The Annals of Applied Probability
2013, Vol. 23, No. 2, 811–833
DOI: 10.1214/12-AAP856

CONVERGENCE ANALYSIS OF SOME MULTIVARIATE MARKOV CHAINS USING STOCHASTIC MONOTONICITY

BY KSHITIJ KHARE AND NABANITA MUKHERJEE

University of Florida and Center for Outcome Research

We provide a nonasymptotic analysis of convergence to stationarity for a collection of Markov chains on multivariate state spaces, from arbitrary starting points, thereby generalizing results in [Khare and Zhou *Ann. Appl. Probab.* **19** (2009) 737–777]. Our examples include the multi-allele Moran model in population genetics and its variants in community ecology, a generalized Ehrenfest urn model and variants of the Pólya urn model. It is shown that all these Markov chains are stochastically monotone with respect to an appropriate partial ordering. Then, using a generalization of the results in [Diaconis, Khare and Saloff-Coste *Sankhya* **72** (2010) 45–76] and [Wilson *Ann. Appl. Probab.* **14** (2004) 274–325] (for univariate totally ordered spaces) to multivariate partially ordered spaces, we obtain explicit nonasymptotic bounds for the distance to stationarity from arbitrary starting points. In previous literature, bounds, if any, were available only from special starting points. The analysis also works for nonreversible Markov chains, and allows us to analyze cases of the multi-allele Moran model not considered in [Khare and Zhou *Ann. Appl. Probab.* **19** (2009) 737–777].

1. Introduction. The theory of Markov chains plays a prominent role in the fields of statistics and applied probability. Markov chains have a wide range of applications in numerous areas from particle transport through finite state machines to the theory of gene expression. Some important applications include modeling scientific phenomena in population genetics, statistical physics and image processing. Another important use is simulating from an intractable probability distribution. It is a well-known fact that, under mild conditions discussed in [1], a Markov chain converges to its stationary distribution. In the applications mentioned above, often it is useful to know exactly how long to run the Markov chain until it reaches sufficiently close to the stationary distribution. Answering this question as accurately as possible, is what obtaining a "nonasymptotic convergence analysis" of Markov chains is all about. The applied probability community has made significant strides in this area in the past three decades. Despite this progress, answering this question still remains a challenging task for various standard Markov chains arising in applied probability and statistics. There are various examples where cur-

Received May 2011; revised February 2012.
MSC2010 subject classifications. Primary 60J10; secondary 60J22, 47H05.
Key words and phrases. Convergence analysis, Markov chains, stochastic monotonicity, partial ordering.

rently available state of the art techniques can give upper bounds that are substantially larger than the correct answer, often by orders of magnitude.

In the current paper, we provide a nonasymptotic analysis of convergence to stationarity for a collection of Markov chains in population genetics. The analysis is based on a generalization of the monotone coupling argument to multivariate state spaces. These Markov chains appear as standard models in population genetics and ecology and include the multi-allele Moran process in population genetics and its variants in community ecology, a generalized Ehrenfest urn model and the Pólya urn process. These Markov chains were analyzed in [9], and the authors provide an exact convergence analysis in terms of the "chi-square distance" by using spectral techniques. But their analysis is somewhat incomplete because it works *only for some natural selected starting points*. Stochastic monotonicity of a Markov chain, along with the knowledge of a monotone eigenfunction (see [3] and [17]), can be used to obtain a nonasymptotic convergence analysis from an arbitrary starting point. Existing results in [3] and [17] require total ordering of the state space, which generally works in the case of univariate state spaces. In multivariate state spaces, however, there often exists a natural partial ordering. We prove that the Markov chains being considered in this paper are stochastically monotone with respect to an appropriate partial ordering; see Theorems 3.1, 3.2, 3.3. *But stochastic monotonicity of a Markov chain with respect to the partial ordering, even with the knowledge of a monotone eigenfunction, is not enough to get desired convergence bounds.* However, an additional condition, satisfied by all the Markov chains under consideration in this paper, enables us to obtain useful convergence bounds; see Theorem 2.1. Another limitation of the spectral techniques used in [9] is that they require reversibility of the Markov chain under consideration. The coupling argument presented in this paper also works for nonreversible Markov chains. Using this, for example, we are able to obtain explicit convergence bounds for generalizations of the standard multi-allele Moran model which are nonreversible.

Another important issue to understand is that out of the three classes of examples considered in this paper, the stationary distribution and the second largest eigenvalue of the Markov chains corresponding to the generalized Ehrenfest urn models and the Pólya urn models are known (*the stationary distribution is unknown for the general multi-allele Moran model*). Hence, for these two models, from a general starting point $\mathbf{x}$, one could potentially consider the crude upper bound $\frac{\lambda^n}{2\sqrt{\pi(\mathbf{x})}}$ for the total variation distance from stationarity after n steps. Here $\pi(\mathbf{x})$ denotes the mass put by the stationary distribution at $\mathbf{x}$, and λ denotes the second largest eigenvalue. However, the upper bounds derived in this paper mostly provide a significant improvement over the crude upper bound. See the remarks in Section 3.2.1 and Section 3.3.1.

Here is an example of our results. The Unified Neutral Theory of Biodiversity and Biogeography (UNTB) is an important theory proposed by ecologist Stephen

Hubbell in his monograph [7] which is used in the study of diversity and species abundances in ecological communities. There are two levels in Hubbell's theory, a metacommunity and a local community.

We concentrate here on the evolution of the local community. The local community has constant population size N with d different species. At each step, one individual is randomly chosen to die and is replaced by a new individual. With probability m, the new individual is chosen randomly from the metacommunity, which has proportion p_i of species i $(i = 1, 2, \ldots, d)$. With probability $1 - m$, the new individual is randomly chosen from the remaining $N - 1$ individuals in the local community. This process is a variant of the so-called multi-allele Moran model in population genetics [5]. The metacommunity evolves at a much larger time scale and is assumed to be fixed during the evolution of the local community.

A very important issue of both practical and theoretical interests is to determine how soon a local community reaches equilibrium (see McGill [11]). Let $K(\cdot, \cdot)$ be the transition density of our local community Markov chain with state space $\mathcal{X}$ and stationary density π. Let $\mathbf{x} \in \mathcal{X}$ be the initial state of the Markov chain. We are interested in answering the following question. For arbitrary $\varepsilon > 0$, how many steps, n, are needed so that the total variation distance between the density of the Markov chain after n steps and the stationary density is less than ε? More precisely, we want to find n such that

$$\| K_{\mathbf{x}}^n - \pi \|_{\mathrm{TV}} = \frac{1}{2} \sum_{\mathcal{X}} |K^n(\mathbf{x}, \mathbf{x}') - \pi(\mathbf{x}')| \leq \varepsilon,$$

where $K_{\mathbf{x}}^n$ denotes the density of the chain started at state $\mathbf{x}$ after n steps.[1]

Khare and Zhou [9] provide an exact answer to this question in terms of the "chi-square distance" by using spectral techniques, *when all individuals belong to the same species to begin with*. So providing any nonasymptotic convergence bounds from an arbitrary starting point was still unresolved. Convergence bounds for the general local community Markov chain are provided in Section 3.1, with an arbitrary starting point. Note that the upper and lower bounds obtained are not exactly matching, but they are within a reasonable range of each other. Considering the fact that no useful analysis was available from an arbitrary starting point, the bounds provided are definitely a significant step forward.

As an illustration, note that under suitable parametrization (see [9]), the local community process by Hubbell is the same as the Pólya down–up model; see Section 3.2. Suppose that the local community has population size $N = 100$ with $d = 5$ species. With probability $m = 0.9$, the new individual is chosen randomly from the meta-community with uniform species frequencies $\mathbf{p} =$

[1]For ease of exposition, if f and g are densities with respect to the counting measure on a finite state space $\mathcal{X}$, $\|f - g\|_{\mathrm{TV}}$ will denote the total variation distance between the probability measures corresponding to f and g.

$(0.2, 0.2, 0.2, 0.2, 0.2)$. Let $\mathbf{X} = (X_1, \ldots, X_d)$ be any (random) count vector of the local community, where X_i is the count of individuals of species i. From Section 3.2, for a starting state $\mathbf{x} = (0, 10, 0, 10, 80)$, the bounds on the total variation distance are obtained as

$$0.375\left(1 - \frac{1}{111}\right)^n \leq \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \leq 100\left(1 - \frac{1}{111}\right)^n. \tag{1.1}$$

For $\varepsilon = 0.01$, (1.1) tells us that at least 401 steps are necessary and at most 1018 steps are sufficient for the total variation distance to be less then 0.01. The crude upper bound for total variation distance is $(2.2186 \times 10^{19})(1 - 1/111)^n$ which gives 5432 steps are sufficient for the total variation distance to be less then 0.01.

The paper is organized in the following way. In Section 2, we provide the necessary background for stochastic monotonicity, and then proceed to prove Theorem 2.1, which generalizes the results in [3] and [17] to multivariate partially ordered finite state spaces to obtain convergence bounds, under appropriate monotonicity assumptions. In Section 3, three classes of Markov chains: multi-allele Moran model, generalized Ehrenfest urn model and generalized Pólya urn model are considered. Each of these Markov chains is shown to be stochastically monotone with respect to an appropriate partial ordering, and also shown to satisfy the other assumptions in Theorem 2.1. All these are combined to provide nonasymptotic convergence bounds for these classes of Markov chains from arbitrary starting points. We conclude the paper with a short discussion in Section 4.

