Exchangeable Markov Processes on $[k]^N$ with Cadlag Sample Paths

Harry Crane · Steven P. Lalley

Received: 24 November 2013 / Revised: 16 June 2014 / Published online: 4 July 2014
© Springer Science+Business Media New York 2014

Abstract Any exchangeable, time-homogeneous Markov processes on $[k]^N$ with cadlag sample paths projects to a Markov process on the simplex whose sample paths are cadlag and of locally bounded variation. Furthermore, any such process has a de Finetti-type description as a mixture of independent, identically distributed copies of time-inhomogeneous Markov processes on $[k]$. In the Feller case, these time-inhomogeneous Markov processes have a relatively simple structure; however, in the non-Feller case, a greater variety of behaviors is possible since the transition law of the underlying Markov process on $[k]^N$ can depend in a nontrivial way on its exchangeable $\sigma$-algebra.

Keywords Exchangeable random partition · de Finetti’s theorem · Hewitt–Savage theorem · Paintbox process · Interacting particle system

Mathematics Subject Classification (2010) Primary 60J25 · 60G09

Harry Crane supported by NSF grant DMS-1308899 and NSA grant H98230-13-1-0299. Steven P. Lalley supported by NSF grant DMS-1106669.

H. Crane
Department of Statistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854, USA

S. P. Lalley
Department of Statistics, University of Chicago, Eckhart Hall, 5734 S. University Ave, Chicago, IL 60637, USA
e-mail: lalley@galton.uchicago.edu

Springer
1 Introduction

We study exchangeable, time-homogeneous Markov processes with cadlag sample paths on $[k]^\mathbb{N}$, the space of $k$-colorings of the natural numbers $\mathbb{N}$. See Sect. 2.1 for a formal definition. The state space is endowed with the product-discrete topology, i.e., the topology of point-wise convergence, under which it is compact and metrizable. In previous work, Crane [5] characterized the class of exchangeable, continuous-time Markov processes having the Feller property on $[k]^\mathbb{N}$ with respect to the product-discrete topology; see Sect. 7 for a résumé of Crane’s results and a discussion of their relation to the main results of this paper. Exchangeable Markov processes with the Feller property are precisely those exchangeable Markov processes that are consistent under subsampling, that is, for each $n \in \mathbb{N}$, the projection onto the first $n$ coordinates is itself Markov. Conversely, by Kolmogorov’s extension theorem ([4], Chapter 7, Section 37), any consistent family of exchangeable Markov processes on the finite state spaces $[k]^n$ extends naturally to an exchangeable Markov process on $[k]^\mathbb{N}$ with the Feller property. Another well-known family of exchangeable partition-valued Feller processes is the class of exchangeable fragmentation-coalescence (EFC) processes, which have been characterized by Berestycki [1], cf. Bertoin [2] and Pitman [12]. An EFC process can be expressed as a superposition of independent, exchangeable coalescent and fragmentation processes. The classical example of an exchangeable partition-valued process, the exchangeable coalescent [10], was originally introduced in connection to certain population genetics models. Processes on $[k]^\mathbb{N}$ also arise somewhat naturally in genetics, as discussed in [5].

In this paper, we characterize exchangeable Markov processes on $[k]^\mathbb{N}$ that do not necessarily satisfy the Feller property, but do have cadlag sample paths. The characterization we give is similar to that in [5], but, because the processes are not assumed to possess the Feller property, we cannot draw general conclusions about the infinitesimal behavior of these processes. In particular, the infinitesimal generator of a non-Feller process need not exist, and therefore, in general, no characterization in terms of infinitesimal jump rates can be given. In Sect. 2.2, we show that non-Feller exchangeable Markov processes on $[k]^\mathbb{N}$ arise naturally as limits of various interesting mean-field models, including the mean-field stochastic Ising model (cf., e.g., [11]) and the Reed-Frost stochastic epidemic (cf., e.g., [6]). These examples show that non-Feller exchangeable Markov processes are by no means pathological; in many circumstances, the non-Feller case is the one of greatest practical interest.

By viewing elements of $\mathbb{N}$ as distinct particles, one can then regard processes on $[k]^\mathbb{N}$ as interacting particle systems in which each particle is in one of $k \geq 1$ internal states at every time. The assumption that such a process has cadlag paths is equivalent to the assumption that, with probability one, each particle spends a strictly positive amount of time in each internal state it visits. When the Feller property fails, the evolution of any subset of particles can depend on the configuration of the entire collection $\mathbb{N}$, but for exchangeable processes, only via the exchangeable $\sigma$-algebra.

The paper is organized as follows. In Sect. 2, we formally define exchangeable Markov processes, present some illustrative examples, and state the main results of the paper. In Sect. 3, we discuss properties of the joint behavior of processes observed at an at most countable collection of times, which includes discrete-time Markov chains.
In Sects. 4–6, we discuss various properties of these processes in continuous time, and prove the main Theorems 2.8, 2.13, and 2.14. Finally, in Sect. 7, we indicate briefly how the representation of Theorem 2.13 specializes when the underlying Markov process on $[k]^\mathbb{N}$ is Feller.

**Notation.** Elements of the state space $[k]^\mathbb{N}$, $[k] := \{1, \ldots, k\}$, will be denoted by $x = (x_n)_{n \in \mathbb{N}}$, with superscripts indicating coordinates. Paths in $[k]^\mathbb{N}$ will be denoted by $(x_t)_{t \in J}$, where $J$ is a subset of $[0, \infty)$; in particular, subscripts will be used to denote time. When convenient, we use the abbreviation $x_J$ for the path $(x_t)_{t \in J}$, or similarly $x^i_J$ for the $i$th coordinate path $(x^i_t)_{t \in J}$. A path $(x_t)_{t \in J}$ in $[k]^\mathbb{N}$ is said to be *cadlag* if each coordinate path $(x^i_t)_{t \in J}$ is piecewise constant and right-continuous.

## 2 Main Results

### 2.1 Exchangeability

Let $S$ be a finite or countable set of sites. A *permutation* of $S$ is a bijection $\sigma : S \rightarrow S$ that fixes all but finitely many elements of $S$. Any permutation of $S$ induces a permutation of $[k]^S$ by relabeling, that is, for any element $x = (x^i)_{i \in S}$ of $[k]^S$, the image $x^\sigma = \sigma x$ is defined by

$$(\sigma x)^i = x^{\sigma(i)}, \quad i \in S.$$ 

We extend this notation to measurable subsets $A \subseteq [k]^S$ by writing $A^\sigma = \sigma A := \{\sigma x : x \in A\}$. We call a $[k]^S$-valued random variable $X$ *exchangeable* (or say that $X$ has an *exchangeable distribution*) if, for every permutation $\sigma$, the distribution of $\sigma X$ coincides with the distribution of $X$.

**Definition 2.1** A Markov process $(X_t)_{t \geq 0}$ in $[k]^S$ (either in discrete or continuous time) is said to be *exchangeable* if

(a) its initial state $X_0$ is exchangeable and

(b) its transition kernel $p_t(x, dy)$ is invariant under permutations $\sigma$ of $S$; that is, for every measurable subset $A \subseteq [k]^S$,

$$p_t(\sigma x, \sigma A) = p_t(x, A) \quad \text{for all } t \geq 0. \quad (1)$$

Recall that a transition kernel $p_t(x, dy)$ has the *Feller property* if, for every bounded, continuous, real-valued function $f$ on the state space, the function

$$T_t f(x) := \int f(y) \ p_t(x, dy)$$

is jointly continuous in $(t, x)$. This property does not involve the initial distribution of the Markov process. Our main results provide a description of Markov processes with cadlag sample paths whose transition kernels are exchangeable but may fail to satisfy the Feller property; however, our description will only apply to those Markov
processes \((X_t)_{t \geq 0}\) whose initial states \(X_0\) are distributed according to an exchangeable probability measure on \([k]^N\).

Except in Sect. 2.2, we will be concerned exclusively with the case \(S = \mathbb{N}\). In this case, for any finite subset or interval \(A \subseteq [0, \infty)\), we denote by \(\mathcal{E}_A\) the exchangeable \(\sigma\)-algebra for the sequence \(X_{\mathcal{A}} := (X^1_{\mathcal{A}}, X^2_{\mathcal{A}}, \ldots)\), and we abbreviate \(\mathcal{E} = \mathcal{E}_{[0, \infty)}\). In particular, \(\mathcal{E}_A\) consists of all events \(E\) that are both measurable with respect to \(X_{\mathcal{A}}\) and satisfy \(E = E^\sigma\), for all permutations \(\sigma : \mathbb{N} \rightarrow \mathbb{N}\) fixing all but finitely many \(n \in \mathbb{N}\).

2.2 Example: Mean-Field Models

If \((X_t)_{t \geq 0}\) is an exchangeable Markov process on a finite configuration space \([k]^S\), then it projects naturally to a Markov process on the \(k\)-simplex \(\Delta_k\), via the mapping

\[
Y^i_t = \frac{1}{|S|} \sum_{s \in S} 1\{X^s_t = i\}, \quad i \in [k].
\]

Moreover, the transition rates for the Markov process \((X_t)_{t \geq 0}\) depend only on the current state of the projection \(Y_t\).

Example 2.2 A mean-field stochastic Ising model (also known as Glauber dynamics) is an exchangeable Markov process on the finite state space \([-1, +1]^n\) whose infinitesimal transition rates are determined by a Hamiltonian function \(H : [-1, +1]^n \rightarrow \mathbb{R}\) that is invariant under permutations \(\sigma : [n] \rightarrow [n]\), that is, \(H \circ \sigma = H\). Only one spin may flip at a time, and for any two configurations \(x, x' \in [-1, +1]^n\) that differ in only one coordinate, the infinitesimal jump rate from \(x\) to \(x'\) is given by

\[
P(X_{t+dt} = x' \mid X_t = x) = e^{-\beta H(x')} \, dt,
\]

where \(\beta > 0\) is the inverse temperature parameter. Since \(H\) is assumed to be invariant under permutations, the Markov process with transition rates (3) is exchangeable. In the case of greatest physical interest, the Hamiltonian function \(H\) is determined by (i) interactions between the spins and an external field and (ii) pair interactions between spins. Moreover, in the mean-field model, all pair interactions are the same.

