K_{n_1, n_2}^0(m_1, m_2) = \sqrt{\binom{n}{n_2} \left(\frac{\pi_1}{\pi_2} \right)^{n_2}} \sum_{l_1 + l_2 = n_2} \binom{m_1}{l_1} \binom{m_2}{l_2} \binom{n}{n_2}^{-1} \left(-\frac{\pi_2}{\pi_1} \right)^{l_1}.$$

These expressions can be simplified by bringing in the hypergeometric function

$$_2F_1(a, b; c; u) = \sum_{l=0}^\infty \frac{a_{(l)} b_{(l)}}{c_{(l)}} \frac{u^l}{l!}$$

and writing

$$\sum_{l_1+l_2=n_2} \binom{m_1}{l_1}\binom{m_2}{l_2}\binom{n}{n_2}^{-1}(-u)^{l_1} = \frac{(m_2)_{[n_2]}}{n_{[n_2]}} {}_2F_1(-n_2,-m_1;m_2-n_2+1;-u)$$

$$= {}_2F_1(-n_2,-m_1;-n;1+u).$$

Here we have applied the standard transformation formula [8],

$$_2F_1(a,b;c;z) = \frac{(b-c)_{[-a]}}{(-c)_{[-a]}} {}_2F_1(a,b;1+a+b-c;1-z),$$

and used the abbreviations $a_{[l]} = a(a-1)\cdots(a-l+1)$ and $a_{(l)} = a(a+1)\cdots(a+l-1)$ to signify falling and rising factorials. The final result,

$$K^0_{n_1,n_2}(m_1,m_2) = \sqrt{\binom{n}{n_2}\left(\frac{\pi_1}{\pi_2}\right)^{n_2}} {}_2F_1\left(-n_2,-m_1;-n;1+\frac{\pi_2}{\pi_1}\right)$$

$$= \sqrt{\binom{n}{n_2}\left(\frac{\pi_1}{\pi_2}\right)^{n_2}} {}_2F_1\left(-n_2,-m_1;-n;\frac{1}{\pi_1}\right),$$

can also be recovered from the classical definition

$$\sum_{k=0}^{n} \binom{n}{k} {}_2F_1\left(-k,-x;-n;\frac{1}{\pi_1}\right) t^k = \left(1-\frac{\pi_2}{\pi_1}t\right)^x (1+t)^{n-x}$$

of the univariate Krawtchouk polynomials by taking $t = y_2\sqrt{\pi_1/\pi_2}$. Ismail [13, p. 184] summarized the properties of the univariate Krawtchouk polynomials.

Although the multivariate Krawtchouk polynomials are complicated, a few simple cases are noteworthy. Let e_j be the standard unit vector with entry j equal to 1 and the remaining $d-1$ entries equal to 0. The values

$$K_{\tilde{n}}(ne_j) = \binom{n}{\tilde{n}} \prod_{i=1}^{d} (w_j^i)^{n_i}$$

$$\text{and} \quad K^0_{\tilde{n}}(ne_j) = \sqrt{\binom{n}{\tilde{n}}} \prod_{i=1}^{d} (w_j^i)^{n_i}$$

are easy to deduce. It is also clear that $K_{ne_1}(\tilde{m}) = K^0_{ne_1}(\tilde{m}) = 1$.

Our approach to the multivariate Krawtchouk polynomials gives a natural proof of their orthogonality under the multinomial distribution and clarifies their role in Markov chain dynamics. Our next proposition showcases the value of the Krawtchouk polynomials in computing the χ^2 distance of a composition chain from equilibrium.

Proposition 3. *When the composition process starts from state $\tilde{m}$, the chi-square distance to stationarity after l steps is*

$$\chi^2_{\tilde{m}}(l) = \sum_{\tilde{n}\neq ne_1} \beta_{\tilde{n}}^{2l} K^0_{\tilde{n}}(\tilde{m})^2. \tag{8}$$

Proof. This is a direct application of Fact 11.

In applying (8) it is helpful to know specific values of the $\beta_{\tilde{n}}$. When $d = 2$, we can express

$$\beta_{\tilde{n}} = \binom{n}{s}^{-1} \sum_{k=0}^{s} \binom{n_1}{s-k}\binom{n_2}{k} \gamma_2^k = {}_2F_1(-n_2, -s; -n; 1 - \gamma_2). \tag{9}$$

The case in which $s = n$ gives

$$\beta_{\tilde{n}} = \prod_{i=1}^{d} \gamma_i^{n_i}.$$

Thus, the second largest eigenvalue in magnitude of the single-particle chain and the composition chain coincide. Furthermore, Proposition 2 shows that the chi-square distance from equilibrium after l steps starting from state ne_j is

$$\chi^2_{ne_j}(l) = \sum_{\tilde{n} \neq ne_1} \binom{n}{\tilde{n}} \prod_{i=1}^{d} (\gamma_i^l w_j^i)^{2n_i}.$$

Finally, the case in which $s = 1$ yields the convex combination

$$\beta_{\tilde{n}} = \sum_{i=1}^{d} \frac{n_i}{n} \gamma_i.$$

If the γ_i are arranged in decreasing order then the reader can check that the second largest eigenvalue in magnitude of the composition chain is either $(n-1)/n + (1/n)\gamma_2$ or γ_d, whichever is larger in magnitude.