2. Monotone Markov chains.

2.1. *Background.* Let $\mathcal{X}$ be a finite state space with total ordering $\leq$. Let $K(\cdot, \cdot)$ be a Markov kernel on $\mathcal{X}$. We say K is stochastically monotone if for all $x \in \mathcal{X}$ and $x' \in \mathcal{X}$ with $x \leq x'$,

$$\sum_{y \leq y'} K(x, y) \geq \sum_{y \leq y'} K(x', y) \qquad \text{for all } y' \in \mathcal{X}.$$

Monotone Markov chains have been thoroughly studied and applied. See Lund and Tweedie [10], Stoyan [15] and the references therein. They are currently popular because of "coupling from the past." See David Wilson's website on perfect sampling, http://research.microsoft.com/en-us/um/people/dbwilson/exact, for extensive references on this subject.

Alternatively, if the state space $\mathcal{X}$ of a Markov chain is totally ordered (e.g., a subset of $\mathbb{Z}$ and $\mathbb{R}$), then the Markov chain with corresponding transition operator K is stochastically monotone if for every monotone function $f : \mathcal{X} \to \mathbb{R}$, the function Kf is also monotone. There is a standard coupling technique available for monotone Markov chains on totally ordered spaces. Wilson [17] uses this coupling technique in the presence of an explicit eigenfunction to provide general convergence bounds for stochastically monotone Markov chains on totally ordered finite

state spaces. Diaconis, Khare and Saloff-Coste [3] provide extensions for general state spaces and use these results to analyze certain two-component Gibbs samplers.

However, for multivariate state spaces, there is often no natural total ordering, but there exists a natural partial ordering. For example, if $\mathcal{X}$ consists of d-dimensional vectors, then entry-wise domination gives rise to a standard partial ordering. A Markov chain with corresponding transition operator K is monotone with respect to a partial ordering, if whenever $f : \mathcal{X} \to \mathbb{R}$ is monotone with respect to the partial ordering, Kf is monotone with respect to the partial ordering. See Fill and Machida [6], Beskos and Roberts [2], Roberts and Rosenthal [14] and the references therein for varied applications. The literature on perfect sampling mainly consists of various techniques for simulating from specific distributions on partially ordered spaces with a unique minimal and maximal element; see Propp and Wilson [13]. *Note that, unlike perfect sampling, our focus is to analyze given Markov chains corresponding to specific models, and not to devise Markov chains to simulate from a specified distribution.*

The theorem listed below generalizes earlier results in Wilson [17] and Diaconis, Khare and Saloff-Coste [3] (for univariate totally ordered spaces) to multivariate partially ordered spaces in order to obtain nonasymptotic convergence results.

2.2. *Convergence of monotone Markov chains*: *General result*.

THEOREM 2.1. *Let K be the transition density of a Markov chain on a finite state space $\mathcal{X}$ equipped with a partial ordering, $\preceq$. Suppose that K has a stationary distribution with density π, and the following conditions are satisfied*:

(a) *K is monotone with respect to the partial ordering, $\preceq$.*

(b) (*Pair-wise dominance property*) *For an arbitrary* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathcal{X}$, *there exists* $\mathbf{z}(\mathbf{x}, \mathbf{y})$ (*depends possibly on* $\mathbf{x}$ *and* $\mathbf{y}$) *such that* $\mathbf{z}$ *either dominates* $\mathbf{x}$ *and* $\mathbf{y}$ *or is dominated by both* $\mathbf{x}$ *and* $\mathbf{y}$ *with respect to* $\preceq$.

(c) $\lambda \in (0, 1)$ *is an eigenvalue of K with strictly monotone eigenfunction f such that*

$$c_1 = \inf_{\mathbf{x}^* \preceq \mathbf{y}^*, \mathbf{x}^* \neq \mathbf{y}^*} \{f(\mathbf{y}^*) - f(\mathbf{x}^*) | \mathbf{x}^*, \mathbf{y}^* \in \mathcal{X}\} > 0, \qquad c_2 = \sup_{\mathbf{x} \in \mathcal{X}} |f(\mathbf{x})| > 0.$$

Then for any starting state $\mathbf{x}$,

$$\frac{\lambda^n}{2c_2} |f(\mathbf{x})| \leq \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \leq \frac{\lambda^n}{c_1} \mathrm{E}|f(\mathbf{Y}) + f(\mathbf{x}) - 2f(\mathbf{z}(\mathbf{x}, \mathbf{Y}))|,$$

where $\mathbf{Y} \sim \pi$.

PROOF. Let $\mathbf{x}^* \in \mathcal{X}$ and $\mathbf{y}^* \in \mathcal{X}$ satisfy $\mathbf{x}^* \preceq \mathbf{y}^*$. It is well known that if a probability distribution μ on $\mathcal{X}$ is stochastically dominated by another probability distribution ν on $\mathcal{X}$, that is, $\int f \, d\mu \leq \int f \, d\nu$ for every monotone function f, then

we can construct random variables $\mathbf{X}$ and $\mathbf{Y}$ such that $\mathbf{X} \sim \mu$, $\mathbf{Y} \sim \nu$ and $\mathbf{X} \preceq \mathbf{Y}$; see for example [8]. Since K is monotone with respect to the partial ordering, $\preceq$, by repeated application of this result, we can construct two coupled Markov chains, $\{\mathbf{X}_n\}_{n\geq 0}$ and $\{\mathbf{Y}_n\}_{n\geq 0}$ such that $\mathbf{X}_0 = \mathbf{x}^*$, $\mathbf{Y}_0 = \mathbf{y}^*$ and $\mathbf{X}_n \preceq \mathbf{Y}_n$ for every $n \geq 1$. Further, if $\mathbf{X}_{n_0} = \mathbf{Y}_{n_0}$, then $\mathbf{X}_n = \mathbf{Y}_n$ for all $n \geq n_0$.

It follows that for any $n \geq 1$,

$$\begin{aligned}\|K^n_{\mathbf{x}^*} - K^n_{\mathbf{y}^*}\|_{\mathrm{TV}} &\leq \mathrm{P}(\mathbf{X}_n \neq \mathbf{Y}_n | \mathbf{X}_0 = \mathbf{x}^*, \mathbf{Y}_0 = \mathbf{y}^*) \\ &\leq \mathrm{E}\left\{ \frac{f(\mathbf{Y}_n) - f(\mathbf{X}_n)}{c_1} \Big| \mathbf{X}_0 = \mathbf{x}^*, \mathbf{Y}_0 = \mathbf{y}^* \right\}.\end{aligned}$$

The previous inequality uses $\mathbf{X}_n \preceq \mathbf{Y}_n$, the strict monotonicity of f and the hypothesis that $f(\mathbf{y}) - f(\mathbf{x}) \geq c_1$ if $\mathbf{x} \preceq \mathbf{y}, \mathbf{x} \neq \mathbf{y}$.

Next, since f is an eigenfunction of K, it follows that

$$\mathrm{E}\{f(\mathbf{Y}_k) - f(\mathbf{X}_k) | \mathbf{X}_{k-1}, \mathbf{Y}_{k-1}\} = \lambda\{f(\mathbf{Y}_{k-1}) - f(\mathbf{X}_{k-1})\},$$

for every $k \geq 1$. Therefore,

$$\begin{aligned}\|K^n_{\mathbf{x}^*} - K^n_{\mathbf{y}^*}\|_{\mathrm{TV}} &\leq \mathrm{E}\left[\mathrm{E}\left\{ \frac{f(\mathbf{Y}_n) - f(\mathbf{X}_n)}{c_1} \Big| \mathbf{X}_{n-1}, \mathbf{Y}_{n-1} \right\} \Big| \mathbf{X}_0 = \mathbf{x}^*, \mathbf{Y}_0 = \mathbf{y}^* \right] \\ &= \frac{\lambda}{c_1} \mathrm{E}\{f(\mathbf{Y}_{n-1}) - f(\mathbf{X}_{n-1}) | \mathbf{X}_0 = \mathbf{x}^*, \mathbf{Y}_0 = \mathbf{y}^*\} \\ &= \frac{\lambda^n}{c_1}\{f(\mathbf{y}^*) - f(\mathbf{x}^*)\}.\end{aligned}$$

Note that the argument above holds for any $\mathbf{x}^* \preceq \mathbf{y}^*$.

