Electronic Journal of Probability

Electron. J. Probab. **18** (2013), no. 61, 1–23.
ISSN: 1083-6489 DOI: 10.1214/EJP.v18-2389

Convergence rates of Markov chains on spaces of partitions*

Harry Crane† Steven P. Lalley‡

Abstract

We study the convergence rate to stationarity for a class of exchangeable partition-valued Markov chains called cut-and-paste chains. The law governing the transitions of a cut-and-paste chain is determined by products of i.i.d. stochastic matrices, which describe the chain induced on the simplex by taking asymptotic frequencies. Using this representation, we establish upper bounds for the mixing times of ergodic cut-and-paste chains; and, under certain conditions on the distribution of the governing random matrices, we show that the "cutoff phenomenon" holds.

Keywords: cut-and-paste chain; mixing time; exchangeability; cutoff phenomenon; Lyapunov exponent.
AMS MSC 2010: Primary 60J05; secondary 60B10 60B20.
Submitted to EJP on October 23, 2012, final version accepted on June 10, 2013.

1 Introduction

A Markov chain $\{X_t\}_{t=0,1,2,\dots}$ on the space $[k]^{\mathbb{N}}$ of $k-$colorings of the positive integers $\mathbb{N}$ is said to be *exchangeable* if its transition law is equivariant with respect to finite permutations of $\mathbb{N}$ (that is, permutations that fix all but finitely many elements of $\mathbb{N}$). Exchangeability does not imply that the Markov chain has the Feller property (relative to the product topology on $[k]^{\mathbb{N}}$), but if a Markov chain is both exchangeable and Feller then it has a simple *paintbox* representation, as proved by Crane [4]. In particular, there exists a sequence $\{S_t\}_{t\geq 1}$ of independent and identically distributed (i.i.d.) $k\times k$ random column-stochastic matrices (the paintbox sequence) such that, conditional on the entire sequence $\{S_t\}_{t\geq 1}$ and on $X_0, X_1, \dots, X_m$, the coordinate random variables $\{X^i_{m+1}\}_{i\in[n]}$ are independent, and X^i_{m+1} has the multinomial distribution specified by the X^i_m column of S_{m+1}. Equivalently (see Proposition 3.3 in Section 3.3), conditional on the paintbox sequence, the coordinate sequences $\{X^i_{m+1}\}_{m\geq 0}$ are independent, time-inhomogeneous Markov chains on the state space $[k]$ with one-step transition probability matrices $S_1, S_2, \dots$. This implies that, for any integer $n\geq 1$, the restriction $X^{[n]}_t$ of X_t to the space $[k]^{[n]}$ is itself a Markov chain. We shall refer to such Markov chains X_t and

*Second author supported by NSF Grant DMS1106669.
†Rutgers University, USA. E-mail: hcrane@stat.rutgers.edu
‡University of Chicago, USA E-mail: lalley@galton.uchicago.edu

$X_t^{[n]}$ as *exchangeable Feller cut-and-paste* chains, or EFCP chains for short. Under mild hypotheses on the paintbox distribution (see the discussion in Section 5) the restrictions of EFCP chains $X_t^{[n]}$ to the finite configuration spaces $[k]^{[n]}$ are ergodic. The main results of this paper, Theorems 1.1–1.2, relate the convergence rates of these chains to properties of the paintbox sequence $S_1, S_2, \dots$.

Theorem 1.1. *Assume that for some $m \geq 1$ there is positive probability that all entries of the matrix product $S_m S_{m-1} \cdots S_1$ are nonzero. Then the EFCP chain $X^{[n]}$ is ergodic, and it mixes in $O(\log n)$ steps.*

Theorem 1.2. *Assume that the distribution of S_1 is absolutely continuous relative to Lebesgue measure on the space of $k \times k$ column-stochastic matrices, with density of class L^p for some $p > 1$. Then the associated EFCP chains $X^{[n]}$ exhibit the* cutoff phenomenon*: there exists a positive constant θ such that for all sufficiently small $\delta, \varepsilon > 0$ the (total variation) mixing times satisfy*

$$(\theta - \delta) \log n \leq t_{\text{mix}}^{(n)}(\varepsilon) \leq t_{\text{mix}}^{(n)}(1 - \varepsilon) \leq (\theta + \delta) \log n \tag{1.1}$$

for all sufficiently large n.

Formal statements of these theorems will be given in due course (see Theorems 5.4 and 5.7 in Section 5), and less stringent hypotheses for the $O(\log n)$ convergence rate will be given. In the special case $k = 2$ the results are related to some classical results for random walks on the hypercube, e.g. the Ehrenfest chain on $\{0, 1\}^{[n]}$; see Examples 5.9 and 5.10.

The key to both results is that the relative frequencies of the different colors are determined by the random matrix products $S_t S_{t-1} \cdots S_1$ (see Proposition 3.3). The hypotheses of Theorem 1.1 ensure that these matrix products contract the $k-$simplex to a point at least exponentially rapidly. The stronger hypotheses of Theorem 1.2 prevent the simplex from collapsing at a faster than exponential rate.

In addition to their own mathematical interest, our main theorems have potential implications in population genetics, as Markov chains on spaces of partitions arise naturally in various contexts. Ewens [5] was first to note the interplay between neutral alleles models and random partitions. Kingman [11] later introduced the coalescent process, a special Markov process on partitions that arises as the scaling limit of both the Wright-Fisher and Moran models in population genetics; see [15]. Since the seminal work of Ewens and Kingman, there has been considerable study of partition-valued Markov chains in the probability literature, mostly involving processes of fragmentation and coagulation. The monographs [1, 14] give an overview of this work from different mathematical perspectives. In this paper, we study Markov chains on $[k]^{\mathbb{N}}$, which (under obvious restrictions on the transition probabilities) project to Markov chainson the space of partitions with a bounded number of blocks (see Section 3.1). When $k = 4$, Markov chains on $[k]^{\mathbb{N}}$ are models of natural interest in problems related to DNA sequencing, where the four colors correspond to the nucleotides (A,C,G,T) that appear in DNA sequences. Our methods draw on the recent work [4], in which the class of exchangeable Feller chains on $[k]^{\mathbb{N}}$ has been characterized in terms of products of i.i.d. stochastic matrices.

The paper is organized as follows. In Section 2, we record some simple and elementary facts about total variation distance, and in Section 3, we define cut-and-paste Markov chains formally and establish the basic relation with the paintbox sequence (Proposition 3.3). In Section 4, we discuss the contractivity properties of products of random stochastic matrices. In Section 5, we prove the main results concerning ergodicity and mixing rates of cut-and-paste chains, and in Section 5.3, we discuss some examples of cut-and-paste chains not covered by our main theorems. Finally, in Section 6, we deduce mixing rate and cutoff for projections of the cut-and-paste chain into the space of ordinary set partitions.

2 Preliminaries: Total Variation Distance

Since the state spaces of interest in our main results are finite, it is natural to use the total variation metric to measure the distance between the law $\mathcal{D}(X_m)$ of the chain X at time $m \geq 1$ and its stationary distribution π. The *total variation distance* $\|\mu - \nu\|_{TV}$ between two probability measures μ, ν on a finite or countable set $\mathcal{X}$ is defined by

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| = \max_{B \subset \mathcal{X}} (\nu(B) - \mu(B)). \tag{2.1}$$

The maximum is attained at $B^* = \{x : \nu(x) \geq \mu(x)\}$ and, since the indicator $\mathbf{1}_{B^*}$ is a function only of the likelihood ratio $d\nu/d\mu$, the total variation distance $\|\mu - \nu\|_{TV}$ is the same as the total variation distance between the $\mu-$ and $\nu-$ distributions of any sufficient statistic. In particular, if $Y = Y(x)$ is a random variable such that $d\nu/d\mu$ is a function of Y, then

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{y} |\nu(Y = y) - \mu(Y = y)|, \tag{2.2}$$

where the sum is over all possible values of $Y(x)$.

Likelihood ratios provide a useful means for showing that two probability measures are close in total variation distance.

Lemma 2.1. *Fix $\varepsilon > 0$ and define*

$$B_\varepsilon = \left\{ x : \left| \frac{\mu(x)}{\nu(x)} - 1 \right| > \varepsilon \right\}.$$

If $\nu(B_\varepsilon) < \varepsilon$, then $\|\mu - \nu\|_{TV} < 2\varepsilon$.

Proof. By definition of B_ε, $B_\varepsilon^{\mathrm{c}} := \{x : |\mu(x) - \nu(x)| \leq \varepsilon \nu(x)\}$ and so $(1 - \varepsilon)\nu(x) \leq \mu(x) \leq (1 + \varepsilon)\nu(x)$ for every $x \in B_\varepsilon^{\mathrm{c}}$ and

$$\mu(B_\varepsilon^{\mathrm{c}}) \geq (1 - \varepsilon)\nu(B_\varepsilon^{\mathrm{c}}).$$

By assumption $\nu(B_\varepsilon) < \varepsilon$, it follows that $\mu(B_\varepsilon^{\mathrm{c}}) \geq (1 - \varepsilon)^2$ and

$$\begin{aligned}
\|\mu - \nu\|_{TV} &= \frac{1}{2} \left[\sum_{x \in B_\varepsilon} |\mu(x) - \nu(x)| + \sum_{x \in B_\varepsilon^{\mathrm{c}}} |\mu(x) - \nu(x)| \right] \\
&\leq \frac{1}{2} \left[\sum_{x \in B_\varepsilon} \mu(x) + \sum_{x \in B_\varepsilon} \nu(x) + \sum_{x \in B_\varepsilon^{\mathrm{c}}} |\mu(x) - \nu(x)| \right] \\
&< 2\varepsilon.
\end{aligned}$$

□

The convergence rates of EFCP chains will be (in the ergodic cases) determined by the contractivity properties of products of random stochastic $k \times k$ matrices on the $(k-1)$-dimensional simplex

$$\Delta_k := \left\{ (s_1, \ldots, s_k)^T : s_i \geq 0 \quad \text{and} \quad \sum_i s_i = 1 \right\}. \tag{2.3}$$

We now record some preliminary lemmas about convergence of probability measures on Δ_k that we will need later. For each $n \in \mathbb{N}$ and each element $s \in \Delta_k$, we define a probability measure ϱ_s^n, the *product-multinomial-s measure*, on $[k]^{[n]}$ by

$$\varrho_s^n(x) := \prod_{j=1}^n s_{x^j} \quad \text{for } \ x = x^1 x^2 \cdots x^n \in [k]^{[n]} \,. \tag{2.4}$$

Observe that the vector $m(x) := (m_1, \ldots, m_k)$ of cell counts defined by $m_j := \sum_{i=1}^n 1_j(x^i)$ is a sufficient statistic for the likelihood ratio $\varrho_s^n(x)/\varrho_{s'}^n(x)$ of any two product-multinomial measures ϱ_s^n and $\varrho_{s'}^n$.

Corollary 2.2. *Fix $\delta, \varepsilon > 0$. If s_n, s_n' are two sequences in Δ_k such that all coordinates of s_n, s_n' are in the interval $[\delta, 1-\delta]$ for every n, and if $\|s_n - s_n'\|_\infty < n^{-1/2-\varepsilon}$, then*

$$\lim_{n\to\infty} \|\varrho_{s_n}^n - \varrho_{s_n'}^n\|_{TV} = 0.$$

Proof. This is a routine consequence of Lemma 2.1, as the hypotheses ensure that the likelihood ratio $d\varrho_{s_n}^n / d\varrho_{s_n'}^n$ is uniformly close to 1 with probability approaching 1 as $n \to \infty$. ☐

A similar argument can be used to establish the following generalization, which is needed in the case of partitions with $k \geq 3$ classes. For $s_1, \ldots, s_k \in \Delta_k$, let $\varrho_{s_1}^{n_1} \otimes \cdots \otimes \varrho_{s_k}^{n_k}$ denote the product measure on $[k]^{n_1 + \cdots + n_k}$, where the first n_1 coordinates are i.i.d. multinomial-s_1, the next n_2 are i.i.d. multinomial-s_2, and so on.