Equation (6) also has some qualitative implications. For instance, it is well known that the number of irreducible classes of a Markov chain equals the multiplicity of the eigenvalue 1 of its kernel [18]. Thus, if the mutation matrix M has r irreducible classes then $\gamma_1 = \cdots = \gamma_r = 1$, and every count vector of the form $\tilde{n} = (n_1, \ldots, n_r, 0, \ldots, 0)$ gives rise to $\beta_{\tilde{n}} = 1$. Hence, the composition chain has at least $\binom{n+r-1}{n}$ irreducible classes. We can similarly employ the fact that an irreducible chain has period r if and only if all roots of unity of order r are eigenvalues of its kernel [18]. For instance, if the mutation matrix M is reversible and periodic, then $\gamma_d = -1$. Equation (6) now shows that $\beta_{ne_1} = 1$ and $\beta_{ne_d} = (-1)^s$. When s is odd, we deduce that the composition chain is periodic, and when s is even, we deduce that the composition chain has at least two irreducible classes. These conclusions can be drawn from more elementary arguments, but it is pleasing that they fall out so easily from a clear understanding of how eigenvalues transfer.

If the full eigenstructure of the single-particle chain is unknown then Fact 4 can be invoked to bound the rate of convergence of the composition chain. For example, suppose that the single-particle chain is close to the stationary distribution π after T steps. Once each particle has been moved at least T times, then both the product chain and the composition chain should be close to equilibrium as well. In many interesting models a strong stationary time T can be constructed.

5. Continuous-time chains

Propositions 1 and 2 continue to hold for continuous-time Markov chains provided that the n identical particles move independently. In this setting we replace the mutation matrix M by a

mutation rate matrix Q. The finite-time transition matrix for a single particle equals the matrix exponential e^{tQ}. If we retain the dictionary ordering of states then the finite-time transition matrix for the product chain equals the tensor product $e^{tQ} \otimes \cdots \otimes e^{tQ}$. The derivative of this matrix at time $t = 0$ equals the infinitesimal rate matrix of the product chain. Because the matrix exponential satisfies the law of exponents for commuting arguments, the product rule of differentiation gives

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{tQ} \otimes \cdots \otimes e^{tQ} \mid_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\prod_{k=1}^{n} I \otimes \cdots \otimes I \otimes e^{tQ} \otimes I \otimes \cdots \otimes I\right]_{t=0}$$

$$= \sum_{k=1}^{n} I \otimes \cdots \otimes I \otimes Q \otimes I \otimes \cdots \otimes I.$$

Except for a missing factor of n^{-1} and the replacement of M by Q, the infinitesimal rate matrix is identical to the product kernel (4) with $s = 1$. Instead of the eigenvalue in (6) for $s = 1$, we now have the eigenvalue

$$\beta_{\tilde{n}} = \sum_{i=1}^{d} n_i \gamma_i. \tag{10}$$

If the eigenvalues γ_i of Q are arranged to have decreasing real parts then $\gamma_1 = 0$ and γ_2 is the eigenvalue with the second largest real part simultaneously for the single-particle chains, the product chain, and the composition chain.

6. Applications

Many contemporary and classical Markov chains fit into our framework.

6.1. Ehrenfest chains

In the classical Ehrenfest model [7], [15], n balls are shuttled between $d = 2$ urns. At each step a ball ($s = 1$) is randomly chosen and shifted to the other urn. The composition (Ehrenfest) chain tracks the number of balls in each urn. In this case the mutation matrix

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{11}$$

has period 2, eigenvalues $\gamma_1 = 1$ and $\gamma_2 = -1$, and equilibrium distribution $\pi = (\frac{1}{2}, \frac{1}{2})$. The product chain is a random walk on the hypercube Z_2^n [4]. The composition chain is reversible with period 2. Its equilibrium distribution is binomial with n trials and success probability $\frac{1}{2}$. Its eigenvalues are $\beta_{n_1, n_2} = 1 - 2n_2/n$ and its column eigenvectors are univariate Krawtchouk polynomials.

Karlin and McGregor [16] generalized the Ehrenfest urn model to $d > 2$ urns. The discrete analog of their continuous-time Markov chain randomly chooses a single ball ($s = 1$) and redistributes it to another urn according to the probability distribution $\pi = (\pi_1, \ldots, \pi_d)$. The composition (Ehrenfest) chain counts the number of balls in each urn. In this setting the mutation matrix M has all rows equal to π, which is obviously the equilibrium distribution. The eigenvalues are $\gamma_1 = 1$ and $\gamma_2 = \cdots = \gamma_d = 0$. Owing to the multiplicity of the eigenvalue 0, any orthonormal basis $w^1, \ldots, w^d$ of ℓ_π^2 with $w^1 = \mathbf{1}$ diagonalizes M.

The composition chain has the eigenvalue $\beta_{\tilde{n}} = n_1/n$ with multiplicity $\binom{d-2+n-n_1}{n-n_1}$. The multivariate Krawtchouk polynomials enter when we make the explicit choice

$$w^i = \frac{1}{\sqrt{(1/\pi_{i-1})(\sum_{j=i}^d \pi_j)^2 + \sum_{j=i}^d \pi_j}}\left(0, \ldots, 0, -\frac{1}{\pi_{i-1}}\sum_{j=i}^d \pi_j, 1, \ldots, 1\right)^\top$$

for $2 \le i \le d$, in diagonalizing the composition chain. These results unify the algebraic perspective of Griffiths [9] with the specific version of the multivariate Krawtchouk polynomials given by Iliev and Xu [14, Theorem 6.2] and implicit in the earlier work of Karlin and McGregor [16].