Note that for any $\mathbf{x} \neq \mathbf{y}$, by the pair-wise dominance assumption, there exists $\mathbf{z}(\mathbf{x}, \mathbf{y})$ (depends possibly on $\mathbf{x}$ and $\mathbf{y}$) such that $\mathbf{z}$ dominates both $\mathbf{x}$ and $\mathbf{y}$ or is dominated by both $\mathbf{x}$ and $\mathbf{y}$. Hence,

$$\begin{aligned}\|K^n_{\mathbf{x}} - K^n_{\mathbf{y}}\|_{\mathrm{TV}} &\leq \|K^n_{\mathbf{x}} - K^n_{\mathbf{z}}\|_{\mathrm{TV}} + \|K^n_{\mathbf{y}} - K^n_{\mathbf{z}}\|_{\mathrm{TV}} \\ &\leq \frac{\lambda^n}{c_1} |f(\mathbf{x}) - f(\mathbf{z})| + \frac{\lambda^n}{c_1} |f(\mathbf{y}) - f(\mathbf{z})| \\ &= \frac{\lambda^n}{c_1} |f(\mathbf{x}) + f(\mathbf{y}) - 2f(\mathbf{z})|.\end{aligned}$$

The previous equality follows from the fact that $\mathbf{z}$ either dominates or is dominated by both $\mathbf{x}$ and $\mathbf{y}$, and f is monotone with respect to $\preceq$, which implies that $f(\mathbf{x}) - f(\mathbf{z})$ and $f(\mathbf{y}) - f(\mathbf{z})$ are either both positive or both negative. Convexity now yields

$$\|K^n_{\mathbf{x}} - \pi\|_{\mathrm{TV}} \leq \sum_{\mathbf{y}\in\mathcal{X}} \pi(\mathbf{y}) \|K^n_{\mathbf{x}} - K^n_{\mathbf{y}}\|_{\mathrm{TV}} \leq \frac{\lambda^n}{c_1} \mathrm{E}_\pi |f(\mathbf{x}) + f(\mathbf{Y}) - 2f(\mathbf{z}(\mathbf{x}, \mathbf{Y}))|.$$

To get the lower bound, note that

$$\|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \geq \frac{1}{2c_2}|\mathrm{E}_{K_{\mathbf{x}}^n}(f(\mathbf{Y})) - \mathrm{E}_\pi(f(\mathbf{Y}))| \geq \frac{\lambda^n}{2c_2}|f(\mathbf{x})|.$$

Hence the theorem is proved. □

REMARK. (1) It is to be noted that Theorem 2.1 works for any arbitrary starting point without requiring the assumption of reversibility. In Section 3.1, we show that the bounds on the total variation distance can be obtained without explicit knowledge of the stationary distribution.

(2) In all our examples, there will be a unique minimal element (and no maximal element), which is clearly sufficient to satisfy the *pair-wise dominance condition*.

We now apply this general result for a variety of Markov chains in population genetics.

3. Applications.

3.1. *The Moran process in population genetics*. The classical Moran process in population genetics models the evolution of a population of constant size by random replacement followed by mutation. Suppose there are d species in a population of size N. At each step, one individual is chosen uniformly to die and independently another is chosen uniformly to reproduce. *They may be the same individual*. If the latter is of species i, the offspring has probability m_{ij}, $1 \leq j \leq d$, to mutate to type j. Let $\mathbf{X}_n = (X_{n1}, \ldots, X_{nd})$ be the vector of counts of species $1, 2, \ldots, d$ at the nth step. Let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Then $\{\mathbf{X}_n\}_{n\geq 0}$ forms a Markov chain on $\mathcal{X}_N^d$, where

$$\mathcal{X}_N^d = \left\{\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{N}_0^d : \sum_{i=1}^d x_i = N\right\}.$$

Let K denote the transition density of this Markov chain. Note that the size of the state space is $|\mathcal{X}_N^d| = \binom{N+d-1}{N}$. The one-step transition probabilities are

$$K(\mathbf{x}, \mathbf{x} + \mathbf{e}_i - \mathbf{e}_j) = \frac{x_j}{N}\left(\sum_{k=1}^d \frac{x_k}{N} m_{ki}\right), \qquad 1 \leq i \neq j \leq d;$$

$$K(\mathbf{x}, \mathbf{x}) = 1 - \sum_{i\neq j} K(\mathbf{x}, \mathbf{x} + \mathbf{e}_i - \mathbf{e}_j); \tag{3.1}$$

$$K(\mathbf{x}, \mathbf{y}) = 0 \qquad \text{otherwise},$$

where $\mathbf{e}_i$ is the unit vector with ith entry equal to 1. *The mutation matrix* $\mathbf{M}$ *is assumed to be irreducible*. This ensures the irreducibility and aperiodicity of the

transition function K; see proof in the Appendix. Hence, the stationary distribution of K exists. Let π denote the density of the stationary distribution with respect to the counting measure.

This model ($d = 2$) is due to Moran [12]. Background and references can be found in the text by Ewens [5]. When $d = 2$, in the continuous-time setting, Donnelly and Rodrigues [4] obtain an upper bound in terms of the separation and total variation distances, when all the individuals belong to the same generation initially. Watkins [16] analyzes the infinite allele Moran model in the discrete-time setting. However, unlike the multi-allele case, the (infinite) vector of species counts does not form a Markov chain. Instead, the N-dimensional vector whose ith entry is the number of species with i individuals at the current stage, forms a Markov chain. It is this fundamentally different Markov chain that is analyzed in Watkins [16] using strong stationary times.

In the multi-allele case, which we analyze, a standard choice of the mutation matrix $\mathbf{M} = \{m_{ij}\}_{1 \le i,j \le d}$ is

$$\mathbf{M} = (1 - m)\mathbf{I} + m\mathbf{P}, \tag{3.2}$$

where $0 < m \le 1$ is the mutation probability of the offspring, and $\mathbf{P}$ is a stochastic matrix with each row $(p_1, \ldots, p_d)$, a probability vector with positive entries. If mutation happens, the offspring will change to species i with probability p_i. It is known from the literature that for this standard choice of the mutation matrix $\mathbf{M}$, the corresponding Markov chain is reversible. Khare and Zhou [9] analyze this Markov chain and provide nonasymptotic convergence bounds in terms of the "chi-square distance" for some natural selected starting points. *In this paper, we generalize this analysis in two directions. First, instead of considering the choice* $\mathbf{M} = (1 - m)\mathbf{I} + m\mathbf{P}$*, we consider a general subclass of mutation matrices described in* (3.3)–(3.5) *which includes this choice as a special case. Second, we provide nonasymptotic convergence bounds from an arbitrary starting point.* Consider the class of mutation matrices $\mathbf{M}$ satisfying one of the monotonicity conditions specified below:

$$m_{dj} < \min_{1 \le k \le d-1} m_{kj} \qquad \text{for every } 1 \le j \le d-1 \tag{3.3}$$

or,

$$m_{dj} \le \min_{1 \le k \le d-1} m_{kj} \qquad \text{for every } 1 \le j \le d-1 \tag{3.4}$$

and $\mathbf{M}^* = \{m^*_{ij}\}_{1 \le i,j \le d-1}$ is irreducible, where $m^*_{ij} = m_{ij} - m_{dj}$ or,

$$m_{dj} \le \min_{1 \le k \le d-1} m_{kj} \qquad \text{for every } 1 \le j \le d-1 \tag{3.5}$$

and $\mathbf{M}^*$ has an eigenvector which has all strictly positive entries.

Each of these conditions essentially says that there is a species, which we call species d without loss of generality, such that the mutation probability from this

species to any species is smaller than the mutation probability from every other species to this species.

It is to be noted that for a general $\mathbf{M}$ satisfying any one of these three conditions, the Markov kernel K is nonreversible, and in this case, often the stationary distribution of K is not known. Note that condition (3.5) is satisfied by the standard choice of $\mathbf{M} = (1-m)\mathbf{I} + m\mathbf{P}$, and hence the analysis of this standard choice will come out as a special case. An example where conditions (3.3) and (3.4) are satisfied would be the following: Suppose $m_{d1} = \delta$ and $m_{dd} = 1-\delta$, that is, the offspring born to species d can possibly mutate only to species 1 with a small probability δ. Suppose $m_{1d} > 0$, that is, species 1 can also mutate to species d with a positive probability. If all the mutation probabilities among species $1, 2, \ldots, d-1$ are larger than δ, that is, $m_{ij} > \delta$ for $1 \le i, j \le d-1$, then conditions (3.3) and (3.4) are satisfied.

Let us introduce a partial ordering on $\mathcal{X}_N^d$. We define $\mathbf{x}, \mathbf{y} \in \mathcal{X}_N^d$ to be partially ordered, that is, $\mathbf{x} \preceq \mathbf{y}$ if $x_i \le y_i, i = 1, 2, \ldots, d-1$. This automatically implies $x_d \ge y_d$. To get bounds on the total variation distance, according to Theorem 2.1, we need an eigenfunction f which is strictly monotone in $\preceq$, that is, if $\mathbf{x}, \mathbf{y} \in \mathcal{X}_N^d$ with $\mathbf{x} \preceq \mathbf{y}$, then $f(\mathbf{x}) \le f(\mathbf{y})$.