Example 2.3 The Reed-Frost epidemic model is a simple mean-field model of an epidemic in which members of a population \([n]\) can be either susceptible, infected, or recovered (henceforth abbreviated \(S, I, R\), respectively). Each susceptible individual becomes infected at rate proportional to the total number infected; each infected individual recovers at a rate \(\varrho\) not depending on the conditions of the other individuals; and recovered individuals are granted permanent immunity from further infection and remain recovered forever. Thus, for \(\beta, \varrho > 0\), for each individual \(i \in [n]\),

\[
P(X_{t+dt}^i = I \mid X_t^i = S \text{ and } N_t^i = m) = \beta m \, dt \quad \text{and}
\]

\[
P(X_{t+dt}^i = R \mid X_t^i = I \text{ and } N_t^i = m, N_t^S = p) = \varrho \, dt,
\]

where
\[ N_j^i = \sum_{i=1}^{n} 1\{X_t^i = j\} \text{ for } j = S, I, R. \]

In both of these examples, the set \([n]\) of sites (or population members) can be of arbitrary size, and in both cases the transition rates extend to continuous functions on the simplex \(\Delta_k\). It is natural in circumstances such as these to inquire about large-\(n\) limits. Clearly, such limits will exist only if the transition rates are properly scaled in order that the flip rates at individual sites remain bounded and nonzero as \(n \to \infty\). Therefore, we shall restrict attention to families of exchangeable Markov processes on configuration spaces \([k]^{[n]}\), with \(k\) fixed, whose transition rates are obtained from a set of common (to all \(n\)) continuous functions \(f_{i,j} : \Delta_k \to \mathbb{R}_+\) on the \(k\)-simplex.

**Definition 2.4** A mean-field model with \(k\) types is a family of exchangeable Markov processes \(X_t = (X_t^i)_{i \in [n]}\) on the finite configuration spaces \([k]^{[n]}\), one for each \(n \in \mathbb{N}\), whose transition rates are related in the following way: for each ordered pair \((i, j) \in [k] \times [k], i \neq j\), there exists a continuous, nonnegative function \(f_{i,j} : \Delta_k \to \mathbb{R}_+\) on the \(k\)-simplex \(\Delta_k\) such that for all \(n \in \mathbb{N}\) and for each site \(s \in [n]\),

\[ P(X_{s+t} = j | X_t) = f_{i,j}(Y_t) \, dt \text{ on the event } X_t^s = i \quad (5) \]

where \(Y_t\) is the projection of \(X_t\) to \(\Delta_k\), defined by (2).

**Proposition 2.5** Consider a mean-field model with transition rates satisfying (5), where the functions \(f_{i,j}\) are nonnegative and Lipschitz continuous on \(\Delta_k\). Assume that the initial states \(X_0\) project to points \(Y_0 \in \Delta_k\) that converge as \(n \to \infty\) to some nonrandom point \(y_0 \in \Delta_k\). Then, as \(n \to \infty\), the Markov processes \(X_t\) on the finite configuration spaces \([k]^{[n]}\) converge in distribution to an exchangeable Markov process \(\tilde{X}_t\) with cadlag sample paths on the infinite configuration space \([k]^{\mathbb{N}}\). The limit process is Feller if and only if the rate functions \(f_{i,j}\) are constant on \(\Delta_k\).

The weak convergence of the Markov processes \(X_t\) under these hypotheses, at least in the case \(k = 2\), follows easily from classical results of Kurtz [8] concerning the large-\(n\) behavior of density-dependent Markov processes. The general case \(k \geq 2\) follows by obvious and routine extensions of Kurtz’s results. Since the important point is not the weak convergence, but rather the fact that the limit processes will, in general, be non-Feller, we shall only sketch the proof in the special case \([k] = \{0, 1\}\).

**Proof of Proposition 2.5** Assume that \([k] = \{0, 1\}\). In this case, the simplex \(\Delta_k\) may be identified with the unit interval \([0, 1]\), the rate functions \(f_{0,1}\) and \(f_{1,0}\) in Definition 2.4 may be viewed as functions on \([0, 1]\), and

\[ Y_t = Y_t^1 = \frac{1}{n} \sum_{s=1}^{n} 1\{X_t^s = 1\}. \]

The Markov process \((Y_t)_{t \in [0,1]}\) is an instance of what Ethier and Kurtz ([7], Chapter 11) call a density-dependent Markov process. By Theorem 2.1 of [7], Chapter 11 (see also
[8]), if the initial states $y_0 \in [0, 1]$ as $n \to \infty$, then the processes $(Y_t)_{t \geq 0}$ converge in probability (on each finite time interval $[0, T]$) to a deterministic trajectory $(y_t)_{t \geq 0}$, where $(y_t)_{t \geq 0}$ is the unique solution to the differential equation

$$\frac{dy}{dt} = f_{0,i}(y_t)(1 - y_t) - f_{1,0}(y_t)y_t \quad (6)$$

with initial condition $y_0$. (Note: The hypothesis that the functions $f_{i,j}$ are Lipschitz continuous is needed to guarantee that solutions to the differential equation exist and are unique—see condition (2.1) of [7], Chapter 11.) Thus, for each $T < \infty$ and $\epsilon > 0$,

$$\lim_{n \to \infty} P \left( \max_{0 \leq t \leq T} \| Y_t - y_t \| \geq \epsilon \right) = 0. \quad (7)$$

Now consider the Markov processes $(X_t)_{t \geq 0}$ on the state spaces $[k]^{[n]}$. These can be constructed using independent auxiliary Poisson processes $N^s_{i,j}(t)$ with rates $\lambda_{i,j} = \max_{y \in [0,1]} f_{i,j}(y)$, one for each $s \in \mathbb{N}$, and i.i.d. sequences of Uniform-$[0,1]$ random variables $U^s_{i,j}(n)$, according to the following rule: at the $m$th occurrence time $\tau^s_{i,j}(m)$ of $N^s_{i,j}$, if site $s$ is currently of color $i$, then it flips to color $j$ if and only if

$$U^s_{i,j}(m) \leq \frac{f_{i,j}(Y^s_{i,j}(m))}{\lambda_{i,j}}.$$

Since the rate functions $f_{i,j}$ are continuous, it follows from (7) that

$$\lim_{n \to \infty} f_{i,j}(Y^s_{i,j}(m)) = f_{i,j}(y^s_{i,j}(m)).$$

Consequently, for each $K < \infty$ and each $T < \infty$, the processes $(X^s_t)_{t \leq T, s \in [K]}$ converge jointly in law to the process $(\tilde{X}^s_t)_{t \leq T, s \in [K]}$ whose coordinate processes $(\tilde{X}^s_t)_{t \leq T}$ are independent, time-inhomogeneous Markov processes governed by the following rule: at the $m$th occurrence time $\tau^s_{i,j}(m)$ of $N^s_{i,j}$, if site $s$ is currently of color $i$ then it flips to color $j$ if and only if

$$U^s_{i,j}(m) \leq \frac{f_{i,j}(y^s_{i,j}(m))}{\lambda_{i,j}}. \quad (8)$$

This proves the weak convergence assertion of the proposition.

It remains to show that the limit process $(\tilde{X}_t)_{t \geq 0}$ has the required properties. It is obvious that the sample paths of $(\tilde{X}_t)_{t \geq 0}$ are cadlag, and since the coordinate processes are conditionally i.i.d. given the initial value $y_0$, it follows that $(\tilde{X}_t)_{t \geq 0}$ is exchangeable. It remains to show that $(\tilde{X}_t)_{t \geq 0}$ is a Markov process. First, observe that the differential equations (6) and the SLLN imply that for each time $t \geq 0$, with probability one, the limiting fraction of coordinates of $\tilde{X}_t$ that have color 1 exists and equals $y_t$, i.e.,

$$\lim_{n \to \infty} n^{-1} \sum_{s=1}^{n} 1\{\tilde{X}^s_t = 1\} = y_t.$$
Since the process \((\tilde{X}_t)_{t \geq 0}\) is built using independent Poisson processes and Uniform-\([0,1]\) random variables, and since (6) defines a flow on \([0, 1]\) (recall that the functions \(f_{i,j}\) are assumed to be Lipschitz, so solutions of (6) are unique), it follows that \((\tilde{X}_t)_{t \geq 0}\) is a Markov process on \([k]^\mathbb{N}\) with time-homogeneous transition kernel. Finally, the Feller property can only hold if the functions \(f_{i,j}\) are constant, because otherwise the transition probabilities will depend in a nontrivial way on the exchangeable \(\sigma\)-algebra.

\[\square\]

Remark 2.6 The structure of the limit process \((\tilde{X}_t)_{t \geq 0}\) in Proposition 2.5 provides an interesting example of a phenomenon that holds generally for exchangeable Markov processes on \([k]^\mathbb{N}\), as we will show in Theorem 2.13: although the Markov process is time-homogeneous, the coordinate processes \(\tilde{X}_t^i\) are conditionally time-inhomogeneous Markov processes, given the exchangeable \(\sigma\)-algebra.

2.3 Failure of the Feller Property

For exchangeable Markov processes without the Feller property, the transition mechanism may depend in a nontrivial way on its exchangeable \(\sigma\)-field, and so there is a greater variety of possible behaviors than for Feller processes. Moreover, in the absence of the Feller property, there need not exist an infinitesimal generator (at least in the usual sense, as a densely defined linear operator on a space of bounded, continuous functions), and so infinitesimal jump rates need not exist. The following example illustrates the difficulty.

Example 2.7 Let \([k] = \{0, 1\}\) and let \(f : [0, 1] \to [0, 1]\) be any increasing homeomorphism of the unit interval. Let \(U_1, U_2, \ldots\) be independent, identically distributed (i.i.d.) random variables with the Uniform distribution on \([0, 1]\). For \(i \in \mathbb{N}\) and \(0 \leq t \leq 1\), define

\[
X_t^i = \begin{cases} 
1 & \text{if } U_i \leq f(t), \\
0 & \text{if } U_i > f(t).
\end{cases}
\]

The process \((X_t)_{0 \leq t \leq 1}\) is Markov and has cadlag sample paths. Furthermore, \((X_t)_{t \geq 0}\) is exchangeable. However, if the homeomorphism \(f\) has the property that its derivative is 0 almost everywhere (as would be the case if \(f\) were the cumulative distribution function of a singular probability measure on \([0, 1]\) with dense support), then the jumps of \(X_t\) would occur only at times in a set of Lebesgue measure 0.