Corollary 2.3. *Fix $\delta, \varepsilon > 0$. For each $i \in [k]$, let $\{s_n^i\}_{n\geq 1}$ and $\{t_n^i\}_{n \geq 1}$ be sequences in Δ_k all of whose entries are in the interval $[\delta, 1-\delta]$, and let $\{K_n^i\}_{n\geq 1}$ be sequences of nonnegative integers such that $\sum_i K_n^i = n$, for every $n \geq 1$. If $\sum_{i=1}^k \|t_n^i - s_n^i\|_\infty < n^{-1/2-\varepsilon}$, then*

$$\lim_{n\to\infty} \left\| \varrho_{s_n^1}^{K_n^1} \otimes \cdots \otimes \varrho_{s_n^k}^{K_n^k} - \varrho_{t_n^1}^{K_n^1} \otimes \cdots \otimes \varrho_{t_n^k}^{K_n^k} \right\|_{TV} = 0.$$

In dealing with probability measures that are defined as mixtures, the following simple tool for bounding total variation distance is useful.

Lemma 2.4. *Let μ, ν be probability measures on a finite or countable space $\mathcal{X}$ that are both mixtures with respect to a common mixing probability measure $\lambda(d\theta)$, that is, such that there are probability measures μ_θ and ν_θ for which*

$$\mu = \int \mu_\theta \, d\lambda(\theta) \quad \text{and} \quad \nu = \int \nu_\theta \, d\lambda(\theta).$$

If $\|\mu_\theta - \nu_\theta\|_{TV} < \varepsilon$ for all θ in a set of $\lambda-$probability $> 1 - \varepsilon$ then

$$\|\mu - \nu\|_{TV} < 2\varepsilon.$$

Lower bounds on total variation distance between two probability measures μ, ν are often easier to establish than upper bounds, because for this it suffices to find a particular set B such that $\mu(B) - \nu(B)$ is large. By (2.2), it suffices to look at sets of the form $B = \{Y \in F\}$, where Y is a sufficient statistic. The following lemma for product Bernoulli measures illustrates this strategy. For $\alpha \in [0,1]$, we write $\nu_\alpha^n := \varrho_s^n$, where $s := (\alpha, 1-\alpha) \in \Delta_2$, to denote the product Bernoulli measure determined by α.

Lemma 2.5. *Fix $\varepsilon > 0$. If α_m, β_m are sequences in $[0,1]$ such that $|\alpha_m - \beta_m| > m^{-1/2+\varepsilon}$, then*

$$\lim_{m\to\infty} \|\nu_{\alpha_m}^m - \nu_{\beta_m}^m\|_{TV} = 1.$$

Proof. Without loss of generality, assume that $\alpha_m < \beta_m$, and let $\gamma_m = (\alpha_m + \beta_m)/2$. Denote by S_m the sum of the coordinate variables. Then, by Chebyshev's inequality,

$$\lim_{m\to\infty} \nu_{\alpha_m}^m \{S_m < m\gamma_m\} = 1 \quad \text{and}$$
$$\lim_{m\to\infty} \nu_{\beta_m}^m \{S_m < m\gamma_m\} = 0.$$

□

Remark 2.6. *Similar results hold for multinomial and product-multinomial sampling. (A) If $s_n, s'_n \in \Delta_k$ are distinct sequences of probability distributions on $[k]$ such that for some coordinate $i \in [k]$ the ith entries of s_n and s'_n differ by at least $n^{-1/2+\varepsilon}$, then*

$$\lim_{n\to\infty} \|\varrho_{s_n}^n - \varrho_{s'_n}^n\|_{TV} = 1.$$

(B) If $s_n^i, t_n^i \in \Delta_k$ are distinct sequences of probability distributions on $[k]$ such that for some pair $i, j \in [k]$ the jth entries of s_n^i and t_n^i differ by at least $n^{-1/2+\varepsilon}$, then for any sequences K_n^i of nonnegative integers such that $\sum_i K_n^i = n$,

$$\lim_{n\to\infty} \left\| \varrho_{s_n^1}^{K_n^1} \otimes \cdots \otimes \varrho_{s_n^k}^{K_n^k} - \varrho_{t_n^1}^{K_n^1} \otimes \cdots \otimes \varrho_{t_n^k}^{K_n^k} \right\|_{TV} = 1.$$

These statements follow directly from Lemma 2.5 by projection on the appropriate coordinate variable.

3 Preliminaries: CP chains and Paintbox Representation

3.1 k-colorings and set partitions

For $k, n \in \mathbb{N}$, a *k-coloring* of $[n] := \{1, \dots, n\}$ is a sequence $x = x^1 \cdots x^n \in [k]^{[n]}$. A *partition* π of $[n]$ is a collection $\{\pi_1, \dots, \pi_r\}$ of non-empty, disjoint subsets (blocks) for which $\bigcup_{i=1}^r \pi_i = [n]$. We denote by $\mathcal{P}_{[n]}$ the space of all partitions of $[n]$ and, in particular, we write $\mathcal{P}_{[n]:k}$ to denote the subspace of partitions of $[n]$ having at most k blocks.

There is an obvious and natural projection $\Pi_n : [k]^{[n]} \to \mathcal{P}_{[n]:k}$ that coincides with $[k]^{[n]} \to [k]^{[n]} / \sim$, where $\sim$ is the equivalence relation

$$x^1 x^2 \cdots x^n \sim x_*^1 x_*^2 \cdots x_*^n$$

if and only if there exists a permutation $\sigma : [k] \to [k]$ such that $x_*^i = \sigma(x^i)$ for every $i = 1, \dots, n$. In particular, for $x \in [k]^{[n]}$, we define $\pi = \Pi_n(x)$ by

$$i \text{ and } j \text{ are in the same block of } \pi \quad \Longleftrightarrow \quad x^i = x^j.$$

In this paper, we primarily study Markov chains on $[k]^{[n]}$; however, some of the Markov chains considered have transition laws that are invariant under permutations of colors $[k]$. In these cases, the image of the Markov chain on $[k]^{[n]}$ by Π_n is a Markov chain on $\mathcal{P}_{[n]:k}$. By elementary properties of total variation distance, the projected chain exhibits the same behavior as the chain on $[k]^{[n]}$ in these cases. We discuss this further in Section 6.

3.2 Matrix operations on $[k]^{\mathbb{N}}$

A key ingredient to our proofs of Theorems 1.1 and 1.2 is the so-called cut-and-paste representation of EFCP chains, proven in [4], which we now introduce. To describe the cut-and-paste Markov chain on $[k]^{[n]}$, it is convenient to regard $x \in [k]^{[n]}$ as an *ordered partition* $(L_1, \dots, L_k)$, where L_i, $i = 1, \dots, k$, is a subset of $[n]$ and $\bigcup_{i=1}^k L_i = [n]$. (Note that, in contrast to the definition of partition above, the L_i need not be non-empty.)

The space $[k]^{[n]}$ is in one-to-one correspondence with the space of ordered partitions through the relation

$$i \in L_j \quad \Longleftrightarrow \quad x^i = j, \tag{3.1}$$

for $i = 1, \dots, n$ and $j = 1, \dots, k$. In particular, L_j consists of those indices colored j in x.

To avoid unnecessary notation, we may regard $x \in [k]^{[n]}$ as a sequence $x^1 \cdots x^n$ of colors or an ordered partition $(L_1, \dots, L_k)$, depending on which is more convenient. The representation as an ordered partition is convenient for characterizing the cut-and-paste Markov chain on $[k]^{[n]}$ by a product of i.i.d. random set-valued partition matrices, which we now define.

Definition 3.1 (Partition matrices)**.** *For any subset $S \subset \mathbb{N}$, define $k \times k$* partition matrix *over S to be a $k \times k$ matrix M whose entries M_{ij} are subsets of S such that every column $M^j := (M_{1j}, \dots, M_{kj})$ corresponds to the ordered partition of some k-coloring of S. For any two $k \times k$ partition matrices M, M', we define the product $M * M = MM'$ by*

$$(MM')_{ij} := \bigcup_{1 \le l \le k} (M_{il} \cap M'_{lj}), \quad \textit{for all } 1 \le i, j \le k. \tag{3.2}$$

We write $\mathcal{M}_{[n]:k}$ to denote the space of $k \times k$ partition matrices of $[n]$. Observe that the matrix product defined by (3.2) makes sense for matrices with entries in any distributive lattice, provided $\cup, \cap$ are replaced by the lattice operations.

As each column of any $M \in \mathcal{M}_{[n]:k}$ is an ordered partition of $[n]$, the set $\mathcal{M}_{[n]:k}$ of $k \times k$ partition matrices over $[n]$ can be identified with $[k]^{[n]} \times \cdots \times [k]^{[n]}$ (k times). Furthermore, a $k \times k$ partition matrix M induces a mapping $M : [k]^{[n]} \to [k]^{[n]}$ by

$$(ML)_i = \bigcup_j (M_{ij} \cap L_j). \tag{3.3}$$

Equivalently, we can define $M : [k]^{[n]} \to [k]^{[n]}$ by $x \mapsto x' := M(x)$, where, for each $i = 1, \dots, n$, x'^i is the color assigned to coordinate i in the x^i-th column of M. As these are equivalent, we use the same notation to describe the map on $[k]^{[n]}$.

In the following lemma, we write L, L' to, respectively, denote the ordered partitions corresponding to $x, x' \in [k]^{[n]}$.

Lemma 3.2. *Let $k, n \in \mathbb{N}$. Then*

(i) for each $x \in [k]^{[n]}$, $ML \in [k]^{[n]}$ for all $M \in \mathcal{M}_{[n]:k}$;

(ii) for any $x, x' \in [k]^{[n]}$, there exists $M \in \mathcal{M}_{[n]:k}$ such that $ML = L'$;

(iii) the pair $(\mathcal{M}_{[n]:k}, *)$ is a monoid (i.e., semigroup with identity) for every $n \in \mathbb{N}$.

The proof is elementary and follows mostly from the definition (3.2) (the semigroup identity is the partition matrix whose diagonal entries are all $[n]$ and whose off-diagonal entries are $\emptyset$). We now describe the role of the semigroup $(\mathcal{M}_{[n]:k}, *)$ in describing the transitions of the cut-and-paste Markov chain.

3.3 Cut-and-paste Markov chains

Fix $n, k \in \mathbb{N}$, let μ be a probability measure on $\mathcal{M}_{[n]:k}$, and let ϱ_0 be a probability measure on $[k]^{[n]}$. The cut-and-paste Markov chain $X = (X_m)_{m \ge 0}$ on $[k]^{[n]}$ with initial distribution ϱ_0 and directing measure μ is constructed as follows. Let $X_0 \sim \varrho_0$ and, independently of X_0, let $M_1, M_2, \dots$ be i.i.d. according to μ. Define

$$X_m = M_m(X_{m-1}) = (M_m \circ M_{m-1} \circ \cdots \circ M_1)(X_0), \quad \text{for } m \ge 1, \tag{3.4}$$

where the operations $M_i(\cdot)$ on $[k]^{[n]}$ are defined in (3.3). We call any Markov chain with the above dynamics a $\mathrm{CP}_n(\mu;\varrho_0)$ chain, or simply a $\mathrm{CP}_n(\mu)$ chain if the initial distribution is unspecified. Henceforth we will use the notation X^i_m to denote the ith coordinate variable in X_m (that is, X^i_m is the color of the site $i \in [n]$ at time $m \geq 0$).