PROPOSITION 3.1. *Let K denote the Moran process specified by* (3.1), *and suppose the mutation matrix $\mathbf{M}$ satisfies any one of conditions* (3.3)–(3.5). *Then K has a linear and strictly monotone eigenfunction f.*

PROOF. Note that

$$\begin{aligned}
\mathrm{E}_{K(\mathbf{x},\cdot)}[\mathbf{X}] &= \sum_{1\le i\ne j\le d} (\mathbf{x} + \mathbf{e}_i - \mathbf{e}_j)\frac{x_j}{N}\left(\sum_{k=1}^{d}\frac{x_k}{N}m_{ki}\right) \\
&\quad + \mathbf{x}\left(1 - \sum_{1\le i\ne j\le d}\frac{x_j}{N}\sum_{k=1}^{d}\frac{x_k}{N}m_{ki}\right) \\
&= \mathbf{x} + \sum_{1\le i\ne j\le d} (\mathbf{e}_i - \mathbf{e}_j)\frac{x_j}{N}\left(\sum_{k=1}^{d}\frac{x_k}{N}m_{ki}\right) \\
&= \mathbf{x} + \sum_{1\le i,j\le d} (\mathbf{e}_i - \mathbf{e}_j)\frac{x_j}{N}\left(\sum_{k=1}^{d}\frac{x_k}{N}m_{ki}\right) \\
&= \mathbf{x} + \sum_{1\le i\le d} \mathbf{e}_i\left(\sum_{k=1}^{d}\frac{x_k}{N}m_{ki}\right) - \sum_{1\le i,j\le d} \mathbf{e}_j\frac{x_j}{N}\left(\sum_{k=1}^{d}\frac{x_k}{N}m_{ki}\right) \\
&= \left\{\left(1-\frac{1}{N}\right)\mathbf{I}_d + \frac{1}{N}\mathbf{M}^T\right\}\mathbf{x}.
\end{aligned}$$

Let $\tilde{\mathbf{a}} = (\tilde{a}_i)_{1\le i\le d}$ be any eigenvector corresponding to an eigenvalue $\tilde{\lambda}$ of $\mathbf{M}$. Then we have

$$\mathrm{E}_{K(\mathbf{x},\cdot)}[\tilde{\mathbf{a}}^T\mathbf{X}] = \left\{\left(1-\frac{1}{N}\right)\tilde{\mathbf{a}}^T + \frac{1}{N}(\mathbf{M}\tilde{\mathbf{a}})^T\right\}\mathbf{x} = \left\{\left(1-\frac{1}{N}\right) + \frac{1}{N}\tilde{\lambda}\right\}\tilde{\mathbf{a}}^T\mathbf{x}.$$

Hence, $f(\mathbf{x}) = \sum_{i=1}^{d}\tilde{a}_i x_i$ is an eigenfunction of K corresponding to the eigenvalue $(1-\frac{1}{N}) + \frac{\tilde{\lambda}}{N}$.

We now show that $\mathbf{M}$ has an eigenvector $\mathbf{a}$ such that $a_i > a_d$ for every $1 \le i \le d-1$. It follows from condition (3.3) that $m^*_{ij} > 0$, and from condition (3.4) that $m^*_{ij} \ge 0$, and $\mathbf{M}^*$ is irreducible. Hence, under condition (3.3) or (3.4), by the Perron–Frobenius theorem, the largest eigenvalue λ^* of $\mathbf{M}^*$ is positive with multiplicity 1, and there exists an eigenvector $\mathbf{a}^* = (a^*_j)_{1\le j\le d-1}$ corresponding to λ^*, such that $\mathbf{a}^*$ has all positive entries. Also, in condition (3.5), we have directly assumed $\mathbf{a}^*$ has all positive entries. Note that

$$\begin{aligned}
\lambda^* &\le \max_{1\le i\le d-1}\sum_{j=1}^{d-1} m^*_{ij} \\
&= \max_{1\le i\le d-1}\sum_{j=1}^{d-1}(m_{ij} - m_{dj}) \\
&= \max_{1\le i\le d-1}(m_{dd} - m_{id}) \\
&\le m_{dd} \\
&< 1,
\end{aligned}$$

since the mutation matrix $\mathbf{M}$ is assumed to be irreducible.

Let c be defined by

$$c = \frac{\sum_{j=1}^{d-1} m_{dj}a^*_j}{\lambda^* - 1},$$

and $\mathbf{a}$ be defined by

$$a_i = \begin{cases} a^*_i + c, & \text{if } 1 \le i \le d-1, \\ c, & \text{if } i = d. \end{cases}$$

Note that, by the definition of c,

$$\sum_{j=1}^{d} m_{dj}a_j = \sum_{j=1}^{d-1} m_{dj}a^*_j + c = (\lambda^* - 1)c + c = \lambda^* c. \tag{3.6}$$

We have

$$\mathbf{M}^*\mathbf{a}^* = \lambda^*\mathbf{a}^* \quad\Longrightarrow\quad \sum_{j=1}^{d-1}(m_{ij} - m_{dj})(a_j - c) = \lambda^*(a_i - c) \qquad \forall 1 \le i \le d-1.$$

Note that $\sum_{j=1}^{d-1}(m_{ij}-m_{dj}) = m_{dd} - m_{id}$ and $a_d = c$. It follows that

$$\sum_{j=1}^{d}(m_{ij}-m_{dj})a_j = \lambda^*(a_i - c). \tag{3.7}$$

Adding (3.6) and (3.7), we get $\mathbf{Ma} = \lambda^*\mathbf{a}$. This shows $\mathbf{a}$ is an eigenvector of $\mathbf{M}$ corresponding to eigenvalue λ^*.

Thus, $f(\mathbf{x}) = \sum_{i=1}^{d} a_i x_i = \sum_{i=1}^{d-1}(a_i - a_d)x_i + Na_d$, which is strictly monotone with respect to $\preceq$, is an eigenfunction of K corresponding to the eigenvalue $\lambda = (1-\frac{1}{N}) + \frac{\lambda^*}{N}$. Since $\lambda < 1$, it follows that $\mathrm{E}_\pi[f(\mathbf{X})] = 0$. □

We now show that for the Moran process, K is monotone with respect to the partial ordering, $\preceq$.

THEOREM 3.1. *Let K denote the Moran process specified by* (3.1), *where the mutation matrix $\mathbf{M}$ satisfies one of the conditions specified in* (3.3)–(3.5). *Then K is monotone with respect to the partial ordering, $\preceq$.*

PROOF. Consider any $\mathbf{x} \in \mathcal{X}_N^d$ and $\mathbf{y} \in \mathcal{X}_N^d$ with $\mathbf{x} \preceq \mathbf{y}$. We construct two random vectors $\mathbf{X}$ and $\mathbf{Y}$ such that $\mathbf{X} \preceq \mathbf{Y}$ with $\mathbf{X} \sim K(\mathbf{x}, \cdot)$ and $\mathbf{Y} \sim K(\mathbf{y}, \cdot)$. This will immediately imply that $Kf(\mathbf{x}) \le Kf(\mathbf{y})$ for any monotone function f and any $\mathbf{x}, \mathbf{y}$ with $\mathbf{x} \preceq \mathbf{y}$.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_d)$. Then by assumption $x_i \le y_i$ for every $1 \le i \le d-1$. We now describe the procedure for obtaining $\mathbf{X}$ and $\mathbf{Y}$.

In order to specify the coupling argument, consider two populations with N individuals each. Population 1 has x_i individuals of species i, and population 2 has y_i individuals of species i, for every $1 \le i \le d$. We label the individuals in the two populations as follows. The individuals of the ith species of population 1 are labeled from $(\sum_{j=1}^{i} x_{j-1} + 1)$ to $\sum_{j=1}^{i} x_j$, $i = 1, 2, \ldots, d$, taking $x_0 = 0$. The labeling of the individuals of population 2 is done in the following way:

- Note that $x_i \le y_i$ for $i = 1, 2, \ldots, d-1$. For the ith species of population 2, where $i = 1, 2, \ldots, d-1$, we give x_i of the individuals the exact same labels as those in species i of population 1. This leaves $y_i - x_i$ "extra individuals" to be labeled later.
- Note that $x_d \ge y_d$. For the dth species of population 2, the y_d individuals of the dth species get exactly same labels as the first y_d individuals of the dth species of population 1.
- Finally, all the $x_d - y_d$ "extra individuals" left over in the first $d-1$ species of population 2 get the $x_d - y_d$ labels in the dth species of population 1 which were not assigned in the previous step.

The following example illustrates the labeling technique of the N individuals in population 1 and population 2. Consider $N = 17$ individuals who belong to $d = 4$

TABLE 1
Labeling of individuals of population 1 *and population* 2

Species	**1st**	**2nd**	**3rd**	**4th**
x	/	/////	///////	////
labels	1	2, 3, 4, 5, 6	7, 8, 9, 10, 11, 12, 13	14, 15, **16**, **17**
y	//	/////	///////	//
labels	1, **16**	2, 3, 4, 5, 6	7, 8, 9, 10, 11, 12, 13, **17**	14, 15

different species type. Also consider $\mathbf{x} = \{1, 5, 7, 4\}$ and $\mathbf{y} = \{2, 5, 8, 2\}$. The table below illustrates the labeling technique.

In Table 1, we label the individuals of population 1 from 1 to 17 based on **x**. For the 1st species of population 2, there are 2 individuals, the first individual gets the label 1, same as the label of the first individual of population 1, and the second individual is an "extra individual," to be labeled later. Now, for the 2nd species, there are the same number of individuals for both the populations, so these individuals get the same labels. For the 3rd species, there is one "extra individual," to be labeled later; other individuals get the same labels. The 4th species has 2 individuals in population 2, who get the same labels as the first 2 individuals of the 4th species in population 1. Last, 2 extra labels 16 and 17 are assigned to the "extra individuals" of species 1 and 3 of population 2, respectively.

Let us return to the general proof, and define $k_1 := \sum_{i=1}^{d-1} x_i$ to be the total number of individuals in the first $d-1$ species of population 1 and $k_2 := y_d$ to be the number of individuals in species d of population 2. We now change the species configuration of population 1 and population 2 in four sub-steps which are described below:

(I) Choose a label uniformly between 1 to N. Call it i_1.

(II) Independently choose another label uniformly between 1 to N. Call it i_2.

(III) Let s_{1,i_2} and s_{2,i_2} denote the species of the individual labeled i_2 in population 1 and population 2, respectively. Add one individual of species s_{1,i_2} to population 1 and one individual of species s_{2,i_2} to population 2.