Although the behavior of the above Markov process is, in certain respects, pathological, the example nevertheless suggests what can happen in general. By the Glivenko–Cantelli Theorem, for each \(t \geq 0\), the limiting frequency of ones in the configuration \(X_t\) exists with probability one. Consequently, for each \(t\), there is a well-defined projection \(Y_t = \pi(X_t)\) of the Markov process \(X_t\) to the 2-simplex \([0, 1]\). In Example 2.7, the projection \((Y_t)_{t \geq 0}\) is a purely deterministic function of \(t\) that follows an increasing, 1

---

1 When \(k = 2\) we write \([k] = \{0, 1\}\) rather than \([k] = \{1, 2\}\).
continuous, but not everywhere differentiable trajectory from 0 to 1. Our first theorem shows that an analogous projection exists in general, and furthermore that it exists simultaneously for all $t$, except possibly on a set of probability zero.

**Theorem 2.8** Let $(X_t)_{t \geq 0}$ be an exchangeable Markov process on $[k]^\mathbb{N}$ with cadlag sample paths. Then, with probability one, for every $t \geq 0$ and every $j \in [k]$, the limiting frequency

$$Y_t^j := \lim_{n \to \infty} n^{-1} \sum_{m=1}^{n} 1\{X_t^m = j\}$$  \hspace{1cm} (9)

exists, and the vector $Y_t = (Y_t^j)_{j \in [k]}$ is an element of the $k$-simplex $\Delta_k$ of probability distributions on $[k]$. The process $(Y_t)_{t \geq 0}$ is Markov and has cadlag sample paths. Furthermore, the sample paths of $(Y_t)_{t \geq 0}$ are of locally bounded variation with respect to the $L^1$ metric on the simplex $\Delta_k$.

Except for the last (and perhaps most interesting) assertion, that the sample paths of $Y_t$ are of locally bounded variation, Theorem 2.8 will be proved in Sects. 3–4: see Corollary 4.4. That the sample paths are of locally bounded variation will be proved in Sect. 5.

**Example 2.9** This example shows that in general, the projection $(Y_t)_{t \geq 0}$ need not be Feller. Let $[k] = \{0, 1\}$ so that the simplex $\Delta_k$ can be identified with the unit interval $[0, 1]$. The transition mechanism of the process $(X_t)_{t \geq 0}$ will involve a sequence $\tau_1, \tau_2, \ldots$ of i.i.d. exponential random variables with mean 1, independent of the initial state, but will also involve the projection $Y_0$ of the initial state. In order that this be well-defined for all elements $x \in [k]^\mathbb{N}$, we extend the definition of the projection as follows:

$$Y_0 := \liminf_{n \to \infty} n^{-1} \sum_{m=1}^{n} 1\{X_0^m = 1\}.$$  \hspace{1cm} (10)

Given the initial state $X_0$ and the auxiliary random variables $\tau_j$, define

$$X_t^j = 1_{[\frac{1}{2}, 1]}(Y_0)$$

if $t \geq \tau_i,$

$$= X_0^j$$

if $t < \tau_j.$

If the initial state $X_0$ is exchangeable, then clearly the process $(X_t)_{t \geq 0}$ is Markov and exchangeable. Its projection $Y_t$ to the simplex $[0, 1]$ follows a deterministic trajectory

$$t \mapsto \gamma(Y_0, t) = \begin{cases} Y_0 + (1 - Y_0)(1 - e^{-t}), & Y_0 \in [\frac{1}{2}, 1] \\ Y_0 e^{-t}, & \text{otherwise} \end{cases}$$

and, for each $Y_0 \in (0, 1)$, this trajectory is continuous. However, the trajectories do not fit together to make a (jointly) continuous flow: in particular, for every $t > 0$ the function $Y_0 \mapsto \gamma(Y_0, t)$ is discontinuous at $Y_0 = 1/2$ and therefore, the projection $(Y_t)_{t \geq 0}$ is not Feller.
Example 2.10 This example shows that in general, the law of the projection \((Y_t)_{t \geq 0}\) does not uniquely determine the law of the covering process \((X_t)_{t \geq 0}\). The process \(Y_t\) will be a two-state chain that jumps back and forth between \(p = 1/3\) and \(q = 2/3\) at the occurrence times \(T_1 < T_2 < \cdots\) of a rate-1 Poisson process. We exhibit two exchangeable Markov processes on \([0, 1]^N\), denoted \(X_t\) and \(Z_t\), both of which project to \(Y_t\). The first process \(X_t\) evolves as follows: for each \(n = 1, 2, \ldots\), (a) if \(Y_{T_n} = 2/3\), then all of the sites \(i\) for which \(X_{T_n}^i = 1\) simultaneously toss fair coins to determine their colors at time \(T_n\), and (b) if \(Y_{T_n} = 1/3\), then all of the sites \(i\) for which \(X_{T_n}^i = 0\) simultaneously toss fair coins to determine their colors at time \(T_n\). The second process \(Z_t\) evolves even more simply: at each time \(T_n\), all sites, regardless of their current colors, toss \((1 - Y_{T_n})\)-coins to determine their colors at time \(T_n\).

Example 2.11 Even if the Markov process \((X_t)_{t \geq 0}\) is Feller, its law need not be determined by its projection \((Y_t)_{t \geq 0}\). Take \([k] = \{0, 1\}\) and consider the degenerate process \(Y_t = (p, 1 - p)\) for all \(t \geq 0\), where \(p < 1/2\). Then \((X_t)_{t \geq 0}\) can evolve either by (a) remaining constant for all \(t \geq 0\), or (b) at the occurrence times \(T_1 < T_2 < \cdots\) of a rate-1 Poisson process, all sites \(i\) for which \(X_{T_n}^i = 0\) flip to \(X_{T_n}^i = 1\), while each site \(i\) with \(X_{T_n}^i = 1\) flips to \(X_{T_n}^i = 0\) with probability \(p/(1 - p)\).

Observe that in both of these evolutions the projection \((Y_t)_{t \geq 0}\) into \(\Delta_k\) is degenerate at \((p, 1 - p)\).

The preceding examples indicate that additional information is required to uniquely determine the law of the covering Markov process. Because the paths of \((Y_t)_{t \geq 0}\) are not necessarily differentiable, this information cannot be encapsulated by infinitesimal rates. Thus, our specification will involve an associated time-inhomogeneous Markov semigroup on \([k]\).

Definition 2.12 Given a Markov process \((Y_t)_{t \geq 0}\) valued in the simplex \(\Delta_k\), a compatible random semigroup is a two-parameter family \((Q_{s,t})_{0 \leq s \leq t}\) of random \(k \times k\) stochastic matrices that satisfy the cocycle relation

\[Q_{r,t} = Q_{r,s}Q_{s,t}\quad\text{for all } r \leq s \leq t\quad\text{a.s.,}\quad (11)\]

and the compatibility relations

\[Y_sQ_{s,t} = Y_t\quad\text{for all } s \leq t\quad\text{a.s.}\quad (12)\]

Theorem 2.13 Let \((X_t)_{t \geq 0}\) be an exchangeable Markov process on \([k]^N\) with cadlag sample paths, and let \((Y_t)_{t \geq 0}\) be its projection to the simplex \(\Delta_k\). Then, there exists a compatible random semigroup \((Q_{s,t})_{0 \leq s \leq t}\) such that (i) for each pair \(s \leq t\), the random matrix \(Q_{s,t}\) is measurable with respect to the exchangeable \(\sigma\)-algebra \(E_{[s,1]}\); and (ii) conditional on the exchangeable \(\sigma\)-algebra \(E_{[0,\infty)}\), the coordinate processes \(X^i\) are independent realizations of a time-inhomogeneous Markov chain on \([k]\) with transition probability matrices \(Q_{s,t}\), that is,

\[P(X^i_s = b \mid X^i_s = a, E_{[0,\infty)}) = Q_{s,t}(a, b)\quad (13)\]
Theorem 2.13 follows from Corollary 3.4 in Sect. 3. We emphasize that although the Markov process \((X_t)_{t \geq 0}\) is time-homogeneous, its conditional distribution given the exchangeable \(\sigma\)-algebra will in general be that of a time-inhomogeneous Markov process. Theorem 2.13 shows that the law of an exchangeable Markov process on \([k]^\mathbb{N}\) is uniquely determined by the law of the two-parameter matrix-valued process \((Q_{s,t})_{0 \leq s < t}\). Example 2.10 shows that in general, the law of \((Q_{s,t})_{0 \leq s < t}\) is not uniquely determined by that of the projection \((Y_t)_{t \geq 0}\). However, if \((Y_t)_{t \geq 0}\) is Markov and has c\-adlag sample paths of bounded variation, then there is at least one compatible semi-group.

**Theorem 2.14** For every Markov process \((Y_t)_{t \geq 0}\) on the \(k\)-simplex \(\Delta_k\) whose sample paths are c\-adlag and of locally bounded variation, there exists an exchangeable Markov process \((X_t)_{t \geq 0}\) on \([k]^\mathbb{N}\) whose projection (9) into \(\Delta_k\) has the same distribution as \((Y_t)_{t \geq 0}\).

Theorem 2.14 will be proved in Sect. 6 by constructing a compatible random semi-group that moves the minimum possible amount of mass subject to the compatibility condition (12).

### 3 Discrete-Time Exchangeable Markov Chains

For the remainder of the paper, we assume that \((X_t)_{t \in T}\), where either \(T = [0, \infty)\) or \(T = \mathbb{Z}_+\), is an exchangeable Markov process on \([k]^\mathbb{N}\) with c\-adlag sample paths.

**Proposition 3.1** For every finite sequence of times \(0 = t_0 < t_1 < \cdots < t_n < \infty\), the vector-valued sequence

\[
\begin{pmatrix}
X_{t_0}^1 \\
X_{t_1}^1 \\
\vdots \\
X_{t_n}^1
\end{pmatrix}, \quad
\begin{pmatrix}
X_{t_0}^2 \\
X_{t_1}^2 \\
\vdots \\
X_{t_n}^2
\end{pmatrix}, \cdots
\]

is exchangeable.