Our main results concern the class of cut-and-paste chains whose directing measures are mixtures of product-multinomial measures μ_S, where S ranges over the set Δ^k_k of $k \times k$ column-stochastic matrices. For any $S \in \Delta^k_k$, the product-multinomial measure μ_S is defined by

$$\mu_S(M) := \prod_{j=1}^{k}\prod_{i=1}^{n} S(M^j(i), j) \quad \text{for } \ M \in \mathcal{M}_{[n]:k}, \tag{3.5}$$

where $M^j(i) = \sum_r r\mathbf{1}\{i \in M_{rj}\}$ denotes the index r of the row such that i is an element of M_{rj}. (In other words, the columns of $M \sim \mu_S$ are independent labeled k-colorings, and, in each column M^j, the elements $i \in [n]$ are independently assigned to rows $r \in [k]$ according to draws from the multinomial-S^j distribution determined by the jth column of S.) For any Borel probability measure Σ on Δ^k_k, we write μ_Σ to denote the Σ-mixture of the measures μ_S on $\mathcal{M}_{[n]:k}$, that is,

$$\mu_\Sigma(\cdot) := \int_{\Delta^k_k} \mu_S(\cdot)\Sigma(dS). \tag{3.6}$$

Crane [4] has shown that every exchangeable, Feller Markov chain on the the space $[k]^{\mathbb{N}}$ of k-colorings of the positive integers is a cut-and-paste chain with directing measure of the form (3.6), and so henceforth, we shall refer to such chains as *exchangeable Feller cut-and-paste chains*, or EFCP chains for short.

An EFCP chain on $[k]^{[n]}$ (or $[k]^{\mathbb{N}}$) with directing measure $\mu = \mu_\Sigma$ can be constructed in two steps as follows. First, choose i.i.d. stochastic matrices $S_1, S_2, \ldots$ with law Σ, all independent of X_0; second, given $X_0, S_1, S_2, \ldots$, let $M_1, M_2, \ldots$ be conditionally independent $k \times k$ partition matrices with laws $M_i \sim \mu_{S_i}$ for each $i = 1, 2, \ldots$, and define the cut-and-paste chain X_m by equation (3.4). This construction is fundamental to our arguments, and so henceforth, when considering an EFCP chain with directing measure μ_Σ, we shall assume that it is defined on a probability space together with a paintbox sequence $S_1, S_2, \ldots$.

For each $m \in \mathbb{N}$, set

$$Q_m := S_m S_{m-1} \cdots S_1. \tag{3.7}$$

Note that Q_m is itself a stochastic matrix. Denote by $\mathcal{S}$ the σ-algebra generated by the paintbox sequence $S_1, S_2, \ldots$.

Proposition 3.3. *Given $\mathcal{G} := \sigma(X_0) \vee \mathcal{S}$, the n coordinate sequences $(X^i_m)_{m\geq 0}$, where $i \in [n]$, are conditionally independent versions of a time-inhomogeneous Markov chain on $[k]$ with one-step transition probability matrices $S_1, S_2, \ldots$. Thus, in particular, for each $m \geq 1$,*

$$\mathbf{P}(X^i_m = x^i_m \text{ for each } i \in [n] \,|\, \mathcal{G}) = \prod_{i=1}^{n} Q_m(x^i_m, X^i_0). \tag{3.8}$$

Proof. We prove that the Markov property holds by induction on m. The case $m = 1$ follows directly by (3.5), as this implies that, conditional on $\mathcal{G}$, the coordinate random variables X^i_1 are independent, with multinomial marginal conditional distributions given by the columns of S_1. Assume, then, that the assertion is true for some $m \geq 1$. Let $\mathcal{F}_m$ be the σ-algebra generated by $\mathcal{G}$ and the random partition matrices $M_1, M_2, \ldots, M_m$. Since the specification (3.4) expresses X_m as a function of $X_0, M_1, M_2, \ldots, M_m$, the random variables X_t^i, where $t \leq m$, are measurable with respect to $\mathcal{F}_m$. Moreover, given $\mathcal{G}$, the random matrix M_{m+1} is conditionally independent of $\mathcal{F}_m$, with conditional distribution (3.5) for $S = S_{m+1}$. Equation (3.5) implies that, conditional on $\mathcal{G}$, the columns M_{m+1}^c of M_{m+1} are independent k-colorings obtained by independent multinomial$-S^c$ sampling. Consequently,

$$\begin{aligned} P(X_{m+1}^i = x_{m+1}^i \ \ \forall\, i \in [n] \mid \mathcal{F}_m) &= P((M_{m+1}X_m)^i = x_{m+1}^i \ \ \forall\, i \in [n] \mid \mathcal{F}_m) \\ &= P((M_{m+1}X_m)^i = x_{m+1}^i \ \ \forall\, i \in [n] \mid \mathcal{G} \vee \sigma(X_m)) \\ &= \prod_{i=1}^n S_{m+1}(x_{m+1}^i, X_m^i), \end{aligned}$$

the second equality by the induction hypothesis and the third by definition of the probability measure $\mu_{S_{m+1}}$. This proves the first assertion of the proposition. Equation (3.8) follows directly. □

Proposition 3.3 shows that, for any $n \geq 1$, a version of the EFCP chain on $[k]^{[n]}$ can be constructed by first generating a paintbox sequence S_m and then, conditional on $\mathcal{S}$, running independent, time-inhomogeneous Markov chains X_m^i with one-step transition probability matrices S_m. From this construction it is evident that a version of the EFCP chain on the infinite state space $[k]^{\mathbb{N}}$ can be constructed by running countably many conditionally independent Markov chains X_m^i, and that, for any $n \in \mathbb{N}$, the projection of this chain to the first n coordinates is a version of the EFCP chain on $[k]^{[n]}$.

4 Random stochastic matrix products

For any EFCP chain $\{X_m\}_{m\geq 0}$, Proposition 3.3 directly relates the conditional distribution of X_m to the product $Q_m = S_m S_{m-1} \cdots S_1$ of i.i.d. random stochastic matrices. Thus, the rates of convergence of these chains are at least implicitly determined by the contractivity properties of the random matrix products Q_m. The asymptotic behavior of i.i.d. random matrix products has been thoroughly investigated, beginning with the seminal paper of Furstenberg and Kesten [6]: see [2] and [7] for extensive reviews. However, the random matrices S_i that occur in the paintbox representation of the $\mathrm{CP}_n(\mu_\Sigma)$ chain are not necessarily invertible, so much of the theory developed in [2] and [7] doesn't apply. On the other hand, the random matrices S_t are *column-stochastic*, and so the deeper results of [2] and [7] are not needed here. In this section, we collect the results concerning the contraction rates of the products Q_m needed for the study of the EFCP chains, and give elementary proofs of these results.

Throughout this section assume that $\{S_i\}_{i\geq 1}$ is a sequence of independent, identically distributed $k \times k$ random column-stochastic matrices with common distribution Σ, and let

$$Q_m := S_m S_{m-1} \cdots S_1, \quad m \geq 1.$$

4.1 Asymptotic Collapse of the Simplex

In the theory of random matrix products, a central role is played by the induced action on projective space. In the theory of products of random *stochastic* matrices an analogous role is played by the action of the matrices on the simplex Δ_k. By definition, the simplex Δ_k consists of all convex combinations of the unit vectors $e_1, e_2, \ldots, e_k$ of $\mathbb{R}^k$; since each column of a $k \times k$ column-stochastic matrix $S \in \Delta_k^k$ lies in Δ_k, the mapping $v \mapsto Sv$ preserves Δ_k. This mapping is *contractive* in the sense that it is Lipschitz (relative to the usual Euclidean metric on $\mathbb{R}^k$) with Lipschitz constant ≤ 1.

The simplex Δ_k is contained in a translate of the $(k-1)$-dimensional vector subspace $V = V_k$ of $\mathbb{R}^k$ consisting of all vectors orthogonal to the vector $\mathbf{1} = (1, 1, \ldots, 1)^T$

(equivalently, the subspace with basis $e_i - e_{i+1}$ where $1 \leq i \leq k-1$). Any stochastic matrix A leaves the subspace V invariant, and hence induces a linear transformation $A|V : V \to V$. Since this transformation is contractive, its *singular values* are all between 0 and 1. (Recall that the singular values of a $d \times d$ matrix S are the square roots of the eigenvalues of the nonnegative definite matrix $S^T S$. Equivalently, they are the lengths of the principal axes of the ellipsoid $S(\mathbb{S}^{d-1})$, where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$.) Denote the singular values of the restriction $Q_n|V$ by

$$1 \geq \lambda_{n,1} \geq \lambda_{n,2} \geq \cdots \geq \lambda_{n,k-1} \geq 0. \tag{4.1}$$

Because the induced mapping $Q_n : \Delta_k \to \Delta_k$ is affine, its Lipschitz constant is just the largest singular value $\lambda_{n,1}$.

Proposition 4.1. *Let $(S_i)_{i\geq 1}$ be independent, identically distributed $k \times k$ column-stochastic random matrices, and let $Q_m = S_m S_{m-1} \cdots S_1$. Then*

$$\lim_{m\to\infty} \text{diameter}(Q_m(\Delta_k)) = 0 \tag{4.2}$$

if and only if there exists $m \geq 1$ such that with positive probability the largest singular value $\lambda_{m,1}$ of $Q_m|V$ is strictly less than 1. In this case,

$$\limsup_{m\to\infty} \text{diameter}(Q_m(\Delta_k))^{1/m} < 1 \quad \textit{almost surely.} \tag{4.3}$$

(Here diameter *refers to the standard Euclidean metric on the simplex.)*

Proof. In order that the asymptotic collapse property (4.2) holds it is necessary that for some m the largest singular value of $Q_m|V$ be less than one. (If not then for each m there would exist points $u_m, v_m \in \Delta_k$ such that the length of $Q_m(u_m - v_m)$ is at least the length of $u_m - v_m$; but this would contradict (4.2).) Conversely, if for some $\varepsilon > 0$ the largest singular of $Q_m|V$ is less than $1-\varepsilon$ with positive probability then with probability 1 infinitely many of the matrix products $S_{mn+m}S_{mn+m-1}\cdots S_{mn+1}$ have largest singular value less than $1-\varepsilon$. Hence, the Lipschitz constant of the mapping on Δ_k induced by Q_{mn} must converge to 0 as $n \to \infty$. In fact even more is true: the asymptotic fraction as $n \to \infty$ of blocks where $S_{mn+m}S_{mn+m-1}\cdots S_{mn+1}$ has largest singular value $< 1-\varepsilon$ is positive, by strong law of large numbers, and so the Lipschitz constant of $Q_{mn} : \Delta_k \to \Delta_k$ decays exponentially. □

Hypothesis 4.2. *For some integer $m \geq 1$ the event that all entries of Q_m are positive has positive probability.*

Corollary 4.3. *Hypothesis 4.2 implies the asymptotic collapse property* (4.2)*.*

Proof. It is well known that if a stochastic matrix has all entries strictly positive then its only eigenvalue of modulus 1 is 1, and this eigenvalue is simple (see, for instance, the discussion of the Perron-Frobenius theorem in the appendix of [9]). Consequently, if Q_m has all entries positive then $\lambda_{m,1} < 1$. □

4.2 The induced Markov chain on the simplex

The sequence of random matrix products $(Q_m)_{m\geq 1}$ induce a Markov chain on the simplex Δ_k in the obvious way: for any initial vector $Y_0 \in \Delta_k$ independent of the sequence $(S_m)_{m\geq 0}$, put

$$Y_m = Q_m Y_0, \quad m \geq 1. \tag{4.4}$$

That the sequence $\{Y_m\}_{m\geq 0}$ is a Markov chain follows from the assumption that the matrices S_i are i.i.d. Since matrix multiplication is continuous, the induced Markov chain is Feller (relative to the usual topology on Δ_k). Consequently, since Δ_k is compact, the induced chain has a stationary distribution, by the usual Bogoliubov-Krylov argument (see, e.g., [13]).