Finally, we can generalize even further by selecting $s > 1$ balls at each step and redistributing each according to the probability vector π. Diaconis *et al.* [5] considered the univariate case, $d = 2$, in studying convergence rates of various Gibbs samplers. Khare and Zhou [20] obtained the explicit diagonalization in the multivariate case. The mutation matrix M and composition chain eigenvectors obviously remain the same. The composition chain eigenvalue $\beta_{\tilde{n}} = \binom{n_1}{s}/\binom{n}{s}$ has multiplicity $\binom{d-2+n-n_1}{n-n_1}$, where $0 \le n_1 \le n$.

6.2. Hoare and Rahman chain

In this example we consider a general urn model combining features of the Hoare–Rahman [12] model and the Karlin–McGregor model. Suppose that there are n balls and d urns. All balls ($s = n$) are moved independently at each step according to the following mechanism. If a ball is in urn i then with probability α_i it remains there; with probability $1 - \alpha_i$ it is redistributed to another urn according to the probability vector $\theta = (\theta_1, \ldots, \theta_d)^\top$. Note that a ball can be redistributed to its current urn. Hoare and Rahman [12] considered the special case in which $d = 3$ and $\alpha_d = 0$. Griffiths [10] commented on the role of the multivariate Krawtchouk polynomials in the restricted Hoare–Rahman model. In the general version of the model, the mutation matrix is reversible with entries

$$m_{ij} = \begin{cases} (1 - \alpha_i)\theta_j, & i \ne j, \\ \alpha_i + (1 - \alpha_i)\theta_i, & i = j, \end{cases}$$

and equilibrium distribution

$$\pi = \frac{1}{\sum_{i=1}^d \theta_i/(1 - \alpha_i)}\left(\frac{\theta_1}{1 - \alpha_1}, \ldots, \frac{\theta_d}{1 - \alpha_d}\right).$$

If we view $\alpha = (\alpha_1, \ldots, \alpha_d)^\top$ and θ as column vectors, then M can be written as the sum of the diagonal matrix $\text{diag}(\alpha)$ plus the outer product matrix $(\mathbf{1} - \alpha)\theta^\top$.

We can calculate the characteristic polynomial $p(\gamma)$ of M by Woodbury's formula [22] for the determinant of a low rank perturbation of a square matrix. The result is

$$\begin{aligned} p(\gamma) &= \det[\gamma I - \text{diag}(\alpha) - (\mathbf{1} - \alpha)\theta^\top] \\ &= \det[\gamma I - \text{diag}(\alpha)]\{1 - \theta^\top[\gamma I - \text{diag}(\alpha)]^{-1}(\mathbf{1} - \alpha)\} \\ &= \prod_{i=1}^d (\gamma - \alpha_i) - \sum_{i=1}^d \theta_i(1 - \alpha_i)\prod_{j\ne i}(\gamma - \alpha_j). \end{aligned}$$

If k of the α_i coincide then their common value is a root of $p(\gamma)$ of multiplicity $k - 1$. For instance, if $\alpha_d = 0$ and $\alpha_1 = \cdots = \alpha_{d-1} = \alpha$, then

$$p(\gamma) = (\gamma - 1)(\gamma - \alpha)^{d-2}(\gamma - \alpha\theta_d),$$

and M has eigenvalues $\gamma_1 = 1$, $\gamma_2 = \cdots = \gamma_{d-1} = \alpha$, and $\gamma_d = \alpha\theta_d$. To generate corresponding column eigenvectors, we take $w^1 = \mathbf{1}$ and $d - 2$ unit vectors $w^2, \ldots, w^{d-1}$ of the form $w = (u^\top, 0)^\top$ orthogonal to w^1 in ℓ_π^2. Prior to normalization, the last eigenvector w^d has its first $d - 1$ entries equal to 1 and its last entry equal to $(1 - \theta_d)/\theta_d(\alpha - 1)$. The reader can check that these choices produce an eigenbasis of ℓ_π^2. Proposition 1 implies that the composition chain is reversible with respect to the multinomial distribution (n, π), and Proposition 2 shows that the eigenvalues of the composition chain reduce to $\beta_{\tilde{n}} = \alpha^{n-n_1}\theta_d^{n_d}$. The appropriate multivariate Krawtchouk polynomials serve as eigenvectors.

6.3. Evolution of DNA compositions

Evolutionary biologists are keenly interested in models of base-pair substitution in DNA. Recall that DNA is a double helix constructed from four nucleotides abbreviated A, G, C, and T. Two of these, A and G, are purines, and two, C and T, are pyrimidines. There is a long tradition of modeling the mutational changes at a single site as a continuous-time Markov chain. For example, Kimura's [21] two-parameter model specifies the infinitesimal generator

$$Q = \begin{pmatrix} -(\alpha + 2\omega) & \alpha & \omega & \omega \\ \alpha & -(\alpha + 2\omega) & \omega & \omega \\ \omega & \omega & -(\alpha + 2\omega) & \alpha \\ \omega & \omega & \alpha & -(\alpha + 2\omega) \end{pmatrix}.$$

This reversible chain has the uniform equilibrium distribution $\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. The chain offers a simple but lovely example of lumping if we take A and G to be equivalent and C and T to be equivalent. From the purine-pyrimidine chain we can immediately deduce the eigenvalues $\gamma_1 = 0$ and $\gamma_2 = -4\omega$ corresponding to the column eigenvectors $w^1 = \mathbf{1}$ and $w^2 = (1, 1, -1, -1)^\top$. The remaining double eigenvalue $\gamma_3 = \gamma_4 = -2(\alpha + \omega)$ has eigenvectors $w^3 = (\sqrt{2}, -\sqrt{2}, 0, 0)^\top$ and $w^4 = (0, 0, -\sqrt{2}, \sqrt{2})^\top$.