**Proof** This is a routine consequence of hypotheses (a) and (b) of Definition 2.1 and the Markov property. \(\square\)

**Corollary 3.2** For every \(t \geq 0\), \((X^n_t)_{n \in \mathbb{N}}\) is an exchangeable sequence of \([k]\)-valued random variables. Consequently, the limiting frequencies \((Y^j_t)_{j \in [k]}\) (9) exist with probability 1. Furthermore, the projection \((Y_t)_{t \geq 0}\) is a Markov process on the simplex \(\Delta_k\).

**Note 3.3** Only the almost sure existence of limiting frequencies \(Y_t\) at fixed times \(t\) is asserted here. The almost sure existence of limiting frequencies at all times \(t\) will be proved later. Thus, the second assertion of the corollary, that the projection \((Y_t)_{t \geq 0}\) is a Markov process, should be interpreted only as a statement about conditional distributions at finitely many time points.
Proof The first assertion is an immediate consequence of de Finetti’s theorem. Denote by \( \mathcal{F}_t := \sigma(X_s)_{0 \leq s \leq t} \), the \( \sigma \)-algebra generated by the random variables \( X_s \) for \( s \leq t \). Since each \( Y_t = \pi(X_t) \) is a measurable function of \( X_t \), the hypothesis that \( (X_t)_{t \geq 0} \) is Markov implies that the conditional distribution of \( Y_{t+r} \) given \( \mathcal{F}_t \) coincides with the conditional distribution of \( Y_{t+r} \) given \( X_t \). The process \( (X_t)_{t \geq 0} \) is exchangeable; therefore, for any permutation \( \sigma : \mathbb{N} \to \mathbb{N} \) fixing all but finitely many elements, the joint distributions of \( (X_{t+1}^\sigma, X_{t+r+1}^\sigma) \) and \( (X_t, X_{t+r}) \) are identical. The projection \( \pi : [k]^\mathbb{N} \to \Delta_k \) is invariant by \( \sigma \), so it follows that the conditional distribution of \( Y_{t+r} \) given \( X_t^\sigma \) is the same as the conditional distribution of \( Y_{t+r} \) given \( X_t \). Thus, the conditional distribution of \( Y_{t+r} \) given \( X_t \) depends only on the exchangeable \( \sigma \)-algebra. By de Finetti’s theorem, this is the conditional distribution of \( Y_{t+r} \) given \( Y_t \); hence, \( (Y_t)_{t \geq 0} \) is Markov. \( \square \)

Corollary 3.4 For every pair of times \( s < t \) and every pair of colors \( i, j \in [k] \),

\[
Q_{s,t}(i, j) := \lim_{n \to \infty} \frac{\sum_{m=1}^{n} 1\{X^m_s = i \text{ and } X^m_t = j\}}{\sum_{m=1}^{n} 1\{X^m_s = i\}}
\]

exists on the event \( Y^i_s > 0 \). With probability one, the matrix \( Q_{s,t} \) is stochastic (with the convention that, for every color \( i \) such that \( Y^i_s = 0 \), the \( i \)th row \( Q_{s,t}(i, \cdot) = \delta_i(\cdot) \)).

Proof By Proposition 3.1, the vector-valued sequence

\[
\left( X^1_s, X^1_t \right), \left( X^2_s, X^2_t \right), \ldots
\]

is exchangeable. Hence, by de Finetti’s theorem, limiting frequencies for all color pairs \( i, j \in [k] \) exist almost surely. Since the number of colors \( k \) is finite, the corresponding matrix \( Q_{s,t} \) is almost surely stochastic. \( \square \)

Proposition 3.5 Let \( A \) be any finite set of times \( 0 < s_0 < s_1 < \cdots < s_n < \infty \). Then, conditional on the exchangeable \( \sigma \)-algebra \( \mathcal{E}_A \), the vectors \( X^1_A, X^2_A, \ldots \) are independent, identically distributed, and for each \( m \in \mathbb{N} \), the sequence \( \{X^m_s\}_{0 \leq s \leq n} \) is a time-inhomogeneous Markov chain on \( [k] \) with transition probabilities \( Q_{s,r,s+1} \) in (14).

Proof That the vectors \( X^1_A, X^2_A, \ldots \) are conditionally i.i.d. given \( \mathcal{E}_A \) follows from Proposition 3.1 and de Finetti’s theorem. It suffices to show that, conditional on \( \mathcal{E}_A \), the sequence \( \{X^m_s\}_{0 \leq s \leq n} \) is Markov with the indicated transition probabilities.

By Proposition 3.1, the sequence of random vectors \( X^1_A, X^2_A, \ldots \) is exchangeable, so de Finetti’s theorem and the strong law of large numbers imply that limiting empirical frequencies exist for all vectors \( j_A \in [k]^A \). We must show that these limiting empirical frequencies are given by

\[
\lim_{n \to \infty} n^{-1} \sum_{r=1}^{n} 1\{X^r_A = j_A\} = P(X^1_{s_0} = j_0 | \mathcal{E}_{s_0}) \prod_{l=1}^{m} Q_{s_{l-1}, s_l}(j_{l-1}, j_l);
\]
that is, we must show that the limiting empirical frequency of the color \( j_{s_n} \) at those sites taking color \( j_{s_n} \) at time \( s_{n-1} \) does not depend on the colors \( j_{s_r} \) taken by these sites at earlier times \( s_r \). However, this is merely a consequence of the Markov property of \( X_{s_0}, X_{s_1}, \ldots \), as we now argue.

Note that, unconditionally, we can generate \( X_{s_0}, X_{s_1}, \ldots, X_{s_n} \) as follows. For each \( 1 \leq r \leq n - 1 \), the law of \( X_{s_{r+1}} \), given \( \sigma(X_{s_r})_{0 \leq j \leq r} \), is a measurable function of \( \pi(X_{s_r}) =: Y_{s_r} \). Moreover, by Corollary 3.4, the matrix \( Q_{s_r, s_{r+1}} \) is a random stochastic matrix with distribution depending only on \( \mathcal{E}_{s_r} \), where \( Q_{s_r, s_{r+1}}(i, j) \) gives the limiting frequency of coordinates for which \( X_{s_r} = j \) among those for which \( X_{s_r} = i \). Hence, given \( \sigma(X_{s_r})_{0 \leq j \leq r} \), we generate a matrix \( Q_{s_r, s_{r+1}} \) randomly and, conditional on \( Q_{s_r, s_{r+1}} \), we generate \( X_{s_{r+1}} \) coordinate-by-coordinate by determining \( X_{s_{r+1}}^l \), for each \( l \in \mathbb{N} \), independently from

\[
\mathbb{P}(X_{s_{r+1}}^l = j \mid Q, X_{s_r}) = Q(X_{s_r}, j).
\]  

Clearly, by the strong law of large numbers, the matrix \( Q_{s_r, s_{r+1}} \), from (14), is identical to \( Q \) for those \( i \) such that \( Y_{s_r}^i > 0 \). Moreover, \( Q_{s_r, s_{r+1}} \) is measurable with respect to \( \mathcal{E}(X_{s_r}, X_{s_{r+1}}) \subset \mathcal{E}_A \). By this and (15), we conclude that \( (X^m_{s_r})_{0 \leq r \leq n} \) is, conditionally on \( \mathcal{E}_A \), a time-inhomogeneous Markov chain on \( [k] \) with transition probabilities (14). □

We now give a complete description of the set of discrete-time exchangeable Markov chains \( (X_t)_{t \in \mathbb{Z}_+} \) on \( [k]^\mathbb{N} \). Recall that \( S_k \) denotes the space of stochastic \( k \times k \) matrices and denote by \( \mathcal{P}(S_k) \) the set of all Borel probability measures on \( S_k \).

**Theorem 3.6** Let \( (X_t)_{t \in \mathbb{Z}_+} \) be a discrete-time exchangeable Markov chain on \( [k]^\mathbb{N} \), and denote by \( Y_t = \pi(X_t) \) the limiting empirical frequency (row) vector at time \( t \) and by \( Q_{n+1} = Q_{n,n+1} \) the limiting empirical transition matrix defined by (14). Then, there exists a measurable mapping \( G : \Delta_k \to \mathcal{P}(S_k) \) such that

\[
\mathcal{D}(Q_{n+1} \mid \sigma(Y_t, Q_t)_{0 \leq t \leq n}) = G(Y_n) \quad \text{and} \quad Y_{n+1} = Y_n Q_{n+1}.
\]  

Moreover, for each \( n \), conditional on \( \sigma(X_t)_{0 \leq t \leq n} \), the coordinate variables \( X^m_{n+1} \) are conditionally independent with marginal distributions

\[
P(X^m_{n+1} = j \mid X_n, Q_{n+1}) = Q_{n+1}(i, j) \quad \text{on} \quad X^m_n = \{i\}.
\]  

Conversely, for any measurable mapping \( G : \Delta_k \to \mathcal{P}(S_k) \) and any initial point \( Y_0 \in \Delta_k \), there is a unique Markov chain \( (X_t)_{t \in \mathbb{Z}_+} \) on \( [k]^\mathbb{N} \) specified by equations (16), (17), (18), and the initial condition that the random variables \( X^m_0 \) are i.i.d. with common distribution \( Y_0 \).