Proposition 4.4. *The stationary distribution of the induced Markov chain on the simplex is unique if and only if the asymptotic collapse property* (4.2) *holds.*

Proof of sufficiency. Let π be a stationary distribution, and let $Y_0 \sim \pi$ and $\tilde{Y}_0$ be random elements of Δ_k that are independent of the sequence $\{Q_m\}_{m\geq 1}$. Define $Y_m = Q_m Y_0$ and $\tilde{Y}_m = Q_m \tilde{Y}_0$. Both sequences $\{Y_m\}_{m\geq 0}$ and $\{\tilde{Y}_m\}_{m\geq 0}$ are versions of the induced chain, and since the distribution of Y_0 is stationary, $Y_m \sim \pi$ for every $m \geq 0$. But the asymptotic collapse property (4.2) implies that as $m \to \infty$,

$$d(Y_m, \tilde{Y}_m) \to 0 \quad \text{a.s.},$$

so the distribution of $\tilde{Y}_m$ approaches π weakly as $m \to \infty$. □

The converse is somewhat more subtle. Recall that the linear subspace $V = V_k$ orthogonal to the vector $\mathbf{1}$ is invariant under multiplication by any stochastic matrix. Define $U \subset V$ to be the set of unit vectors u in V such that $\|Q_m u\| = \|u\|$ almost surely for every $m \geq 1$. Clearly, the set U is a closed subset of the unit sphere in V, and it is also invariant, that is, $Q_m(U) \subset U$ almost surely.

Lemma 4.5. *The set U is empty if and only if the asymptotic collapse property* (4.2) *holds.*

Proof. If (4.2) holds then $\lim_{m\to\infty} \lambda_{m,1} = 0$, and so $\|Q_m u\| \to 0$ a.s. for every unit vector $u \in V$. Thus, $U = \emptyset$.

To prove the converse statement, assume that the asymptotic collapse property (4.2) *fails.* Then by Proposition 4.1, for each $m \geq 1$ the largest singular value of $Q_m|V$ is $\lambda_{m,1} = 1$, and consequently there exist (possibly random) unit vectors $v_m \in V$ such that $\|Q_m v_m\| = 1$. Since each matrix S_i is contractive, it follows that $\|Q_m v_{m+n}\| = 1$ for all $m, n \geq 1$. Hence, by the compactness of the unit sphere and the continuity of the maps $Q_m|V$, there exists a possibly random unit vector u such that $\|Q_m u\| = 1$ for every $m \geq 1$.

We will now show that there exists a *non-random* unit vector u such that $\|Q_m u\| = 1$ for every m, almost surely. Suppose to the contrary that there were no such u. For each unit vector u, let $p_m(u)$ be the probability that $\|Q_m u\| < 1$. Since the matrices S_m are weakly contractive, for any unit vector u the events $\|Q_m u\| = 1$ are decreasing in m, and so $p_m(u)$ is non-decreasing. Hence, by a subsequence argument, if for every $m \geq 1$ there were a unit vector u_m such that $p_m(u_m) = 0$, then there would be a unit vector u such that $p_m(u) = 0$ for every m. But by assumption there is no such u; consequently, there must be some finite $m \geq 1$ such that $p_m(u) > 0$ for every unit vector.

For each fixed m, the function $p_m(u)$ is lower semi-continuous (by the continuity of matrix multiplication), and therefore attains a minimum on the unit sphere of V. Since p_m is strictly positive, it follows that there exists $\delta > 0$ such that $p_m(u) \geq \delta$ for every unit vector u. But if this is the case then there can be no random unit vector u such that $\|Q_m u\| = 1$ for every $m \geq 1$, because for each m the event $\|Q_{m+1} u\| < \|Q_m u\|$ would have conditional probability (given $S_1, S_2, \ldots, S_m$) at least δ. □

Proof of necessity in Proposition 4.4. If the asymptotic collapse property (4.2) fails, then by Lemma 4.5 there exists a unit vector $u \in V$ such that $\|Q_m u\| = 1$ for all $m \geq 1$, almost surely. Hence, since Δ_k is contained in a translate of V, there exist distinct $\mu, \nu \in \Delta_k$ such that $\|Q_m(\mu - \nu)\| = \|\mu - \nu\|$ for all $m \geq 1$, a.s. By compactness, there exists such a pair $(\mu, \nu) \in \Delta_k^2$ for which $\|\mu - \nu\|$ is maximal. Fix such a pair (μ, ν), and let $A \subset \Delta_k^2$ be the set of all pairs (y, z) such that

$$\|S_1 y - S_1 z\| = \|\mu - \nu\| \quad \text{a.s.}$$

Note that the set A is closed, and consequently compact. Furthermore, because μ, ν have been chosen so that $\|\mu - \nu\|$ is maximal, for any pair $(y, z) \in A$ the points y and z must both lie in the boundary $\partial\,\Delta_k$ of the simplex.

Define $Y_m = Q_m \mu$, $Z_m = Q_m \nu$, and $R_m = (Y_m + Z_m)/2$. By construction, for each $m \geq 0$, the pair (Y_m, Z_m) lies in the set A. The sequence (Y_m, Z_m, R_m) is a Δ_k^3-valued Markov chain, each of whose projections on Δ_k is a version of the induced chain. Since Δ_k^3 is compact, the Bogoliubov-Krylov argument implies that the Markov chain (Y_m, Z_m, R_m) has a stationary distribution λ whose projection $\lambda_{Y,Z}$ on the first two coordinates is supported by A. Each of the marginal distributions λ_Y, λ_Z, and λ_R is obviously stationary for the induced chain on the simplex, and both λ_Y and λ_Z have supports contained in $\partial\,\Delta_k$. Clearly, if $(Y, Z, R) \sim \lambda$ then $R = (Y + Z)/2$.

We may assume that $\lambda_Y = \lambda_Z$, for otherwise there is nothing to prove. We claim that $\lambda_R \neq \lambda_Y$. To see this, let D be the minimal integer such that λ_Y is supported by the union $\partial_D\,\Delta_k$ of the $D-$dimensional faces of Δ_k. If $(Y, Z, R) \sim \lambda$, then $Y \neq Z$, since $\lambda_{Y,Z}$ has support in A. Consequently, $(Y + Z)/2$ is contained in the interior of a $(D + 1)-$dimensional face of Δ_k. It follows that $\lambda_R \neq \lambda_Y$.

□

Remark 4.6. Recurrence Times. *Assume that the asymptotic collapse property* (4.2) *holds, and let ν be the unique stationary distribution for the induced chain on the simplex. Say that a point v of the simplex is a* support point *of ν if ν gives positive probability to every open neighborhood of v. Fix such a neighborhood U, and let τ be the first time $m \geq 1$ that $Y_m \in U$. Then there exists $0 < r = r_U < 1$ such that for all $m \geq 1$,*

$$P\{\tau > m\} \leq r^m,$$

regardless of the initial state Y_0 of the induced chain. To see this, observe that because $\nu(U) > 0$ there exists m such that the event $Q_m(\Delta_k) \subset U$ has positive probability. Consequently, because the matrices S_i are i.i.d., the probability that $Q_{mn}(\Delta_k) \not\subset U$ for all $n = 1, 2, \ldots, N$ is exponentially decaying in N.

Remark 4.7. Relation between the induced chain on Δ_k and the EFCP. *Let $\{X_m\}_{m \geq 0}$ be a version of the EFCP chain on $[k]^{\mathbb{N}}$ with paintbox sequence $\{S_m\}_{m \geq 1}$. By Proposition 3.3, the individual coordinate sequences $\{X_m^i\}_{m \geq 0}$ are conditionally independent given $\mathcal{G} = \sigma(X_0, S_1, S_2, \ldots)$, and for each i the sequence $\{X_m^i\}_{m \geq 0}$ evolves as a time-inhomogeneous Markov chain with one-step transition probability matrices S_m. Consequently, by the strong law of large numbers, if the initial state X_0 has the property that the limiting frequencies of all colors $r \in [k]$ exist with probability one (as would be the case if the initial distribution is exchangeable), then this property persists for all times $m \geq 1$. In this case, the sequence $\{Y_m\}_{m \geq 0}$, where Y_m is the vector of limiting color frequencies in the mth generation, is a version of the induced Markov chain on the simplex Δ_k. Moreover, the jth column of the stochastic matrix S_m coincides with the limit frequencies of colors in X_m among those indices $i \in \mathbb{N}$ such that $X_{m-1}^i = j$. Thus, the paintbox sequence can be recovered (as a measurable function) from the EFCP chain.*

4.3 Asymptotic Decay Rates

Lebesgue measure on Δ_k is obtained by translating Lebesgue measure on V (the choice of Lebesgue measure depends on the choice of basis for V, but for any two choices the corresponding Lebesgue measures differ only by a scalar multiple). The $k-$fold product of Lebesgue measure on Δ_k will be referred to as *Lebesgue measure* on Δ_k^k.

Hypothesis 4.8. *The distribution Σ of the random stochastic matrix S_1 is absolutely continuous with respect to Lebesgue measure on $\mathcal{S}_k$ and has a density of class L^p for some $p > 1$.*

Hypothesis 4.8 implies that the conditional distribution of the ith column of S_1, given the other $k-1$ columns, is absolutely continuous relative to Lebesgue measure on Δ_k. Consequently, the conditional probability that it is a linear combination of the other $k-1$ columns is 0. Therefore, the matrices S_t are almost surely nonsingular, and so the Furstenberg theory ([2], chapters 3–4) applies. Furthermore, under Hypothesis 4.8, all entries of S_1 are almost surely positive. Thus, Hypothesis 4.8 implies Hypothesis 4.2.

Proposition 4.9. *Under Hypothesis 4.8,*

$$E|\log|\det S_1|| < \infty, \tag{4.5}$$

and consequently

$$\lim_{n\to\infty} (\det(Q_n|V))^{1/n} = e^{\kappa} \quad \textit{where} \quad \kappa = E \log \det S_1. \tag{4.6}$$

Remark 4.10. *The determinant of S_1 is the volume of the polyhedron $S_1[0,1]^k$, which is $\sqrt{k}$ times the volume of the $(k-1)-$dimensional polyhedron with vertices $S_1 e_i$, where $1 \le i \le k$. The volume of this $(k-1)-$dimensional polyhedron is the determinant of the restriction $S_1|V$. Consequently,*

$$\det S_1|V = \prod_{i=1}^{k-1} \lambda_{1,i}.$$

Proof. The assertion (4.6) follows from (4.5), by the strong law of large numbers, since the determinant is multiplicative. It remains to prove (4.5). Fix $\varepsilon > 0$, and consider the event $\det S_1 < \varepsilon$. This event can occur only if the smallest singular value of S_1 is less than $\varepsilon^{1/k}$, and this can happen only if one of the vectors $S_1 e_i$ lies within distance $\varepsilon^{1/k}$ (or so) of a convex linear combination of the remaining $S_1 e_j$.