As a first approximation, most biologists are willing to assume that different sites evolve independently. Under independence and Kimura's model, the composition chain counting the number of bases of the various types at n sites approaches a multinomial distribution with parameters (n, π). Equations (3) and (10) imply that

$$\Pr(\tilde{X}_t = \tilde{y} \mid \tilde{X}_0 = \tilde{x}) = \binom{n}{\tilde{y}}\frac{1}{4^n}\sum_{\tilde{n}}\exp\left\{t\sum_i n_i\gamma_i\right\}K_{\tilde{n}}^0(\tilde{x})K_{\tilde{n}}^0(\tilde{y}),$$

where the $K_{\tilde{n}}^0$ are the normalized Krawtchouk polynomials defined from the w^i.

6.4. The light bulb problem

This problem of Rao *et al.* [24] served as the impetus for the current paper. They were motivated by a certain pharmaceutical experiment. For our purposes, it is clearer to adopt their alternative formulation involving light bulbs. There are n light bulbs, and each bulb is either on or off. At epoch t we randomly select s_t light bulbs. If a selected bulb is on then we switch it off, and if it is off then we switch it on. The composition chain follows the number of on

bulbs. We will temporarily limit our attention to the homogeneous case, $s_t = s$. The reversible mutation matrix

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

has the uniform distribution $\pi = (\frac{1}{2}, \frac{1}{2})$ as its equilibrium. The eigenvalues of M are $\gamma_1 = 1$ and $\gamma_2 = -1$ with corresponding column eigenvectors $w^1 = (1, 1)^\top$ and $w^2 = (-1, 1)^\top$. The discussion at the end of Section 4 indicates that the composition chain has period 2 when s is odd and two irreducible classes when s is even. In the former case, the chain alternates between having an odd and an even number of bulbs on. In the latter case, the even and odd parity states do not communicate. According to Proposition 1, the composition chain is reversible with the binomial distribution $(n, \frac{1}{2})$ as its equilibrium. Proposition 2 determines the eigenvalues as

$$\beta_{n_1,n_2} = {}_2F_1(-n_2, -s; -n; 2)$$

and the eigenvectors as univariate Krawtchouk polynomials.

The commutativity of the kernels M_s is the key to understanding the inhomogeneous version of the light bulb problem. Commutativity implies that the product chains share a common basis of row eigenvectors; this basis projects to a common basis of row eigenvectors for the composition chains and from there to a common basis of column eigenvectors for the composition chains. It is well known that we can construct a martingale Y_t relative to a Markov chain X_t by taking a column eigenvector w with eigenvalue λ and setting $Y_t = \lambda^{-t} w_{X_t}$ [17, p. 242]. For a time-inhomogeneous chain X_t with a shared eigenvector w but differing eigenvalues λ_r at each epoch r, the standard argument demonstrates that $Y_t = (\prod_{r=1}^t \lambda_r)^{-1} w_{X_t}$ is a martingale. The basic property $E(Y_t) = E(Y_0)$ gives us $E(Y_t)$ for free whenever Y_0 is constant.

The martingale construction makes it possible to calculate the moments of the number of on bulbs $X_t = m_2$ at epoch t. We will rest content with the first two moments. For the mean of the composition chain, the relevant eigenvalue and eigenvector are

$$\beta_{n-1,1} = 1 - \frac{2s}{n}, \qquad K_{n-1,1}(m_1, m_2) = 2m_2 - n.$$

For the second moment of the composition chain, these become

$$\beta_{n-2,2} = 1 - \frac{4s}{n} + \frac{4s(s-1)}{n(n-1)}, \qquad K_{n-2,2}(m_1, m_2) = \frac{n(n-1)}{2} + 2m_2(m_2 - n).$$

With these ideas in mind, we state and prove a proposition generalizing various results of Rao *et al.* [24].

Proposition 4. *Suppose that $s_1, s_2, \ldots$ is a switching pattern in the light bulb problem with n total bulbs. If X_t is the number of on bulbs after t steps then*

$$E(X_t \mid X_0) = \frac{n}{2}\left(1 - \left(1 - \frac{2X_0}{n}\right)\prod_{r=1}^t \left(1 - \frac{2s_r}{n}\right)\right),$$

$$E(X_t^2 \mid X_0) = \frac{n(n-1)}{4}\left\{\left(1 - \frac{4X_0(n - X_0)}{n(n-1)}\right)\prod_{r=1}^t \left(1 - \frac{4s_r(n - s_r)}{n(n-1)}\right) - 1\right\}$$
$$+ n\, E(X_t \mid X_0).$$

The distribution of X_t given X_0 can be expressed as

$$\Pr(X_t = y \mid X_0 = x)$$

$$= 2^{-n} \binom{n}{y} \sum_{k=0}^{n} \left(\prod_{r=1}^{t} {}_2F_1(-k, -s_r; -n; 2) \right) K_{n-k,k}^0(n-x, x) K_{n-k,k}^0(n-y, y).$$

Proof. Given our identification of the appropriate martingales, the displayed formulae for the first two moments are straightforward to derive. The identity for the discrete density is a consequence of the following general considerations. Let T_r be the kernel of an inhomogeneous Markov chain at step r. Suppose that the T_r are reversible and possess a common basis $\{v^i\}$ of row eigenvectors and a corresponding basis of column eigenvectors $\{w^i\}$. The biorthogonal decomposition $T_r = \sum_i \gamma_{ri} w^i v^i$ then leads to the t-step kernel

$$\prod_{r=1}^{t} \left(\sum_i \gamma_{ri} w^i v^i \right) = \sum_i \left(\prod_{r=1}^{t} \gamma_{ri} \right) w^i v^i$$

$$= \sum_i \left(\prod_{r=1}^{t} \gamma_{ri} \right) w^i (w^i)^\top \operatorname{diag}(\pi),$$

where π is the common equilibrium distribution and the γ_{ri} are the eigenvalues. To apply this result to our specific chain, we merely collect the eigenvalues from (9) and equate π to the binomial distribution $(n, \frac{1}{2})$.