**Proof** Let \( (X_t)_{t \in \mathbb{Z}_+} \) be an exchangeable Markov chain on \( [k]^\mathbb{N} \), and then, by (b) of Definition 2.1, the law of \( X_{n+1} \) given \( \sigma(X_t)_{0 \leq t \leq n} \) depends only on \( Y_n = \pi(X_n) \), which exists almost surely by Corollary 3.4. Also implied by Corollary 3.4 is that the conditional law of \( Q_{n+1} \) given \( \sigma(Y_t, Q_t)_{0 \leq t \leq n} \) depends only on \( Y_n \), which implies (16). Proposition 3.5 now implies (17) and (18).
The converse is immediate from the following construction. To begin, we choose \( X_0 \), given \( Y_0 \), by sampling its coordinates \( X_1^0, X_2^0, \ldots \) i.i.d. from \( Y_0 \). Subsequently, for each \( n \geq 1 \), given \( X_n \) has limiting frequency \( Y_n = \pi(X_n) \), we draw a random stochastic matrix \( S \) from \( G(Y_n) \), which acts as the transition probability matrix for each coordinate, as in Proposition 3.5.

\[ \square \]

4 The Color-Swatch Process

Assume now that \( (X_t)_{t \geq 0} \) is an exchangeable, continuous-time Markov process on \([k]^N\). By Proposition 3.1, for any finite set \( A \subset \mathbb{R}_+ \), the vector-valued sequence \( X_A^1, X_A^2, \ldots \) is exchangeable. Consequently, by the Hewitt–Savage theorem, for any fixed \( t \geq 0 \), the sequence \( X_t^1, X_t^2, \ldots \) has limiting empirical frequencies with probability one. Since \( \mathbb{R}_+ \) is uncountable, it does not directly follow that limiting empirical frequencies exist simultaneously for all \( t \geq 0 \). In this section, we prove that, under the additional hypothesis that the sample paths are cadlag, limiting empirical frequencies exist for all times \( t \) and that the induced projection of \( (X_t)_{t \geq 0} \) to the \((k-1)\)-dimensional simplex is a Markov process with cadlag sample paths.

4.1 Color Swatches

Viewing \([k]\) as a collection of \( k \) distinct colors, we may regard any cadlag path \( \{x(t)\}_{t \in [0,1]} \) in \([k]\) as a color swatch, that is, a concatenation of finitely many nonoverlapping colored intervals \( J_i \) whose union is \([0,1]\). In general, a color swatch is parametrized by

(i) a positive integer \( m + 1 \) (the number of intervals \( J_i \));
(ii) a sequence \( 0 = t_0 < t_1 < \cdots < t_{m+1} = 1 \), the endpoints of the intervals \( J_i \); and
(iii) a list \((\kappa_1, \kappa_2, \ldots, \kappa_{m+1})\) of colors (elements of \([k]\)) subject to the restriction that no two successive colors in the list are the same.

The space \( \mathcal{W} \) of all color swatches is naturally partitioned as \( \mathcal{W} = \bigcup_{m=0}^{\infty} \mathcal{W}_m \), where \( \mathcal{W}_m \) is the set of all color swatches with \( m + 1 \) distinctly colored subintervals. Furthermore, \( \mathcal{W} \) is equipped with the Skorohod metric \( d \), which is defined as follows: for any pair \( f, g \in \mathcal{W} \),

\[
d(f, g) = \inf_h \max(\|h - id\|_{\infty}, \|f - g \circ h\|_{\infty}).
\]

Here \( \| \cdot \|_{\infty} \) denotes the sup norm and the infimum is over all increasing homeomorphisms \( h : [0,1] \to [0,1] \). The elements \( f, g \in \mathcal{W} \) are viewed as functions on \([0,1]\) with range \([k]\), which inherits the Euclidean norm on \( \mathbb{R} \). The topology on \( \mathcal{W} \) induced by the Skorohod metric is separable [3], and \( \mathcal{W} \) can be completed by adjoining points to \( \mathcal{W}_m \) for which ties \( t_i = t_{i+1} \) are allowed. Since \( \mathcal{W} \) is an open subset of this completion, it, along with its Borel sets, is a Borel space. Thus, in particular, the Hewitt–Savage extension of de Finetti’s theorem [9] applies to exchangeable sequences of \( \mathcal{W} \)-valued random variables.
Proposition 4.1 If \((X_t)_{t \in \mathbb{R}^+}\) is an exchangeable Markov process on \([k]^{\mathbb{N}}\) with cadlag sample paths, then the sequence

\[ X^1_{[0,1]}, X^2_{[0,1]}, \ldots \]  

of color swatches induced by the coordinate processes \((X^i_t)_{t \in \mathbb{R}^+}\) is an exchangeable sequence of \(W\)-valued random variables. Consequently, with probability one the empirical distributions \(m^{-1} \sum_{i=1}^m \delta_{X^i_{[0,1]}}\) converge weakly to a random Borel probability measure \(\Theta\) on \(W\). The random measure \(\Theta\) generates the exchangeable \(\sigma\)-algebra \(\mathcal{E}_{[0,1]}\) of the sequence \((20)\), and, conditional on \(\mathcal{E}_{[0,1]}\), the color swatches \(X^1_{[0,1]}, X^2_{[0,1]}, \ldots\) are independent, identically distributed with distribution \(\Theta\).

Proof By assumption, each \(X^i_{[0,1]}, i = 1, 2, \ldots\) is cadlag, and so \(X^i_{[0,1]} \in W\) for every \(i \in \mathbb{N}\) almost surely. That the sequence \((20)\) is exchangeable follows from Proposition 3.1, because the \(\sigma\)-algebra \(\sigma(X^i_{[0,1]}, i \in \mathbb{N})\) generated by the coordinate variables \(x_q\), where \(q\) is rational, generates the Borel \(\sigma\)-algebra on \(W\). The rest follows from the Hewitt–Savage and de Finetti theorems. \(\square\)

Corollary 4.2 If \((X_t)_{t \in \mathbb{R}^+}\) is an exchangeable Markov process on \([k]^{\mathbb{N}}\) with cadlag sample paths, then, conditional on \(\mathcal{E}_{[0,1]}\), the color swatches \(X^m_{[0,1]}\) are i.i.d. and \(X^m_{[0,1]}\) has the same conditional law as the path of an inhomogeneous, continuous-time Markov chain on the finite state space \([k]\) with transition probabilities \(Q_{s,t}\); that is, for all \(0 \leq s \leq t \leq 1\),

\[ P(X^m_t = j \mid \mathcal{E}_{[0,1]} \uplus \sigma(X^m_{s \leq r}) = Q_{s,t}(i, j) \text{ on } X^m_s = i. \]  

Proof We have already shown, in Proposition 4.1, that the color swatches \(X^m_{[0,1]}\) are conditionally i.i.d. given the exchangeable \(\sigma\)-algebra \(\mathcal{E}_{[0,1]}\). That the individual color swatches are conditionally inhomogeneous Markov processes with transition probabilities \((21)\) follows routinely from Proposition 3.5, because the exchangeable \(\sigma\)-algebra \(\mathcal{E}_{[0,1]}\) is generated by \(\bigvee_{n=1}^\infty \mathcal{E}_{D_n}\), where \(D_n\) is the set of \(n\)th-level dyadic rationals \(m/2^n\) in \([0,1]\). \(\square\)

4.2 Existence of Limiting Empirical Color Frequencies

Proposition 4.1 implies that if \((X_t)_{t \geq 0}\) is an exchangeable Markov process on \([k]^{\mathbb{N}}\) with cadlag sample paths, then the joint distribution of the sequence \((X^i_{[0,1]})_{i \in \mathbb{N}}\) of color swatches is a mixture of product measures. Thus, to prove that this sequence has well-defined limiting empirical color frequencies at all times \(t \in [0, 1]\), it suffices to show that this is the case for any sequence of i.i.d. color swatches.

Proposition 4.3 Let \(Z^1, Z^2, \ldots\) be a sequence of independent, identically distributed \(W\)-valued random variables. Then, with probability one, for every \(t \in [0, 1]\) the empirical distributions of the sequence \((Z^j_t)_{i \in \mathbb{N}}\) converge to a nonrandom probability distribution \(\pi(Z_t) := (\pi(Z^j_t))^j_{j \in [k]}\) on \([k]\); that is, for each \(j \in [k]\) and \(t \in [0, 1]\),
\[
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} 1\{Z_i^t = j\} = \pi(Z_t)^j.
\] (22)

**Proof**  To prove that the sequence of empirical frequencies converges, we need only show that, for any \(\varepsilon > 0\), the upper and lower limits of the sequence of averages (22) differ by no more than \(\varepsilon\). Fix \(j \in [k]\) and, for each \(t \in [0, 1]\), define

\[
L_t^+ = \limsup_{n \to \infty} n^{-1} \sum_{i=1}^{n} 1\{Z_i^t = j\}
\]

and

\[
L_t^- = \liminf_{n \to \infty} n^{-1} \sum_{i=1}^{n} 1\{Z_i^t = j\}.
\]

For any fixed \(t \in [0, 1]\), the strong law of large numbers implies that \(L_t^+ = L_t^-\) almost surely. We will show that with probability one, \(L_t^+ - L_t^- \leq 2\varepsilon\) for all \(t \in [0, 1]\).

By hypothesis, the random functions \(Z^1, Z^2, \ldots\) have cadlag sample paths, and in particular, each \(Z^i\) has at most finitely many discontinuities in the time interval \([0, 1]\). Define

\[
D = \{t \in [0, 1] : P\{Z_t^i \neq Z_t^{i-}\} > 0\}.
\] (23)

We claim that this set is at most countable. For suppose not: Then for some \(\varepsilon > 0\), the subset \(D_\varepsilon\) consisting of times \(t \in D\) such that \(P\{Z_t^i \neq Z_t^{i-}\} \geq \varepsilon\) would be uncountable. But by the SLLN, for every \(t \in D_\varepsilon\), the limiting fraction of paths \(Z^i\) that are discontinuous at \(t\) is at least \(\varepsilon\). As a consequence, the limiting fraction of paths \(Z^i\) that have infinitely many discontinuities in \([0,1]\) must be positive, contradicting the hypothesis that each all paths \(Z^i\) are cadlag. It follows that \(D\) is at most countable. Furthermore, for each \(\varepsilon > 0\), the subset \(D_\varepsilon \subset D\) must be finite to ensure each \(Z^i\) is cadlag.

Fix \(\varepsilon > 0\) and consider the set \(R_\varepsilon = [0, 1] \setminus D_\varepsilon\). For each time \(t \in R_\varepsilon\), the probability that \(Z^i\) has a discontinuity at time \(t\) is less than \(\varepsilon\). We claim that there is a countable partition \(J_1, J_2, \ldots\) of \(R_\varepsilon\) into nonoverlapping intervals \(J_l\) such that, for each interval \(J_l\),

\[
P\{Z^i is discontinuous at some \(t \in J_l\} < \varepsilon.
\] (24)

As \(D_\varepsilon\) is finite, we construct such a partition by first dividing the constituent intervals of \(R_\varepsilon\) into halves; and then dividing these halves into quarters; the quarters into eighths, and so on. We stop the subdivision process whenever an interval \(J_l\) is small enough that (24) holds. This must happen eventually for each interval not abutting one of the points of \(D_\varepsilon\), because otherwise there would be a nested sequence of closed intervals \(J_n \subset R_\varepsilon\) that shrink to a point \(t \in R_\varepsilon\), and for each \(n \geq 1\)

\[
P\{Z^i has a discontinuity in \(J_n\} \geq \varepsilon.
\]

But this would imply that \(t \in D_\varepsilon\), a contradiction.