The vectors $S_1 e_i$, where $i \in [k]$, are the columns of S_1, whose distribution is assumed to have a L^p density $f(M)$ with respect to Lebesgue measure dM on $\mathcal{S}_k$. Fix an integer $m \ge 1$, and consider the subset B_m of $\mathcal{S}_k$ consisting of all $k \times k$ stochastic matrices M such that the ith column Me_i lies within distance e^{-m} of the set of all convex combinations of the remaining columns Me_j. Elementary geometry shows that the set B_m has Lebesgue measure $\le Ce^{-m}$, for some constant $C = C_k$ depending on the dimension but not on m or i. Consequently, by the Hölder inequality, for a suitable constant $C' = C'_k < \infty$,

$$\begin{aligned}
E|\log|\det S_1|| &\le C' \sum_{m=0}^{\infty} (m+1) \int_{B_m} f(M)\, dM \\
&\le C' \sum_{m=0}^{\infty} (m+1) \left\{ \int_{B_m} 1\, dM \right\}^{1/q} \left\{ \int f(M)^p\, dM \right\}^{1/p} \\
&\le C' \sum_{m=0}^{\infty} (m+1) e^{-m/q} \left\{ \int f(M)^p\, dM \right\}^{1/p} < \infty
\end{aligned}$$

where $1/p + 1/q = 1$. In fact, this also shows that $\log|\det S_1|$ has finite moments of all orders, and even a finite moment generating function in a neighborhood of 0. $\square$

Proposition 4.11. *Under Hypotheses 4.8,*

$$\lim_{n\to\infty} \lambda_{n,1}^{1/n} := \lambda_1 \quad \textit{exists a.s.} \tag{4.7}$$

Moreover, the limit λ_1 is constant and satisfies $0 < \lambda_1 < 1$.

Remark 4.12. *It can be shown that the Lyapunov exponents of the sequence Q_m are the same as those of $Q_m|V$, but with one additional Lyapunov exponent 0. Thus, $\log \lambda_1$ is the* second *Lyapunov exponent of the sequence Q_m.*

Remark 4.13. *Hypothesis 4.8 implies that the distribution of S_1 is* strongly irreducible *(cf. [2], ch. 3), and so a theorem of Furstenberg implies that the top two Lyapunov exponents of the sequence Q_m are distinct. However, additional hypotheses are needed to guarantee that $\lambda_1 > 0$. This is the main point of Propositions 4.9–4.11.*

Proof of Proposition 4.11. The almost sure convergence follows from the Furstenberg-Kesten theorem [6] (or alternatively, Kingman's subadditive ergodic theorem [10]), because the largest singular value of $Q_n|V$ is the matrix norm of $Q_n|V$, and the matrix norm is sub-multiplicative. That the limit λ_1 is constant follows from the Kolmogorov 0-1 law, because if the matrices S_j are nonsingular (as they are under the hypotheses on the distribution of S_1) the value of λ_1 will not depend on any initial segment $S_m S_{m-1} \cdots S_1$ of the matrix products.

That $\lambda_1 < 1$ follows from assertion (4.3) of Proposition 4.1, because Hypothesis 4.2 implies that there is a positive probability $\eta > 0$ that all entries of S_1 are at least $\varepsilon > 0$, in which case S_1 is strictly contractive on Δ_k – and hence also on V – with contraction factor $\theta = \theta(\varepsilon) < 1$ ([8], Proposition 1.3).

Finally, the assertion that $\lambda_1 > 0$ follows from Proposition 4.9, because for any stochastic matrix each singular value is bounded below by the determinant. □

Corollary 4.14. *Under Hypothesis 4.8,*

$$\lim_{n\to\infty} \max_{i\neq j} \|Q_n e_i - Q_n e_j\|^{1/n} = \lambda_1 \quad \textit{almost surely.}$$

Proof. The lim sup of the maximum cannot be greater than λ_1, because for each n the singular value $\lambda_{n,1}$ of $Q_n|V$ is just the matrix norm $\|Q_n\|$. To prove the reverse inequality, assume the contrary. Then there is a subsequence $n = n_m \to \infty$ along which

$$\limsup_{m\to\infty} \max_{i\neq j} \|Q_n e_i - Q_n e_j\|^{1/n} < \lambda_1 - \varepsilon$$

for some $\varepsilon > 0$. Denote by $u = u_n \in V$ the unit vector that maximizes $\|Q_n u\|$. Because the vectors $e_i - e_{i+1}$ form a basis of V, for each n the vector u_n is a linear combination $u_n = \sum_i a_{ni}(e_i - e_{i+1})$, and because each u_n is a unit vector, the coefficients a_{ni} are uniformly bounded by (say) C in magnitude. Consequently,

$$\|Q_n u_n\| \leq C \sum_i \|Q_n(e_i - e_{i+1})\|.$$

This implies that along the subsequence $n = n_m$ we have

$$\limsup_{m\to\infty} \|Q_n u_n\|^{1/n} < \lambda_1 - \varepsilon.$$

But this contradicts the fact that $\|Q_n|V\|^{1/n} \to \lambda_1$ from Proposition 4.11.

□

Remark 4.15. *It can be also be shown that*

$$\lim_{n\to\infty} \min_{i\neq j} \|Q_n e_i - Q_n e_j\|^{1/n} = \lambda_1.$$

This, however, will not be needed for the results of Section 5.

Remark 4.16. *The argument used to prove that $\lambda_1 < 1$ in the proof of Proposition 4.11 also proves that even if Hypothesis 4.8 fails, if the distribution of S_1 puts positive weight on the set of stochastic matrices with all entries at least ε, for some $\varepsilon > 0$, then*

$$\limsup_{n\to\infty} \max_{i\neq j} \|Q_n e_i - Q_n e_j\|^{1/n} < 1. \tag{4.8}$$

Hypothesis 4.8 guarantees that the sequence $\|Q_n e_i - Q_n e_j\|^{1/n}$ has a limit, and that the limit is positive. When Hypothesis 4.8 fails, the convergence in (4.8) *can be super-exponential (i.e., the limsup in* (4.8) *can be* 0*). For instance, this is the case if for some rank-1 stochastic matrix A with all entries positive there is positive probability that $S_1 = A$.*

5 Convergence to stationarity of EFCP chains

Assume throughout this section that $\{X_m\}_{m\geq 1}$ is an EFCP chain on $[k]^{[n]}$ or $[k]^{\mathbb{N}}$ with directing measure μ_Σ, as defined by (3.6). Let $S_1, S_2, \ldots$ be the associated paint-box sequence: these are i.i.d. random column-stochastic matrices with distribution Σ. Proposition 3.3 shows that the joint distribution of the coordinate variables X_m^i of an EFCP chain with paintbox sequence $\{S_i\}_{i\geq 1}$ is controlled by the random matrix products $Q_m = S_m S_{m-1} \cdots S_1$. In this section, we use this fact together with the results concerning random matrix products recounted in Section 4 to determine the mixing rates of the restrictions $\{X_m^{[n]}\}_{m\geq 1}$ of EFCP chains to the finite configuration spaces $[k]^{[n]}$.

5.1 Ergodicity

An EFCP chain need not be ergodic: for instance, if each S_i is the identity matrix then every state is absorbing and $X_m^i = X_0^i$ for every $m \geq 1$ and every $i \in \mathbb{N}$. More generally, if the random matrices S_i are all permutation matrices then the *unlabeled* partitions of $\mathbb{N}$ induced by the labeled partitions X_m do not change with m, and so the restrictions $X_m^{[n]}$ cannot be ergodic. The failure of ergodicity in these examples stems from the fact that the matrix products Q_m do not contract the simplex Δ_k.

Proposition 5.1. *Let λ be any stationary distribution for the induced Markov chain on the simplex. Then for each $n \in \mathbb{N} \cup \{\infty\}$ the $\lambda-$mixture ϱ_λ^n of the product-multinomial measures on $[k]^{[n]}$ is stationary for the EFCP chain on $[k]^{[n]}$.*

Remark 5.2. *Recall that the product-multinomial measures ϱ_s^n are defined by equation* (2.4)*; the $\lambda-$mixture is defined to be the average*

$$\varrho_\lambda^n = \int_{\Delta_k} \varrho_s^n \, \lambda(ds).$$

Thus, a random configuration $X \in [k]^{[n]}$ with distribution ϱ_λ^n can be obtained by first choosing $s \sim \lambda$, then, conditional on s, independently assigning colors to the coordinates $i \in [n]$ by sampling from the ϱ_s^n distribution.

Proof. This is an immediate consequence of Proposition 3.3. □

Proposition 5.3. *Assume that with probability one the random matrix products Q_m asymptotically collapse the simplex Δ_k, that is,*

$$\lim_{m\to\infty} \text{diameter}(Q_m(\Delta_k)) = 0. \tag{5.1}$$

Then for each $n \in \mathbb{N}$ the corresponding EFCP chain $\{X_m^{[n]}\}_{m\geq 0}$ on $[k]^{[n]}$ is ergodic, i.e., has a unique stationary distribution. Conversely, if, for some $n \geq 1$, the EFCP chain $\{X_m^{[n]}\}_{m\geq 0}$ is ergodic, then the asymptotic collapse property (5.1) *must hold.*

Proof. Fix $n \geq 1$. By Propositions 4.4 and 5.1, there exists at least one stationary distribution π. Let $\{X_m\}_{m\geq 0}$ and $\{\tilde{X}_m\}_{m\geq 0}$ be conditionally independent versions of the EFCP chain given the (same) paintbox sequence $(S_i)_{i\geq 1}$, with $\tilde{X}_0 \sim \pi$ and $X_0 \sim \nu$ arbitrary. Then, for any time $m \geq 1$, the conditional distributions of X_m and $\tilde{X}_m$, given the paintbox sequence, can be recovered from the formula (3.8) by integrating out over the distributions of X_0 and $\tilde{X}_0$, respectively. But under the hypothesis (5.1), for large m the columns of Q_m are, with high probability, nearly identical, and so for large m the products

$$\prod_{i=1}^{n} Q_m(x_m^i, X_0^i) \quad \text{and} \quad \prod_{i=1}^{n} Q_m(x_m^i, \tilde{X}_0^i)$$

will be very nearly the same. It follows, by integrating over all paintbox sequences, that the unconditional distributions of X_m and $\tilde{X}_m$ will be nearly the same when m is large. This proves that the stationary distribution π is unique and that as $m \to \infty$ the distribution of X_m converges to π.