Note that if $1 \le i_2 \le k_1 + k_2$, then $s_{1,i_2} = s_{2,i_2} := s_{i_2}$. In this case the newly added individual in both the populations mutates in the following way: Generate $U \sim \text{Uniform}[0, 1]$. If $0 \le U < m_{s_{i_2}1}$, the added individual mutates to species 1. If $m_{s_{i_2}1} \le U < m_{s_{i_2}1} + m_{s_{i_2}2}$, the added individual mutates to species 2, and so on. Finally, if $m_{s_{i_2}1} + m_{s_{i_2}2} + \cdots + m_{s_{i_2}(d-1)} \le U \le 1$, the added individual mutates to species d. Hence, after the mutation, both populations have an individual of the same species added, which therefore preserves the partial ordering between their species configurations.

Next, suppose $k_1 + k_2 + 1 \le i_2 \le N$, then $s_{1,i_2} = d$ and s_{2,i_2} is one of the first $d-1$ species. Note that $m_{s_{2,i_2}j} \ge m_{dj}$ for every $j = 1, 2, \ldots, d-1$. The

newly added individual in population 1 mutates in the following way: Generate $U \sim \text{Uniform}[0,1]$. If $0 \le U < m_{d1}$, the added individual mutates to species 1. If $m_{d1} \le U < m_{d1} + m_{d2}$, the added individual mutates to species 2, and so on. Finally, if $m_{d1} + m_{d2} + \cdots + m_{d(d-1)} \le U \le 1$, the added individual mutates to species d. Now, in population 2, the newly added individual mutates in the following way: Choose the same U as for population 1. If $0 \le U < m_{d1}$ or $m_{d1} + m_{d2} + \cdots + m_{d(d-1)} \le U < m_{s_{2,i_2}1} + m_{d2} + \cdots + m_{d(d-1)}$, the individual mutates to species 1. If $m_{d1} \le U < m_{d1} + m_{d2}$ or $m_{s_{2,i_2}1} + m_{d2} + \cdots + m_{d(d-1)} \le U < m_{s_{2,i_2}1} + m_{s_{2,i_2}2} + m_{d3} + \cdots + m_{d(d-1)}$, the individual mutates to species 2, and so on. Finally, if $m_{s_{2,i_2}1} + m_{s_{2,i_2}2} + \cdots + m_{s_{2,i_2}(d-1)} \le U \le 1$, the individual mutates to species d. Hence, when $0 \le U \le m_{d1} + m_{d2} + \cdots + m_{d(d-1)}$ or when $m_{s_{2,i_2}1} + m_{s_{2,i_2}2} + \cdots + m_{s_{2,i_2}(d-1)} \le U \le 1$, the newly added individual in both the populations mutate to the same species, which preserves the partial ordering between their species configurations. Alternatively, if $m_{d1} + m_{d2} + \cdots + m_{d(d-1)} \le U \le m_{s_{2,i_2}1} + m_{s_{2,i_2}2} + \cdots + m_{s_{2,i_2}(d-1)}$, then after mutation the newly added individual in the population 1 is in species d, but the newly added individual in the population 2 is in any of the first $d-1$ species. This again preserves the partial ordering between the species configurations in population 1 and population 2.

(IV) Finally, the individual corresponding to the label i_1 dies for both the populations. If $1 \le i_1 \le k_1 + k_2$, then the individual belongs to the same species for both the populations. If $k_1 + k_2 + 1 \le i_1 \le N$, then the individual corresponding to the label i_1 belongs to species d for population 1 and is an "extra individual" in the first $d-1$ species of population 2. In either case, the partial ordering is preserved.

Let $\mathbf{X}$ and $\mathbf{Y}$ be the resulting species configurations of population 1 and population 2, respectively. Note that marginally the movement from both $\mathbf{x}$ to $\mathbf{X}$ and $\mathbf{y}$ to $\mathbf{Y}$ follows the transition mechanism of K, and $\mathbf{X} \preceq \mathbf{Y}$. This completes the proof. □

3.1.1. *Bounds on total variation distance.* For the partial ordering, $\preceq$, discussed above, applying Theorem 2.1 in the case of the Moran model, provides us with bounds on the total variation distance. We have shown that for the Moran process, K is monotone with respect to the partial ordering, $\preceq$; see Theorem 3.1. It is easily seen that $\mathbf{0}$ (with first $d-1$ entries equal to zero, and the dth entry equal to N) is dominated by $\mathbf{x}$ for every $\mathbf{x} \in \mathcal{X}_N^d$. Hence, the pair-wise dominance property is satisfied. Recall that by Proposition 3.1, there exists an eigenfunction $f(\mathbf{x}) = \sum_{i=1}^d a_i x_i = \sum_{i=1}^{d-1} (a_i - a_d) x_i + N a_d$ of K corresponding to the eigenvalue $\lambda = 1 - \frac{1}{N} + \frac{\lambda^*}{N}$, such that f is strictly monotone with respect to the partial ordering, $\preceq$. Hence, the conditions of Theorem 2.1 are satisfied, and the bounds on

the total variation distance are obtained as

$$\frac{\lambda^n}{2c_2}|f(\mathbf{x})| \le \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \le \frac{\lambda^n}{c_1}\mathrm{E}_\pi\{f(\mathbf{Y}) + f(\mathbf{x}) - 2f(\mathbf{0})\}$$

$$\implies \quad \frac{\lambda^n}{2c_2}|f(\mathbf{x})| \le \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \le \frac{\lambda^n}{c_1}\{f(\mathbf{x}) - 2f(\mathbf{0})\}$$

$$\implies \quad \frac{\lambda^n}{2c_2}\left|\sum_{i=1}^{d-1} a_i^* x_i + N a_d\right| \le \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \le \frac{\lambda^n}{c_1}\left\{\sum_{i=1}^{d-1} a_i^* x_i - N a_d\right\},$$

where $a_i^* = a_i - a_d > 0$ for every $1 \le i \le d-1$ (by the monotonicity of f), $c_1 = \min_{1\le i\le d-1} a_i^* > 0$ and $c_2 = \max\{-Na_d, N(\max_{1\le i\le d-1} a_i^* + a_d)\}$. Note that $a_d < 0$. Note again that the stationary distribution π in not known in general, but the analysis above leads to upper and lower bounds which do not depend on the stationary distribution, and are reasonably close to each other.

3.1.2. *Bounds on total variation distance in the special case.* We now provide a nonasymptotic convergence analysis for the special choice of $\mathbf{M} = (1-m)\mathbf{I} + m\mathbf{P}$. It has been proved earlier in Khare and Zhou [9] that the Markov chain K corresponding to the multi-allele Moran model with $\mathbf{M} = (1-m)\mathbf{I} + m\mathbf{P}$ has second largest eigenvalue $\lambda = 1 - \frac{|\boldsymbol{\alpha}|}{N(N+|\boldsymbol{\alpha}|)}$, where $|\boldsymbol{\alpha}| := \sum_{i=1}^d \alpha_i$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, where $\alpha_i = \frac{Nmp_i}{1-m}$, with the eigenspace given by the space of centered linear functions of $x_1, x_2, \ldots, x_{d-1}$. After simplification, we obtain $\lambda = 1 - \frac{m}{N}$. It is known that the stationary distribution in this case is the Dirichlet-multinomial distribution with parameters N and $\boldsymbol{\alpha}$. The Dirichlet-multinomial distribution with parameters $N > 0$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, $\alpha_i > 0$, has probability mass function given by

$$\mathcal{DM}(\mathbf{x}|N, \boldsymbol{\alpha}) = \frac{\prod_{i=1}^d \binom{x_i+\alpha_i-1}{x_i}}{\binom{N+|\boldsymbol{\alpha}|-1}{N}}, \qquad \mathbf{x} \in \mathcal{X}_N^d.$$

Since $\mathbf{M} = (1-m)\mathbf{I} + m\mathbf{P}$, it follows that $\mathbf{M}^* = (1-m)\mathbf{I}_{d-1}$. Hence, any $(d-1)$-dimensional vector with positive entries is an eigenvector of $\mathbf{M}^*$. Suppose we choose the eigenvector $\mathbf{a}^*$ of $\mathbf{M}^*$ such that $a_i^* = 1$ for $i < d$. Then, for the Markov chain K, we get the eigenfunction $f(\mathbf{x}) = \sum_{i=1}^{d-1} x_i - N(1-p_d)$ corresponding to the eigenvalue $\lambda = 1 - \frac{m}{N}$. Note that f is strictly monotone with respect to the partial ordering, $\preceq$. As in the case of the general multi-allele Moran model, here also it is easily seen that $\mathbf{0}$ is dominated by $\mathbf{x}$ for every $\mathbf{x} \in \mathcal{X}_N^d$. We have $c_1 = 1$ and $c_2 = \max\{Np_d, N(1-p_d)\}$. Thus, bounds on total variation distance are obtained as

$$\|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \ge \frac{(1-m/N)^n}{2\max\{Np_d, N(1-p_d)\}}\left|\sum_{i=1}^{d-1} x_i - N(1-p_d)\right|, \tag{3.8}$$

$$\|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \le \left(1 - \frac{m}{N}\right)^n \left(\sum_{i=1}^{d-1} x_i + N(1 - p_d)\right). \tag{3.9}$$

EXAMPLE 3.1.1. Consider the multi-allele Moran model in the special case when the mutation matrix $\mathbf{M} = (1 - m)\mathbf{I} + m\mathbf{P}$. Suppose the population size $N = 100$, with $d = 5$ species and mutation probability $m = 0.7$. When mutation occurs, the individual mutates to the ith species with probability $p_i = 1/5$. Using (3.8) and (3.8), for a starting state $\mathbf{x} = (0, 10, 0, 10, 80)$, the bounds on the total variation distance are obtained as

$$0.375\left(1 - \frac{7}{1000}\right)^n \le \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \le 100\left(1 - \frac{7}{1000}\right)^n. \tag{3.10}$$

For $\varepsilon = 0.01$, (3.10) tells us that 516 steps are necessary and 1312 steps are sufficient for the total variation distance to be less then 0.01. The crude upper bound for the total variation distance is $(2.1665 \times 10^{15})(1 - \frac{7}{1000})^n$, which gives 5683 steps are sufficient for the total variation distance to be less then 0.01.