6.5. Coalescence times for a multi-person random walk on a graph

It is well known that any reversible Markov chain can be interpreted as a random walk on a graph. This fact suggests rephrasing our process as a multi-person random walk. Thus, we envision n people distributed on the vertices of a graph whose arcs are assigned weights according to the entries of the mutation kernel M. At each step, we randomly choose s people and move them independently to neighboring vertices. The composition chain records the number of people located at each vertex at each epoch.

Coalescent times are among the most interesting random variables arising in a composition chain. By a coalescent time we mean the number of steps until all n people meet at a given vertex. Recently, Tian and Liu [25] have attacked the problem of calculating mean coalescent times for continuous-time walks using the machinery of tensor products. Their exact results are limited to unweighted regular graphs. We now tackle the same problem for discrete-time walks on weighted graphs for general s. Our strategy is to derive the probability generating functions of the coalescent times and extract from them their moments and approximate probabilities.

We begin with a general Markov chain X_m with kernel K. For two states x and y, we define the first hitting time $T_{xy} = \inf\{j \geq 1 : X_j = y, X_0 = x\}$. When $x = y$, T_{xx} is the first return time to x. To T_{xy} and K we associate the generating functions

$$G_{T_{xy}}(z) = \sum_{j=1}^{\infty} \Pr(T_{xy} = j) z^j, \qquad G_{K_{xy}}(z) = \sum_{j=0}^{\infty} (K^j)_{xy} z^j.$$

It is well known that $G_{T_{xy}}(z)$ and $G_{K_{xy}(z)}$ are intimately related [4], [7]. Indeed, the two

convolution relations

$$(K^j)_{xx} = 1_{\{j=0\}} + \sum_{i=1}^{j} \Pr(T_{xx} = i)(K^{j-i})_{xx},$$

$$(K^j)_{xy} = \sum_{i=1}^{j} \Pr(T_{xy} = i)(K^{j-i})_{yy}, \qquad x \neq y,$$

immediately translate into the identities

$$G_{K_{xx}}(z) = 1 + G_{T_{xx}}(z)G_{K_{xx}}(z), \qquad G_{K_{xy}}(z) = G_{T_{xy}}(z)G_{K_{yy}}(z), \qquad x \neq y.$$

These in turn entail

$$G_{T_{xx}}(z) = 1 - \frac{1}{G_{K_{xx}}(z)}, \qquad G_{T_{xy}}(z) = \frac{G_{K_{xy}}(z)}{G_{K_{yy}}(z)}, \qquad x \neq y.$$

If K is reversible, we can exploit its biorthogonal decomposition $K = \sum_i \gamma_i w^i v^i$ to calculate $G_{K_{xy}}(z)$ and hence $G_{T_{xy}}(z)$. Viewing $G_{K_{xy}}(z)$ as an entry of a matrix, $G_K(z)$ easily leads to the representation

$$G_K(z) = \sum_{j=0}^{\infty} \sum_i \gamma_i^j z^j w^i v^i = \sum_i \frac{1}{1 - \gamma_i z} w^i v^i.$$

Here $w^1 = \mathbf{1}$, $\gamma_1 = 1$, and $(v^j)^\top = \mathrm{diag}(\pi)w^j$ for all j, where π is the equilibrium distribution. To calculate the mean $\mathrm{E}(T_{xx})$, we impose ergodicity and take the limit of

$$\frac{1 - G_{T_{xx}}(z)}{1 - z} = \frac{1}{(1-z)G_{K_{xx}}(z)} = \frac{1}{\sum_i (1-z)w_x^i v_x^i/(1 - \gamma_i z)}$$

as z tends to 1 from the left. In the limit, $\mathrm{E}(T_{xx}) = \pi_x^{-1}$, one of the most basic identities of Markov chain theory. For the case in which $x \neq y$, the difference quotient defining $\mathrm{E}(T_{xy})$ amounts to

$$\frac{1 - G_{T_{xy}}(z)}{1 - z} = \frac{1 - G_{K_{xy}}(z)/G_{K_{yy}}(z)}{1 - z} = \frac{\sum_i (w_y^i v_y^i - w_x^i v_y^i)/(1 - \gamma_i z)}{\sum_i (1-z)w_y^i v_y^i/(1 - \gamma_i z)}.$$

Because $w_y^1 v_y^1 = w_x^1 v_y^1$, the limit of the difference quotient reduces to

$$\mathrm{E}(T_{xy}) = \frac{\sum_{i>1}(w_y^i v_y^i - w_x^i v_y^i)/(1 - \gamma_i)}{w_y^1 v_y^1}$$

$$= \sum_{i>1} \frac{1}{1 - \gamma_i}(w_y^i w_y^i - w_x^i w_y^i).$$

The next proposition summarizes our conclusions about hitting times for composition chains.