By Proposition 4.4, if the asymptotic collapse property (5.1) fails then the induced Markov chain on the simplex has at least two distinct stationary distributions μ, ν. By Proposition 5.1, these correspond to different stationary distributions for the EFCP. □

5.2 Mixing rate and cutoff for EFCP chains

We measure distance to stationarity using the total variation metric (2.1). Write $\mathcal{D}(X_m)$ to denote the distribution of X_m. In general, the distance $\|\mathcal{D}(X_m) - \pi\|_{TV}$ will depend on the distribution of the initial state X_0. The $\varepsilon-$*mixing time* is defined to be the number of steps needed to bring the total variation distance between $\mathcal{D}(X_m)$ and π below ε for *all* initial states x_0:

$$t_{\text{mix}}(\varepsilon) = t_{\text{mix}}^{(n)}(\varepsilon) = \min\{m \geq 1 \,:\, \max_{x_0} \|\mathcal{D}(X_m) - \pi\|_{TV} < \varepsilon\}. \tag{5.2}$$

Theorem 5.4. *Assume that with probability one the random matrix products $Q_m = S_m S_{m-1} \cdots S_1$ asymptotically collapse the simplex Δ_k, that is, relation* (4.2) *holds. Then for a suitable constant $K = K_\Sigma < \infty$ depending only on the distribution Σ of S_1, the mixing times of the corresponding EFCP chains on the finite state spaces $[k]^{[n]}$ satisfy*

$$t_{\text{mix}}^{(n)}(\varepsilon) \leq K \log n. \tag{5.3}$$

Remark 5.5. *In some cases the mixing times will be of smaller order of magnitude than $\log n$. Suppose, for instance, that for some $m \geq 1$, the event that the matrix Q_m is of rank 1 has positive probability. (This would be the case, for instance, if the columns of S_1 were independently chosen from a probability distribution on Δ_k with an atom.) Let T be the least m for which this is the case; then $T < \infty$ almost surely, since matrix rank is sub-multiplicative, and $Q_m(\Delta_k)$ is a singleton for any $m \geq T$. Consequently, for any elements $a, b, c \in [k]$,*

$$Q_m(a, b) = Q_m(a, c) \quad \textit{if } T \leq m.$$

Hence, if $\{X_m\}_{m\geq 0}$ *and* $\{\tilde{X}_m\}_{m\geq 0}$ *are versions of the EFCP chain with different initial conditions* X_0 *and* $\tilde{X}_0$*, but with the same paintbox sequence* S_m*, then, by Proposition 3.3,* X_m *and* $\tilde{X}_m$ *have the same conditional distribution, given* $\sigma(S_i)_{i\geq 1}$*, on the event* $T \leq m$*. It follows that the total variation distance between the* unconditional *distributions of* X_m *and* $\tilde{X}_m$ *is no greater than* $P\{T > m\}$*. Thus, for any* $n \in \mathbb{N}$*, the EFCP chain mixes in* $O(1)$ *steps, that is, for any* $\varepsilon > 0$ *there exists* $K_\varepsilon < \infty$ *such that for all* n*,*

$$t_{\text{mix}}^{(n)}(\varepsilon) \leq K_\varepsilon.$$

Proof of Theorem 5.4. (A) Consider first the special case where for some $\delta > 0$ every entry of S_1 is at least δ, with probability one. It then follows that no entry of Q_m is smaller than δ. By Proposition 4.1, if (4.2) holds then the diameters of the sets $Q_m(\Delta_k)$ shrink exponentially fast: in particular, for some (nonrandom) $\varrho < 1$,

$$\text{diameter}(Q_m(\Delta_k)) < \varrho^m \tag{5.4}$$

eventually, with probability 1.

Let $\{X_m\}_{m\geq 0}$ and $\{\tilde{X}_m\}_{m\geq 0}$ be versions of the EFCP on $[k]^{[n]}$ with different initial conditions X_0 and $\tilde{X}_0$, but with the same paintbox sequence S_m. By Proposition 3.3, the conditional distributions of X_m and $\tilde{X}_m$ given the paintbox sequence are product-multinomials:

$$\begin{aligned} P(X_m^i = x^i \ \text{ for each } i \in [n] \,|\, \mathcal{S}) &= \prod_{i=1}^n Q_m(x_m^i, X_0^i) \quad \text{and} \\ P(\tilde{X}_m^i = x^i \ \text{ for each } i \in [n] \,|\, \mathcal{S}) &= \prod_{i=1}^n Q_m(x_m^i, \tilde{X}_0^i). \end{aligned} \tag{5.5}$$

Since the multinomial distributions $Q_m(\cdot,\cdot)$ assign probability at least $\delta > 0$ to every color $j \in [k]$, Corollary 2.3 implies that for any $\varepsilon > 0$, if $m = K\log n$, where $K > -1/(2\log\varrho)$, then for all sufficiently large n the total variation distance between the conditional distributions of X_m and $\tilde{X}_m$ will differ by ε on the event (5.4) holds. Since (5.4) holds eventually, with probability one, the inequality (5.3) now follows by Lemma 2.4.

(B) The general case requires a bit more care, because if the entries of the matrices Q_m are not bounded below then the product-multinomial distributions (5.5) will not be bounded away from $\partial\,\Delta_k$, as required by Corollary 2.3.

Assume first that for some $m \geq 1$ there is positive probability that $Q_m(\Delta_k)$ is contained in the interior of Δ_k. Then for some $\delta > 0$ there is probability at least δ that every entry of Q_m is at least δ. Consequently, for any $\alpha > 0$ and any $K > 0$, with probability converging to one as $n \to \infty$, there will exist $m \in [K\log n, K(1+\alpha)\log n]$ (possibly random) such that every entry of Q_m is at least δ. By (5.4) the probability that the diameter of $Q_m(\Delta_k)$ is less than ϱ^m converges to 1 as $m \to \infty$. It then follows from Corollary 2.3, by the same argument as in (A), that if $K > -1/(2\log\varrho)$ then the total variation distance between the conditional distributions of X_m and $\tilde{X}_m$ will differ by a vanishingly small amount. Since total variation distance decreases with time, it follows that the total variation distance between the conditional distributions of $X_{K+K\alpha}$ and $\tilde{X}_{K+K\alpha}$ are also vanishingly small. Consequently, the distance between the unconditional distributions is also small, and so (5.3) follows, by Lemma 2.4.

(C) Finally, consider the case where $Q_m(\Delta_k)$ intersects $\partial\,\Delta_k$ for every m, with probability one. Recall (Proposition 5.1) that if the asymptotic collapse property (4.2) holds then the induced Markov chain Y_m on the simplex has a unique stationary distribution ν. If there is no $m \in \mathbb{N}$ such that $Q_m(\Delta_k)$ is contained in the interior of Δ_k, then the support of ν must be contained in the boundary $\partial\,\Delta_k$. Fix a support point v, and let m be sufficiently large that (5.4) holds. Since $Q_m(\Delta_k)$ must intersect $\partial\,\Delta_k$, it follows that for any coordinate $a \in [k]$ such that $v_a = 0$ (note that there must be at least one such a, because $v \in \partial\,\Delta_k$), the ath coordinate $(Q_m y)_a$ of any point in the image $Q_m(\Delta_k)$ must be smaller than ϱ^m. If K is chosen sufficiently large and $m \geq K \log n$, then $\varrho^m < n^{-2}$; hence, by Proposition 3.3,

$$P(X^i_m = a \text{ for some } i \in [n] \,|\, \sigma(S_l)_{l\geq 1}) \leq n \cdot n^{-2} = n^{-1} \to 0,$$

and similarly for $\tilde{X}_m$. Therefore, the contribution to the total variation distance between the conditional distributions of X_m and $\tilde{X}_m$ from states $x^1 x^2 \cdots x^n$ in which the color a appears at least once is vanishingly small. But for those states for which no such color appears, the factors $Q_m(a, b)$ in (5.5) will be bounded below by the minimum nonzero entry of v, and the result will follow by a routine modification of the argument in (B) above. □

Parts (A)-(B) of the foregoing proof provide an explicit bound in the special case where $Q_m(\Delta_k)$ is contained in the interior of Δ_k with positive probability.

Corollary 5.6. *Assume that with probability one the random matrix products* $Q_m = S_m S_{m-1} \cdots S_1$ *asymptotically collapse the simplex* Δ_k*, so that for some* $0 < \varrho < 1$*,*

$$\text{diameter}(Q_m(\Delta_k)) < \varrho^m$$

for all sufficiently large m*, with probability* 1*. Assume also that with positive probability* $Q_m(\Delta_k)$ *is contained in the interior of* Δ_k*, for some* $m \geq 1$*. Then for any* $K > -1/(2 \log \varrho)$ *the bound* (5.3) *holds for all sufficiently large* n*.*

Theorem 5.7. *Assume that the paintbox distribution* Σ *satisfies Hypothesis 4.8. Then the corresponding EFCP chains exhibit the* cutoff phenomenon*, that is, for all* $\varepsilon, \delta \in (0, 1/2)$*, if* n *is sufficiently large, then*

$$(\theta - \delta) \log n \leq t^{(n)}_{\text{mix}}(1 - \varepsilon) \leq t^{(n)}_{\text{mix}}(\varepsilon) \leq (\theta + \delta) \log n, \tag{5.6}$$

where

$$\theta = -1/(2 \log \lambda_1) \tag{5.7}$$

and λ_1 *is the second Lyapunov exponent of the sequence* Q_m*, that is, as in Proposition* (4.11).

Proof of the Upper Bound $t_{\text{mix}}(\varepsilon) \leq (\theta + \delta) \log n$*.* Because the distribution of S_1 is absolutely continuous with respect to Lebesgue measure, there is positive probability that all entries of $S_1 = Q_1$ are positive, and so there is positive probability that $Q_1(\Delta_k)$ is contained in the interior of Δ_k. Therefore, Corollary 5.6 applies. But Proposition 4.11 and Corollary 4.14 implies that, under Hypothesis 4.8, that $\varrho = \lambda_1$.

□

Proof of the Lower Bound $t_{\text{mix}}(\varepsilon) \geq (\theta - \delta) \log n$*.* It suffices to show that there exist initial states $x_0, \tilde{x}_0$ such that if $\{X_t\}_{t\geq 0}$ and $\{\tilde{X}_t\}_{t\geq 0}$ are versions of the EFCP chain with initial states $X_0 = x_0$ and $\tilde{X}_0 = \tilde{x}_0$, respectively, then the distributions of X_m and $\tilde{X}_m$ have total variation distance near 1 when $m \leq (\theta - \delta) \log n$. The proof will rely on Corollary 4.14, according to which there is a (possibly random) pair of indices $i \neq j$ for which

$$\lim_{m\to\infty} \|Q_m e_i - Q_m e_j\|^{1/m} = \lambda_1. \tag{5.8}$$

Consider first, to fix ideas, the special case $k = 2$. In this case (5.8) holds with $i = 1$ and $j = 2$. Assume that $n = 2n'$ is even (if n is odd, project onto the first $n-1$ coordinates), and let

$$x_0 = 11\cdots 111\cdots 1 \quad \text{and} \quad \tilde{x}_0 = 111\cdots 122\cdots 2$$

be the elements of $[k]^n$ such that x_0 has all coordinates colored 1, while $\tilde{x}_0$ has its first n' colored 1 but its second n' colored 2. We will show that the distributions of X_m and $\tilde{X}_m$ remain at large total variation distance at time $m = (\theta - \alpha)\log n$. Without loss of generality, assume that both of the chains $\{X_t\}_{t\geq 0}$ and $\{\tilde{X}_t\}_{t\geq 0}$ have the same paintbox sequence $S_1, S_2, \ldots$. Then by Proposition 3.3, the conditional distributions of X_m and $\tilde{X}_m$ given $\mathcal{S} = \sigma(S_t)_{t\geq 1}$ are product-multinomials; in particular, for any state $x = (x^l)_{l\in[n]} \in [k]^{[n]}$,

$$P(X_m^l = x^l \text{ for all } l \in [n] \,|\, \sigma(S_t)_{t\geq 1}) = \prod_{l=1}^{n} Q_m(x^l, 1) \quad \text{and}$$
$$P(\tilde{X}_m^l = \tilde{x}^l \text{ for all } l \in [n] \,|\, \sigma(S_t)_{t\geq 1}) = \prod_{l=1}^{n'} Q_m(\tilde{x}^l, 1) \prod_{l=n'+1}^{2n'} Q_m(\tilde{x}^l, 2).$$