Suppose, then, that \(J_l\) is a countable partition of \(R_\varepsilon\) into nonoverlapping intervals \(J_l\) such that for each interval, \(J_l\) condition (24) holds. Then the SLLN implies that for
each interval $J_l$ of the partition,

$$
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} 1\{Z^i \text{ is discontinuous at some } t \in J_l\} < \varepsilon.
$$

It follows that neither $L^+_t$ nor $L^-_t$ can vary by more than $\varepsilon$ over the interval $J_i$. Since $L^+_s = L^-_s$ almost surely for each endpoint $s$ of $J_i$ (this set is countable!), we conclude that for each interval $J_i$, with probability one,

$$
\sup_{t \in J_i} L^+_t - L^-_t \leq 2\varepsilon.
$$

Since $[0, 1]$ is covered by the intervals $J_i$ and the nonrandom finite set $D_\varepsilon$ (on which $L^+_t = L^-_t$ a.s.), it follows that for every $\varepsilon > 0$,

$$
\sup_{t \in [0, 1]} L^+_t - L^-_t \leq 2\varepsilon \text{ a.s.}
$$

Since $\varepsilon > 0$ is arbitrary, the limits (22) exist for all $t \in [0, 1]$ almost surely. Since, for any finite $n \in \mathbb{N}$, the limits (22) do not depend on the first $n$ elements of $(Z^i_{[0,1]})_{i \in \mathbb{N}}$, Kolmogorov’s 0–1 law implies the limits are deterministic. \hfill \qed

**Corollary 4.4** If $(X_t)_{t \geq 0}$ is an exchangeable, continuous-time Markov process on $[k]^{\mathbb{N}}$ with cadlag sample paths, then for every $t \geq 0$, the empirical limiting frequency vector $Y_t = \pi(X_t)$ exists, and the projection $(Y_t)_{t \geq 0}$ is a Markov process with cadlag sample paths in the simplex $\Delta_k$.

**Proof** The existence of limiting empirical frequencies for all $t \geq 0$ follows from Propositions 4.1 and 4.3 and the Hewitt–Savage extension of de Finetti’s theorem [9]. The Markov property of $(Y_t)_{t \geq 0}$ follows from Corollary 3.2. Finally, the sample paths of $(Y_t)_{t \geq 0}$ are cadlag since, by Proposition 4.3, $(Y_t)_{t \geq 0}$ is continuous at every $t \notin D$ almost surely, where $D$ is defined by (23), and $D$ is a countable, nonrandom subset of $[0, 1]$. \hfill \qed

### 4.3 Characterization of Discontinuities

**Proposition 4.5** Let $(X_t)_{t \in \mathbb{R}_+}$ be an exchangeable Markov process on $[k]^{\mathbb{N}}$ with cadlag sample paths. Then, with probability one, the sample path $t \mapsto X_t$ has at most countably many discontinuities. Furthermore, with probability one, there are precisely two possible types of discontinuity at $s$: either

- (I) there exists a unique $i \in \mathbb{N}$ such that $X^i_s$ is discontinuous at $s$, or
- (II) $P(X^i_s \text{ is discontinuous at } t = s | \mathcal{E}) > 0$, and for each pair, $j_1 \neq j_2 \in [k]$.

$$
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} 1\{X^i_{s-} = j_1 \text{ and } X^i_s = j_2\} = P(X^i_{s-} = j_1 \text{ and } X^i_s = j_2 | \mathcal{E}).
$$

(25)
Moreover, the projection $Y_t = \pi(X_t)$ to the simplex has cadlag paths, with discontinuities only at the times of type-(II) discontinuities of $(X_t)_{t \in \mathbb{R}^+}$.

**Proof** It suffices to prove the corresponding assertions for the restriction $X_{[0,1]}$, because the case $X_{[0,T]}$ for arbitrary $T$ follows by rescaling, and the case $X_{[0,\infty)}$ follows by exhaustion. By Proposition 4.1, it suffices to prove the corresponding statement for sequences of i.i.d. color swatches. Thus, we assume that $Z^{1}, Z^{2}, \ldots$ are independent, identically distributed $\mathcal{S}$-valued random variables, and as in the proof of Proposition 4.3, we define $D_{k}$ to be the set of all $s \in [0,1]$ such that $P(Z^{1}_{t} \text{ discontinuous at } t = s) \geq \varepsilon$ and $D := \bigcup_{\varepsilon > 0} D_{\varepsilon}$. By the same argument as in the proof of Proposition 4.3, $D$ is at most countable because $Z^{1}$ has at most finitely many discontinuities in $[0,1]$ almost surely. By the strong law of large numbers, for each $s \in D$ and any pair $j_1 \neq j_2 \in [k]$, \[
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} 1\{Z_{s-}^{i} = j_1 \text{ and } Z_{s}^{i} = j_2\} = P(Z_{s-}^{1} = j_1 \text{ and } Z_{s}^{1} = j_2). \]

Next, we show that for $s \notin D$, there can be at most one index $i \in \mathbb{N}$ for which $Z_{s}^{i}$ has a discontinuity at $t = s$. For any $i \in \mathbb{N}$, let $D_{i}$ be the set of discontinuities of $Z_{i}$. Because $(Z_{i})_{i \in \mathbb{N}}$ is an i.i.d. collection of color swatches, \[
P\left(Z_{i} \text{ discontinuous at some } t \in D_{i} \setminus D \mid D_{i}\right) = 0, \quad \text{for all } j \neq i, \]

since for any $t \notin D$, the unconditional probability that $Z_{i}$ is discontinuous at $t$ is 0. Hence, unconditionally, the probability that $Z_{i}$ and $Z_{j}$ share a common point of discontinuity outside $D$ is 0. Since there are only countably many pairs $i, j$, it follows that there is zero probability that some pair has a common discontinuity outside of $D$.

Finally, by the proof of Proposition 4.3, for each $\varepsilon > 0$, there is a countable partition of the open set $[0,1] \setminus D_{\varepsilon}$ into intervals $J_{t}$ such that (24) holds. It follows by the SLLN that the limiting empirical frequencies $\pi(Z_{t})$ cannot vary by more than $\varepsilon$ in any of the intervals $J_{t}$. Since $\varepsilon > 0$ can be made arbitrarily small, it follows that the only discontinuities of the projection $\pi(Z_{t})$ must be at points $t \in D$. \[\square\]

5 Bounded Variation of Sample Paths in $\Delta_{k}$

5.1 Mass Transfer

Let $(X_{t})_{t \geq 0}$ be a continuous-time, exchangeable Markov process on $[k]^{\mathbb{N}}$ with cadlag sample paths, and let $Y_{t} = \pi(X_{t})$ be its projection to the simplex. Corollary 3.4 implies that for every pair of colors $i, j \in [k]$ and any two times $s \leq t$, the limiting fraction $Q_{s,t}(i,j)$ of sites with color $i$ at time $s$ that flip to color $j$ at time $t$ exists, and the matrix $Q_{s,t}$ is stochastic. The results of Sect. 4 imply that the process $Q_{s,t}$ can be extended to a two-parameter process $\{Q_{s,t}\}_{0 \leq s \leq t}$ valued in the space $\mathcal{S}_{k}$ of $k \times k$ stochastic matrices. For each fixed $s$, the sample paths $\{Q_{s,t}\}_{s \leq t}$ are cadlag, with
discontinuities only at the type-(II) discontinuities of \(X_t\). Furthermore, the matrices \(Q_{s,t}\) satisfy the cocycle equations (11) and the compatibility condition (12). We define the total mass transfer \(T_{s,t}\) between times \(s\) and \(t\) to be the limiting fraction of sites \(m \in \mathbb{N}\) that have different colors at times \(s\) and \(t\), that is,

\[
T_{s,t} := \sum_{a \neq b} Y_t^a Q_{s,t}(a, b) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n 1\{X^i_s \neq X^i_t\}. \tag{26}
\]

Observe that the variation \(\|Y_t - Y_s\|_1 := \sum_{i=1}^k |Y_t^i - Y_s^i|\) in the frequency vector between times \(s\) and \(t\) is bounded above by \(T_{s,t}\). Moreover, the mass transfer process \((T_{s,t})_{0 \leq s \leq t}\) satisfies

\[
T_{r,t} \leq T_{r,s} + T_{s,t} \quad \text{for all } r \leq s \leq t. \tag{27}
\]

**Proposition 5.1** With probability one, the mass transfer process \((T_{s,t})_{0 \leq s \leq t}\) has bounded variation on finite intervals, that is,

\[
\sup_{0 = s_0 \leq s_1 \leq s_2 \leq \ldots \leq s_n = 1} \sum_{i=0}^{n-1} T_{s_i, s_{i+1}} < \infty \quad \text{a.s.} \tag{28}
\]

### 5.2 Proof of Proposition 5.1

We first prove Proposition 5.1 in the special case \(k = 2\). This case is more straightforward than the general case because if the mass transfer process has unbounded variation, then any transfer of mass from 0 to 1 must eventually be matched by a corresponding transfer from 1 to 0; whereas, in the general case, mass can be transported along incomplete cycles in the complete graph \(K_{\lfloor k \rfloor}\).