Proposition 5. *For a reversible composition chain, the generating functions of the hitting times* $T_{\tilde{x}\tilde{x}}$ *and* $T_{\tilde{x}\tilde{y}}$ *are*

$$G_{T_{\tilde{x}\tilde{x}}}(z) = 1 - \left[\binom{n}{\tilde{x}} \pi^{\tilde{x}} \sum_{\tilde{n}} \frac{K_{\tilde{n}}^0(\tilde{x}) K_{\tilde{n}}^0(\tilde{x})}{1 - \beta_{\tilde{n}} z} \right]^{-1},$$

$$G_{T_{\tilde{x}\tilde{y}}}(z) = \frac{\sum_{\tilde{n}} K_{\tilde{n}}^0(\tilde{x}) K_{\tilde{n}}^0(\tilde{y}) / (1 - \beta_{\tilde{n}} z)}{\sum_{\tilde{n}} K_{\tilde{n}}^0(\tilde{y}) K_{\tilde{n}}^0(\tilde{y}) / (1 - \beta_{\tilde{n}} z)}, \qquad \tilde{x} \neq \tilde{y},$$

where the eigenvalues $\beta_{\tilde{n}}$ *are defined by (6) and the column eigenvectors* $K_{\tilde{n}}^0(\tilde{x})$ *are the normalized multivariate Krawtchouk polynomials. If the mutation kernel M is ergodic in addition to being reversible, then the hitting times are finite with probability 1 and have means*

$$\mathrm{E}(T_{\tilde{x}\tilde{x}}) = \left(\binom{n}{\tilde{x}} \pi^{\tilde{x}} \right)^{-1},$$

$$\mathrm{E}(T_{\tilde{x}\tilde{y}}) = \sum_{\tilde{n} \neq ne_1} \frac{K_{\tilde{n}}^0(\tilde{y}) [K_{\tilde{n}}^0(\tilde{y}) - K_{\tilde{n}}^0(\tilde{x})]}{1 - \beta_{\tilde{n}}}, \qquad \tilde{x} \neq \tilde{y}.$$

When $\tilde{x} = ne_i$ *and* $\tilde{y} = ne_j$, *the hitting times are examples of coalescent times, and we have the simpler formulae:*

$$G_{T_{ne_i,ne_i}}(z) = 1 - \left(\pi_i^n \sum_{\tilde{n}} \binom{n}{\tilde{n}} \frac{\prod_{k=1}^d (w_i^k)^{2n_k}}{1 - \beta_{\tilde{n}} z} \right)^{-1},$$

$$G_{T_{ne_i,ne_j}}(z) = \sum_{\tilde{n}} \binom{n}{\tilde{n}} \frac{\prod_{k=1}^d (w_i^k w_j^k)^{n_k}}{1 - \beta_{\tilde{n}} z} \bigg/ \sum_{\tilde{n}} \binom{n}{\tilde{n}} \frac{\prod_{k=1}^d (w_j^k)^{2n_k}}{1 - \beta_{\tilde{n}} z}, \qquad i \neq j,$$

and

$$\mathrm{E}(T_{ne_i,ne_i}) = \frac{1}{\pi_i^n},$$

$$\mathrm{E}(T_{ne_i,ne_j}) = \sum_{\tilde{n} \neq ne_1} \binom{n}{\tilde{n}} \frac{\prod_{k=1}^d (w_j^k)^{2n_k} - \prod_{k=1}^d (w_i^k w_j^k)^{n_k}}{1 - \beta_{\tilde{n}}}, \qquad i \neq j.$$

Example 1. (*Multi-person random walk on a circle.*) Tian and Liu [25] considered a symmetric random walk on the vertices of a regular polygon. At each step the person executing the chain moves clockwise or counterclockwise one vertex. The chain is clearly reversible with the uniform distribution as its equilibrium distribution. Because the mutation matrix M for arbitrary d is a circulant matrix, its eigenvalues and eigenvectors can be easily extracted via the finite Fourier transform. The eigenvalue $\lambda_1 = 1$ corresponds to the column eigenvector $w^1 = \mathbf{1}$. When d is even, -1 is an eigenvalue with the eigenvector $w^{d/2+1}$ having entries $w_j^{d/2+1} = (-1)^{j-1}$. The remaining eigenvalues have multiplicity 2. For each k between 2 and $\lfloor (d+1)/2 \rfloor$, we have the shared eigenvalue and paired eigenvectors

$$\lambda_k = \lambda_{d-k+2} = \cos\left[\frac{2\pi(k-1)}{d} \right],$$

$$w_j^k = \sqrt{2} \cos\left[\frac{2\pi(j-1)(k-1)}{d} \right], \qquad w_j^{d-k+2} = \sqrt{2} \sin\left[\frac{2\pi(j-1)(k-1)}{d} \right].$$

For comparison with the results of Tian and Liu [25], we consider the two cases characterized by $n = d = 3$ and $n = d = 4$. If all n people start at different vertices, and we move one person at a time ($s = 1$), then application of Proposition 5 shows that it takes on average 31 steps for three people to meet at a specific vertex of a triangle and 305.371 steps for four people to meet at a specific vertex of a square.