But relation (5.8) implies that, for some $\alpha = \alpha(\delta) > 0$, if $m = (\theta - \delta)\log n$ then the $\ell^\infty-$distance between the ith and jth columns of Q_m is at least $n^{-1/2+\alpha}$, with probability approaching 1 as $n \to \infty$. Consequently, the first n' and second n' coordinates of $\tilde{X}_m$ are (conditional on $\mathcal{S}$) independent samples from Bernoulli distributions whose parameters differ by at least $n^{-1/2+\alpha}$, but the $2n'$ coordinates of X_m are (conditional on $\mathcal{S}$) a single sample from the same Bernoulli distribution. It follows, by Lemma 2.5 (see Remark 2.6, statement (B)), that the *unconditional* distributions of X_m and $\tilde{X}_m$ are at large total variation distance, because in $\tilde{X}_m$ the first and second blocks of n' coordinates are distinguishable whereas in X_m they are not. Thus, if $m = (\theta - \delta)\log n$ then as $n \to \infty$,

$$\|\mathcal{D}(X_m) - \mathcal{D}(\tilde{X}_m)\|_{TV} \longrightarrow 1.$$

The general case is proved by a similar argument. Let $n = 2k(k-1)n'$ be an integer multiple of $2k(k-1)$. Break the coordinate set $[n]$ into $k(k-1)$ non-overlapping blocks of size $2n'$, one for each ordered pair (i,j) of distinct colors. In the block indexed by (i,j) let x_0 take the value i, and let $\tilde{x}_0$ take the value i in the first half of the block and the value j in the second half. Let $\{X_t\}_{t\geq 0}$ and $\{\tilde{X}_t\}_{t\geq 0}$ be versions of the EFCP chain with initial states x_0 and $\tilde{x}_0$, respectively. Then by an argument similar to that used in the binary case $k = 2$, if $m = (\theta - \delta)\log n$, then for large n, in *some* block (i,j) of $\tilde{X}_m$ the first n' and second n' coordinates of $\tilde{X}_m$ will be distinguishable, but in X_m they will not. Therefore, the unconditional distributions of X_m and $\tilde{X}_m$ will be at total variation distance near 1.

□

Example 5.8 (*Self-similar cut-and-paste chains*)**.** *Self-similar cut-and-paste chains were introduced in [3]. These are EFCP chains for which the paintbox measure $\Sigma = \Sigma_\nu$ is such that the columns of $S_1 \sim \Sigma$ are i.i.d. with common distribution ν on Δ_k. If $S_1, S_2, \ldots$ are i.i.d. with distribution Σ_ν then the random matrix products $Q_m = S_m S_{m-1} \cdots S_1$ asymptotically collapse the simplex, and so Theorem 5.4 applies. If, in addition, the measure ν has a density of class L^p relative to Lebesgue measure on Δ_k, then Theorem 5.7 applies.*

A particular example of the self-similar cut-and-paste chain is determined by the symmetric Dirichlet distribution on Δ_k. For any choice of $\alpha > 0$, let

$$\nu_\alpha^{(k)}(dx) := \frac{\Gamma(k\alpha)}{\Gamma(\alpha)^k} x_1^{\alpha-1} \cdots x_k^{\alpha-1} dx, \quad x \in \Delta_k,$$

the density of the Dirichlet$(\alpha, \dots, \alpha)$ *distribution on* Δ_k*. In this case, the transition probabilities of the corresponding EFCP chain can be written down explicitly and it is easy to show that this chain is reversible with respect to a version of the Pitman-Ewens two-parameter distribution on* $[k]^{[n]}$*; see [3], Section 6 for more discussion. In the case* $k = 2$*,* $\nu_\alpha^{(k)}$ *corresponds to the Beta distribution with parameter* (α, α)*, and the transition probabilities on* $[k]^{[n]}$ *are*

$$q_n(x, x') := \frac{\alpha^{\uparrow n_{00}} \alpha^{\uparrow n_{01}} \alpha^{\uparrow n_{10}} \alpha^{\uparrow n_{11}}}{(2\alpha)^{\uparrow n_0} (2\alpha)^{\uparrow n_1}}, \quad x, x' \in [k]^{[n]},$$

where $n_{ij} := \sum_{l=1}^n \mathbf{1}(x^l = i \text{ and } x'^l = j)$*,* $n_i := n_{i0} + n_{i1}$*,* $i, j = 0, 1$*, and* $\alpha^{\uparrow j} := \alpha(\alpha + 1) \cdots (\alpha + j - 1)$*. In this case,* $S_1 \sim \Sigma_\nu$ *has the form*

$$S_1 := \begin{pmatrix} \Theta & \Theta' \\ 1 - \Theta & 1 - \Theta' \end{pmatrix},$$

where Θ, Θ' *are independent Beta*(α, α) *random variables. Hence, the determinant of* S_1 *is* $\Theta - \Theta'$*. The singular values of a* 2×2 *stochastic matrix are determined by its determinant, and so the exact value of the constant* θ *in Theorem 5.7 can be derived from the quantity*

$$E|\log|\det S_1|| = E|\log|\Theta - \Theta'|| = \int_{[0,1]\times[0,1]} |\log|\theta - \theta'||\nu_\alpha^{(2)}(d\theta)\nu_\alpha^{(2)}(d\theta'),$$

which can be computed explicitly for specific values of α*.*

5.3 Examples

We now discuss some examples of Markov chains on $[k]^{[n]}$ whose transitions are governed by an i.i.d. sequence of random partition matrices $M_1, M_2, \dots$ with law μ, but which are not EFCP chains because μ does not coincide with μ_Σ for some probability measure Σ on Δ_k^k. As a result, the examples we show are not covered by Theorems 5.4 or 5.7. We are, however, able to establish upper bounds and, in some cases, cutoff using different techniques. All of the chains in these examples are reversible and ergodic relative to the uniform distribution on $[k]^{[n]}$.

Example 5.9 (*Ehrenfest chain on the hypercube*)**.** *For* $k = 2$*, we regard* $x \in \{0, 1\}^{[n]}$ *as an element of* n*-dimensional hypercube. For each* $i = 1, \dots, n$ *and* $a \in \{0, 1\}$*, we define* $M_{a,i}$ *as the* 2×2 *partition matrix with entries*

$$M_{0,i} := \begin{pmatrix} [n]\backslash\{i\} & \emptyset \\ \{i\} & [n] \end{pmatrix} \quad \textit{or} \quad M_{1,i} := \begin{pmatrix} [n] & \{i\} \\ \emptyset & [n]\backslash\{i\} \end{pmatrix}.$$

Let $x_0 \in \{0, 1\}^n$ *be an initial state and first choose* $a_1, a_2, \dots$ *i.i.d. Bernoulli(1/2) and, independently of* (a_m)*, choose* $i_1, i_2, \dots$ *i.i.d. from the uniform distribution on* $[n]$*. Then the chain* $X = (X_m)_{m \geq 0}$ *is constructed by* $X_0 = x_0$ *and, for* $m = 1, 2, \dots$*,* $X_m = M_{a_m, i_m}(X_{m-1})$*, as defined in* (3.3)*. This corresponds to the usual Ehrenfest chain on the hypercube, which is known to exhibit the cutoff phenomenon at* $(1/2)n \log n$*; e.g. see [12], Example 18.2.2.*

Example 5.10 (*General Ehrenfest chain*)**.** *A more general form of the Ehrenfest chain in the previous example is described as follows. Fix* $n \in \mathbb{N}$*, take* $\alpha \in (0, 1)$ *and choose a random subset* $A \subset [n]$ *uniformly among all subsets of* $[n]$ *with cardinality* $\lfloor \alpha n \rfloor := \max\{r \in \mathbb{N} : r \leq \alpha n\}$*, the* floor *of* αn*. For* $i \in \{0, 1\}$ *and* $A \subset [n]$*, we define the partition matrix* $M(A, i)$ *by either*

$$M(A, 0) := \begin{pmatrix} [n]\backslash A & \emptyset \\ A & [n] \end{pmatrix} \quad \textit{or} \quad M(A, 1) := \begin{pmatrix} [n] & A \\ \emptyset & [n]\backslash A \end{pmatrix}.$$

Let $A = (A_1, A_2, \dots)$ *be an i.i.d. sequence of uniform subsets of size* $\lfloor \alpha n \rfloor$*, let* $I = (I_1, I_2, \dots)$ *be i.i.d. Bernoulli*$(1/2)$ *and let* $x_0 \in \{0,1\}^n$*. Conditional on* A *and* I*, we construct* $X = (X_m)_{m \geq 0}$ *by putting* $X_0 = x_0$ *and, for each* $m \geq 1$*, define* X_m *as in* (3.4) *from the sequence* $M(A_1, I_1), M(A_2, I_2), \dots$*. We call* X *an* Ehrenfest(α) *chain.*

Define the coupling time T *by*

$$T := \min\left\{ t \geq 1 : \bigcup_{j=1}^{t} A_j = [n] \right\}.$$

Any two Ehrenfest(α*) chains* X *and* X' *constructed from the same sequences* A *and* I *will be coupled by time* T*.*

An upper bound on the distance to stationarity of the general Ehrenfest(α*) chain is obtained by standard properties of the hypergeometric distribution. In particular, let* $R_t := \#\left([n] \backslash \bigcup_{j=1}^{t} A_j\right)$ *be the number of indices that have not appeared in one of* $A_1, \dots, A_t$*. By definition,* $\{T \leq t\} = \{R_t = 0\}$ *and standard calculations give*

$$\mathbb{P}(R_{t+1} = j | R_t = r) = \binom{r}{r-j}\binom{n-r}{j}\binom{n}{\lfloor \alpha n \rfloor}^{-1}, \; j = 0, 1, \dots, r,$$

$$\mathbb{E}(R_t) = n\left(1 - \frac{\lfloor \alpha n \rfloor}{n}\right)^t.$$

For fixed $\alpha \in (0,1)$*, the* ε*-mixing time is bounded above by*

$$\|\mathcal{D}(X_t) - \pi\|_{TV} \leq n\left(1 - \frac{\lfloor \alpha n \rfloor}{n}\right)^t \leq n \exp\{-\lfloor \alpha n \rfloor t/n\} \tag{5.9}$$

and it immediately follows, for $\beta > 0$ *and* $t = \left(\frac{n}{2\lfloor \alpha n \rfloor} \log n + \beta \frac{n}{\lfloor \alpha n \rfloor}\right)$*, that*

$$\|\mathcal{D}(X_t) - \pi\|_{TV} \leq n^{-1/2} \exp(-\beta) \to 0 \text{ as } \beta \to \infty.$$

When $\alpha \in (0, 1/2]$*, we can use Proposition 7.8 from [12] and some standard theory for coupon collecting to obtain the lower bound*

$$\|\mathcal{D}(X_t) - \pi\|_{TV} \geq 1 - 8\exp\{-2\beta + 1\},$$

when $t = \left(\frac{n}{2\lfloor \alpha n \rfloor} \log n - \beta \frac{n}{\lfloor \alpha n \rfloor}\right)$*. Hence, these chains exhibit cutoff at* $n/(2\lfloor \alpha n \rfloor) \log n$*.*

Note that the standard Ehrenfest chain (Example 5.9) corresponds to $\alpha = 1/n$*.*

Example 5.11 (*A* $\log\log n$ *upper bound on mixing time*)**.** *For the general Ehrenfest chains described above, the upper bound* (5.9) *on mixing time can be applied more generally to sequences* $\alpha := (\alpha_1, \alpha_2, \dots)$ *in* $(0,1)$*. For each* $n \in \mathbb{N}$*, let* $\alpha_n = 1 - \exp\{-\log n / \log\log n\}$ *and let* X^n *be an Ehrenfest(*α_n*) chain. By* (5.9)*, for* $t \geq (1 + \beta) \log\log n$*,* $\beta > 0$*, we have*

$$\|\mathcal{D}(X_t^n) - \pi\|_{TV} \leq n^{-\beta},$$

which converges to 0 as $n \to \infty$*.*

In general, for any function $f(n)$ *of* $n \in \mathbb{N}$*, we can obtain an upper bound of* $(1 + \beta) f(n)$ *by the relation*

$$\alpha_n = 1 - \exp\left\{-\frac{\log n}{f(n)}\right\}.$$

The space $[k]$ is a group under addition modulo k defined by

$$x +^* x' := x + x' - 2 \text{ (mod } k) + 1.$$

(To avoid unnecessary formalism, we henceforth write $x + x'$ to denote the above addition operation.)