**Proof for the case \(k=2\)** Assume that (28) does not hold; then, there is a positive probability that the supremum in (28) is \(+\infty\). Note that this event is in the exchangeable \(\sigma\)-algebra \(\mathcal{E}_{[0,1]}\). Furthermore, on this event, for every \(M \geq 1\), we can partition \([0,1]\) into nonoverlapping subintervals \(J_1, \ldots, J_M\) such that the total variation of the mass transfer process within each subinterval exceeds \(M\). The mass transfer from color \(a\) to color \(b\) in any time interval \((s, t]\) is bounded above by \(Q_{s,t}(a, b)\); and so, to each \(J_i\), \(i = 1, \ldots, M\), there must be a further partition into \(N\) nonoverlapping subintervals \(J_{ij} = (s_{ij}, t_{ij})\) such that

\[
\sum_{j=1}^N Q_{s_{ij}, t_{ij}}(0, 1) \geq M \quad \text{and} \quad \sum_{j=1}^N Q_{s_{ij}, t_{ij}}(1, 0) \geq M. \tag{29}
\]

Recall that by Corollary 4.2, the coordinate process \((X^m_t)_{t \in [0,1]}\) is, conditional on the exchangeable \(\sigma\)-algebra \(\mathcal{E}_{[0,1]}\), an inhomogeneous continuous-time Markov chain on \([k]\) with transition probabilities \(Q_{s,t}\). Consequently, the conditional probability that \(X^1_t\) has no discontinuity in \(J_i\), given the event (29) and the exchangeable \(\sigma\)-algebra, is bounded above by
\[
\max \left( \prod_{j=1}^{N} (1 - Q_{s,j,t_j}(0, 1)), \prod_{j=1}^{N} (1 - Q_{s,j,t_j}(1, 0)) \right) \leq e^{-M}.
\]

Hence,

\[
P(X_{[0,1]}^1 \text{ has at least } M \text{ discontinuities } | \mathcal{E}_{[0,1]}) \geq 1 - Me^{-M},
\]
on the event that there exist nonoverlapping intervals \(J_{ij}\) satisfying (29). Since \(M\) is arbitrary, it follows that \(X_{[0,1]}^1\) must have infinitely many discontinuities, contradicting the hypothesis that the sample paths \((X_t)_{t \geq 0}\) are cadlag. This completes the proof for \(k = 2\).

The proof of Proposition 5.1 in the general case \(k \geq 2\) is a consequence of the same ideas as the \(k = 2\) case above. The key distinction is that mass can be transferred along incomplete cycles of the complete graph \(K_{[k]}\), and these cycles can vary measurably with the exchangeable \(\sigma\)-algebra of the process. However, since there are only finitely many states, unbounded variation and the pigeonhole principle demands that for every \(M \geq 1\), there must be at least one starting state \(j \in [k]\) for which \(Y_j^0 > 0\) and the conditional probability that \(X_{[0,1]}^1\) has at least \(M\) discontinuities given \(X_0^1 = j\) is at least \(1 - Me^{-M}\).

**Proof of Proposition 5.1 for \(k \geq 2\)**

Again, we suppose that there is positive probability that the supremum in (28) is \(+\infty\). Then the total variation must exceed every positive integer, and thus, it must exceed \(M^2k^3\), for every \(M \geq 1\). Therefore, there is some sequence of times \(0 = s_0 < s_1 < \cdots < s_N = 1\) such that \(\sum_{i=0}^{N-1} T_{s_i,s_{i+1}} \geq M^2k^3\). It follows that we can specify a collection \(J_1, \ldots, J_{M^2k}\) of nonoverlapping subintervals so that the total variation within each subinterval exceeds \(Mk^2\). In this way, for each \(i = 1, \ldots, Mk\), there is a further subpartition \(J_{i1}, \ldots, J_{iN}\) of \(J_i\) and a color \(\kappa_i^* \in [k]\) such that

\[
\sum_{j=1}^{N} Y_{s_{ij}}^\kappa^* Q_{s_{ij},t_j}(\kappa_i^*, [k] \setminus \{\kappa_i^*\}) \geq Mk.
\]

Furthermore, we have

\[
Y_{s_{ij}}^\kappa^* := \sum_{l=1}^{k} Y_0^l Q_{0,s_{ij}}(l, \kappa_i^*),
\]

and so there must be at least one \(\kappa_i^0 \in [k]\) for which

\[
\sum_{j=1}^{N} Y_0^{\kappa_i^0} Q_{0,s_{ij}}(\kappa_i^0, \kappa_i^*) Q_{s_{ij},t_j}(\kappa_i^*, [k] \setminus \{\kappa_i^*\}) \geq M.
\]
Since $k < \infty$, any such $\kappa_i^0$ must have $Y_0^{\kappa_i^0} > 0$. Consequently, within each subinterval, the conditional probability that $X_{[0,1]}^1$ experiences no discontinuities in $J_i$, given $X_0^1 = \kappa_i^0$ and the exchangeable $\sigma$-algebra, satisfies

$$P(X_{[0,1]}^1 \text{ has no discontinuity in } J_i \mid \mathcal{E}_i, X_0^1 = \kappa_i^0) \leq \prod_{j=1}^N \left(1 - Q_{0,s_{ij}}(\kappa_i^0, \kappa_i^*) Q_{s_{ij},s_{ij}}(\kappa_i^*, \{k \setminus \{\kappa_i^*\})\right) \leq e^{-M}.$$ 

By the pigeonhole principle, there must be some $r \in [k]$ such that $r = \kappa_i^0$ for at least $M$ indices $i = 1, \ldots, Mk$. In this case, let $1 \leq i_1 < \cdots < i_M \leq Mk$ be a subset of those indices for which $r = \kappa_i^0$. Then, we have

$$P(X_{[0,1]}^1 \text{ has less than } M \text{ discontinuities } \mid \mathcal{E}_{[0,1]}, X_0^1 = r) \leq \sum_{j=1}^M P(X_{[0,1]}^1 \text{ has no discontinuity in } J_{i_j} \mid \mathcal{E}_{[0,1]}, X_0^1 = r) \leq M e^{-M}.$$ 

Therefore,

$$P \left( X_{[0,1]}^1 \text{ has at least } M \text{ discontinuities } \mid \mathcal{E}_{[0,1]}, X_0^1 = r \right) \geq 1 - M e^{-M},$$

for every $M \geq 1$. Since $Y_0^r$ must be strictly positive for any such color, we conclude that there is positive probability that $(X_t^1)_{t \in [0,1]}$ has more than $M$ discontinuities in $[0,1]$, for every $M \geq 1$, which contradicts the assumption that $X_t^1$ has cadlag sample paths. We conclude that the projection $Y_t = \pi(X_t)$ must have locally bounded variation almost surely. \hfill \qed

6 Construction of Associated Chain on $[k]^N$

In this section, we prove Theorem 2.14, which states that for every Markov process $(Y_t)_{t \geq 0}$ on the $k$-simplex $\Delta_k$ whose sample paths are cadlag and of locally bounded variation, there exists an exchangeable Markov process $(X_t)_{t \geq 0}$ on $[k]^N$ whose projection (9) into $\Delta_k$ has the same distribution as $(Y_t)_{t \geq 0}$. The strategy is as follows. First,
we will show that the process $(Y_t)_{t \geq 0}$ uniquely determines a two-parameter process $(Q_{s,t})_{0 \leq s \leq t}$ taking values in the space of $k \times k$ stochastic matrices such that $(Q_{s,t})_{0 \leq s \leq t}$ satisfies the cocycle relations (11), the compatibility condition (12), and a minimality condition (32) spelled out below. Then, given the two-parameter process $(Q_{s,t})_{0 \leq s \leq t}$, we will construct an i.i.d. sequence of time-inhomogeneous Markov chains $X^i_t$ with transition probability matrices $Q_{s,t}$. Finally, we will show that $[k]^{\mathbb{N}}$-valued process

$$X_t = X^{i_1}_t X^{i_2}_t \cdots$$

is Markov, exchangeable, has cadlag paths and projects to $Y_t$.

**Proposition 6.1** If $(Y_t)_{t \in [0,a]}$ is a cadlag path of bounded variation in $\Delta_k$, then there exists a two-parameter family $(Q_{s,t})_{0 \leq s \leq t \leq a}$ of $k \times k$ stochastic matrices satisfying (11) and (12) such that for every subinterval $[c, d] \subseteq [0, a]$,

$$\sup_{c=s_0 \leq s_1 \leq s_2 \leq \cdots \leq s_n = d} \sum_{i=0}^{n-1} T_{s_i, s_{i+1}} = \| (Y_t)_{t \in [c,d]} \|_{TV}$$

(32)

where $\| \cdot \|_{TV}$ denotes the total variation norm on $\Delta_k$-valued paths and $T_{s,t}$ is the mass transfer function defined by (26). The mapping that sends the path $(Y_t)_{t \in [0,a]}$ to the two-parameter function $(Q_{s,t})_{0 \leq s \leq t \leq a}$ is measurable, and for each pair $c < d$, the section $(Q_{s,t})_{c \leq s \leq t \leq d}$ depends only on $(Y_t)_{t \in [c,d]}$.

**Proof** First, we consider the special case where the path $t \mapsto Y_t$ is smooth, and then, we will indicate the modifications necessary for the more general case where the path is not necessarily smooth, but is continuous and has (locally) bounded variation. Finally, we will indicate how to augment the construction to account for the possibility of jump discontinuities. Assume that $(Y_t)_{t \in [0,a]}$ is smooth, and set $DY_t = dY_t/dt$. Because each $Y_s$ is a probability distribution on the set $[k]$, its entries must sum to 1, and so the entries of the vector $DY_s$ must sum to 0 for almost every $s$. Let $J_+$, $J_-$, and $J_0$ be the sets of indices $i \in [k]$ such that $DY_s^i > 0$, $DY_s^i < 0$, and $DY_s^i = 0$, respectively, and define for each $s \in [0, a]$ a $k \times k$ rate matrix $R_s$ as follows:

$$R_s(i,j) = +DY_s^i DY_s^j / \sum_{i' \in J_-} DY_s^{i'} \quad \text{if } i \in J_- \text{ and } j \in J_+;$$

$$R_s(i,j) = -DY_s^i DY_s^j / \sum_{i' \in J_+} DY_s^{i'} \quad \text{if } i \in J_+ \text{ and } j \in J_-;$$

$$R_s(i,i) = DY_s^i; \quad \text{and}$$

$$R_s(i, j) = 0 \quad \text{otherwise.}$$

This matrix determines the instantaneous rates of mass flow between indices $i, j \in [k]$ following the MARX-ENGELS protocol (to each according to his needs; from each according to his abilities.) The two-parameter stochastic matrix-valued process $Q_{s,t}$ is then defined by the matrix exponential

$$e^{DS}.$$
\text{That this is in fact a stochastic matrix follows because the row sums of each } R_u \text{ are 0, and the compatibility condition (12) follows by a routine differentiation of (33). The total variation identity (32) follows from the fact that for any smooth function } f : [a, b] \to \mathbb{R}, \text{ the total variation of } f \text{ on the interval } [a, b] \text{ is}

\| f \|_{TV} = \int_a^b |Df(s)| \, ds.