3.2. *Sequential Pólya urn models.* Choose d urns with N balls distributed in them. Suppose the inherent weight of urn i is α_i, $i = 1, 2, \ldots, d$, and let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ denote the vector of urn weights and $|\boldsymbol{\alpha}| = \sum_{i=1}^d \alpha_i$ denote the total inherent weight of d urns. Suppose that each ball has unit weight.

(1) Pólya level model [9]: Consider the Markov chain whose one-step movement consists of the following sub-steps:

(i) Randomly choose s balls out of N balls and mark them.
(ii) Draw an urn with probability proportional to its weight (inherent weight + weight of balls) and add a ball (of unit weight) to the chosen urn. Repeat this s times.
(iii) Remove the s marked balls from the respective urns.

(2) Pólya up–down model [9]: These are variations of Pólya level models, where the three steps are performed in the following order (ii), (i) (with $N + s$ total balls) and (iii).

(3) Pólya down–up model [9]: These are variations of Pólya level models, where the three steps are performed in the following order (i), (iii) and (ii).

We first analyze the Markov chain corresponding to the Pólya level model. Let X_{ni} denote the number of balls in the ith urn at the nth step of the Pólya level model. Then $\{\mathbf{X}_n = (X_{n1}, X_{n2}, \ldots, X_{nd}), n = 0, 1, 2, \ldots\}$ forms a multivariate Markov chain on $\mathcal{X}_N^d$. Let K denote the transition density of this Markov chain. Let $\preceq$ be the partial ordering on $\mathcal{X}_N^d$ as in the multi-allele Moran model.

THEOREM 3.2. *K is monotone with respect to the partial ordering, $\preceq$.*

PROOF. Consider any $\mathbf{x} \in \mathcal{X}_N^d$ and $\mathbf{y} \in \mathcal{X}_N^d$ with $\mathbf{x} \preceq \mathbf{y}$. We construct two random vectors $\mathbf{X}$ and $\mathbf{Y}$ such that $\mathbf{X} \preceq \mathbf{Y}$ with $\mathbf{X} \sim K(\mathbf{x}, \cdot)$ and $\mathbf{Y} \sim K(\mathbf{y}, \cdot)$. This will immediately imply that $Kf(\mathbf{x}) \leq Kf(\mathbf{y})$ for any monotone function f and any $\mathbf{x}, \mathbf{y}$ with $\mathbf{x} \preceq \mathbf{y}$.

In order to specify the coupling argument, we consider two populations of N balls each, with N balls distributed in d urns based on $\mathbf{x}$ and $\mathbf{y}$, respectively. We use the same labeling technique for both the populations as discussed in Theorem 3.1 (regarding species as urns and individuals as balls).

We now change the urn configuration of population 1 and population 2 in three sub-steps which are described below:

(I) Choose s labels without replacement from 1 to N.

(II) This sub-step will consist of s sequential urn draws, and after each draw, an extra ball will be added to the chosen urn for both the populations as described below. Repeat the following for $j = 1, 2, \ldots, s$.

Generate $U_j \sim \text{Uniform}[0, 1]$. Now, at the beginning of the jth draw in this sub-step, there are, in total, $N + j - 1$ balls each in both the populations. Hence the total weight of the urns (with balls) in both the populations is $|\boldsymbol{\alpha}| + N + j - 1$. Let $\mathbf{X}^{j-1} := (x_1^{j-1}, x_2^{j-1}, \ldots, x_d^{j-1})$ be the configuration of the balls in the d urns of population 1 at the beginning of the jth draw, and $\mathbf{Y}^{j-1} := (y_1^{j-1}, y_2^{j-1}, \ldots, y_d^{j-1})$ be the configuration of the balls in the d urns of population 2 at the beginning of the jth draw. Let us denote the normalized probability vector of the urn weights for population 1 by $\mathbf{p}^{j-1} = (p_1^{j-1}, p_2^{j-1}, \ldots, p_d^{j-1})$, and the normalized probability vector of urn weights for population 2 by $\mathbf{q}^{j-1} = (q_1^{j-1}, q_2^{j-1}, \ldots, q_d^{j-1})$, where $p_i^{j-1} = \frac{\alpha_i + x_i^{j-1}}{|\boldsymbol{\alpha}| + N + j - 1}$ and $q_i^{j-1} = \frac{\alpha_i + y_i^{j-1}}{|\boldsymbol{\alpha}| + N + j - 1}$.

Procedure to choose an urn for population 1 at the jth draw:

If $0 \leq U_j < p_1^{j-1}$, choose urn 1. If $p_1^{j-1} \leq U_j < p_1^{j-1} + p_2^{j-1}$, choose urn 2 and so on. Finally, if $p_1^{j-1} + p_2^{j-1} + \cdots + p_{(d-1)}^{j-1} \leq U_j \leq 1$, choose urn d. Add a ball to the chosen urn.

The following is the procedure to choose urn for population 2 at the jth draw:

If $0 \leq U_j < p_1^{j-1}$ or, $p_1^{j-1} + p_2^{j-1} + \cdots + p_{(d-1)}^{j-1} \leq U_j < q_1^{j-1} + p_2^{j-1} + \cdots + p_{(d-1)}^{j-1}$, choose urn 1. If $p_1^{j-1} \leq U_j < p_1^{j-1} + p_2^{j-1}$ or, $q_1^{j-1} + p_2^{j-1} + \cdots + p_{(d-1)}^{j-1} \leq U_j < q_1^{j-1} + q_2^{j-1} + \cdots + p_{(d-1)}^{j-1}$, choose urn 2 and so on. Finally, if $q_1^{j-1} + q_2^{j-1} + \cdots + q_{(d-1)}^{j-1} \leq U_j \leq 1$, choose urn d. Add a ball to the chosen urn.

(III) Remove the balls corresponding to the s labels in sub-step (I) from both the populations.

It is to be noted that in sub-step (II), assuming $\mathbf{X}^{j-1} \preceq \mathbf{Y}^{j-1}$ (and hence $\mathbf{p}^{j-1} \preceq \mathbf{q}^{j-1}$), the mechanism for drawing urns is such that either the same urn is chosen for both the populations or when the dth urn is chosen for population 1, then any

of the first $d-1$ urns is chosen for population 2. Hence, $\mathbf{X}^j \preceq \mathbf{Y}^j$ (and hence $\mathbf{p}^j \preceq \mathbf{q}^j$). Since $\mathbf{X}^0 = \mathbf{x}$ and $\mathbf{Y}^0 = \mathbf{y}$, it follows by induction (on j) that $\mathbf{X}^j \preceq \mathbf{Y}^j$ for $j = 1, 2, \ldots, s$. In sub-step (III), the balls with the same s labels are removed from both the populations. Based on the labeling procedure, either balls with the same label lie in the same urn for both the populations, or the ball lies in the dth urn for population 1 and is an "extra ball" in the first $d-1$ urns for population 2. In either case, removing balls with the same label from both the populations does not change the partial ordering of the urn configurations.

Let $\mathbf{X}$ and $\mathbf{Y}$ be the resulting urn configurations of population 1 and population 2, respectively. It follows from the discussion above that $\mathbf{X} \preceq \mathbf{Y}$. Note that marginally the movement from both $\mathbf{x}$ to $\mathbf{X}$ and $\mathbf{y}$ to $\mathbf{Y}$ follows the transition mechanism of K. To see this, note that the probability of choosing the ith urn at the jth draw in sub-step (II) for population 1 is $\mathrm{P}(\sum_{\ell=1}^{i-1} p_\ell^j \le U_j \le \sum_{\ell=1}^{i} p_\ell^j) = p_i^j$; and the corresponding probability for population 2 is $\mathrm{P}(\sum_{\ell=1}^{i-1} p_\ell^j \le U_j \le \sum_{\ell=1}^{i} p_\ell^j) + \mathrm{P}(\sum_{\ell=1}^{i-1} q_\ell^j + \sum_{\ell=i}^{d-1} p_\ell^j \le U_j \le \sum_{\ell=1}^{i} q_\ell^j + \sum_{\ell=i+1}^{d-1} p_\ell^j) = q_i^j$. This completes the proof. □

We can similarly argue that the Markov chain corresponding to the Pólya up–down model and the Pólya down–up model are stochastically monotone with respect to the partial ordering, $\preceq$ in $\mathcal{X}_N^d$.

3.2.1. *Bounds on total variation distance.* In case of the Pólya level model, the second largest eigenvalue $\lambda = 1 - \frac{s|\boldsymbol{\alpha}|}{N(N+|\boldsymbol{\alpha}|)}$. We know that the stationary distribution of the Pólya level model is the Dirichlet-multinomial distribution with parameters N and $\boldsymbol{\alpha}$. The eigenfunction $f(\mathbf{x}) = \sum_{i=1}^{d-1} x_i - N(1 - \frac{\alpha_d}{|\boldsymbol{\alpha}|})$ corresponding to λ is strictly monotone in $\preceq$. Let $\mathbf{0}$ be a d-dimensional vector such that the first $d-1$ entires are zero, and the dth entry is N. It is easily seen that $\mathbf{0}$ is dominated by $\mathbf{x}$ for every $\mathbf{x} \in \mathcal{X}_N^d$. Hence, the conditions of Theorem 2.1 are satisfied, with $c_1 = 1$ and $c_2 = \max\{N\frac{\alpha_d}{|\boldsymbol{\alpha}|}, N(1 - \frac{\alpha_d}{|\boldsymbol{\alpha}|})\}$. Let $p_i = \frac{\alpha_i}{|\boldsymbol{\alpha}|}, i = 1, 2, \ldots, d$. Thus, the bounds on the total variation distance are obtained as

$$\|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \ge \frac{\lambda^n}{2\max\{Np_d, N(1-p_d)\}} \left| \sum_{i=1}^{d-1} x_i - N(1-p_d) \right|, \tag{3.11}$$

$$\|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \le \lambda^n \left(\sum_{i=1}^{d-1} x_i + N(1-p_d) \right). \tag{3.12}$$

Similarly, in the case of Pólya down–up models, the second largest eigenvalue is given by $\lambda = (1 - \frac{s}{N})(1 - \frac{s}{N+|\boldsymbol{\alpha}|})^{-1}$ and in the case of Pólya up–down models, the second largest eigenvalue is given by $\lambda = (1 + \frac{s}{N})^{-1}(1 + \frac{s}{N+|\boldsymbol{\alpha}|})$. These can be substituted in (3.11) and (3.12) to get the corresponding total variation bounds for these models.

REMARK. Note that the coefficient of λ^n in the upper bound derived in (3.12) is at most $2N$. Let us try and compare it to the coefficient of λ^n in the crude upper bound, which is given by

$$\frac{1}{2\sqrt{\pi(\mathbf{x})}} = \frac{1}{2}\sqrt{\frac{\binom{N+|\boldsymbol{\alpha}|-1}{N}}{\prod_{i=1}^{d}\binom{x_i+\alpha_i-1}{x_i}}}.$$

At one possible extreme, when all entries of $\mathbf{x}$ except the ith one are zero, the coefficient is essentially a polynomial in N of degree $\frac{|\boldsymbol{\alpha}|-\alpha_i}{2}$. At the other possible extreme, when all the entries of $\mathbf{x}$ are equal to $\frac{N}{d}$ (assuming $\frac{N}{d}$ is an integer), the coefficient is essentially a polynomial in N of degree $\frac{d-1}{2}$. The main fact is that the coefficient of λ^n in the upper bound derived in (3.12) is linear in N, whereas the coefficient of λ^n in the crude upper bound almost always behaves like a polynomial of a higher degree in N.

EXAMPLE 3.2.1. Consider the Pólya level model where $N = 100$ balls are distributed in $d = 5$ urns. Suppose $s = 2$ balls are chosen and each urn has inherent weight $\alpha_i = 180$ for every $1 \le i \le 5$. Using (3.11) and (3.12), for a starting state $\mathbf{x} = (0, 20, 0, 20, 60)$, the bounds on the total variation distance are obtained as

$$0.25\left(1 - \frac{9}{500}\right)^n \le \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \le 120\left(1 - \frac{9}{500}\right)^n. \tag{3.13}$$

For $\varepsilon = 0.01$, (3.13) tells us that 178 steps are necessary and 518 steps are sufficient for the total variation distance to be less than 0.01. The crude upper bound for total variation distance is $(6.1094 \times 10^{13})(1 - \frac{9}{500})^n$ which would have implied 2002 steps are sufficient for the total variation distance to be less then 0.01.

3.3. *A generalized Ehrenfest urn model.* There are N indistinguishable balls to be distributed to d urns. At each step, s balls are chosen at random from the total of N balls, and each of them is redistributed independently according to the same probability $\mathbf{p} = (p_1, p_2, \ldots, p_d)$. Let X_{ni} be the number of balls in the ith urn at the nth step of the Markov chain. Then $\{\mathbf{X}_n = (X_{n1}, X_{n2}, \ldots, X_{nd}), n = 0, 1, 2, \ldots\}$ forms a multivariate Markov chain on $\mathcal{X}_N^d$. Let K denote the transition density of this Markov chain.

Consider the same partial ordering, $\preceq$, as defined in the case of the Moran process. We now show that K is a monotone Markov chain with respect to the partial ordering, $\preceq$.

THEOREM 3.3. *K is monotone with respect to the partial ordering, $\preceq$.*

PROOF. Consider any $\mathbf{x} \in \mathcal{X}_N^d$ and $\mathbf{y} \in \mathcal{X}_N^d$ with $\mathbf{x} \preceq \mathbf{y}$. We construct two random vectors $\mathbf{X}$ and $\mathbf{Y}$ such that $\mathbf{X} \preceq \mathbf{Y}$ with $\mathbf{X} \sim K(\mathbf{x}, \cdot)$ and $\mathbf{Y} \sim K(\mathbf{y}, \cdot)$. This will

immediately imply that $Kf(\mathbf{x}) \leq Kf(\mathbf{y})$ for any monotone function f and any $\mathbf{x}, \mathbf{y}$ with $\mathbf{x} \preceq \mathbf{y}$.

In order to specify the coupling argument, we consider two populations of N balls each, with N balls distributed in d urns based on $\mathbf{x}$ and $\mathbf{y}$, respectively. We use the same labeling technique for both the populations as discussed in Theorem 3.1 (regarding species as urns and individuals as balls).

We now change the urn configuration of population 1 and population 2 in five sub-steps which are described below:

(I) Choose s labels without replacement from 1 to N.
(II) Remove the balls with the chosen labels from both $\mathbf{x}$ and $\mathbf{y}$.
(III) Choose an urn, such that urn i is chosen with probability p_i for every $i = 1, 2, \ldots, d$.
(IV) Add a ball to the chosen urn for both the current $\mathbf{X}$ and $\mathbf{Y}$ configurations.
(V) Repeat steps (III) and (IV) s times independently.

Let $k_1 := \sum_{i=1}^{d-1} x_i$ be the total number of balls in the first $d-1$ urns of population 1, and $k_2 := y_d$ be the number of balls in the dth urn of population 2. Consider sub-steps (I) and (II). Without loss of generality, let us assume out of s labels chosen, r labels are between 1 and $k_1 + k_2$ and $s - r$ labels are between $k_1 + k_2 + 1$ and N.

- Each of the r balls corresponding to labels 1 to $k_1 + k_2$ lies in exactly the same urn for both the populations. Removing these does not change the partial ordering between the urn configurations.
- Each of the $s - r$ balls corresponding to labels $k_1 + k_2 + 1$ to N lie in urn d for population 1, and are "extra balls" lying in the first $d - 1$ urns for population 2. Hence, removing them does not change the partial ordering between the urn configurations of population 1 and population 2.

Consider sub-steps (III), (IV) and (V). Since the balls are put in the same urn for both the populations, adding the new balls does not change the partial ordering between the urn configurations.

Let $\mathbf{X}$ and $\mathbf{Y}$ be the resulting urn configurations of population 1 and population 2, respectively. Note that marginally the movement from both $\mathbf{x}$ to $\mathbf{X}$ and $\mathbf{y}$ to $\mathbf{Y}$ follows the transition mechanism of K, and $\mathbf{X} \preceq \mathbf{Y}$. This completes the proof. □

The following example illustrates the one-step movement of the above construction in population 1 and population 2 for Theorem 3.3.

EXAMPLE 3.3.1. Consider the same $\mathbf{x}$ and $\mathbf{y}$ as in Table 1. Suppose the 4 balls chosen in sub-step (I) are with labels 6, 8, 14 and 16. It is evident that the removal of the balls with the chosen labels in sub-step (II) does not alter the partial ordering

between the urn configurations of the two populations. Since the urn chosen in sub-step (III) is same for both the populations, adding a ball to the urn in sub-step (IV) does not change the partial ordering between the urn configurations of the two populations.

3.3.1. *Bounds on total variation distance.* For the partial ordering, $\preceq$, discussed above, applying Theorem 2.1 in the case of the generalized Ehrenfest urn model, provides us with bounds on the total variation distance.