Write $\mathbb{N}_k^n$ to denote the group $[k]^{[n]}$ together with the operation $+$, which we define by componentwise addition modulo k of the coordinates of $x \in [k]^{[n]}$. That is, for any $x, x' \in [k]^{[n]}$, we define

$$(x + x')^i := x^i + x'^i - 2 \text{ (mod } k) + 1, \quad 1 \le i \le n.$$

This action makes the space $[k]^{[n]}$ into a group with a corresponding action, also denoted $+$. In fact, for $x, x' \in [k]^{[n]}$, $x + x'$ equals $M_x(x')$, where M_x is the partition matrix whose jth column is the jth cyclic shift of the classes of x; that is, for $L = (L_1, \ldots, L_k)$, the ordered partition associated to x through (3.1), we define

$$M_x := \begin{pmatrix} L_1 & L_k & L_{k-1} & \cdots & L_2 \\ L_2 & L_1 & L_k & \cdots & L_3 \\ L_3 & L_2 & L_1 & \cdots & L_4 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_k & L_{k-1} & L_{k-2} & \cdots & L_1 \end{pmatrix}. \tag{5.10}$$

Then, for every $x, x' \in [k]^{\mathbb{N}}$, we have $x + x' = M_x(x')$.

Example 5.12. *For $n \in \mathbb{N}$, let ϱ_n be a probability measure on $[k]^{[n]}$ and let $x_0 \in [k]^{[n]}$. A $\mathrm{CP}_n(\varrho_n)$ chain X with initial state $X_0 = x_0$ can be constructed as follows. First, generate $Y_1, Y_2, \ldots$ i.i.d. from ϱ_n. Conditional on $Y_1, Y_2, \ldots$, put*

$$X_m = Y_m + Y_{m-1} + \cdots + Y_1 + X_0, \quad m \ge 1.$$

Under the definition (5.10)*, this is a cut-and-paste chain; however, the columns of each matrix are a deterministic function of one another.*

Consider the case where ϱ_n is a product measure of a probability measure λ on $[k]$ which is symmetric, i.e.

$$\lambda(j) = \lambda(k - j + 1) > 0, \quad j = 1, \ldots, k.$$

In this case, it is easy to see that the $\mathrm{CP}_n(\varrho_n)$ chain is reversible and hence has the uniform distribution as its unique stationary distribution.

For this construction of X, the directing measure μ on $\mathcal{M}_{[n]:k}$ induced by λ cannot be represented as μ_Σ for some measure Σ on $\mathcal{S}_k$. Nonetheless, the mixing time of X is bounded above by $K \log n$ for some constant $K \le 2/\min_j \lambda(j) < \infty$.

6 Projected cut-and-paste chains

Recall from Section 3.1 that there is a natural projection $\Pi_n : [k]^{[n]} \to \mathcal{P}_{[n]:k}$. If $\{X_m\}_{m \ge 0}$ is a Markov chain on $[k]^{[n]}$ whose transition probability matrix is invariant under permutations of the colors $[k]$, then the projection $\{\Pi_n(X_m)\}_{m \ge 0}$ is also a Markov chain. Assume henceforth that this is the case.

Following is a simple sufficient condition for the law of an EFCP chain to be invariant under permutations of the label set $[k]$. Say that a probability measure Σ on the space Δ_k^k of column-stochastic matrices is *row-column exchangeable* if the distribution of $S_1 \sim \Sigma$ is invariant under independent permutations of the rows or the columns.

Lemma 6.1. *If $\{X_m\}_{m\geq 0}$ is an EFCP chain on $[k]^{[n]}$ whose paintbox measure Σ is row-column exchangeable then its transition probability matrix is invariant under permutations of the colors $[k]$.*

Proof. For any permutation γ of $[k]$, define the *recoloring* of $x \in [k]^{[n]}$ by γ by $x_\gamma = (x^i_\gamma)_{i\in[n]}$, where $x^i_\gamma = \gamma(x^i)$ for $i = 1, \dots, n$. For $x, x' \in [k]^{[n]}$, let $P(x, x')$ denote the transition probability from x to x' under the operation (3.4) with directing measure μ_Σ. By row-column exchangeability of Σ, we have, for all permutations γ, γ' of $[k]$,

$$P(x, x') = P(x, x'_\gamma) = P(x_\gamma, x') = P(x_\gamma, x'_\gamma),$$

for every $x, x' \in [k]^{[n]}$. It follows immediately that the transition probability $Q = P\Pi_n^{-1}$ of the projected chain $\Pi_n(X)$ is given by

$$Q(\Pi_n(x), \Pi_n(x')) = k^{\downarrow \#\Pi_n(x')} P(x, x'), \quad \text{for every } x, x' \in [k]^{[n]},$$

where $\#\Pi_n(x')$ denotes the number of blocks of the partition $\Pi_n(x')$. ☐

Following Crane [4], we call the induced chain $\Pi := \Pi_\infty(X)$ of an EFCP chain with RCE directing measure Σ a *homogeneous cut-and-paste chain*.

If the chain $\{X_m\}_{m\geq 0}$ is ergodic, then its unique stationary distribution is invariant under permutations of $[k]$, since its transition probability matrix is, and therefore projects via Π_n to a stationary distribution for the projected chain $\{\Pi_n(X_m)\}_{m\geq 0}$. The sufficiency principle (equation (2.2)) for total variation distance (see also Lemma 7.9 of [12]) implies that the rate of convergence of the projected chain $\{\Pi_n(X_m)\}_{m\geq 0}$ is bounded by that of the original chain $\{X_m\}_{m\geq 0}$. Theorem 5.4 provides a bound for this convergence when the chain $\{X_m\}_{m\geq 0}$ is an EFCP chain.

Corollary 6.2. *Assume that $\{X_m = X_m^{[n]}\}_{m\geq 0}$ is an EFCP chain on $[k]^{[n]}$ whose paintbox measure Σ is RCE and satisfies the hypothesis of Theorem 5.4 (in particular, the random matrix products Q_m asymptotically collapse the simplex Δ_k). Then, for a suitable constant $K = K_\Sigma < \infty$ depending only on the distribution Σ of S_1 and for any $\varepsilon > 0$, the mixing times $t^{(n)}_{\text{mix}}(\varepsilon)$ of the projected chain $\{\Pi_n(X_m)\}_{m\geq 0}$ satisfy*

$$t^{(n)}_{\text{mix}}(\varepsilon) \leq K \log n$$

for all sufficiently large n.

Theorem 6.3. *Suppose Σ is a row-column exchangeable probability measure on $\mathcal{S}_k$. Let X be a $\mathrm{CP}_n(\mu_\Sigma)$ chain and let $Y = \Pi_n(X)$ be its projection into $\mathcal{P}_{[n]:k}$. Let $t_X(\varepsilon)$ and $t_Y(\varepsilon)$ denote the ε-mixing times of X and Y respectively. Then*

$$t_X(\varepsilon) = t_Y(\varepsilon).$$

In particular, if $l(\varepsilon, n) \leq t_X(\varepsilon) \leq L(\varepsilon, n)$ are upper and lower bounds on the ε-mixing times of X, then

$$l(\varepsilon, n) \leq t_Y(\varepsilon) \leq L(\varepsilon, n),$$

and vice versa. Moreover, X exhibits the cutoff phenomenon if and only if Y exhibits the cutoff phenomenon.

Proof. If π is the stationary distribution for X, then $\pi\Pi_n^{-1}$ is the stationary distribution of Y. The rest follows by the proceeding discussion regarding sufficiency of $\Pi_n(X)$ and the sufficiency principle (2.2). ☐

Corollary 6.4. *Assume that the paintbox measure Σ is row-column exchangeable and satisfies Hypothesis 4.8, and let $\{X_m\}_{m\geq 0}$ be the EFCP chain on $[k]^{[n]}$ with associated paintbox measure Σ. Then the homogeneous cut-and-paste chain $\Pi_n(X)$ exhibits the cutoff phenomenon at time $\theta \log n$, where $\theta = -1/(2 \log \lambda_1)$.*

References

[1] Jean Bertoin. *Random coagulation and fragmentation processes*. Volume 102 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2006. MR-2253162

[2] P. Bougerol and J. Lacroix. *Products of random matrices with applications to Schrödinger operators*. Birkhäuser, Boston, MA, 1985. MR-0886674

[3] Harry Crane. A consistent Markov partition process generated from the paintbox process. *J. Appl. Probab.*, 43(3):778–791, 2011. MR-2884815

[4] Harry Crane. Homogeneous cut-and-paste processes. *Preprint*, 2012. MR-3054889

[5] Warren J. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3:87–112, 1972. MR-0325177

[6] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960. MR-0121828

[7] I. Ya. Golʹdsheĭd and G. A. Margulis. Lyapunov exponents of a product of random matrices. *Uspekhi Mat. Nauk*, 44(5(269)):13–60, 1989. MR-1040268

[8] M. Golubitsky, E. Keeler, and M. Rothschild. Convergence of the age structure: Applications of the projective metric. *Th. Pop. Biol.*, 7:84–93, 1975. MR-0371433

[9] Samuel Karlin and Howard M. Taylor. *A first course in stochastic processes*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, second edition, 1975. MR-0356197

[10] J. F. C. Kingman. Subadditive ergodic theory. *Ann. Probability*, 1:883–909, 1973. With discussion by D. L. Burkholder, Daryl Daley, H. Kesten, P. Ney, Frank Spitzer and J. M. Hammersley, and a reply by the author. MR-0356192

[11] J. F. C. Kingman. The coalescent. *Stochastic processes and their applications*, 13:235–248, 1982. MR-0671034

[12] D. Levin, Y. Peres, and E. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009. MR-2466937

[13] Karl Petersen. *Ergodic theory*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1983. MR-0833286

[14] Jim Pitman. *Combinatorial stochastic processes*. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Springer, New York, 2005. MR-2245368

[15] Simon Tavaré. Ancestral Inference in Population Genetics. Lectures from the 31st Summer School on Probability Theory held in Saint-Flour, 2001. Springer-Verlag, Berlin, 2004. MR-2071630