Since we have at our disposal the generating functions of the coalescent times, we can say more [11], [22]. Any probability generating function

$$P(s) = \sum_{k=0}^{\infty} p_k s^k$$

can be extended to the boundary of the unit circle in the complex plane via the equation

$$P(e^{2\pi \sqrt{-1}t}) = \sum_{k=0}^{\infty} p_k e^{2\pi \sqrt{-1}tk}.$$

This creates a periodic function in t whose kth Fourier coefficient p_k can be recovered via the finite Riemann sum

$$p_k \approx \frac{1}{m} \sum_{j=0}^{m-1} P(e^{2\pi \sqrt{-1}j/n}) e^{-2\pi \sqrt{-1}kj/m}.$$

In practice, we evaluate this finite Fourier transform via the fast Fourier transform algorithm for some large power m of 2. For sufficiently large m, all of the coefficients $p_0, \ldots, p_{m-1}$ can be computed accurately. Accuracy can be checked by comparing the numerically computed mean of $P(s)$ with its theoretical mean. Figures 1 and 2 display the discrete densities computed by the Fourier method for the two coalescent times.
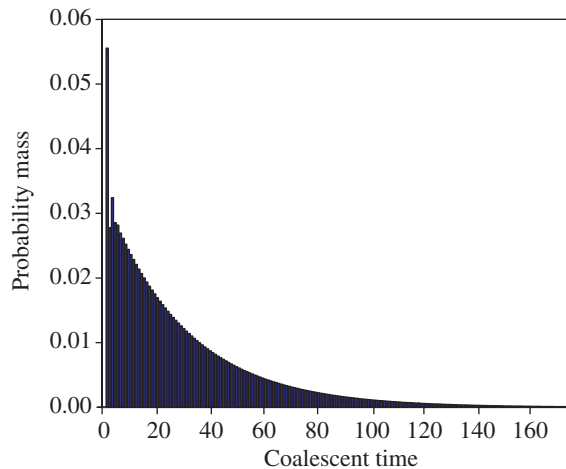

FIGURE 1: Distribution of the coalescent time for a three-person random walk on a triangle. All people start from distinct vertices.
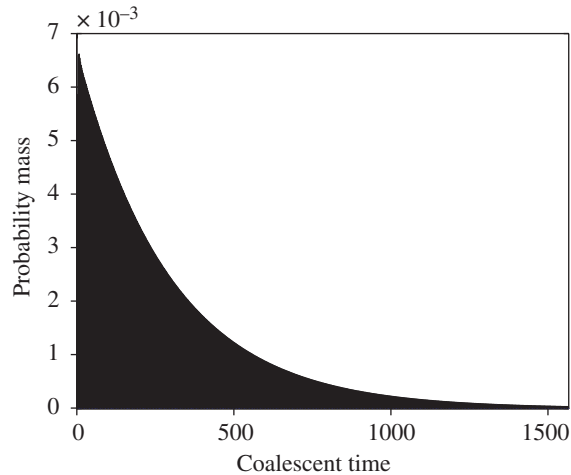
FIGURE 2: Distribution of the coalescent time for a four-person random walk on a square. All people start from distinct vertices.

Example 2. (*Multi-person random walk on the flower graph F_{m3}.*) Consider the flower graph F_{mn} with m petals and n vertices per petal. All edges have weight 1. Each petal communicates to other petals through a shared hub vertex. Erdős *et al.* [6] considered the special case F_{m3}. A single-person random walk on F_{mn} is reversible with equilibrium distribution putting mass $1/n$ on the hub vertex and mass $1/mn$ on all other vertices. Boyd *et al.* [2] described in detail the spectral decomposition of the kernel of this random walk. As a second illustration of mean coalescent times, we consider a three-person random walk on the flower graph F_{33} depicted in Figure 3. Table 1 lists the mean coalescent times $E(T_{ne_i, ne_j})$ for $i, j = 1, 2, 3, 7$. It is interesting that these means decrease as the number s of walkers moved per epoch increases. This phenomenon is not universal.
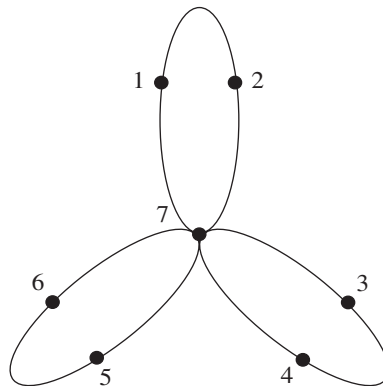

FIGURE 3: Random walk on the flower graph F_{33}.

TABLE 1: Mean coalescent times for a three-person random walk on the flower graph F_{33}. The number of vertices d is 7, the number of walkers moved per epoch is s, and the starting and ending states are ne_i for $i = 1, 2, 3, 7$. See Figure 3 for the labels of the seven vertices.

$s = 1$			
729.0000	811.0286	880.5214	870.8571
811.0286	729.0000	880.5214	870.8571
880.5214	880.5214	729.0000	870.8571
33.0000	33.0000	33.0000	27.0000

$s = 2$			
729.0000	749.2154	772.5758	768.2703
749.2154	729.0000	772.5758	768.2703
772.5758	772.5758	729.0000	768.2703
28.2000	28.2000	28.2000	27.0000

$s = 3$			
729.0000	660.1714	758.7429	754.0000
660.1714	729.0000	758.7429	754.0000
758.7429	758.7429	729.0000	754.0000
26.0000	26.0000	26.0000	27.0000

7. Discussion

Symmetry is one of the central themes of mathematics. In the current paper we have exploited symmetry in the guise of state lumping and reversibility. This perspective clarifies the mysterious nature of Griffiths' multivariate Krawtchouk polynomials. The emergence of these polynomials as eigenfunctions emphasizes their practical importance and opens a window of understanding on the convergence rate of several classical and modern Markov chains. Equally important are the applications of the Krawtchouk polynomials in forming martingales and calculating mean coalescence times.

In our examples the permutation group drives symmetry. We can imagine other groups playing the same role. In conversation with us, Persi Diaconis suggested a novel variant of the light bulb chain that he calls the potato chain. A chef bakes a circle of n potatoes. Each potato has two sides. At each step, the chef randomly forks a span of s potatoes and flips them. In our tensor product notation, the product chain has the kernel

$$M_s = \frac{1}{n} \sum_{k=1}^{n} I \otimes \cdots \otimes I \otimes M \otimes \cdots \otimes M \otimes I \otimes \cdots \otimes I.$$

Here M is the mutation matrix (11), and the kth summand has M inserted into slots $k, \ldots, k + s - 1$ of the tensor product. When $k + s - 1 > n$, we wrap the extra kernels on the left of the tensor product. Although the product chain is reducible, it can be diagonalized by our previous arguments. In this case, the permutation group must be replaced by the group generated by addition modulo n. With this proviso, the group action of entry shifting is consistent with lumping. Now, however, we must lump by orbits rather than by counts, and the multinomial distribution no longer prevails. For instance, if n is even and $s = 2$, the two nonequivalent states $(0, 1, 0, 1, \ldots, 0, 1)$ and $(1, 0, 0, 1, \ldots, 0, 1)$ have the same counts.

In some cases, the eigenstructure of the lower lumped chain is known, but the eigenstructure of the upper chain is unknown. Fact 8 allows us to lift eigenvectors from the lower chain to the upper chain. For reversible chains, the analog of Fact 12 says that lifting preserves norms and orthogonality. These results pose the challenge of filling in the missing eigenvalues and eigenvectors of the upper chain. We know of no systematic way of achieving this, but it is possible in special circumstances. The purine-pyrimidine lumping in Kimura's chain is a case in point.

Another class of open problems is to extend our results to interacting particles. Finding the complete eigenstructure of the Moran process in population genetics is a good place to start. Here the particles are identical, and the count process is well defined, but the replacement of one particle is tied to the state of another particle. Although resolution of this and similar open problems is intrinsically more difficult than the simpler problems faced here, we believe that our analysis offers helpful clues.

Acknowledgements

References

[1] BARR, D. R. AND THOMAS, M. U. (1977). An eigenvector condition for Markov chain lumpability. *Operat. Res.* **25,** 1028–1031.
[2] BOYD, S., DIACONIS, P., PARRILO, P. AND XIAO, L. (2005). Symmetry analysis of reversible Markov chains. *Internet Math.* **2,** 31–71.
[3] COBB G. W. AND CHEN, Y-P. (2003). An application of Markov chain Monte Carlo to community ecology. *Amer. Math. Monthly* **110,** 265–288.
[4] DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA.
[5] DIACONIS, P., KHARE, K. AND SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.* **23,** 151–200.
[6] ERDŐS, P., RÉNYI, A. AND SÓS, V. T. (1966). On a problem of graph theory. *Studia Sci. Math. Hung.* **1,** 215–235.
[7] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd edn. John Wiley, New York.
[8] GRAHAM, R. L., KNUTH, D. E. AND PATASHNIK, O. (1989). *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, Reading, MA.
[9] GRIFFITHS, R. C. (1971). Orthogonal polynomials on the multinomial distribution. *Austral. J. Statist.* **13,** 27–35.
[10] GRIFFITHS, R. C. (2007). Notes on 'A probabilistic origin for a new class of bivariate polynomials' by M. R. Hoare and M. Rahman. Unpublished notes.
[11] HENRICI, P. (1979). Fast Fourier methods in computational complex analysis. *SIAM Rev.* **21,** 481–527.
[12] HOARE, M. R. AND RAHMAN, M. (2008). A probabilistic origin for a new class of bivariate polynomials. *SIGMA* **4,** 089.
[13] ISMAIL, M. E. H. (2005). *Classical and Quantum Orthogonal Polynomials in One Variable* (Encyclopaedia Math. Appl. **98**). Cambridge University Press.
[14] ILIEV, P. AND XU, Y. (2007). Discrete orthogonal polynomials and difference equations of several variables. *Adv. Math.* **212,** 1–36.
[15] KAC, M. (1947). Random walk and the theory of Brownian motion. *Amer. Math. Monthly* **54,** 369–391.
[16] KARLIN, S. AND McGREGOR, J. (1965). Ehrenfest urn models. *J. Appl. Prob.* **2,** 352–376.
[17] KARLIN, S. AND TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*. Academic Press, New York.
[18] KARLIN, S. AND TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic Press, New York.
[19] KEMENY, J. G. AND SNELL, J. L. (1983). *Finite Markov Chains*. Springer, New York.
[20] KHARE, K. AND ZHOU, H. (2009). Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. To appear in *Ann. Appl. Prob.*
[21] KIMURA, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Molec. Evol.* **16,** 111–120.

[22] LANGE, K. (1999). *Numerical Analysis for Statisticians*. Springer, New York.
[23] LAZZERONI, L. C. AND LANGE, K. (1997). Markov chains for Monte Carlo tests of genetic equilibrium in multidimensional contingency tables. *Ann. Statist.* **25,** 138–168.
[24] RAO, C. R., RAO, M. B. AND ZHANG, H. (2007). One bulb? Two bulbs? How many bulbs light up—a discrete probability problem involving dermal patches. *Sankhyā* **69,** 137–161.
[25] TIAN, J. P. AND LIU, Z. (2007). Coalescent random walks on graphs. *J. Comput. Appl. Math.* **202,** 144–154.