It has been proved earlier in Khare and Zhou [9] that the generalized Ehrenfest urn model has second largest eigenvalue $\lambda = 1 - \frac{s}{N}$, with the eigenspace given by the space of linear functions of $x_1, x_2, \ldots, x_{d-1}$. It is known that the stationary distribution is the multinomial distribution with parameters N and $\mathbf{p}$. The eigenfunction $f(\mathbf{x}) = p_d \sum_{i=1}^{d-1} x_i - (1-p_d)x_d = \sum_{i=1}^{d-1} x_i - N(1-p_d)$ corresponding to the eigenvalue λ is strictly monotone in $\preceq$. Again, it is easily seen that $\mathbf{0}$ is dominated by $\mathbf{x}$, for every $\mathbf{x} \in \mathcal{X}_N^d$. Hence, the conditions of Theorem 2.1 are satisfied. We have $c_1 = 1$ and $c_2 = \max\{Np_d, N(1-p_d)\}$. Thus, the bounds on total variation distance are

$$\|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \geq \frac{(1-s/N)^n}{2\max\{Np_d, N(1-p_d)\}} \left| \sum_{i=1}^{d-1} x_i - N(1-p_d) \right|, \tag{3.14}$$

$$\|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \leq \left(1 - \frac{s}{N}\right)^n \left(\sum_{i=1}^{d-1} x_i + N(1-p_d) \right). \tag{3.15}$$

REMARK. Note that the coefficient of $(1 - \frac{s}{N})^n$ in the upper bound derived in (3.15) is at most $2N$. We compare it to the coefficient of $(1 - \frac{s}{N})^n$ in the crude upper bound, which is given by

$$\frac{1}{2\sqrt{\pi(\mathbf{x})}} = \frac{1}{2\sqrt{\binom{N}{\mathbf{x}}}} \prod_{i=1}^{d} \left(\frac{1}{\sqrt{p_i}} \right)^{x_i}.$$

At one possible extreme, when all entries of $\mathbf{x}$ except the ith one are zero, the coefficient is $\frac{1}{2}(\frac{1}{\sqrt{p_i}})^N$. At the other possible extreme, when all the entries of $\mathbf{x}$ are equal to $\frac{N}{d}$ (assuming $\frac{N}{d}$ is an integer), using Stirling's approximation for large N,[2] the coefficient is

$$\frac{(2\pi N)^{(d-1)/4}}{2d^{d/4}} \left(\frac{1}{\sqrt{d(\prod_{i=1}^d p_i)^{1/d}}} \right)^N.$$

Since $\sum_{i=1}^d p_i = 1$, it follows by the AM-GM inequality that $d(\prod_{i=1}^d p_i)^{1/d} < 1$, unless all the entries of $\mathbf{p}$ are equal. Hence, if all the entries of $\mathbf{x}$ are the same and

[2]Note that N is the notation for the total number of balls in the urns, not the number of steps.

all entries of p_i are not the same, the coefficient of $(1 - \frac{s}{N})^n$ in the crude upper bound is exponential in N. If all the p_i are same, the coefficient is of the order $N^{(d-1)/4}$.

The main fact is that the coefficient of $(1 - \frac{s}{N})^n$ in the upper bound derived in (3.15) is linear in N, whereas the coefficient of $(1 - \frac{s}{N})^n$ in the crude upper bound is almost always exponential in N.

EXAMPLE 3.3.2. Consider the generalized Ehrenfest urn model where $N = 100$ balls are distributed in $d = 5$ urns. Suppose $s = 1$ ball is chosen and each urn is chosen with probability $p_i = 1/5, i = 1, 2, \ldots, 5$. Using (3.14) and (3.15), for a starting state $\mathbf{x} = (0, 20, 0, 20, 60)$, the bounds on the total variation distance are obtained as

$$0.25\left(1 - \frac{1}{100}\right)^n \leq \|K_{\mathbf{x}}^n - \pi\|_{\mathrm{TV}} \leq 120\left(1 - \frac{1}{100}\right)^n. \tag{3.16}$$

For $\varepsilon = 0.01$, (3.16) tells us that 321 steps are necessary and 935 steps are sufficient for the total variation distance to be less then 0.01. The crude upper bound for total variation distance $\frac{1}{2\sqrt{\pi(\mathbf{x})}}(1 - 1/100)^n = (1.02 \times 10^{15})(1 - 1/100)^n$ would have implied that 3897 steps are sufficient for the total variation distance to be less then 0.01.

4. Discussion. We use a probabilistic technique based on a monotone coupling argument for analyzing all the examples in this paper. We obtain reasonable upper and lower bounds for the total variation distance for any arbitrary starting point of the Markov chain, significantly broadening previous results in [9]. This analysis is very simple to implement, requiring the knowledge of a single eigenfunction and its corresponding eigenvalue. In addition, the analysis does not require the assumption of reversibility. As an illustration, we provide the nonreversible Moran model in Section 3.1. The next goal is to sharpen the bounds to obtain matching upper and lower bounds, and to generalize the techniques developed in this paper for continuous state spaces.

APPENDIX

LEMMA 1. *If the mutation matrix* $\mathbf{M}$ *is irreducible, then the transition density* K *in* (3.1) *is irreducible and aperiodic.*

PROOF. We first show irreducibility. Let $\mathbf{x} \in \mathcal{X}_N^d$ be arbitrarily chosen. Let $i \neq j$ be such that $1 \leq i, j \leq d$ and $x_i > 0$. By the irreducibility of $\mathbf{M}$, there exists $n \in \mathbb{N}$ such that $(\mathbf{M}^n)_{ij} > 0$. As a result, there exist $i = k_0, k_1, k_2, \ldots, k_{n-1}, k_n = j$ such that $\prod_{l=0}^{n-1} m_{k_l k_{l+1}} > 0$. Let $\mathbf{x}^0 = \mathbf{x}$, and $\mathbf{x}^l = \mathbf{x}^{l-1} + \mathbf{e}_{k_l} - \mathbf{e}_{k_{l-1}}$ for $1 \leq l \leq n$.

Note that by construction, $x^l_{k_l} > 0$, which implies $\mathbf{x}^l \in \mathcal{X}^d_N$ for every $1 \le l \le n$. Hence,

$$\begin{aligned} K^n(\mathbf{x}, \mathbf{x} + \mathbf{e}_j - \mathbf{e}_i) &= K^n(\mathbf{x}^0, \mathbf{x}^n) \\ &\ge \prod_{l=0}^{n-1} K(\mathbf{x}^l, \mathbf{x}^{l+1}) \\ &\ge \prod_{l=0}^{n-1} \frac{x^l_{k_l}}{N} \frac{x^l_{k_l}}{N} m_{k_l k_{l+1}} \\ &> 0. \end{aligned}$$

We have thus shown that if $\mathbf{x}$ and $\mathbf{y}$ are neighbors in $\mathcal{X}^d_N$, that is, if $\mathbf{y}$ can be obtained from $\mathbf{x}$ by removing an individual in one species and adding an individual in another, then there exists $n \in \mathbb{N}$ such that $K^n(\mathbf{x}, \mathbf{y}) > 0$. Since any two elements of $\mathcal{X}^d_N$ are connected by a path such that successive elements in the path are neighbors, it follows that K is irreducible.

We now show aperiodicity. Since $\mathbf{M}$ is irreducible, there exist i, j such that $1 \le i \ne j \le d$ and $m_{ij} > 0$. If $\mathbf{x} \in \mathcal{X}^d_N$ is such that $x_i, x_j > 0$, then

$$K(\mathbf{x}, \mathbf{x}) \ge \frac{x_j}{N} \frac{x_i}{N} m_{ij} > 0.$$

Since K is irreducible, and there exists at least one $\mathbf{x} \in \mathcal{X}^d_N$ such that $K(\mathbf{x}, \mathbf{x}) > 0$, it follows that K is aperiodic. □

REFERENCES

[1] ATHREYA, K. B., DOSS, H. and SETHURAMAN, J. (1996). On the convergence of the Markov chain simulation method. *Ann. Statist.* **24** 69–100. MR1389881

[2] BESKOS, A. and ROBERTS, G. O. (2005). One-shop CFTP; application to a class of truncated Gaussian densities. *Methodol. Comput. Appl. Probab.* **7** 407–437. MR2235153

[3] DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2010). Gibbs sampling, conjugate priors and coupling. *Sankhya A* **72** 136–169. MR2658168

[4] DONNELLY, P. and RODRIGUES, E. R. (2000). Convergence to stationarity in the Moran model. *J. Appl. Probab.* **37** 705–717. MR1782447

[5] EWENS, W. J. (2004). *Mathematical Population Genetics. I: Theoretical Introduction*, 2nd ed. *Interdisciplinary Applied Mathematics* **27**. Springer, New York. MR2026891

[6] FILL, J. A. and MACHIDA, M. (2001). Stochastic monotonicity and realizable monotonicity. *Ann. Probab.* **29** 938–978. MR1849183

[7] HUBBELL, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography. Monographs in Population Biology* **32**. Princeton Univ. Press, Princeton, NJ.

[8] KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912. MR0494447

[9] KHARE, K. and ZHOU, H. (2009). Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *Ann. Appl. Probab.* **19** 737–777. MR2521887

[10] LUND, R. B. and TWEEDIE, R. L. (1996). Geometric convergence rates for stochastically ordered Markov chains. *Math. Oper. Res.* **21** 182–194. MR1385873

[11] MCGILL, B. J. (2003). A test of the unified neutral theory of biodiversity. *Nature* **422** 881–885.
[12] MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54** 60–71. MR0127989
[13] PROPP, J. G. and WILSON, D. B. (1998). How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *J. Algorithms* **27** 170–217. MR1622393
[14] ROBERTS, G. O. and ROSENTHAL, J. S. (1999). Convergence of slice sampler Markov chains. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 643–660. MR1707866
[15] STOYAN, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*. Wiley, Chichester. MR0754339
[16] WATKINS, J. C. (2010). Convergence time to the Ewens sampling formula in the infinite alleles Moran model. *J. Math. Biol.* **60** 189–206. MR2552725
[17] WILSON, D. B. (2004). Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.* **14** 274–325. MR2023023

DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611
USA
E-MAIL: kdkhare@stat.ufl.edu

CENTER FOR OUTCOME RESEARCH
THE CHILDREN'S HOSPITAL OF PHILADELPHIA
3535 MARKET STREET, SUITE 1005
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: mukherjeen@email.chop.edu