\text{Next, consider the more general case where } Y_t \text{ is continuous and of locally bounded variation, but not necessarily smooth. The preceding construction may fail because, although } t \mapsto Y_t \text{ is differentiable at almost every } t \text{ (by the Lebesgue differentiation theorem), } Y_t - Y_s \text{ is not always equal to the integral of the derivative } DY_r \text{ over the interval } r \in [s, t], \text{ and this is needed for the compatibility condition (12). To circumvent this difficulty, we differentiate not with respect to } t \text{ but instead with respect to the total variation of the path. In particular, let}

T_t = \|(Y_s)_{s \in [0, t]}\|_{TV}

\text{be the total variation of the path } (Y_s)_{s \geq 0} \text{ on the time interval } [0, t], \text{ and set}

DY_t = \frac{dY_t}{dT_t};

\text{that is, } DY_t^i \text{ is the Radon-Nikodym derivative of the signed measure } [a, b] \mapsto Y_b^i - Y_a^i \text{ with respect to the positive measure } [a, b] \mapsto T_b - T_a. \text{ Since the signed measure is absolutely continuous with respect to the positive measure, it follows that for every } t < \infty,

Y_t - Y_0 = \int_0^t DY_s \, dT_s.

\text{Now define the rate matrices } R_s \text{ as above and set}

Q_{s,t} := \exp \left\{ \int_s^t R_u \, dT_u \right\}.

\text{The rest of the argument now proceeds as above. Observe that this is really a time-change argument and could be reformulated as such. In particular, if one defines } \tilde{Y}_{T_t} = Y_t, \text{ then the process } \tilde{Y}_s \text{ has differentiable paths with bounded derivative,}
and the argument for the differentiable case applies, producing a two-parameter process \((\tilde{Q}_{s,t})_{0 \leq s \leq t}\). Reversing the time change then gives the desired process \((Q_{s,t})_{0 \leq s \leq t}\).

Finally, suppose that \(t \mapsto Y_t\) has jump discontinuities at times \(r \in C\), where \(C\) is a countable set. Because \((Y_t)_{t \geq 0}\) has locally bounded variation, the jump sizes (measured by the \(L^1\)-norm on the simplex \(\Delta_k\)) are the sizes \(\Delta T_r\) of the jumps of the total variation function \(T_t\), which are summable over any bounded time interval. Define a new path \(\tilde{Y}_t\) by “opening” each jump discontinuity of \(Y_t\), that is, at each \(r \in C\) insert an interval of length \(\Delta T_r\), and let \(\tilde{Y}_t\) follow the straight-line path from \(Y_r^-\) to \(Y_r^+\) (at speed 1) during this inserted interval. The new path \(\tilde{Y}_t\) will still be of locally bounded variation, and it will be continuous, so the construction of the preceding paragraph will now yield a two-parameter family \(\tilde{Q}_{s,t}\) satisfying the cocycle identity (11) and the compatibility condition (12) for the path \(\tilde{Y}_t\). By “closing” all of the intervals opened in creating \(\tilde{Y}_t\), we obtain a two-parameter family \(Q_{s,t}\) that satisfies the compatibility condition (12) for the path \(Y_t\).

\[\square\]

**Proof of Theorem 2.14** Let \((Y_t)_{t \geq 0}\) be a Markov process on the simplex \(\Delta_k\) whose sample paths are cadlag and of locally bounded variation. By Proposition 6.1, there is a compatible random semigroup \((Q_{s,t})_{0 \leq s \leq t}\) such that for any \(c < d\), the section \((Q_{s,t})_{c \leq s \leq d}\) is a measurable function of the path \((Y_t)_{t \in [c,d]}\). Given the realization of \((Q_{s,t})_{0 \leq s \leq t}\), let \((X_t^i)_{t \geq 0}\) be i.i.d. copies of a time-inhomogeneous Markov chain on \([k]\) with transition probability matrices \(Q_{s,t}\) (cf. equation (13)) and initial distribution \(Y_0\), and let

\[X_t = X_1^1 X_1^2 \cdots .\]

This process is, by construction, exchangeable. It is also Markov because, for any \(s \geq 0\), the evolution of \(X_t\) for \(t \geq s\) depends, by construction, only on the state \(X_s\) and the section \((Q_{r,t})_{s \leq r \leq t}\) of the random semigroup, which in turn depends only on \((Y_t)_{t \geq s}\). That the projection of \(X_t\) to the simplex \(\Delta_k\) is \(Y_t\) follows by the strong law of large numbers and the compatibility condition (12): in particular,

\[
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \mathbf{1}(X_i^j = b) = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{a=1}^{k} \mathbf{1}(X_0^j = a \text{ and } X_i^j = b) \\
= \sum_{a=1}^{k} Y_0^a Q_{0,t}(a,b) \\
= Y_t^b.
\]

Finally, the fact that the paths of \((X_t)_{t \geq 0}\) are cadlag follows from the minimality condition (32), because this implies that the total mass transfer in any bounded time interval is finite, which implies that the number of jumps in any of the component chains \((X_i^j)_{t \geq 0}\) in a bounded time interval is finite.  

\[\square\]
7 The Feller Case

Theorem 2.13 associates to every exchangeable, cadlag process \((X_t)_{t \geq 0}\) on \([k]^N\) a compatible random semigroup \((Q_{s,t})_{0 \leq s \leq t}\). The random semigroup \((Q_{s,t})_{0 \leq s \leq t}\) is a two-parameter process on \(S_k\) that gives the transition probabilities of a time-inhomogeneous Markov process on \([k]\). The outcome of Theorem 2.13 can be viewed as a de Finetti-type characterization of exchangeable Markov processes \((X_t)_{t \geq 0}\) on \([k]^N\) with cadlag sample paths. In the special case when \((X_t)_{t \geq 0}\) is a Feller process, the cut-and-paste representation of \((X_t)_{t \geq 0}\), as proven in Theorem 2.6 of [5], implies a description of the compatible random semigroup as a Lévy matrix flow.

**Definition 7.1** A Lévy matrix flow is a collection \((Q_{s,t})_{0 \leq s \leq t}\) of stochastic matrices such that

1. for every \(s < t < u\), \(Q_{s,u} = Q_{s,t}Q_{t,u}\) a.s.,
2. for \(s < t\), the law of \(Q_{s,t}\) depends only on \(t - s\),
3. for \(s_1 < s_2 < \cdots < s_n\), the matrices \(Q_{s_1,s_2}, Q_{s_2,s_3}, \ldots, Q_{s_{n-1},s_n}\) are independent, and
4. \(Q_{0,0} = I_k\), the \(k \times k\) identity matrix, and \(Q_{0,t} \xrightarrow{p} I_k\) as \(t \downarrow 0\).

In Definition 7.1(iv), convergence is with respect to the total variation metric on \(S_k\):

\[
\|Q - Q'\|_{TV} := \sum_{i,j=1}^k |Q_{ij} - Q'_{ij}|, \quad Q, Q' \in S_k.
\]

Under the product-discrete topology on \([k]^N\), the combination of exchangeability and the Feller property is equivalent to the combination of exchangeability and consistency under subsampling; see discussion in Sect. 1 and reference to [4]. Hence, under the additional assumption that \((X_t)_{t \geq 0}\) is Feller, each restriction \((X^{[n]}_t)_{t \geq 0}\) of \((X_t)_{t \geq 0}\) to \([k]^{[n]}\) is unconditionally a time-homogeneous Markov chain on \([k]^{[n]}\), for every \(n \in \mathbb{N}\). When \((X_t)_{t \geq 0}\) is Feller, so is its projection \((Y_t)_{t \geq 0}\) into \(\Delta_k\). The following observation follows directly from the discussion in [5].

**Proposition 7.2** Let \((X_t)_{t \geq 0}\) be an exchangeable Feller process on \([k]^N\). Then its compatible random semigroup \((Q_{s,t})_{0 \leq s \leq t}\) is a Lévy matrix flow corresponding to a unique Feller process on \(S_k\). That is, there exists a unique Feller process \((S_t)_{t \geq 0}\) on \(S_k\) such that for all \(0 \leq s \leq t\),

- \(Q_{0,t} = S_t\),
- \(Q_{s,t} = \mathcal{L} S_{t-s}\), and
- \(Q_{s,t}\) is measurable with respect to \(\sigma (S_r)_{s \leq r \leq t}\).

**References**

1. Berestycki, J.: Exchangeable fragmentation-coalescence processes and their equilibrium measures. Electron. J. Probab. 9(25), 770–824 (2004). (electronic)
2. Bertoin, J.: Random Fragmentation and Coagulation Processes, volume 102 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (2006)
3. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
4. Billingsley, P.: Probability and Measure. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 3rd edn, A Wiley-Interscience Publication (1995)
5. Crane, H.: The cut-and-paste process. Ann. Probab. (2014, in press)
6. Dolgoarshinnykh, R., Lalley, S.: Critical scaling for the SIS stochastic epidemic. J. Appl. Probab. 43, 892–898 (2006)
7. Ethier, S., Kurtz, T.: Markov Processes. Probability and Mathematical Statistics. Wiley, New York, Wiley Series in Probability and Mathematical Statistics (1986)
8. Kurtz, T.: Solutions of ordinary differential equations as limits of pure jump Markov processes. J. Appl. Probab. 7, 49–58 (1970)
9. Hewitt, E., Savage, L.J.: Symmetric measures on cartesian products. Trans. Am. Math. Soc. 80, 470–501 (1955)
10. Kingman, J.F.C.: The coalescent. Stoch. Process. Appl. 13(3), 235–248 (1982)
11. Levin, D., Luczak, M., Peres, Y.: Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. Probab. Theory Relat. Fields 146, 223–265 (2010)
12. Pitman, J.: Combinatorial Stochastic Processes, volume 1875 of Lecture Notes in Mathematics. Springer, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard