STICK-BREAKING PROCESSES, CLUMPING, AND MARKOV CHAIN OCCUPATION LAWS

ZACH DIETZ, WILLIAM LIPPITT, SUNDER SETHURAMAN

Abstract. We connect the empirical or 'occupation' laws of certain discrete space time-inhomogeneous Markov chains, related to simulated annealing, to a novel class of 'stick-breaking' processes, a 'nonexchangeable' generalization of the Dirichlet process used in nonparametric Bayesian statistics. To make this unexpected correspondence, we examine an intermediate 'clumped' structure in both the time-inhomogeneous Markov chains and the stick-breaking processes, perhaps of its own interest, which records the sequence of different states visited and the scaled proportions of time spent on them. By matching the associated intermediate structures, we identify the limits of the empirical measures of the time-inhomogeneous Markov chains as types of stick-breaking processes.

1. Introduction

We consider the connections among discrete space time-inhomogeneous Markov chains related to simulated annealing, 'clumped' residual allocation models (RAMs), and a novel, general class of 'nonexchangeable' stick-breaking processes, which includes the Dirichlet process in a case. In particular, with respect to these time-inhomogeneous chains, we identify the empirical occupation law limits with stick-breaking processes in this class. The method to derive the correspondence involves a notion of intermediate structure found via a type of 'clumping' procedure, perhaps of its own interest.

On the one hand, the time-inhomogeneous Markov chains that we consider are stylized models of simulated annealing and Gibbs samplers; see [5], [12], [17], [19]. [50]. On the other hand, RAMs, Dirichlet processes, and stick-breaking processes have wide application in population genetics, ecology, combinatorial stochastic processes, and Bayesian nonparametric statistics; see books and surveys [9], [10], [20], [21], [29], [45] and references therein. A main purpose of the paper is to develop what seems to be an unexpected connection between these apriori different subjects.

In the next three subsections, we discuss first some of the relevant background on the time-inhomogeneous Markov chains considered, next RAMs, and stick-breaking processes, and last an informal summary of our main results on the empirical or 'occupation' laws of the Markov chains and stick-breaking processes via their 'clumped' intermediate structures.

1.1. **Time-inhomogeneous Markov chains studied.** We concentrate in this article on discrete spaces $\mathscr{X} \subseteq \mathbb{N}$, that is those composed of either a finite or a countably infinite number of elements. Let G be a generator kernel on $\mathscr{X}$, that

2010 *Mathematics Subject Classification.* 60E99, 60G57, 60J10.

Key words and phrases. RAM, GEM, Dirichlet, inhomogeneous, Markov, stick-breaking, empirical, clumping.

is $G_{i,j} \geq 0$ for $i \neq j \in \mathscr{X}$, and $G_{i,i} = -\sum_{j\neq i} G_{i,j}$. Suppose the entries of G are suitably bounded so that the kernel

$$K_n = I + \frac{G}{n} \tag{1.1}$$

is a stochastic kernel for all n large enough, and set $K_n = I$ otherwise. Suppose also that each K_n is irreducible and positive recurrent for sufficiently large n.

We will be interested in the time-inhomogeneous Markov chain $\{M_n\}_{n\geq 1}$ on $\mathscr{X}$ associated to kernels $\{K_n\}_{n\geq 1}$. In such chains, every point in $\mathscr{X}$ represents a valley from which the chain rarely but almost surely exits to enter another point valley. In this way, a certain 'landscape' is explored and the chains can be considered as simplified models of simulated annealing.

Interestingly, for finite $\mathscr{X}$, it was noted in [19] and [50] that the empirical distributions of these chains converge weakly, but not a.s. or in probability, as would be the case for a homogeneous Markov chain. Moreover, for generators G without zero entries, this weak convergence limit,

$$\nu_G(\cdot) \stackrel{d}{=} \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \delta_{M_j}(\cdot), \tag{1.2}$$

was identified in [12] by computing its moments. When $|\mathscr{X}| = 2$, one may see that ν_G is a Dirichlet distribution with parameters $(G(2,1), G(1,2))$. However, curiously, when G is of the form $G = \theta(Q - I)$ for $\theta > 0$ and Q is a stochastic 'constant' matrix with equal rows μ, it was also shown in [12] that ν_G has Dirichlet distribution with parameters $\{\theta\mu(i)\}_{i=1}^{k}$ by matching the moments.

In this context, a main goal is to understand more constructively, by considering an intermediate 'clumped' structure of the chain, the limit (1.2), and its generalization to countably infinite state space. Importantly, the construction allows to represent the limit as a 'stick-breaking' process. These processes are novel, and in particular non-Dirichlet, unless G is of form $G = \theta(Q - I)$ where Q is a 'constant' stochastic kernel (cf. Theorem 2.15). Given that the 'stick-breaking' apparatus is seminal in Bayesian statistics, the general representation provided here may be of potential use in applications.

We remark that the form of the scaling factor n^{-1} in the definition of K_n is not rigid–it can be say b_n where $nb_n \to 1$ (cf. Remark 2.14). However, if the factor were of the form $n^{-\gamma}$ for $\gamma \neq 1$, other phenomena occur: For instance, for finite $\mathscr{X}$, if $\gamma > 1$, the associated time-inhomogeneous Markov chain would fixate a.s. and, when $0 < \gamma < 1$, the empirical distributions would converge to a constant probability vector in probability (cf. [5], [12]).

Generalizations of the time-inhomogeneous Markov chains have been considered, such as coin-turning chains on two states [17], and freezing Markov chains on finitely many states [5]. In particular, among other results in [5], through stochastic approximation techniques, the limit law (1.2) was characterized as the stationary distribution of a certain continuous-time piecewise-deterministic Markov process; see also related results in [27] in the context of mRNA model applications. From a different view, related chains have been studied in terms of 'metastability' in the time-inhomogeneous context in [8], [38] and references therein, and also in the setting of arrays of stationary Markov chains (cf. [6], [33], [40]). However, the detailed connections of $\{M_n\}_{n\geq 1}$ and ν_G to 'stick breaking', the main focus of this article described more in Subsection 1.3, appears new.

We now turn to a discussion of RAMs and stick-breaking processes.

1.2. **GEM, Dirichlet, and stick-breaking measures.** Consider the infinite-dimensional simplex Δ_∞ of all discrete (probability) distributions on $\mathbb{N} = \{1, 2, \ldots\}$. A residual allocation model is a distribution on Δ_∞, introduced in the 1940's [26] as a means to address problems of apportionment: Let $\{X_n\}_{n\geq 1}$ be independent $[0, 1]$-valued random variables, called 'residual fractions'. Consider the associated process $\langle P_n : n \geq 1\rangle \in [0, 1]^{\mathbb{N}}$, given by $P_1 = X_1$ and

$$P_n = \left(1 - \sum_{j=1}^{n-1} P_j\right) X_n = (1 - X_1)\cdots(1 - X_{n-1})X_n \quad \text{for } n \geq 2;$$

see Lemma 3.1 for the induction leading to the last equality. If $\sum_{n\geq 1} P_n \overset{a.s.}{=} 1$, the distribution $\langle P_n : n \geq 1\rangle \in \Delta_\infty$ is the associated RAM. In general, $\langle P_n : n \geq 1\rangle$ need not sum to 1 for a given realization. We note a simple condition equivalent to $\sum_{n\geq 1} P_n \overset{a.s.}{=} 1$ is that $\prod_{j=1}^{\infty}(1 - X_j) \overset{a.s.}{=} 0$, the case for nontrivial, independent, identically distributed (iid) fractions (cf. Lemma 3.1).

The RAM when the fractions $\{X_n\}_{n\geq 1}$ are iid Beta$(1, \theta)$ random variables is the well-known Griffiths-Engen-McCloskey (GEM) model with parameter θ. There are many characterizations and studies of the GEM sequence and its variants in recent years. For instance, the GEM model is the unique RAM with iid fractions that is invariant in law under size-biased permutation. Also, the GEM sequence is the unique invariant measure of 'split and merge' dynamics. In addition, there are important connections with Poisson-Dirichlet models. See for instance, among others, [1], [2], [11], [16], [22], [30], [31], [32], [37], [41], [43], [44], [46], and references therein.

Moreover, the GEM sequence is a fundamental building block of Dirichlet processes, which often serve as a measure on priors in Bayesian nonparametric statistics [20], [21]. With respect to a measurable space $(\mathscr{Y}, \mathscr{B})$, consider the space of probability measures $\mathbb{P}_{\mathscr{Y}}$ endowed with σ-field generated by the sets $\{P : P(A) < r\}$ for $A \in \mathscr{B}$ and $r > 0$. We say that D is a random probability sample from the Dirichlet process, with 'parameters' $\theta > 0$ and probability measure μ on $\mathscr{Y}$, if for any finite partition $\{A_i\}_{i=1}^m$ the vector $\langle D(A_1), \ldots, D(A_m)\rangle$ has the Dirichlet distribution with parameters $\langle\theta\mu(A_i) : 1 \leq i \leq m\rangle$.

The 'stick-breaking' representation of the Dirichlet process with parameters (θ, μ), in terms of a GEM(θ) sequence $\langle P_i : i \geq 1\rangle$, and an independent sequence of iid random variables $\{W_i\}_{i\geq 1}$ with common distribution μ, is given by

$$D(\cdot; \theta, \mu) = \sum_{i=1}^{\infty} P_i \delta_{W_i}(\cdot). \tag{1.3}$$

There is a large literature on Dirichlet processes stemming from the seminal works [4], [18]. See [44], [49] with respect to the 'stick-breaking' construction, and books [20], [21], [39], [45] for more on their history, other representations including that with respect to the 'Chinese restaurant process', and their use in practice.

We note, when $\mathscr{Y} = \{1, \ldots k\}$ is a finite space, $\mu = \langle\mu(1), \ldots, \mu(k)\rangle$ and $A_i = \{i\}$ for $1 \leq i \leq k$, the property that $\langle D(A_1), \ldots, D(A_k)\rangle$ is given by a Dirichlet distribution was first stated in a population genetics context in [14]; see also [28].

In this article, we focus on a class of generalized stick-breaking processes on $\mathscr{X} \subseteq \mathbb{N}$. Let $\langle P_i : i \geq 1\rangle$ be a GEM(θ) sequence and let $\{T_i\}_{i\geq 1}$ be an independent stationary Markov chain with irreducible transition kernel Q and stationary distribution μ on $\mathscr{X}$. The random measures, in stick-breaking form,

$$\nu(\cdot;\theta,\mu,Q) = \sum_{i=1}^{\infty} P_i \delta_{T_i}(\cdot), \tag{1.4}$$

are natural, and what seems to be novel generalizations of the stick-breaking representation of the Dirichlet process, here with respect to stationary Markovian samples $\{T_i\}_{i\geq 1}$ instead of the iid ones in (1.3). We also note that other generalizations of Dirichlet processes have been considered, among them, Polya tree [35], Pitman-Yor [44], [47], and Beta processes [7].

In particular, we will show that ν satisfies a 'self-similarity' equation (cf. Theorem 2.19), uniquely characterizing its distribution. This equation is reminiscent of the regenerative structure present in the stick-breaking representation of the Dirichlet process [49], in integral constructions of the Dirichlet processs [34], [48], and in other related settings [24], [23].

We note however that ν is not 'permutation exchangeable' when the Markov chain $\{T_i\}_{i\geq 1}$ is not an iid sequence in the sense that the GEM sequence $\langle P_i : i \geq 1\rangle$ may not be replaced by an arbitrary finite permutation without changing the measure (cf. Theorem 2.21). In contrast, when $\{T_i\}_{i\geq 1}$ is iid and ν is the Dirichlet process, such an exchangeability property holds. For example, the Poisson-Dirichlet order statistics $\langle \hat{P}_i : i \geq 1\rangle$ of $\langle P_i : i \geq 1\rangle$ may be used instead without changing the Dirichlet process (cf. [44]).

Given this 'nonexchangeability', tools and representations, standard with respect to the exchangeable Dirichlet process, such as Poisson-Dirichlet statistics and Gamma subordinators, do not seem readily applicable. We will mostly rely on the nuts-and-bolts definition of the GEM measure, the structure of the Markov chain $\{T_i\}_{i\geq 1}$, and the 'clumping' procedure to analyze ν.

1.3. **Clumping in time-inhomogeneous Markov chains and stick-breaking processes.** Returning to the time-inhomogeneous Markov chain $\mathbf{M} = \{M_n\}_{n\geq 1}$ with kernels $\{K_n\}_{n\geq 1}$ (1.1), starting from initial distribution π, consider the random empirical occupation measure on $\mathscr{X}$,

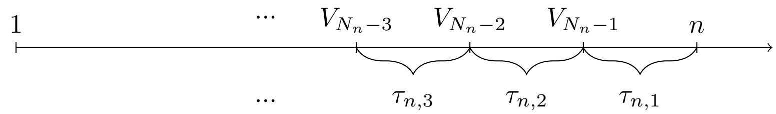
$$\nu_n(\cdot) = \frac{1}{n}\sum_{j=1}^{n} \delta_{M_i}(\cdot).$$

One might feel there is some resemblance between ν_n and ν in (1.4) as both express sums of point masses with weights adding to 1. To make this more precise, we implement a 'reverse' clumping procedure to gather local occupations of the same state, or clumped occupations, of the empirical measure of $\mathbf{M}$ up to time n.

In a Markov chain with kernels $\{K_n\}_{n\geq 1}$, later clumps of the chain on a state are typically larger than earlier clumps. To keep the clump sizes from tending to zero after normalization, we consider the clumps in reverse chronological order, starting from time n, so that the clumped occupations may converge nontrivially in distribution.

Formally, let $1 = V_1 < V_2 < \cdots$ be the successive times when the Markov chain changes state, and let $N_n = \min\{i : V_i > n\}$. Going backwards from time n, let $\tau_{n,1}$ be the length $n + 1 - V_{N_n - 1}$ of the last visit to state $Y_{n,1} = M_{V_{N_n-1}}$, $\tau_{n,2}$ be the length $V_{N_n-1} - V_{N_n-2}$ of the visit to state $Y_{n,2} = M_{V_{N_n-2}}$, and $\tau_{n,k}$ be the length $V_{N_n-(k-1)} - V_{N_n-k}$ of the visit to $Y_{n,k} = M_{V_{N_n-k}}$ for $1 < k < N_n$. Let also $\tau_{n,k} = 0$ and $Y_{n,k} = M_1$ for $k \geq N_n$. In addition, define $P_{n,k} = \tau_{n,k}/n$ for $k \geq 1$.

The figure below depicts, in a realization, the clumping boundaries V_j marked in forward times, and the lengths of local occupations $\tau_{n,j} = nP_{n,j}$ given backwards in time starting from time n.



Then, ν_n is written as

$$\nu_n(\cdot) = \sum_{j=1}^{N_n-1} P_{n,j}\delta_{Y_{n,j}}(\cdot) = \sum_{j=1}^{\infty} P_{n,j}\delta_{Y_{n,j}}(\cdot).$$

We now identify an intermediate structure in the time-inhomogeneous Markov chain. We show (cf. Theorem 2.10) that as $n \to \infty$, $\mathbf{Y}_n = \{Y_{n,j}\}_{j\geq 1}$ converges to a stationary homogeneous Markov chain $\mathbf{Z} = \{Z_j\}_{j\geq 1}$ whose transition probability of moving from i to a different state j is $\mu_j G_{j,i}/[\mu_i(-G_{i,i})]$, in terms of G and μ, where μ is the stationary eigenvector of G. Also, conditionally on the values $\{Y_{n,j}\}_{j\geq 1}$, the distributions of $\mathbf{P}_n = \langle P_{n,j} : j \geq 1\rangle$ converge as $n \to \infty$ to a 'disordered' GEM $\mathbf{R} = \langle R_j : j \geq 1\rangle$, a RAM with independent fractions distributed as $\{\text{Beta}(1, -G_{Z_i,Z_i})\}_{i\geq 1}$. In particular, the joint law of $\langle P_{n,j} : j \geq 1\rangle$ and $\{Y_{n,j}\}_{j\geq 1}$ converges as $n \to \infty$ to a joint distribution, characterized in terms of G' where $G'_{i,j} = \frac{\mu_j}{\mu_i} G_{j,i}$ for $i, j \in \mathscr{X}$, and dubbed the MCcGEM(G') distribution with respect to μ.

Importantly, however, to match the limit of ν_n to ν in (1.4), we will need to 'clump' ν and consider its clumped intermediate structure and correspondence to the intermediate structure MCcGEM distribution arising above from the time-inhomogeneous Markov chain.

To this end, analogous to the switching times with respect to $\mathbf{M}$, suppose $\{V_i\}_{i\geq 1}$ are the times when the stationary Markov chain $\mathbf{T} = \{T_i\}_{i\geq 1}$ changes its state with the convention $V_1 = 1$. With respect to $\mathbf{P} = \langle P_i : i \geq 1\rangle$, consider $P_i^V = \sum_{j=V_i}^{V_{i+1}-1} P_j$ for $i \geq 1$. We show that (cf. Theorems 2.4 and 2.7) the law of $\mathbf{Y} = \{Y_i = T_{V_i}\}_{i\geq 1}$ can be computed as another Markov chain on $\mathscr{X}$ with a transition kernel found in terms of Q, where the transition probability of moving from i to a different state j is $Q_{i,j}/(1 - Q_{i,i})$. Also, conditional on the locations $\{Y_i\}_{i\geq 1}$, the sequence $\mathbf{P^V} = \langle P_i^V : i \geq 1\rangle$ is a RAM whose associated fractions are $\{\text{Beta}\big(1, \theta(1 - Q_{Y_i,Y_i})\big)\}_{i\geq 1}$, another 'disordered' GEM. Indeed, the joint law $\big(\langle P_i^V : i \geq 1\rangle, \{Y_i\}_{i\geq 1}\big)$ is a MCcGEM$\big(\theta(Q - I)\big)$ distribution with respect to μ.

In terms of this clumped structure with respect to ν, we may write that

$$\nu(\cdot) = \sum_{i=1}^{\infty} P_i^V \delta_{Y_i}(\cdot). \tag{1.5}$$

We are now in position to identify the intermediate structure and limit of the empirical measures ν_n of the time-inhomogeneous Markov chain in terms of the clumped structure of a stick-breaking measure ν. In particular, when $G' = \theta(Q - I)$, these intermediate structures match.

We state in Theorem 2.12 that ν_n converges to a matched random measure ν either in 'stick-breaking' or 'clumped' forms (1.4), (1.5), and vice versa, in Corollary 2.13, starting from the stick-breaking process ν, we identify it as the limit of the empirical measure of a matched time-inhomogeneous Markov chain.

We show in Theorem 2.15 that ν is *not* a Dirichlet process unless $G' = \theta(Q - I)$ and Q is a constant stochastic matrix with identical rows μ. In this case, the matched Markov chain $\{T_i\}_{i\geq 1}$ is actually composed of iid random variables with distribution μ, and hence the limit ν is seen as a Dirichlet process (cf. remark after (1.2)). See Subsection 2.3 for further remarks, and what appears to be new representations of the Dirichlet process.

In the figure below, the main ideas, relationships and identifications between the empirical measure of the time-inhomogeneous Markov chain, the stick-breaking process, and their clumped intermediate structures, are summarized.

$$\begin{array}{lll}
\sum_{j=1}^\infty \frac{1}{n}\delta_{M_j} = & \sum_{j=1}^\infty P_{n,j}\delta_{Y_{n,j}} & \qquad \mathbf{M} \xrightarrow{\text{rev. cl.}} (\mathbf{P}_n, \mathbf{Y}_n) \\
& & \qquad \qquad \qquad \quad \downarrow \mathrm{d} \\
\qquad \overset{\mathrm{d}}{\to} & \sum_{j=1}^\infty R_j \delta_{Z_j} & \qquad \qquad \qquad (\mathbf{R}, \mathbf{Z}) \\
& & \qquad \qquad \qquad \quad \overset{\mathrm{d}}{=} \\
\qquad \overset{\mathrm{d}}{=} & \sum_{j=1}^\infty P_j^V \delta_{Y_j} = \sum_{j=1}^\infty P_j \delta_{T_j} & \qquad \qquad \qquad (\mathbf{P^V}, \mathbf{Y}) \quad \xleftarrow{\text{cl.}} (\mathbf{P}, \mathbf{T})
\end{array}$$

Finally, we mention that extensions of our results to a class of reducible generators G is possible, with more notation and some modifications of the proofs, are available in the extended arXiv version [13].

Organization of the paper. The main results, Theorems 2.4, 2.7, 2.10, 2.12, 2.13, 2.15, 2.19, and 2.21 are stated in Section 2. Proofs are then given in Section 3.

2. Statement of results

We now formalize notation and state our main results, and related remarks about them, in several subsections. It will be convenient to develop notions from the bottom to the top with respect to schema figure at the end of Section 1.

We will use the convention that empty sums equal 0, and empty products are 1. Also, $1/0 = \infty$, $0/0 = 0$, and $0^0 = 1$. The notation v^t signifies that the vector v is in row form.

2.1. **RAMs, GEMs and MCcGEM laws.**

A residual allocation model (RAM) is a way of defining a random probability measure on $\mathbb{N}$ by iteratively assigning a random portion of the unassigned probability remaining to the next integer.

Definition 2.1 (Residual Allocation Model - RAM)**.** *Let $\mathbf{X} = \{X_j\}_{j\geq 1}$ be a collection of independent $[0,1]$-valued random variables. Define*

$$P_1 = X_1 \quad \text{and} \quad P_j = X_j\left(1 - \sum_{i=1}^{j-1} P_i\right) \quad \text{for} \quad j \geq 2. \tag{2.1}$$

Then, if $\mathbf{P} = \langle P_j : j \geq 1\rangle$ is a.s. a probability measure on $\mathbb{N}$, that is if $\sum_{j=1}^\infty P_j \overset{a.s.}{=} 1$, we say $\mathbf{P}$ is a RAM. If $\mathbf{X}$ consists of iid fractions, and the associated $\mathbf{P}$ is a RAM, we say $\mathbf{P}$ is a self-similar RAM.

Consider now the following identity, verified in Lemma 3.1: For an arbitrary sequence of numbers $\{a_j\}_{j\geq 1}$ and $k \geq 1$,

$$\prod_{j=1}^{k}(1-a_j) + \sum_{j=1}^{k} a_j \prod_{i=1}^{j-1}(1-a_i) = 1. \tag{2.2}$$

Then, the sequence in (2.1) satisfies $P_j = X_j \prod_{i=1}^{j-1}(1-X_i)$ for $j \geq 1$ (cf. Proposition 3.2). Accordingly, we have the useful observation that $\mathbf{P}$ is a RAM exactly when $\prod_{j\geq 1}(1-X_j) \overset{a.s.}{=} 0$.

A specific and well-known example of a RAM is the Griffiths-Engen-McCloskey (GEM) sequence.

Definition 2.2 (GEM)**.** *Fix $\theta > 0$. Let $\mathbf{X} = \{X_j\}_{j\geq 1}$ be a sequence of iid variables with common distribution $Beta(1, \theta)$. Then, the self-similar RAM $\mathbf{P}$, constructed from $\mathbf{X}$, is said to be a GEM(θ) distribution.*

Also, consider a sequence $\{\theta_j\}_{j\geq 1}$ of positive numbers, and let $\mathbf{X}$ be a sequence of independent random variables where $X_j \sim \mathrm{Beta}(1, \theta_j)$ for $j \geq 1$. When the measure $\mathbf{P}$, found in terms of $\mathbf{X}$, is a RAM, we will say it is a disordered GEM sequence with parameters $\{\theta_j\}_{j\geq 1}$.

Now, in a RAM $\mathbf{P}$, one can clump adjacent probabilities with respect to an increasing sequence $\mathbf{u}$, marking boundaries of clumps, to form a new probability measure $\mathbf{P}^{\mathbf{u}}$ on $\mathbb{N}$.

Definition 2.3 (Clumped measure)**.** *Let $\mathbf{u} = \{u_j\}_{j\geq 1}$ be an increasing sequence in $\mathbb{N}$ with $u_1 = 1$ and let $\mathbf{P}$ be a RAM. We clump $\mathbf{P}$ according to $\mathbf{u}$ to construct a new probability measure $\mathbf{P}^{\mathbf{u}} = \langle P_j^u : j \geq 1\rangle$ on $\mathbb{N}$ where, for $j \geq 1$,*

$$P_j^u = \sum_{i=u_j}^{u_{j+1}-1} P_i$$

A natural question is when $\mathbf{P}^{\mathbf{u}}$ is also a RAM. Although it is not difficult to see that $\mathbf{P}^{\mathbf{u}}$ is always a RAM when $\mathbf{u}$ is deterministic, the situation is more involved when a random sequence is used for the clumping.

Specifically, we will be interested in a random clumping sequence $\mathbf{V}$ constructed from a stationary, irreducible Markov chain $\mathbf{T} = \{T_i\}_{i\geq 1}$ on the discrete space $\mathscr{X} \subseteq \mathbb{N}$. The sequence $\mathbf{V}$ keeps track of the times when $\mathbf{T}$ switches values between repeated values in $\mathbf{T}$.

For example, if $\mathbf{T} = (1, 1, 2, 2, 2, 2, 4, 1, 1, 5, \ldots)$ is observed, we define $\mathbf{V} = (1, 3, 7, 8, 10, \ldots)$. More formally, let $V_1 = 1$ and, for $j \geq 1$, set

$$V_{j+1} = \inf\{v > V_j : T_v \neq T_{v-1}\} \tag{2.3}$$

Since $\mathbf{T}$ is irreducible, $\mathbf{V}$ is almost surely well-defined, consisting of an increasing sequence of integers.

Define now $\mathbf{Y} = \{Y_j\}_{j\geq 1}$ by $Y_j = T_{V_j}$ for $j \geq 1$. We think of $\mathbf{Y}$ as the sequence of values taken by $\mathbf{T}$ without repetition.

In what follows, we will say that a sequence $\mathbf{z}$ is a 'possible' sequence for a Markov chain $\mathbf{Z}$ on $\mathscr{X}$ if the event $\{Z_i = z_i : 1 \leq i \leq n\}$ has positive probability for each $n \geq 1$.

Theorem 2.4 (Clumped RAMs)**.** *Let $\mathbf{P}$ be a RAM. Fix an increasing sequence $\mathbf{u} = \{u_j\}_{j\geq 1}$ in $\mathbb{N}$ with $u_1 = 1$. Then,*

(1) $\mathbf{P^u}$ *is a RAM with respect to fractions* $\mathbf{X^u} = \{X_j^u\}_{j\geq 1}$ *where*

$$X_j^u = \sum_{i=u_j}^{u_{j+1}-1} X_i \prod_{l=u_j}^{i-1}(1-X_l) = 1 - \prod_{i=u_j}^{u_{j+1}-1}(1-X_i)$$

Let now $\mathbf{T} = \{T_j\}_{j\geq 1}$ *be an irreducible, positive-recurrent Markov chain, independent of* $\mathbf{P}$ *and with homogeneous transition kernel* Q *and stationary initial distribution* μ.

(2) *Then, the sequence* $\mathbf{Y} = \{T_{V_j}\}_{j\geq 1}$ *is a Markov chain with homogeneous transition kernel* K *given by*

$$K(z,w) = \frac{Q_{z,w}}{1-Q_{z,z}}\mathbf{1}(z \neq w),$$

with initial distribution μ.

Let $\mathbf{t}$ *be a possible sequence in* $\mathscr{X}$ *with respect to* $\mathbf{T}$*. Let* $\mathbf{y}$ *be a possible sequence in* $\mathscr{X}$ *with respect to* $\mathbf{Y}$*. Define* $\mathbf{V}$ *with respect to* $\mathbf{T}$ *as in* (2.3).

(3) *Then,* $\mathbf{P^V}\big|\mathbf{T} = \mathbf{t}$ *is a RAM.*
(4) *Also, if* $\mathbf{P}$ *is self-similar, then* $\mathbf{P^V}\big|\mathbf{Y} = \mathbf{y}$ *is a RAM.*

We remark that the specifications of the fractions and their distributions in items (4) are given in the proof of Theorem 2.4. These specifications, in the case when $\mathbf{P}$ is a GEM(θ) sequence, are part of Theorem 2.7.

Also, in item (4) above, we note that the self-similarity of $\mathbf{P}$ is important to deduce that $\mathbf{P^V}|\mathbf{Y}$ is a RAM; see for instance Example 2.9.

We now consider the clumping procedures with respect to a GEM distribution $\mathbf{P}$. It will be convenient to define the notion of a generator kernel or matrix, these terms used interchangeably.

Definition 2.5 (Generator kernel)**.** *Let* $G = \{G_{i,j} : i,j \in \mathscr{X}\}$ *be a square matrix on* $\mathscr{X}$*. We say that* G *is a generator kernel if it satisfies* $G_{i,j} \geq 0$ *for* $i \neq j$ *and* $G_{i,i} = -\sum_{j\neq i} G_{i,j}$*. In addition, we will assume a boundedness condition,* $\sup_i |G_{i,i}| < \infty$*.*

Every matrix of the form $G = \theta(Q - I)$, where $\theta > 0$ and Q is a stochastic kernel on $\mathscr{X}$, is a generator matrix. Moreover, we claim that every generator matrix can be (non-uniquely) decomposed in this fashion: The final condition in Definition 2.5 ensures that all entries are bounded, $\sup_{l,k}|G_{l,k}| \leq \sup_i |G_{i,i}| < \infty$, so that a normalizing θ and $Q = I + G/\theta$ can be found.

We will say that G is irreducible, positive-recurrent when an associated Q is an irreducible, positive-recurrent kernel on $\mathscr{X}$. We also say G has a stationary distribution μ when μ is a left-eigenvector of G with eigenvalue 0, or equivalently when μ is a stationary distribution of an associated Q. We note, in an irreducible generator G, the diagonal entries $G_{w,w} \neq 0$, that is there cannot be zero rows in G, and also any stationary distribution μ is unique, with positive entries, $\mu_i > 0$ for $i \in \mathscr{X}$.

We now formally define the notion of a Markov Chain conditional GEM (MC-cGEM) joint distribution on the space $[0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$, endowed with the product topology and product σ-field formed in terms of the Borel σ-fields on $[0,1]$ and $\mathscr{X}$. This topology is discussed more in Subsection 3.4.

Definition 2.6 (MCcGEM distribution)**.** *With respect to an irreducible generator matrix G, let $\mathbf{Z}$ be a homogeneous Markov chain with initial distribution μ and transition kernel K_G on $\mathscr{X}$ given by*

$$K_G(w,z) = \frac{G_{w,z}}{-G_{w,w}}\mathbf{1}(z \neq w) \tag{2.4}$$

Consider variables $\mathbf{X} = \{X_j\}_{j\geq 1}$, on the same probability space as $\mathbf{Z}$, such that $X_j\big|\mathbf{Z} = \mathbf{z} \sim$ Beta$(1, -G_{z_j,z_j})$ and $\{X_j\big|\mathbf{Z} = \mathbf{z}\}_{j\geq 1}$ are independent. Define $\mathbf{R}$ where $R_j = X_j \prod_{i=1}^{j-1}(1 - X_i)$ for $j \geq 1$, and observe that $\mathbf{R}\big|\mathbf{Z} = \mathbf{z}$ is a disordered GEM with parameters $\{-G_{z_j,z_j}\}_{j\geq 1}$ (see below).

We say that the pair $(\mathbf{R}, \mathbf{Z})$ has MCcGEM(G) distribution with respect to μ.

To see that $\mathbf{R}\big|\mathbf{Z} = \mathbf{z}$ is a disordered GEM, we need only observe that $\mathbf{R}\big|\mathbf{Z} = \mathbf{z}$ is a probability distribution on $\mathbb{N}$. Here, $\prod_{n\geq 1}(1 - X_n)\big|\big(\mathbf{Z} = \mathbf{z}\big) = 0$ a.s. exactly when $\sum_{n\geq 1} X_n\big|\mathbf{Z} = \mathbf{z}$ diverges a.s. As the tail σ-field is trivial, the opposite is the summability $\sum_{n\geq 1} X_n\big|\big(\mathbf{Z} = \mathbf{z}\big) < \infty$ a.s. By Kolmogorov's 3-series theorem, and that $\mathbf{X}|\mathbf{Z} = \mathbf{z}$ is composed of Beta random variables on $[0, 1]$ with means $\{(1 - G_{z_j,z_j})^{-1}\}_{j\geq 1}$ and variances dominated by the means, almost sure summability holds exactly when $\sum_{j\geq 1} |G^{-1}_{z_j,z_j}| < \infty$. For a generator matrix G, this is never the case as the terms $\{|G_{x,x}|\}_{x\in\mathscr{X}}$ are uniformly bounded above.

We now describe a relation between GEM distributions and MCcGEM laws through clumping with respect to a homogeneous Markov chain.

Theorem 2.7 (GEM to MCcGEM)**.** *Let $\theta > 0$ and $\mathbf{P}$ be GEM(θ) distribution. Let also $\mathbf{T} = \{T_j\}_{j\geq 1}$ be an independent, irreducible homogeneous Markov chain with kernel Q and initial distribution μ. Recall the associated switch times $\mathbf{V}$, the clumped distribution $\mathbf{P}^{\mathbf{V}}$, and the Markov chain $\mathbf{Y}$ near* (2.3)*.*

Then, $\mathbf{Y}$ is a homogeneous Markov chain with kernel

$$K_{\theta(Q-I)}(w,z) = \frac{Q_{w,z}}{1 - Q_{w,w}}1(w \neq z),$$

and initial distribution μ. Also, $\mathbf{P}^{\mathbf{V}}|\mathbf{Y} = \mathbf{y}$ is a disordered GEM with parameters $\{\theta(1 - Q_{y_j,y_j})\}_{j\geq 1}$. Hence, $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y})$ has MCcGEM$(\theta(Q - I))$ distribution with respect to μ.

Some cases of interest are developed in the following examples.

Example 2.8. Suppose $\mathbf{P} \sim$ GEM(θ) and that $\mathbf{T}$ is a stationary homogeneous Markov chain with stochastic kernel Q where Q has constant diagonal entries, $Q_{i,i} = q$ for $i \in \mathscr{X}$. By Theorem 2.7, $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y}$ is a disordered GEM sequence with parameters $\{\theta(1-Q_{Y_i,Y_i})\}_{i\geq 1}$. However, since $Q_{Y_i,Y_i} \equiv q$, we conclude $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y} = \mathbf{P}^{\mathbf{V}}$ does not depend on $\mathbf{Y}$ and is actually a GEM$(\theta(1 - q))$ sequence. In this case, the pair $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y})$ consists of independent sequences.

Example 2.9. We now consider a RAM $\mathbf{P}$ constructed from independent fractions $X_j \sim$ Beta(1/2, 1+j/2) for $j \geq 1$. Such a RAM is a member of the well-known 2-parameter GEM(α, θ) family, here with $\mathbf{P} \sim$ GEM(1/2,1). Let $\mathbf{T}$ be a sequence of iid Bernoulli(1/2) variables. Thought of as a Markov chain on the 2-state space $\mathscr{X} = \{1, 2\}$, every entry of the stochastic kernel Q of $\mathbf{T}$ equals 1/2. Here, the increments $\{V_{i+1} - V_i : i \geq 1\}$ do not depend on $\mathbf{Y}$, and hence $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y} = \mathbf{P}^{\mathbf{V}}$.

We now observe that $\mathbf{P}^{\mathbf{V}}$ is not a RAM with respect to natural fractions $\mathbf{X}^{\mathbf{V}}$: If it were a RAM, consider the fractions $\mathbf{X}^{\mathbf{V}}$ (cf. given in part (1) of Theorem 2.4 by substituting $\mathbf{V}$ for $\mathbf{u}$). One may compute the covariance $\mathrm{Cov}[1-X_1^V, 1-X_2^V] \approx -.005391$, and so the fractions X_1^V and X_2^V are not independent, and $\mathbf{P}^{\mathbf{V}}$ cannot be an associated RAM. See the extended arXiv version [13] for more details.

2.2. **Clumping in Markov chains.** The notion of clumping can be applied to random probability measures on $\mathbb{N}$ which are not RAMs. In particular, to capture the empirical occupation law limit of a Markov chain, we study its local occupations, or clumps of the sequence indexed in time, as it explores the space $\mathscr{X}$. As noted in the introduction, we will look at these local occupations in reverse order.

Let $\mathbf{M} = \{M_j\}_{j\geq 1}$ be a Markov chain on the discrete space $\mathscr{X}$, which changes states infinitely often a.s. Recall the definition of the switching times $\mathbf{V}$ (cf. (2.3)). Let $N_n = \min\{i : V_i > n\}$ index the first switch after time n, and note that $N_n \uparrow \infty$ a.s. For $1 < k < N_n \leq i$ and $j \geq 1$, define

$$\tau_{n,1} = n + 1 - V_{N_n-1}, \quad \tau_{n,k} = V_{N_n-(k-1)} - V_{N_n-k}, \quad \text{and} \quad \tau_{n,i} = 0.$$

Also, set

$$Y_{n,1} = M_n = M_{V_{N_n-1}}, \quad Y_{n,k} = M_{V_{N_n-k}}, \quad \text{and} \quad Y_{n,i} = M_1, \tag{2.5}$$

and $P_{n,j} = \tau_{n,j}/n$. Consider the sequences $\mathbf{P}_n = \langle P_{n,j} : j \geq 1\rangle$ and $\mathbf{Y}_n = \{Y_{n,j}\}_{j\geq 1}$.

As a concrete example, consider an observation

$$\mathbf{M} = (1, 1, 1, 6, 6, 1, 3, 3, 3, 5, \ldots).$$

Then for $n = 4$, the local occupations are summarized by eventually constant sequences $\mathbf{P}_4 = (1/4, 3/4, 0, 0, 0, 0, \ldots)$ and $\mathbf{Y}_4 = (6, 1, 1, 1, 1, 1, \ldots)$. Similarly, when $n = 7$, we have $\mathbf{P}_7 = (1/7, 1/7, 2/7, 3/7, 0, 0, 0, \ldots)$ and $\mathbf{Y}_7 = (3, 1, 6, 1, 1, 1, 1, \ldots)$. For a more general depiction, please refer to the figure in Section 1.3.

Hence, for $l \in \mathscr{X}$, we have generally that

$$\nu_n(l) := \frac{1}{n}\sum_{j=1}^{n} \delta_{M_j}(l) = \sum_{j=1}^{\infty} P_{n,j}\delta_{Y_{n,j}}(l).$$

In the middle of the display, we see the average Markov chain $\mathbf{M}$ occupation of state l in the first n steps. On the right-hand side, the sum is over local occupations, or clumps, of states seen in the chain $\mathbf{M}$ through n steps. The notion suggested by this relation is that we may study the limit average occupation law of $\mathbf{M}$ by investigating the limit of the pair $(\mathbf{P}_n, \mathbf{Y}_n)$ describing local occupations.

We now focus on a class of time-inhomogeneous Markov chains for which the limits of $(\mathbf{P}_n, \mathbf{Y}_n)$ have succinct representation. Specifically, we consider inhomogeneous Markov chains $\mathbf{M}$ with transition kernels of the form $I + G/n$, where G is an irreducible positive-recurrent generator matrix. A finite space $\mathscr{X}$ case where G was taken to have no zero entries at all was studied in [12]. See also [5], [17] for related developments.

In these chains, states change infinitely often a.s. and the clump lengths $V_k - V_{k-1}$ are typically growing with k, unlike for homogeneous Markov chains. In particular, rather than an ergodic theorem, it was shown in [12] (cf. (1.2)) that the occupation laws converge weakly to a nontrivial distribution. Here, we consider a countable space generalization and formulate a characterization of these occupation limits through the reversed clumping device described above.

In the following statement, we say that a matrix is non-negative if all its entries are non-negative. Additionally, weak convergences here are in the sense of finite-dimensional distributions, the natural sense associated to the product space $[0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$ endowed with the product topology.

Theorem 2.10 (Time-inhomogenous MC to MCcGEM)**.** *Let G be an irreducible positive-recurrent generator matrix on $\mathscr{X}$ with stationary distribution μ. Let $\theta > 0$ and $c \in \mathbb{N}$ be such that both $c, \theta > \inf\{r > 0 : I + r^{-1}G$ is non-negative$\}$, and define $Q = I + G/\theta$.*

Define kernels $\{K_n\}_{n\geq 1}$ by

$$K_n = I + \frac{G}{n}\mathbb{1}(n > c), \tag{2.6}$$

and let $\mathbf{M}$ be the inhomogeneous Markov chain with initial distribution π and transition kernels $\{K_n\}_{n\geq 1}$. Define $(\mathbf{P}_n, \mathbf{Y}_n)$ as above with respect to $\mathbf{M}$, and also define the irreducible generator matrix G' by

$$G'_{ij} = \frac{\mu_j}{\mu_i} G_{ji}. \tag{2.7}$$

Then, $\mathbf{Y}_n$ converges weakly to the homogeneous Markov chain $\mathbf{Z}$ with kernel

$$K_{G'}(w,z) = \frac{\mu_z G_{z,w}}{-\mu_w G_{w,w}} 1(w \neq z) = \frac{\mu_z Q_{z,w}}{\mu_w(1 - Q_{w,w})} 1(w \neq z),$$

and initial distribution μ. Also, we have that the conditional law of $\mathbf{P}_n \big| \mathbf{Y}_n$ converges weakly to the conditional law of $\mathbf{R}|\mathbf{Z}$, where $\mathbf{R}|\mathbf{Z}$ is a disordered GEM sequence with parameters $\{-G_{Z_j,Z_j} = -G'_{Z_j,Z_j}\}_{j\geq 1}$. Therefore, the associated pairs $(\mathbf{P}_n, \mathbf{Y}_n)$ converge weakly to $(\mathbf{R}, \mathbf{Z})$ having MCcGEM(G') distribution with respect to μ.

Example 2.11. In the context of Example 2.8, suppose G has constant diagonal entries g. Then, the local occupations of the inhomogeneous Markov chain $\mathbf{P}_n$ would converge to a GEM$(-g)$ distribution, not just conditionally in terms of a MCcGEM distribution.

We now characterize the limit occupation law of $\mathbf{M}$ in a 'stick-breaking' form with respect to either a MCcGEM distribution, or a paired GEM distribution and homogeneous Markov chain. In the following, weak convergence of ν_n is with respect to the product topology on $\Delta_{\mathscr{X}}$, the space of probability measures on $\mathscr{X}$ (cf. Subsection 3.4.1).

Theorem 2.12 (Occupation laws to MCcGEM and stick-breaking measures)**.** *Consider the setting and assumptions of Theorem 2.10. Observe that μ is a stationary distribution of $Q' = I + G'/\theta$, and let $\mathbf{T}$ be the homogeneous Markov chain with kernel Q' and stationary, initial distribution μ. Let $\mathbf{P}$ be a GEM(θ) sequence independent of $\mathbf{T}$.*

Then, $\nu_n = \left\langle \frac{1}{n}\sum_{j=1}^n \delta_{M_j}(l) : l \in \mathscr{X} \right\rangle \xrightarrow{d} \nu$, where

$$\nu \stackrel{d}{=} \left\langle \sum_{j=1}^{\infty} R_j \delta_{Z_j}(l) : l \in \mathscr{X} \right\rangle \stackrel{d}{=} \left\langle \sum_{j=1}^{\infty} P_j \delta_{T_j}(l) : l \in \mathscr{X} \right\rangle. \tag{2.8}$$

Reversing the procedure, starting from the stick-breaking process $\sum_{j\geq 1} P_j \delta_{T_j}$, we may identify it as the limit of the occupation measure of a matched time-inhomogeneous Markov chain, an immediate corollary of Theorem 2.12.

Corollary 2.13 (Stick-breaking measures to Occupation laws)**.** *Let $\theta > 0$ and $\mathbf{P}$ be a GEM(θ) sequence. Let also $\tilde{Q}$ be an irreducible stochastic matrix with stationary distribution μ. Suppose $\mathbf{T}$ is an independent homogeneous Markov chain with kernel $\tilde{Q}$ starting from stationary initial distribution μ.*

Then,

$$\left\langle \sum_{j=1}^{\infty} P_j \delta_{T_j}(l) : l \in \mathscr{X} \right\rangle \stackrel{d}{=} \nu,$$

where $\nu \stackrel{d}{=} \lim_{n\to\infty} \nu_n$ is the occupation law defined with respect to an inhomogeneous Markov chain $\mathbf{M}$, as in the setting of Theorem 2.10, with respect to generator matrix $\tilde{G}'$, where $\tilde{G} = \theta(\tilde{Q} - I)$ and $\tilde{G}'_{ij} = \big(\mu_j/\mu_i\big)\tilde{G}_{ji}$ for $i, j \in \mathscr{X}$.

Remark 2.14. We comment that the logarithmic divergence of the partial sums of the scaling factor n^{-1} in the definition of K_n is important for the results Theorems 2.10 and 2.12 , although as their proofs show, the form of the factor may be relaxed to b_n where $nb_n \to 1$.

2.3. **Dirichlet and non-Dirichlet process limits.** In a particular case of Theorem 2.12, we observe that we may recover Dirichlet processes. Suppose $\mu_i > 0$ for all $i \in \mathscr{X}$. When Q has constant rows equal to μ^t, the Markov chain $\mathbf{T}$ has transition kernel $Q' = Q$, and therefore $\mathbf{T}$ is an iid sequence with common distribution μ. Then, $\nu = \sum_{j\geq 1} P_j \delta_{T_j}$, formed from a GEM($\theta$) sequence $\mathbf{P}$ and an independent sequence of iid random variables $\mathbf{T}$, is the 'stick-breaking' representation of a Dirichlet process with parameters θ and measure μ on the discrete space $\mathscr{X}$ (cf. [44], [49]).

However, since the distribution of ν is determined by G, there is a degree of freedom in specifying G via a pair (θ, Q). Write G in two forms: (1) $G = \theta(Q - I)$ where $\theta > 0$ and Q is stochastic with constant rows μ^t, and also (2) $G = \tilde{\theta}(\tilde{Q} - I)$ where $\tilde{\theta} > 0$, $\theta \neq \tilde{\theta}$, and $\tilde{Q}$ is stochastic. Then again, $\tilde{Q} = \tilde{Q}'$ and via Theorem 2.12, we recover a different stick-breaking representation, $\sum_{j=1}^{\infty} P_j^{\tilde{\theta}} \delta_{T_j^{\tilde{Q}}}$, of the Dirichlet process with parameters θ and μ, in terms of GEM($\tilde{\theta}$) sequence $\mathbf{P}^{\tilde{\theta}}$ and an independent homogeneous Markov chain $\mathbf{T}^{\tilde{\mathbf{Q}}}$ with $T_1^{\tilde{Q}} \sim \mu$ and kernel $\tilde{Q}$.

Here, $\tilde{Q} = \frac{\theta}{\tilde{\theta}} Q + (1 - \frac{\theta}{\tilde{\theta}}) I$ is the weighted average of Q and I. Since $\tilde{Q}$ no longer has constant rows, $\mathbf{T}^{\tilde{\mathbf{Q}}}$ no longer consists of iid variables. The chain $\mathbf{T}^{\tilde{\mathbf{Q}}}$ is, in a sense, a more or less 'sticky' version of an iid $\sim \mu$ sequence depending on the weight of I in the weighted average relation for $\tilde{Q}$. We remark these different representations of the Dirichlet process appear new.

Finally, we state that the above formulation of G is the only case when ν is a Dirichlet process; in particular, we observe for generic G that ν is *not* a Dirichlet process!

Theorem 2.15 (Non-Dirichlet processes)**.** *Consider the setting of Theorem 2.12. The measure ν is a Dirichlet process exactly when G is of form $G = \alpha(Q - I)$ for an $\alpha > 0$ and Q is a 'constant' stochastic kernel whose rows all equal μ^t.*

2.4. **Self-similarity of the occupation laws.** At this point, it is natural to ask for other ways to understand the laws in Theorem 2.12. Consider the general random measure

$$\nu \stackrel{d}{=} \left\langle \sum_{j=1}^{\infty} P_j \delta_{T_j}(l) : l \in \mathscr{X} \right\rangle, \tag{2.9}$$

where $\mathbf{P}$ is a self-similar RAM composed of fractions $\mathbf{X}$, and $\mathbf{T}$ is an independent homogeneous, irreducible, positive-recurrent Markov chain with transition kernel Q and initial stationary distribution μ. We remark that ν reduces to the measure in Theorem 2.12 when $\mathbf{P} \sim \mathrm{GEM}(\theta)$. We first discuss an example.

Example 2.16. As we have noted earlier, if $\mathbf{P} \sim \mathrm{GEM}(\theta)$ and $\mathbf{T}$ is an independent sequence of iid variables with distribution μ, the measure ν is the 'stick-breaking' representation of the Dirichlet process with parameters θ and measure μ on $\mathscr{X}$. Following [49], a self-similarity relation can be deduced:

$$\nu \stackrel{d}{=} X_1 \delta_{T_1} + (1 - X_1)\tilde{\nu},$$

where $\tilde{\nu} \stackrel{d}{=} \nu$ is another random measure, and $X_1 \sim \mathrm{Beta}(1, \theta)$, $T_1 \sim \mu$ and $\tilde{\nu}$ are independent. From such an equation, the Dirichlet process characterization of ν with parameters θ and measure μ on $\mathscr{X}$ follows from classical considerations. Moreover, this relation is central in calculation of a posterior distribution, given say X_1, when ν is thought of as a law on priors. See also the recent work [34] and [48] on related integral characterizations.

We now define a more general notion of self-similarity. This notion is well known (cf. [25] among other references). With respect to a measurable space $(\mathscr{A}, \mathscr{B}_{\mathscr{A}})$, let $\mathbb{P}_{\mathscr{A}}$ be the space of probability measures on $(\mathscr{A}, \mathscr{B}_{\mathscr{A}})$. Let $\mathbb{F}_{\mathscr{A}}$ be the smallest σ-field generated by sets of the form $\Big\{ \{\chi : \chi(A) < r\} : A \in \mathscr{B}_{\mathscr{A}}, r \in [0,1] \Big\}$.

Definition 2.17 (Self-similar random measure)**.** *We say that the law of a random distribution χ on $(\mathbb{P}_{\mathscr{A}}, \mathbb{F}_{\mathscr{A}})$ is self-similar with respect to (η, X) if it satisfies*

$$\chi(\cdot) \stackrel{d}{=} X\eta(\cdot) + (1 - X)\tilde{\chi}(\cdot), \tag{2.10}$$

where X is a $[0,1]$-valued random variable, η is a random distribution on $\mathbb{P}_{\mathscr{A}}$, and $\tilde{\chi}$ is random measure with the same distribution as χ and independent of (η, X), defined on the space $[0,1] \times \mathbb{P}_{\mathscr{A}} \times \mathbb{P}_{\mathscr{A}}$.

The key is that such self-similarity may uniquely identify a distribution. The following is part of Lemma 3.3 in [49]; see also [25] for more involved statements. For the convenience of the reader, a proof is given in Subsection 3.5.

Lemma 2.18. *There exists a unique in law self-similar random measure χ on $(\mathbb{P}_{\mathscr{A}}, \mathbb{F}_{\mathscr{A}})$ with respect to (η, X) when $\mathscr{P}(X = 0) < 1$.*

We now state that ν, defined in (2.9), is self-similar in a certain way. Let $\mathbf{X} = \{X_j\}_{j \geq 1}$ be the iid fractions from which $\mathbf{P}$ is constructed. For each (positive-recurrent) state i of Q, let $\mathbf{T}^i$ be a Markov chain with transition kernel Q and initial value $T_1^i = i$, independent of $\mathbf{X}$ and (ν, T_1). Define the finite cycle length and associated clumped residual fraction,

$$W^i := \inf\{j > 1 : T_j^i = i\} \quad \text{and} \quad X^i := \sum_{j=1}^{W^i - 1} X_j \prod_{l=1}^{j-1} (1 - X_l).$$

Set

$$\eta^i := \left(X^i\right)^{-1} \sum_{j=1}^{W^i-1} \left[X_j \prod_{l=1}^{j-1} (1 - X_l) \right] \delta_{T^i_j} \quad \text{and}$$
$$\nu^i := \nu \big| T_1 = i.$$

Theorem 2.19 (Type of self-similarity)**.** *The law of (ν, T_1) uniquely satisfies the following: Marginally, $T_1 \sim \mu$ and, for each state i of Q,*

$$\nu^i \stackrel{d}{=} X^i \eta^i + (1 - X^i)\tilde{\nu}^i, \tag{2.11}$$

where $\tilde{\nu}^i$ is random measure with the same law as ν^i, such that $\tilde{\nu}^i$ and (η^i, X^i) are independent.

2.5. **On 'nonexchangeability'.** Let $\pi : \mathbb{N} \to \mathbb{N}$ be a permutation of $\mathbb{N}$, that is π is $1 : 1$ and onto. We say the π is a finite permutation when all but a finitely many values in $\mathbb{N}$ are fixed points. Let also $\mathbf{P} = \langle P_i : i \geq 1 \rangle$ be a GEM distribution with parameter θ. Suppose that $\mathbf{T} = \{T_i\}_{i \geq 1}$ is a positive-recurrent, irreducible Markov chain, starting from stationary distribution μ.

Definition 2.20 (Permutation exchangeability)**.** *We say that the representation of the stick-breaking measure $\nu(\cdot) = \sum_{i \geq 1} P_i \delta_{T_i}(\cdot)$ given by $(\mathbf{P}, \mathbf{T})$ is permutation exchangeable when, for each finite permutation π, we have that*

$$\nu(\cdot) \stackrel{d}{=} \sum_{i \geq 1} P_{\pi(i)} \delta_{T_i}(\cdot).$$

Theorem 2.21 (Type of nonexchangeability)**.** *The representation of ν given by $(\mathbf{P}, \mathbf{T})$ is permutation exchangeable if and only if $\mathbf{T}$ is a sequence of iid random variables, in which case ν is a Dirichlet process on $\mathbb{N}$ with parameters (μ, θ).*

We remark that the Dirichlet process, in view of the discussion in Subsection 2.3, has both permutation exchangeable and permutation non-exchangeable representions!

3. Proofs

After some preliminaries, we prove in the succeeding subsections Theorems 2.4, 2.7, 2.10, 2.12, 2.13, 2.19, and then 2.15 and 2.21. We first note a standard algebraic identity, leading to useful formulas for RAMs. Recall our conventions specified at the beginning of section 2.

Lemma 3.1. *For any sequence of numbers a_j and integer $k \geq 1$, we have*

$$\prod_{j=1}^{k} (1 - a_j) + \sum_{j=1}^{k} a_j \prod_{i=1}^{j-1} (1 - a_i) = 1. \tag{3.1}$$

Proof. We proceed by an induction. Equation (3.1) is trivially true for $k = 1$: $(1 - a_1) + a_1 = 1$. If it is true for $k - 1$, then the left-hand side of (3.1) equals

$$\begin{aligned}
&\prod_{j=1}^{k-1} (1 - a_j) - a_k \prod_{j=1}^{k-1} (1 - a_j) + \sum_{j=1}^{k-1} a_j \prod_{i=1}^{j-1} (1 - a_i) + a_k \prod_{j=1}^{k-1} (1 - a_j) \\
&= \prod_{j=1}^{k-1} (1 - a_j) + \sum_{j=1}^{k-1} a_j \prod_{i=1}^{j-1} (1 - a_i) = 1.
\end{aligned}$$

□

Proposition 3.2. *Consider a distribution* $\mathbf{P} = \langle P_j : j \geq 1\rangle$ *on* $\mathbb{N}$ *and factors* $\mathbf{X} = \{X_j\}_{j\geq 1}$ *with*

$$X_j = \begin{cases} P_j\left(1 - \sum_{i=1}^{j-1} P_i\right)^{-1} & \text{when } \sum_{i=1}^{j-1} P_i < 1 \\ 1 & \text{otherwise.} \end{cases}$$

Then, $P_j = X_j \prod_{i=1}^{j-1}(1 - X_i)$ *for* $j \geq 1$.

In particular, if $\mathbf{P}$ *is a RAM constructed from* $\mathbf{X} = \{X_j\}_{j\geq 1}$*, for* $1 \leq k \leq r$*, we have*

$$\sum_{j=1}^{r} P_j = 1 - \prod_{j=1}^{r}(1 - X_j), \text{ and } \sum_{j=k}^{r} P_j = \prod_{j=1}^{k-1}(1 - X_j)\left[1 - \prod_{j=k}^{r}(1 - X_j)\right]. \quad (3.2)$$

Proof. Part (I) follows from (3.1) by an induction: Trivially, $P_1 = X_1$. Suppose $P_k = X_k \prod_{i=1}^{k-1}(1 - X_i)$ for $k \leq j$ and so, by (3.1), we have $\prod_{k=1}^{j}(1 - X_k) = 1 - \sum_{k=1}^{j} P_k$. Then, $P_{j+1} = X_{j+1}\left(1 - \sum_{k=1}^{j} P_k\right) = X_{j+1}\prod_{k=1}^{j}(1 - X_k)$.

For Part (II), the lines in (3.2) follow from Part (I) and (3.1). □

3.1. **Proof of Theorem 2.4: Clumped RAMs.** Let $\mathbf{P}$ be a RAM, and let $\mathbf{X} = \{X_j\}_{j\geq 1}$ be the independent proportions from which $\mathbf{P}$ is constructed. From Proposition 3.2, for $j \geq 1$, we have $P_j = X_j \prod_{i=1}^{j-1}(1 - X_i)$.

Let $\mathbf{u} = \{u_j\}_{j\geq 1}$ be an increasing sequence in $\mathbb{N}$ with $u_1 = 1$. Define new proportions $\mathbf{X^u} = \{X_j^u\}_{j\geq 1}$ from $\mathbf{X}$, using Proposition 3.2 again: For $j \geq 1$,

$$X_j^u = \sum_{i=u_j}^{u_{j+1}-1} X_i \prod_{l=u_j}^{i-1}(1 - X_l) = 1 - \prod_{i=u_j}^{u_{j+1}-1}(1 - X_i) \quad (3.3)$$

Recall that $P_j^u = \sum_{i=u_j}^{u_{j+1}-1} P_i$ for $j \geq 1$ and $\mathbf{P^u} = \{P_j^u\}_{j\geq 1}$.

We now proceed to the proofs of Parts (1)-(4).

3.1.1. *Proof of Part (1).* We now verify that $\mathbf{P^u}$ is a RAM with respect to fractions $\mathbf{X^u}$: For $1 \leq j$, noting (3.3), write

$$\begin{aligned} P_j^u &= \sum_{i=u_j}^{u_{j+1}-1} P_i = \sum_{i=u_j}^{u_{j+1}-1} X_i \prod_{l=1}^{i-1}(1 - X_l) \\ &= \left[\sum_{i=u_j}^{u_{j+1}-1} X_i \prod_{l=u_j}^{i-1}(1 - X_l)\right] \prod_{l=1}^{u_j-1}(1 - X_l) \\ &= X_j^u \prod_{l=1}^{u_j-1}(1 - X_l) \\ &= X_j^u \left[\prod_{l=u_1}^{u_2-1}(1 - X_l)\right] \cdots \left[\prod_{l=u_{j-1}}^{u_j-1}(1 - X_l)\right] = X_j^u \prod_{i=1}^{j-1}(1 - X_i^u). \end{aligned}$$

Since $\mathbf{X} = \{X_j\}_{j\geq 1}$ is composed of independent variables, so is $\mathbf{X^u} = \{X_j^u\}_{j\geq 1}$. Hence, as $\sum_{j\geq 1} P_j^u = \sum_{j\geq 1} P_j \overset{a.s.}{=} 1$, by definition, $\mathbf{P^u}$ is a RAM constructed from independent proportions $\mathbf{X^u}$. □

3.1.2. *Proof of Part (2).* Let $\mathbf{y} = \{y_i\}_{i\geq 1}$ be a possible sequence for $\mathbf{Y}$ in $\mathscr{X}$. For $1 \leq n$, the event that $Y_i = y_i$ for $1 \leq i \leq n$ means the chain $\mathbf{T}$ starts in y_1, staying there until time V_2, when it switches to y_2, remaining there until time V_3, and so on up to time V_n when it moves into y_n. Write for $1 \leq n$ that

$$\begin{aligned}
&\mathscr{P}\left(Y_i = y_i : 1 \leq i \leq n\right) \\
&= \sum_{l_1=1}^{\infty} \cdots \sum_{l_{n-1}=1}^{\infty} \mathscr{P}\left(Y_i = y_i : 1 \leq i \leq n;\ V_{i+1} - V_i = l_i : 1 \leq i \leq n-1\right) \\
&= \sum_{l_1=1}^{\infty} \cdots \sum_{l_{n-1}=1}^{\infty} \mathscr{P}(T_1 = y_1) \prod_{i=1}^{n-1} Q_{y_i y_i}^{l_i - 1} Q_{y_i y_{i+1}} \\
&= \mathscr{P}(T_1 = y_1) \prod_{i=1}^{n-1} \frac{Q_{y_i y_{i+1}}}{1 - Q_{y_i y_i}} \;=\; \mathscr{P}(T_1 = y_1) \prod_{i=1}^{n-1} K(y_i, y_{i+1}). \qquad (3.4)
\end{aligned}$$

We conclude therefore that $\mathbf{Y}$ is a Markov chain with kernel K. $\square$

3.1.3. *Proof of Part (3).* Recall the definition of the increasing random sequence $\mathbf{V}$ with $V_1 = 1$ (cf. (2.3)), and $\mathbf{P^V}$. For each realization, $\mathbf{V}$ is a function of the Markov sequence $\mathbf{T}$. Therefore, conditional on $\mathbf{T}$ given the possible trajectory $\mathbf{t}$ with respect to $\mathbf{T}$, it follows immediately from the proved Part (1) that $\mathbf{P^V}\big|\mathbf{T} = \mathbf{t}$ is a RAM. $\square$

3.1.4. *Proof of Part (4).* If $\mathbf{P}$ is a RAM, we have $\sum_{i\geq 1} P_i^V = \sum_{i \geq 1} P_i = 1$ a.s. Hence, we need only show the associated fractions $\mathbf{X^V}$ are conditionally independent to deduce that $\mathbf{P^V}\big|\mathbf{Y} = \mathbf{y}$ is a RAM.

Let $\mathbf{y}$ be a possible sequence with respect to $\mathbf{Y}$. For $1 \leq m \leq n$, consider the fixed times $V_{i+1} - V_i = l_i \in \mathbb{N}$ for $1 \leq i \leq m$. Noting (3.4), and summing over $V_{i+1} - V_i = l_i$ for $m+1 \leq i \leq n$, we have

$$\begin{aligned}
&\mathscr{P}\left(Y_i = y_i : 1 \leq i \leq n+1, \text{ and } V_{i+1} - V_i = l_i : 1 \leq i \leq m\right) \\
&= \mathscr{P}\left(T_1 = y_1\right) \left[\prod_{i=1}^{m} Q_{y_i,y_i}^{l_i-1} Q_{y_i,y_{i+1}}\right] \sum_{l_{m+1}=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \prod_{i=m+1}^{n} Q_{y_i,y_i}^{l_i-1} Q_{y_i,y_{i+1}} \\
&= \mathscr{P}\left(T_1 = y_1\right) \left[\prod_{i=1}^{m} Q_{y_i,y_i}^{l_i-1} Q_{y_i,y_{i+1}}\right] \prod_{i=m+1}^{n} \sum_{l_i=1}^{\infty} Q_{y_i,y_i}^{l_i-1} Q_{y_i,y_{i+1}} \\
&= \mathscr{P}\left(T_1 = y_1\right) \left[\prod_{i=1}^{n} \frac{Q_{y_i,y_{i+1}}}{1 - Q_{y_i,y_i}}\right] \prod_{i=1}^{m} Q_{y_i,y_i}^{l_i-1}(1 - Q_{y_i,y_i}) \\
&= \mathscr{P}\left(Y_i = y_i : 1 \leq i \leq n+1\right) \prod_{i=1}^{m} Q_{y_i,y_i}^{l_i-1}(1 - Q_{y_i,y_i}). \qquad (3.5)
\end{aligned}$$

Recall (3.3), and consider the variables $\mathbf{X^V} = \{X_j^V\}_{j\geq 1}$ where

$$X_j^V = \sum_{i=V_j}^{V_{j+1}-1} X_i \prod_{l=V_j}^{i-1} (1 - X_l) \;=\; 1 - \prod_{i=V_j}^{V_{j+1}-1} (1 - X_i) \qquad (3.6)$$

When $\mathbf{X}$ is composed of iid variables, that is $\mathbf{P}$ is a self-similar RAM, we will argue now that the fractions $\mathbf{X^V}\big|\mathbf{Y} = \mathbf{y}$ form a conditionally independent sequence, and therefore $\mathbf{P^V}\big|\mathbf{Y} = \mathbf{y}$ is RAM.

Let $r \geq n \geq 1$, and $\langle \alpha_i : 1 \leq i \leq n \rangle \in (0,1)^n$. Write

$$\begin{aligned}
&\mathscr{P}\left(1 - X_j^V \leq \alpha_j : 1 \leq j \leq n \middle| Y_j = y_j : 1 \leq j \leq r+1\right) \\
&= \sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \mathscr{P}\Big(1 - X_j^V \leq \alpha_j : 1 \leq j \leq n \Big| \\
&\qquad\qquad \{V_{i+1} - V_i = l_i, 1 \leq i \leq n\} \cap \{Y_i = y_i : 1 \leq i \leq r+1\}\Big) \\
&\qquad \times \mathscr{P}\left(V_{i+1} - V_i = l_i, 1 \leq i \leq n \middle| Y_j = y_j : 1 \leq j \leq r+1\right).
\end{aligned} \tag{3.7}$$

Relative to $\{l_j\}_{j=1}^n$, define the sequence $\mathbf{u} = \{u_j\}_{j=1}^{n+1}$ where $u_1 = 1$ and $u_j = 1 + \sum_{k=1}^{j-1} l_k$ for $2 \leq j \leq n+1$, which marks the first n times when $\mathbf{T}$ changes states. In particular, on the event $\{V_{i+1} - V_i = l_i, 1 \leq i \leq n\}$, we have $V_j = u_j$ for $1 \leq j \leq n+1$. Given this event, from (3.3), the fractions $\{X_j^V\}_{j=1}^n$ satisfy $1 - X_j^V = \prod_{k=u_j}^{u_{j+1}-1}(1 - X_k)$ for $1 \leq j \leq n$ and are independent, no longer depending on $\mathbf{Y}$. The last display (3.7), noting (3.5), equals

$$\begin{aligned}
&\sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \mathscr{P}\left(\prod_{k=u_j}^{u_{j+1}-1}(1 - X_k) \leq \alpha_j : 1 \leq j \leq n\right) \left[\prod_{j=1}^{n} Q_{y_j y_j}^{l_j - 1}(1 - Q_{y_j y_j})\right] \\
&= \sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \prod_{j=1}^{n} \mathscr{P}\left(\prod_{k=u_j}^{u_{j+1}-1}(1 - X_k) \leq \alpha_j\right) Q_{y_j y_j}^{l_j - 1}(1 - Q_{y_j y_j}) \\
&\quad = \sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \prod_{j=1}^{n} \mathscr{P}\left(\prod_{k=1}^{l_j}(1 - X_k) \leq \alpha_j\right) Q_{y_j y_j}^{l_j - 1}(1 - Q_{y_j y_j}) \\
&\quad = \prod_{j=1}^{n} \sum_{l_j=1}^{\infty} \mathscr{P}\left(\prod_{k=1}^{l_j}(1 - X_k) \leq \alpha_j\right) Q_{y_j y_j}^{l_j - 1}(1 - Q_{y_j y_j}),
\end{aligned} \tag{3.8}$$

in factored form. Therefore, the fractions $\mathbf{X^V}$ are conditionally independent as desired and $\mathbf{P^V}\big|\mathbf{Y} = \mathbf{y}$ is a RAM. $\square$

3.2. **Proof of Theorem 2.7: GEM to MCcGEM.** Let $\mathbf{P} = \langle P_i : i \geq 1 \rangle$ be a GEM(θ) sequence with respect to corresponding iid Beta$(1, \theta)$ proportions $\mathbf{X} = \{X_j\}_{j \geq 1}$. Also, let $\mathbf{T}$ be an independent Markov chain on $\mathscr{X}$, starting from stationary distribution μ, with homogeneous, irreducible kernel Q.

In Part (2) of Theorem 2.4, we showed that the associated sequence $\mathbf{Y}$ is a Markov chain with transition kernel K on $\mathscr{X}$ such that

$$K(z, w) = \frac{Q_{z,w}}{1 - Q_{z,z}} \mathbf{1}(z \neq w)$$

By inspection, the kernel $K = K_G$, in the definition of the MCcGEM distribution (2.4), where $G = \theta(Q - I)$.

Recall now the switch times $\mathbf{V}$ with respect to the chain $\mathbf{T}$ (cf. (2.3)). In Part (4) of Theorem 2.4, as $\mathbf{P}$ is a self-similar RAM, we proved that $\mathbf{P^V}$, conditional on $\mathbf{Y}$, is a RAM. In particular, we showed that the associated fractions $\mathbf{X^V} = \{X_j^V\}_{j \geq 1}$, given $\mathbf{Y}$, are independent variables. Hence, to identify the joint distribution of

$(\mathbf{P^V}, \mathbf{Y})$, we need only find the conditional distribution of each fraction $X_j^V \big| \mathbf{Y}$, for $j \geq 1$.

To this end, let $\mathbf{y}$ be a possible sequence for $\mathbf{Y}$. Recall from (3.6) that $X_j^V = 1 - \prod_{k=V_j}^{V_{j+1}-1}(1 - X_k)$. Write, for $j < n$ and $m \geq 1$,

$$\begin{aligned}
&\mathscr{E}\left[(1 - X_j^V)^m \middle| Y_i = y_i : 1 \leq i \leq n\right] \\
=&\mathscr{E}\left[\mathscr{E}\left[\prod_{k=V_j}^{V_{j+1}-1}(1 - X_k)^m \middle| Y_i = y_i, V_{i+1} - V_i : 1 \leq i \leq n\right] \middle| Y_i = y_i : 1 \leq i \leq n\right].
\end{aligned}$$

Note now, if Z is a Beta$(1, \alpha)$ random variable, then $\mathscr{E}[(1 - Z)^m] = \frac{\alpha}{\alpha+m}$. Then, by the independence of $\mathbf{X}$ and $\mathbf{T}$, noting from (3.5) that $\mathscr{P}(V_{j+1} - V_j = \ell | Y_i = y_i : 1 \leq i \leq n) = Q_{y_i,y_i}^{\ell-1}(1 - Q_{y_i,y_i})$, the above display equals

$$\begin{aligned}
&\mathscr{E}\left[\left(\frac{\theta}{\theta + m}\right)^{V_{j+1}-V_j} \middle| Y_i = y_i : 1 \leq i \leq n\right] \\
&= \sum_{l=1}^{\infty}\left(\frac{\theta}{\theta + m}\right)^l Q_{y_j,y_j}^{l-1}(1 - Q_{y_j,y_j}) \;=\; \frac{\theta(1 - Q_{y_j,y_j})}{\theta(1 - Q_{y_j,y_j}) + m}.
\end{aligned}$$

Thus, for all $j \geq 1$, we see that $X_j^V \Big| \mathbf{Y} = \mathbf{y}$ is a Beta$(1, \theta(1 - Q_{y_j,y_j}))$ random variable. Hence $\mathbf{P^V} \big| \mathbf{Y} = \mathbf{y}$ is a disordered GEM with parameters $\theta(1 - Q_{y_j,y_j}) = -G_{y_j,y_j}$ for $j \geq 1$. Therefore, we conclude that $(\mathbf{P^V}, \mathbf{Y})$ has a MCcGEM$(\theta(Q - I))$ distribution with respect to μ. □

3.3. **Proof of Theorem 2.10: Time inhomogeneous MC to MCcGEM.** We first specify certain asymptotics which will be helpful, before going to the main body of the proof in Subsection 3.3.1.

Lemma 3.3. *For $\gamma > 0$ and integers $1 \leq m \leq n$, let*

$$f_m^n(\gamma) = \prod_{j=m+1}^{n}\left(1 - \frac{\gamma}{j}\right).$$

Then, for $0 < a < b$ and integers $c \geq 0$, we have

$$\lim_{n\to\infty} f_c^n(\gamma)n^\gamma = \frac{\Gamma(c+1)}{\Gamma(c+1-\gamma)} \quad \textit{and} \quad \lim_{n\to\infty} f_{\lfloor an\rfloor}^{\lfloor bn\rfloor}(\gamma)\left(\frac{b}{a}\right)^\gamma = 1.$$

Proof. Write

$$f_l^n(\gamma) = \prod_{j=l+1}^{n}\left(1 - \frac{\gamma}{j}\right) = \frac{\prod_{j=l}^{n}(j - \gamma)}{\prod_{j=l}^{n} j} = \frac{\Gamma(n+1-\gamma)\Gamma(l+1)}{\Gamma(n+1)\Gamma(l+1-\gamma)}.$$

By Stirling's approximation, for $u, v \in \mathbb{R}$, we have $\frac{\Gamma(n+u)}{\Gamma(n+v)}n^{v-u} \to 1$ as $n \to \infty$, from which the desired asymptotics follow immediately. □

Proposition 3.4. *Let $r \geq 1$ be an integer. Let also $\{a_i\}_{j=1}^r$, $\{b_i\}_{i=1}^r$, and $\{\gamma_i\}_{i=1}^r$ be collections of positive numbers such that $a_j < b_j$ for $1 \leq j \leq r$. Then,*

$$\lim_{s_0\to\infty}\sum_{s_1=\lceil a_1s_0\rceil}^{\lceil b_1s_0\rceil-1}\cdots\sum_{s_r=\lceil a_rs_{r-1}\rceil}^{\lceil b_rs_{r-1}\rceil-1}\left[\prod_{j=1}^{r} s_j{}^{-1}f_{s_j}^{\lfloor b_js_{j-1}\rfloor-1}(\gamma_j)\right]$$

$$= \prod_{j=1}^{r} \gamma_j^{-1}\Big(1 - \Big(\frac{a_j}{b_j}\Big)^{\gamma_j}\Big).$$

Proof. The argument follows by inputting the asymptotics in Lemma 3.3. We show only the case $r = 1$, as the extension to $r > 1$ is straightforward.

Again, by Stirling's approximation, $\lim_{n\to\infty} \frac{\Gamma(n+u)}{\Gamma(n+v)} n^{v-u} = 1$ for each $u, v \in \mathbb{R}$. Then, for $\epsilon > 0$ and all large n, we have

$$(1-\epsilon)n^{u-v} \leq \frac{\Gamma(n+u)}{\Gamma(n+v)} \leq (1+\epsilon)n^{u-v}.$$

Hence, for $\epsilon, a, b, \gamma > 0$ with $a < b$, and sufficiently large n, we estimate

$$\begin{aligned}
&(1-\epsilon)^2 \sum_{s=\lfloor an \rfloor}^{\lfloor bn \rfloor -1} \lfloor bn \rfloor^{-\gamma} s^{\gamma-1} \\
&\leq \sum_{s=\lfloor an \rfloor}^{\lfloor bn \rfloor -1} \frac{\Gamma(\lfloor bn \rfloor - \gamma)\Gamma(s)}{\Gamma(\lfloor bn \rfloor)\Gamma(s+1-\gamma)} \ = \ \sum_{s=\lfloor an \rfloor}^{\lfloor bn \rfloor -1} s^{-1} f_s^{\lfloor bn \rfloor -1}(\gamma) \\
&\leq \ (1+\epsilon)^2 \sum_{s=\lfloor an \rfloor}^{\lfloor bn \rfloor -1} \lfloor bn \rfloor^{-\gamma} s^{\gamma-1}.
\end{aligned} \tag{3.9}$$

Now, by the monotonicity of $s^{\gamma-1}$, we have for $n > 2/a$ that $\sum_{s=\lfloor an \rfloor}^{\lfloor bn \rfloor -1} s^{\gamma-1}$ is between the integrals $\int_{\lfloor an \rfloor -1}^{\lfloor bn \rfloor -1} s^{\gamma-1} ds$ and $\int_{\lfloor an \rfloor}^{\lfloor bn \rfloor} s^{\gamma-1} ds$. We may compute

$$\begin{aligned}
&\lim_{n\to\infty} \lfloor bn \rfloor^{-\gamma} \int_{\lfloor an \rfloor -1}^{\lfloor bn \rfloor -1} s^{\gamma-1} ds \\
&\quad = \lim_{n\to\infty} \lfloor bn \rfloor^{-\gamma} \int_{\lfloor an \rfloor}^{\lfloor bn \rfloor} s^{\gamma-1} ds \ = \ \frac{1}{\gamma}\left(1 - \left(\frac{a}{b}\right)^{\gamma}\right).
\end{aligned}$$

Then, inserting into (3.9), the proposition follows for $r = 1$. □

We now show a form of 'weak ergodicity' for the Markov chain $\mathbf{M}$.

Lemma 3.5. *For an irreducible positive-recurrent generator matrix G, let $\theta > 0$, and $c \geq 1$ be an integer, such that $c, \theta > \inf\{r > 0 : I + r^{-1}G$ is non-negative$\}$. Suppose μ is a stationary distribution of Q, and let π be a stochastic vector. Recall that $K_n = I + \frac{G}{n}\mathbb{1}(n > c)$ for $n \geq 1$ (cf. (2.6)).*

Then, as $n \to \infty$, both (a) $(\mu^n)^t := \pi^t \prod_{i=1}^n K_i \to \mu^t$, and (b) $\big(\mu^n\big)^t Q \to \mu^t$, hold entry-wise.

Proof. We separate into four steps. Recall that since G is irreducible, positive-recurrent that Q is irreducible, positive-recurrent.

Step 1. Fix an integer $m \geq \max(c, \theta)$ and write the stochastic matrix,

$$\begin{aligned}
\prod_{j=m+1}^{n} K_j &= \prod_{j=m+1}^{n} \left[\left(1 - \frac{\theta}{j}\right) I + \frac{\theta}{j} Q\right] \\
&= \left[\prod_{j=m+1}^{n} \left(1 - \frac{\theta}{j}\right)\right] \left(I + \sum_{i=1}^{n-m} Q^i \sum_{m<j_1<\cdots<j_i\leq n} \prod_{l=1}^{i} \frac{\theta}{j_l - \theta}\right),
\end{aligned}$$

as a polynomial in Q with positive coefficients.

Step 2. We now show that any fixed degree coefficient of the polynomial vanishes as $n \to \infty$. For each i, denote the nth coefficient of Q^i by $[Q^i]_n$. By Lemma 3.3, $[Q^0]_n = f_m^n(\theta) \to 0$ as $n \to \infty$. Also, as $f_m^n(\theta) \sim n^{-\theta}$ by Lemma 3.3, we have for $i \geq 1$ that

$$\begin{aligned}
[Q^i]_n &= \left[\prod_{j=m+1}^{n}\left(1-\frac{\theta}{j}\right)\right] \sum_{m<j_1<\ldots<j_i\leq n} \prod_{l=1}^{i} \frac{\theta}{j_l-\theta} \\
&= \theta^i f_m^n(\theta) \sum_{j_1=m+1}^{n-i+1} \frac{1}{j_1-\theta} \sum_{j_2=j_1+1}^{n-i+2} \frac{1}{j_2-\theta} \cdots \sum_{j_i=j_{i-1}+1}^{n} \frac{1}{j_i-\theta} \\
&\leq \theta^i f_m^n(\theta) \left[\ln\left(\frac{n-\theta}{m+1-\theta}\right) + \frac{1}{m+1-\theta}\right]^i \\
&\leq C(\theta,m) n^{-\theta} \left[\ln\left(\frac{n-\theta}{m+1-\theta}\right)\right]^i \xrightarrow{n\to\infty} 0.
\end{aligned}$$

Step 3. For each $x \in \mathscr{X}$, let e_x denote the vector in $\mathbb{R}^{\mathscr{X}}$ with a 1 in the entry corresponding to state x and 0's elsewhere. Since Q is a stochastic kernel, observe for each $x \in \mathscr{X}$ and $n \geq m$ that

$$1 = \sum_{z\in\mathscr{X}} e_x^t \left[\prod_{j=m+1}^{n} K_j\right] e_z = \sum_{i=0}^{n-m} [Q^i]_n \sum_{z\in\mathscr{X}} e_x^t Q^i e_z = \sum_{i=0}^{n-m} [Q^i]_n.$$

Also, as μ is a stationary eigenvector of Q, note that μ is also a stationary eigenvector of $\{K_n\}_{n\geq 1}$.

Recall that $\mu^m = \pi^t \prod_{i=1}^m K_i$. We have that Q is irreducible, positive-recurrent, and since $\theta > \inf\{r > 0 : I + r^{-1}G \text{ is non-negative}\}$ we note $Q(a,a) > 0$ for $a \in \mathscr{X}$, and so Q is aperiodic also. Then, we have the convergence $(\mu^m - \mu)^t Q^n \to 0$.

With these observations, for each $x \in \mathscr{X}$ and positive integers n and $R < n-m$, we may bound

$$\begin{aligned}
|\mu_l^n - \mu_l| &= \left|(\mu^m-\mu)^t \left[\prod_{j=m+1}^{n} K_j\right] e_l\right| \\
&= \Big|\sum_{i=0}^{n-m} [Q^i]_n (\mu^m-\mu)^t Q^i e_l\Big| \\
&\leq \sum_{i=0}^{R} [Q^i]_n + \Big|\sum_{i=R+1}^{n-m} [Q^i]_n (\mu^m-\mu)^t Q^i e_l\Big| \\
&\leq \sum_{i=0}^{R} [Q^i]_n + \max_{r>R} \left|(\mu^m-\mu)^t Q^r e_l\right|.
\end{aligned}$$

As $n \to \infty$, the last display converges to $\max_{r>R} \left|(\mu^m-\mu)^T Q^r e_l\right|$, by the calculation in Step 2, and in turn vanishes as $R \to \infty$. Hence, the limit (a) follows.

Step 4. Finally, by Fatou's lemma, the proved first limit (a), and that μ is a stationary vector of Q, we have for each $j \in \mathscr{X}$ that

$$\liminf_{n\to\infty} \left(\mu^n\right)^t Q e_j = \liminf_{n\to\infty} \sum_{i\in\mathscr{X}} \mu_i^n Q_{i,j} \geq \sum_{i\in\mathscr{X}} \mu_i Q_{i,j} = \mu_j. \tag{3.10}$$

Now, suppose for a particular $k \in \mathscr{X}$ that $\limsup_{n\to\infty} \left(\mu^n\right)^t Q e_k = L > \mu_k$. Then, as $\left(\mu^n\right)^t Q$ is a stochastic vector, we would have for each $n \geq 1$ that

$$1 = \limsup_{n\to\infty} \sum_{l\in\mathscr{X}} \left(\mu^n\right)^t Q e_l \geq L + \liminf_{n\to\infty} \sum_{l\neq k} \left(\mu^n\right)^t Q e_l.$$

But, as μ is a stochastic vector and noting (3.10), we have by Fatou's lemma again that the last display is larger than $L + \sum_{l\neq k} \mu_l > 1$, a contradiction, and the second limit (b) holds. $\square$

3.3.1. *Completion of the proof of Theorem 2.10.* We will argue in steps.

Step 1. Recall the definition of kernel G' (cf. (2.7)). We now argue that G' is a generator matrix: As μ is a stationary vector of Q and $G = \theta(Q - I)$, we have $\mu^t G = 0$ is the zero vector. Since G is a generator matrix, we have $G'_{i,j} = (\mu_j/\mu_i) G_{j,i} \geq 0$ for $i \neq j$, and $\sum_j G'_{i,j} = \frac{1}{\mu_i}\sum_j \mu_j G_{j,i} = 0$. Moreover,

$$\sup_i |G'_{i,i}| \ = \ \sup_i |G_{i,i}| \ < \ \infty.$$

Step 2. Recall the Markov chain $\mathbf{M}$, with transition kernels $\{K_n = I + \frac{G}{n} 1(n > c)\}_{n\geq 1}$ (cf. (2.6)), starting from π. Recall the associated variable N_n and sequence $\mathbf{P}_n$.

Now, for $i \geq N_n > j \geq 1$ define

$$X_{n,j} = P_{n,j} \Big/ \left(1 - \sum_{i=1}^{j-1} P_{n,i}\right) \qquad \text{and} \qquad X_{n,i} = 1. \tag{3.11}$$

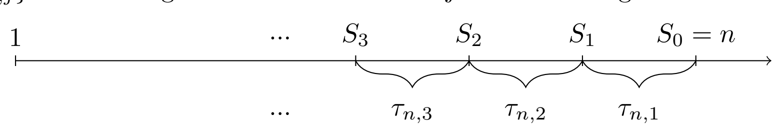
The variables $\mathbf{X}_n = \{X_{n,i}\}_{i\geq 1}$ are the associated fractions to the distribution $\mathbf{P}_n$ on $\mathbb{N}$ and, by Proposition 3.2, for $j \geq 1$,

$$P_{n,j} = X_{n,j} \prod_{i=1}^{j-1} (1 - X_{n,i}) \ \text{ and } \ 1 - \sum_{i=1}^{j-1} P_{n,i} = \prod_{i=1}^{j-1} \left(1 - X_{n,i}\right). \tag{3.12}$$

For $j \geq 0$, also define

$$S_j = n \left(1 - \sum_{i=1}^{j} P_{n,i}\right) = n \prod_{i=1}^{j} \left(1 - X_{n,i}\right). \tag{3.13}$$

In terms of the switching times $\mathbf{V}$, and the first time N_n that the chain $\mathbf{M}$ switches after time n, we have $S_0 = n$, $S_j = V_{N_n - j} - 1$ for $1 \leq j \leq N_n - 1$, and $S_j = 0$ for $j \geq N_n$. Recall also that $\tau_{n,j} = nP_{n,j}$ for $j \geq 1$. In words, $\{S_j\}$ are the times before time n at which the chain switches states when considered in reverse order, and $\{\tau_{n,j}\}$ are the lengths of the associated sojourns in the figure below.



Step 3. Recall the sequence $\mathbf{Y}_n$ given in (2.5), where $Y_{n,j} = M_{V_{N_n-j}}$ for $1 \le j \le N_n - 1$ and $Y_{n,i} = M_1$ for $i \ge N_n$. We now aim to compute the finite dimensional distributions of $(\mathbf{P}_n, \mathbf{Y}_n)$ or equivalently of $(\mathbf{X}_n, \mathbf{Y}_n)$. To this end, fix the integer $r \ge 1$, and consider numbers $\{\beta_j\}_{j=1}^r \in (0,1)^r$ such that $s_j := n\prod_{i=1}^j (1-\beta_i) \in \mathbb{N}$, for $1 \le j \le r$, are all integers. Set also $s_0 = n$ and recall $S_0 = n$.

Note from (3.12) and (3.13) that

$$\begin{aligned} X_{n,j} = \beta_j \text{ for } 1 \le j \le r \quad &\iff \quad S_j = s_j = n\prod_{i=1}^j (1-\beta_i) \text{ for } 1 \le j \le r \\ &\iff \quad \tau_{n,j} = nP_{n,j} = s_{j-1} - s_j \text{ for } 1 \le j \le r. \end{aligned}$$

Then, with respect to a possible sequence $\mathbf{y}$, we have

$$\begin{aligned} &\mathscr{P}\left(X_{n,j} = \beta_j, Y_{n,j} = y_j : 1 \le j \le r\right) \qquad (3.14)\\ =&\mathscr{P}\left(\tau_{n,j} = s_{j-1} - s_j,\ Y_{n,j} = y_j : 1 \le j \le r\right) \\ =&\sum_{\substack{y_{r+1}\in\mathscr{X}: \\ y_{r+1}\ne y_r}} \mathscr{P}(M_{s_r} = y_{r+1}) \prod_{j=1}^r \mathscr{P}\left(M_{s_{j-1}} = \cdots = M_{s_j+1} = y_j \middle| M_{s_j} = y_{j+1}\right). \end{aligned}$$

Note the computation for $c \le l < n$ and $z \ne y$,

$$\begin{aligned} &\mathscr{P}(M_n = \cdots = M_{l+1} = y | M_l = z) \\ &\qquad = \frac{G_{z,y}}{l} \prod_{j=l+1}^{n-1} \left(1 + \frac{G_{y,y}}{j}\right) = \frac{G_{z,y}}{l} f_l^{n-1}\left(-G_{y,y}\right). \end{aligned}$$

Recall also that $\mu_y^s = P\left(M_s = y\right)$. Since $G = \theta(Q-I)$, we observe

$$\begin{aligned} \sum_{\substack{y\in\mathscr{X}: \\ y\ne z}} \mu_y^s G_{y,z} &= \theta \sum_{\substack{y\in\mathscr{X}: \\ y\ne z}} \mu_y^s \left(Q-I\right)_{y,z} \\ &= \theta \sum_{\substack{y\in\mathscr{X}: \\ y\ne z}} \mu_y^s Q_{y,z} = \theta\left[\left(\mu^s\right)^t Q e_z - \mu_z^s Q_{z,z}\right]. \end{aligned}$$

Then, (3.14) equals

$$\begin{aligned} &\sum_{\substack{y_{r+1}\in\mathscr{X}: \\ y_{r+1}\ne y_r}} \mu_{y_{r+1}}^{s_r} \prod_{j=1}^r \frac{G_{y_{j+1},y_j}}{s_j} f_{s_j}^{s_{j-1}-1}\left(-G_{y_j,y_j}\right) \\ =&\left[(\mu^{s_r})^t Q e_{y_r} - \mu_{y_r}^{s_r} Q_{y_r,y_r}\right] \frac{\theta}{s_r} f_{s_r}^{s_{r-1}-1}\left(-G_{y_r,y_r}\right) \qquad (3.15)\\ &\times \prod_{j=1}^{r-1} \frac{G_{y_{j+1},y_j}}{s_j} f_{s_j}^{s_{j-1}-1}\left(-G_{y_j,y_j}\right) \end{aligned}$$

Step 4. We now sum the display (3.15) over all appropriate values of $\{s_j\}_{j=1}^r$ such that $0 < X_{n,j} \le \beta_j$ for $1 \le j \le r < N_n$, where we recall N_n is the time the chain switches after time n. Then, we have from (3.13) that

$$1 \le \tau_{n,j} = nP_{n,j} = S_{j-1} - S_j = X_{n,j} S_{j-1}. \qquad (3.16)$$

Moreover, also from (3.13), we have $s_r \ge n\prod_{j=1}^r (1-\beta_j)$ diverges to infinity as $n \to \infty$.

Recall $s_0 = n$ and $\lim_{n\to\infty} N_n = \infty$ a.s. Then, with equation (3.16) in hand,

$$
\begin{aligned}
&\mathscr{P}\left(0 < X_{n,j} \le \beta_j,\ Y_{n,j} = y_j : 1 \le j \le r\right)\\
&= \mathscr{P}\left(1 \le \tau_{n,j} = S_{j-1} - S_j \le S_{j-1}\beta_j,\ Y_{n,j} = y_j : 1 \le j \le r\right)\\
&= \mathscr{P}\left(S_{j-1}(1-\beta_j) \le S_j \le S_{j-1} - 1,\ Y_{n,j} = y_j : 1 \le j \le r\right)\\
&= \sum_{s_1=\lceil s_0(1-\beta_1)\rceil}^{s_0-1} \cdots \sum_{s_r=\lceil s_{r-1}(1-\beta_r)\rceil}^{s_{r-1}-1} \left[(\mu^{s_r})^t Q e_{y_r} - \mu^{s_r}_{y_r} Q_{y_r,y_r}\right]\\
&\qquad \times \frac{\theta}{s_r} f_{s_r}^{s_{r-1}-1}\left(-G_{y_r,y_r}\right) \prod_{j=1}^{r-1} \frac{G_{y_{j+1},y_j}}{s_j} f_{s_j}^{s_{j-1}-1}\left(-G_{y_j,y_j}\right).
\end{aligned}
$$

Step 5. From (3.13), the sum index $s_r \ge n\prod_{j=1}^r (1-\beta_j)$ diverges to infinity as $n \to \infty$. Also, by Lemma 3.5, we have $\lim_{s\to\infty} \mu^s_y = \mu_y$ and $\lim_{s\to\infty} \left(\mu^s\right)^t Q e_y = \mu_y$ for each $y \in \mathscr{X}$. Therefore, as $n \to \infty$, we have

$$
\theta\left[\left(\mu^{s_r}\right)^t Q e_{y_r} - \mu^{s_r}_{y_r} Q_{y_r,y_r}\right] \to \theta\mu_{y_r}(1 - Q_{y_r,y_r}) = \mu_{y_r}\big(-G_{y_r,y_r}\big).
$$

Note that $-G_{i,i} > 0$ for each $i \in \mathscr{X}$ since by assumption G is irreducible and so has no zero rows. Thus, by Proposition 3.4, we have

$$
\begin{aligned}
&\lim_{n\to\infty} \mathscr{P}\left(0 < X_{n,j} \le \beta_j,\ Y_{n,j} = y_j : 1 \le j \le r\right)\\
=&\mu_{y_r}(-G_{y_r,y_r}) \prod_{j=1}^{r-1} G_{y_{j+1},y_j}\\
&\qquad \times \lim_{n\to\infty} \sum_{s_1=\lceil s_0(1-\beta_1)\rceil}^{s_0-1} \cdots \sum_{s_r=\lceil s_{r-1}(1-\beta_r)\rceil}^{s_{r-1}-1} \prod_{j=1}^{r} {s_j}^{-1} f_{s_j}^{s_{j-1}-1}\left(-G_{y_j,y_j}\right)\\
=&\mu_{y_r}(-G_{y_r,y_r}) \left[\prod_{j=1}^{r-1} G_{y_{j+1},y_j}\right] \left[\prod_{j=1}^{r} \left(-G_{y_j,y_j}\right)^{-1} \left(1 - (1-\beta_j)^{-G_{y_j,y_j}}\right)\right]\\
=&\mu_{y_1} \left[\prod_{j=1}^{r-1} \frac{\mu_{y_{j+1}}}{\mu_{y_j}} \frac{G_{y_{j+1},y_j}}{-G_{y_j,y_j}}\right] \left[\prod_{j=1}^{r} \left(1 - (1-\beta_j)^{-G_{y_j,y_j}}\right)\right]\\
=&\mu_{y_1} \left[\prod_{j=1}^{r-1} \frac{G'_{y_j,y_{j+1}}}{-G'_{y_j,y_j}}\right] \left[\prod_{j=1}^{r} \left(1 - (1-\beta_j)^{-G'_{y_j,y_j}}\right)\right],
\end{aligned}
$$

decomposed as a product of i) the transition probability of the chain $\mathbf{Z}$, with kernel $K_{G'}$ (cf. (2.4)) and initial distribution μ, running through states $\{y_j\}_{j=1}^r$, and ii) the distribution functions of independent Beta$(1, -G'_{y_j,y_j})$ random variables for $1 \le j \le r$. Hence, the finite dimensional distributional convergence of $(\mathbf{P}_n, \mathbf{Y}_n)$ as $n \to \infty$ is established. □

3.4. **Proof of Theorem 2.12: Occupation laws to MCcGEM and stick-breaking measures.** In the setting of Theorems 2.10 and 2.12, consider the pairs $\{(\mathbf{P}_n, \mathbf{Y}_n)\}_{n\ge1}$, $(\mathbf{R}, \mathbf{Z})$, and $(\mathbf{P}, \mathbf{T})$. These objects belong to $[0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$. We now discuss the topology on this space and its relatives, before going to the proof of (2.8) in Subsection 3.4.2.

3.4.1. *Topology.* We endow the space $[0,1]^{\mathbb{N}}$ with a standard product metric ρ^1 and σ-field, generated in terms of this metric, which yields the usual product σ-field built from the Borel σ-fields on copies of $[0,1]$: For $p, p' \in [0,1]^{\mathbb{N}}$,

$$\rho^1(p,p') = \sum_{n=1}^{\infty} 2^{-n}|p_j - p'_j|.$$

Consider now the metric ρ on $[0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$ defined as follows:
For $(p,y),(p',y') \in [0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$,

$$\rho((p,y),(p',y')) = \sum_{n=1}^{\infty} 2^{-n}\left[|p_j - p'_j| + |y_j - y'_j|\right].$$

The corresponding σ-field on $[0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$, generated by ρ, is the usual product σ-field formed from the Borel σ-fields on copies of $[0,1]$ and $\mathscr{X}$. Importantly, weak convergence of probability measures on $[0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$ translates to finite dimensional convergence of these laws. Moreover, $([0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}, \rho)$ is a complete, separable metric space.

Recall that Δ_∞ is the collection of all probabilities on $\mathbb{N}$:

$$\Delta_\infty = \left\{ p \in [0,1]^{\mathbb{N}} : \sum_{j=1}^{\infty} p_j = 1 \right\}.$$

Since

$$\Delta_\infty = \bigcap_{n=1}^{\infty} \bigcap_{M=1}^{\infty} \bigcup_{m=M}^{\infty} \left\{ p \in [0,1]^{\mathbb{N}} : 1 - \frac{1}{n} \leq \sum_{j=1}^{m} p_j \leq 1 + \frac{1}{n} \right\},$$

$\Delta_\infty \times \mathscr{X}^{\mathbb{N}}$ is a measurable set in $[0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}}$. We may endow $\Delta_\infty \times \mathscr{X}^{\mathbb{N}}$ with the restriction of the metric ρ and the σ-field generated from the associated metric topology.

For a fixed point $(p',y') \in \Delta_\infty \times \mathscr{X}^{\mathbb{N}}$, the projection map $f : [0,1]^{\mathbb{N}} \times \mathscr{X}^{\mathbb{N}} \to \Delta_\infty \times \mathscr{X}^{\mathbb{N}}$, given by

$$f(p,y) = \begin{cases} (p,y) & : (p,y) \in \Delta_\infty \times \mathscr{X}^{\mathbb{N}} \\ (p',y') & : (p,y) \notin \Delta_\infty \times \mathscr{X}^{\mathbb{N}} \end{cases},$$

is measurable, and also continuous on the subset $\Delta_\infty \times \mathscr{X}^{\mathbb{N}}$. Now, denote the collection of probabilities on $\mathscr{X}$,

$$\Delta_{\mathscr{X}} = \left\{ p \in [0,1]^{\mathscr{X}} : \sum_{l \in \mathscr{X}} p_l = 1 \right\},$$

and endow it with the metric $\rho^2(p,p') = \sum_{x \in \mathscr{X}} 2^{-x}|p_x - p'_x|$, and the associated Borel σ-field. Define $g : \Delta_\infty \times \mathscr{X}^{\mathbb{N}} \to \Delta_{\mathscr{X}}$ by

$$g((p,y)) = \left\langle \sum_{j=1}^{\infty} p_j \mathbb{1}_l(y_j) : l \in \mathscr{X} \right\rangle.$$

Then, g is a continuous and therefore measurable function on $\Delta_\infty \times \mathscr{X}^{\mathbb{N}}$: Indeed, if $\{(p^n, y^n)\}_{n \geq 1}$ and (p,y) belong to $\Delta_{\mathscr{X}} \times \mathscr{X}^{\mathbb{N}}$, and the finite dimensional convergence $(p^n, y^n) \to (p,y)$ holds, for each $l \in \mathscr{X}$, we have $\sum_{j \geq A} p^n_j \mathbb{1}_l(y^n_j) \leq \sum_{j \geq A} p^n_j = 1 -$

$\sum_{j<A} p_j^n \overset{n\to\infty}{\longrightarrow} 1-\sum_{j<A} p_j$. The claim now follows since (1) $\sum_{j<A} p_j^n \mathbb{1}_l(y_j^n) \xrightarrow{n\to\infty} \sum_{j<A} p_j \mathbb{1}_l(y_j) \xrightarrow{A\to\infty} g((p,y))$, and (2) $\sum_{j\geq A} p_j \xrightarrow{A\to\infty} 1$.

3.4.2. *Proof of* (2.8). First, we verify that $\{(\mathbf{P}_n, \mathbf{Y}_n)\}_{n\geq 1}$, $(\mathbf{R}, \mathbf{Z})$ and $(\mathbf{P}, \mathbf{T})$ belong almost surely to $\Delta_\infty \times \mathscr{X}^N$. Clearly, $\{(\mathbf{P}_n, \mathbf{Y}_n)\}_{n\geq 1}$ surely lives in $\Delta_\infty \times \mathscr{X}^{\mathbb{N}}$ by construction. Also, $(\mathbf{R}, \mathbf{Z})$ and $(\mathbf{P}, \mathbf{T})$ lie almost surely in $\Delta_\infty \times \mathscr{X}^{\mathbb{N}}$ since, by Theorem 2.10 and the assumptions of Theorem 2.12, we have that $\mathbf{R}$ and $\mathbf{P}$ are RAMs, and so $\sum_{j=1}^\infty R_j \overset{d}{=} \sum_{j=1}^\infty P_j \overset{a.s.}{=} 1$.

Now, from the finite dimensional or in other words weak convergence of $(\mathbf{P}_n, \mathbf{Y}_n)$ to $(\mathbf{R}, \mathbf{Z})$ in Theorem 2.10, we have $\nu_n = g\big((\mathbf{P}_n, \mathbf{Y}_n)\big) = g\circ f\big((\mathbf{P}_n, \mathbf{Y}_n)\big)$ converges weakly to $\nu = g\circ f\big((\mathbf{R}, \mathbf{Z})\big)$ by the continuous mapping theorem, and so the left equality in (2.8) holds.

On the other hand, with respect to $(\mathbf{P}, \mathbf{T})$, define $\mathbf{P}^{\mathbf{V}}$ and $\mathbf{Y}$ as in the setting of Theorem 2.7. Recall that $\mathbf{T}$ is a Markov chain with kernel $Q' = I + G'/\theta$ and initial stationary distribution μ. Then, by Theorem 2.7, noting that $G' = \theta(Q' - I)$, we have that $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y})$ has a MCcGEM(G') distribution. Hence, $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y}) \overset{d}{=} (\mathbf{R}, \mathbf{Z})$. Since almost surely, by 'unclumping',

$$g\circ f\big((\mathbf{P}^{\mathbf{V}}, \mathbf{Y})\big) = g\circ f\big((\mathbf{P}, \mathbf{T})\big) = \left\langle \sum_{j\geq 1} P_j 1_l(T_j) : l \in \mathscr{X} \right\rangle,$$

we have $g\circ f\big((\mathbf{R}, \mathbf{Z})\big) \overset{d}{=} g\circ f\big((\mathbf{P}, \mathbf{T})\big)$, and the right equality of (2.8) holds. □

3.5. **Proof of Theorem 2.19: Type of self-similarity.** We first give a proof of Lemma 2.18, before going to the main argument in Subsection 3.5.2

3.5.1. *Proof of Lemma 2.18.* Let $\{(\eta_j, X_j)\}_{j\geq 1}$ be i.i.d. copies of (η, X), independent of (η, X), all on a common probability space.

Existence: Let $\chi(\cdot) = \sum_{j=1}^\infty \eta_j(\cdot) X_j \prod_{i=1}^{j-1}(1 - X_i)$. Since $\mathscr{P}(X = 0) < 1$, we have $\prod_{j\geq 1}(1 - X_j) = 0$ a.s., and so $\big\langle X_j \prod_{i=1}^{j-1}(1 - X_i) : j \geq 1\big\rangle$ is a RAM. Hence, χ is a random probability measure on $\mathscr{A}$ as $\chi(\mathscr{A}) = \sum_{j=1}^\infty X_j \prod_{i=1}^{j-1}(1 - X_i) \overset{a.s.}{=} 1$. Moreover, (2.10) holds straightfowardly:

$$\chi = X_1\eta_1 + (1 - X_1)\left[\sum_{j=2}^\infty \eta_j(\cdot) X_j \prod_{i=2}^{j-1}(1 - X_i)\right] \overset{d}{=} X_1\eta_1 + (1 - X_1)\tilde{\chi},$$

where $\tilde{\chi}$ has the same law as χ and is independent of (X_1, η_1).

Uniqueness: Suppose χ^a and χ^b both satisfy the self-similarity equation (2.10). On a probability space, where $\{(\eta_j, X_j)\}_{j\geq 1}$, χ^a and χ^b are independent, define a sequence of measures: $\chi_1^a = \chi^a$, $\chi_1^b = \chi^b$ and, for $j \geq 1$,

$$\chi_{j+1}^a = X_j\eta_j + (1 - X_j)\chi_j^a \quad \text{and} \quad \chi_{j+1}^b = X_j\eta_j + (1 - X_j)\chi_j^b.$$

By construction, $\{\chi_j^a\}_{j\geq 1}$ and $\{\chi_j^b\}_{j\geq 1}$ are two sequences of identically distributed random measures distributed as χ^a and χ^b respectively.

We note again that $\prod_{j\geq 1}(1-X_j)=0$ a.s. as $\mathscr{P}(X=0)<1$. Then, in terms of the variational norm $\|\cdot\|$,

$$\begin{aligned}\left\|\chi^a_{j+1}-\chi^b_{j+1}\right\| &= |1-X_j|\left\|\chi^a_j-\chi^b_j\right\| \\ &= \left[\prod_{i=1}^{j}|1-X_i|\right]\left\|\chi^a_1-\chi^b_1\right\| \leq \prod_{i=1}^{j}|1-X_i|,\end{aligned}$$

which vanishes a.s. as $j\to\infty$. Hence, $\chi^a \overset{d}{=} \chi^b$. □

3.5.2. *Completion of the proof of Theorem 2.19.* Recall our conventions at the beginning of Section 2 and that $\mathbf{X}=\{X_j\}_{j\geq 1}$ is a collection of iid variables, and $\mathbf{T}$ is the homogeneous, irreducible Markov chain with kernel Q and initial stationary distribution μ. Let $\mathbf{P}=\langle P_j : j\geq 1\rangle$ be the RAM constructed from $\mathbf{X}$. For each state i of Q, let $\mathbf{T}^i=\mathbf{T}\big|T_1=i$ be the Markov chain with transition kernel Q and initial value $T^i_1=i$. Since i is recurrent, the time $W^i=\inf\{l>1 : T^i_l=i\}$ is a.s. finite. Recall

$$X^i=\sum_{l=1}^{W^i-1}X_l\prod_{n=1}^{l-1}(1-X_n)=\sum_{l=1}^{W^i-1}P_l=1-\prod_{l=1}^{W^i-1}(1-X_l). \tag{3.17}$$

Recall also $\eta^i=\left(X^i\right)^{-1}\sum_{l=1}^{W^i-1}\left[X_l\prod_{n=1}^{l-1}(1-X_n)\right]\delta_{T^i_l}$.

We now rewrite the measure $\nu^i=\nu\big|T_1=i$ as follows:

$$\begin{aligned}\nu^i=\sum_{l\geq 1}P_l\delta_{T^i_l} &= \sum_{l=1}^{W^i-1}P_l\delta_{T^i_l}+\sum_{l\geq W^i}P_l\delta_{T^i_l} \\ &= X^i\eta^i+\left(1-X^i\right)\sum_{l\geq W^i}\frac{P_l}{1-X^i}\delta_{T^i_l}.\end{aligned} \tag{3.18}$$

Then, by (3.17) and Proposition 3.2 for $j\geq 1$ we have

$$\begin{aligned}\frac{P_{j-1+W^i}}{1-X^i} &= \frac{X_{j-1+W^i}\prod_{l=1}^{j-1+W^i-1}(1-X_l)}{\prod_{l=1}^{W^i-1}(1-X_l)} \\ &= X_{j-1+W^i}\prod_{l=W^i}^{j-1+W^i-1}(1-X_l) \;=\; X_{j-1+W^i}\prod_{l=1}^{j-1}(1-X_{l-1+W^i}).\end{aligned}$$

Hence, as $\mathbf{X}$ is composed of iid variables, independent of $\mathbf{T}^i$ and therefore W^i, we see that

$$\begin{aligned}&\left\langle\frac{P_{j-1+W^i}}{1-X^i}=X_{j-1+W^i}\prod_{l=1}^{j-1}(1-X_{l-1+W^i}) : j\geq 1\right\rangle \\ &\overset{d}{=}\left\langle X_j\prod_{l=1}^{j-1}(1-X_l) : j\geq 1\right\rangle=\mathbf{P}.\end{aligned}$$

Clearly, as the chain starts over again at location i, $\{T^i_l\}_{l\geq W^i}\overset{d}{=}\mathbf{T}^i$.

Moreover, by conditioning on the value of W^i and noting that $\mathbf{X}$ and $\mathbf{T}^i$ are independent, the sequences $\left\langle\frac{P_{j-1+W^i}}{1-X^i} : j\geq 1\right\rangle$ and $\{T^i_l\}_{l\geq W^i}$ are independent. Similarly, we see that the sum $\sum_{l\geq W^i}\frac{P_l}{1-X^i}\delta_{T^i_l}$, which depends only on variables

$\{X_k\}_{k\geq W^i}$ and $\{T^i_k\}_{k\geq W^i}$ indexed beyond the first cycle, is independent of the pair (X^i,η^i). In particular, the sum $\tilde{\nu}^i := \sum_{l\geq W^i}\frac{P_l}{1-X^i}\delta_{T^i_l} \stackrel{d}{=} \nu^i$.

Hence, from these observations, (3.18) represents the sought after self-similarity equation (2.11).

Finally, a distribution ν^i satisfying (2.11) is unique by Lemma 2.18 since $X^i_1 \in (0,1]$ a.s. Also, by assumption, $T_1 \sim \mu$ where $\mu_i > 0$ for each $i \in \mathscr{X}$ and each i is recurrent. Therefore, as T_1 necessarily is a recurrent state, the distribution of the pair (ν, T_1) is also unique. □

3.6. **Proof of Theorem 2.15: Non-Dirichlet processes.** One way is immediate by the initial remarks in Subsection 2.3: When $G = \alpha(Q-I)$ for an $\alpha > 0$ and Q is a 'constant' kernel with rows equal to μ^t, then ν is a Dirichlet (α,μ) process.

We now show this is a necessary condition. Recall that $G'_{x,y} = \big(\mu_y/\mu_x\big)G_{j,i}$, and so $G' = D^{-1}G^tD$ where D is a diagonal kernel with diagonal entries $\{\mu_i\}$ and G^t is the transpose of G. From the stick-breaking representation of $\nu = \sum_{i\geq 1}P_i\delta_{T_i}$, we may compute first and second moments of $\nu(A)$ for $A \subset \mathscr{X}$. Indeed, recall that $P_i = X_i\prod_{\ell=1}^{i-1}(1-X_\ell)$ where $\{X_i\}$ are iid Beta$(1,\theta)$ r.v.'s, and $\mathbf{T}$ is an independent Markov chain with transition kernel $Q' = I + G'/\theta$ starting from stationary distribution μ.

Then, $E[\nu(A)] = \mu(A)$. Also, after a computation in terms of a Beta$(1,\theta)$ r.v. X where $E[X] = (1+\theta)^{-1}$, $E[(1-X)] = \theta(1+\theta)^{-1}$, $E[X^2] = 2(1+\theta)^{-1}(2+\theta)^{-1}$, $E[(1-X)^2] = \theta(2+\theta)^{-1}$, and $E[X(1-X)] = \theta(1+\theta)^{-1}(2+\theta)^{-1}$, and noting $\mu(A) = \sum_{x,y\in A}\mu_x I_{x,y}$ and $E[\delta_{T_i}(A)\delta_{T_jj}(A)] = \sum_{x,y\in A}\mu_x(Q')^{j-i}_{x,y}$, we have

$$\begin{aligned}
E\big[\nu(A)^2\big] &= \sum_{i\geq 1}E[P_i^2]\mu(A) + 2\sum_{i\geq 1}\sum_{j>i}E[P_iP_j]E[\delta_{T_i}(A)\delta_{T_j}(A)]\\
&= \sum_{i\geq 1}E[(1-X)^2]^{i-1}E[X^2]\mu(A)\\
&\quad + 2\sum_{i\geq 1}\sum_{j>i}E[(1-X)^2]^{i-1}E[X(1-X)]E[(1-X)]^{j-i-1}E[X]E[\delta_{T_i}(A)\delta_{T_j}(A)]\\
&= \sum_{x,y\in A}\mu_x(I-G')^{-1}_{x,y}
\end{aligned}$$

where $(I-G')^{-1}$ is interpreted as $(1+\theta)^{-1}\big(I-\frac{\theta}{1+\theta}Q'\big)^{-1} = (1+\theta)^{-1}\sum_{\ell\geq 0}\big(\frac{\theta}{1+\theta}Q'\big)^{\ell}$.

By choosing $A = \{x\}$ and $A = \{x,y\}$ for $x \neq y$, we conclude with some algebra that

$$\begin{aligned}
E[\nu(x)^2] &= \mu_x(I-G')^{-1}_{x,x}\\
E[\nu(x)\nu(y)] &= \frac{1}{2}\mu_x(I-G')^{-1}_{x,y} + \frac{1}{2}\mu_y(I-G')^{-1}_{y,x}\\
&= \frac{\mu_x}{2}\big\{(I-G')^{-1}_{x,y} + (I-G)^{-1}_{x,y}\big\}.
\end{aligned}$$

We remark that these moment formulas match those in the finite dimensional setting in [12].

Now, if ν is a Dirichlet process with parameters $(\kappa_x)_{x\in\mathscr{X}}$ and $\alpha = \sum_{x\in\mathscr{X}}\kappa_x$, that is with parameters α and base measure $\alpha^{-1}\kappa.$, by matching means, we must have, for $x \neq y$ that $\mu_x = \alpha^{-1}\kappa_x$ and by matching moments of order two,

$$\mu_x(I-G')^{-1}_{x,x} = \frac{\kappa_x(\kappa_x+1)}{\alpha(\alpha+1)}$$
$$\frac{\mu_x}{2}\left\{(I-G')^{-1}_{x,y} + (I-G)^{-1}_{x,y}\right\} = \frac{\kappa_x\kappa_y}{\alpha(\alpha+1)}.$$

Let S be the stochastic kernel with constant rows μ^t. Then, we have

$$\frac{1}{2}(I-G')^{-1} + \frac{1}{2}(I-G)^{-1} = \frac{1}{\alpha+1}(I+\alpha S). \tag{3.19}$$

Consider now complex Hilbert space $\mathscr{H} = L^2(\mathscr{X},\mu)$ and the closed subspace $\mathscr{K} = \{g \in \mathscr{H} : \sum_{x\in\mathscr{X}} g(x)\mu_x = 0\}$ thought of also as a Hilbert space, with Hermitian inner-product $\langle f, g\rangle = \sum \bar{f}(x)g(x)\mu(x)$ where $\bar{f}$ is the complex conjugate. A kernel M on $\mathscr{X}$ is seen as on operator on these Hilbert spaces, where $(Mf)(x) = \sum_{y\in\mathscr{X}} M_{x,y}f(y)$.

Note that $Gf \in \mathscr{K}$, since $\mu^t G = \mathbf{0}^t$, and $Sf = \langle 1, f\rangle\mathbf{1}$, where $\mathbf{0}$ and $\mathbf{1}$ are constant functions taking values 0 and 1 respectively. Then, we observe that $SG = GS = SG' = G'S$ all vanish.

Hence, multiplying (3.19) through by $I-G$ and $I-G'$, we have

$$\alpha(I-S) = GG' + \frac{\alpha-1}{2}(G'+G). \tag{3.20}$$

Since G is a generator ($-G_{x,x} = \sum_{y\neq x} G_{x,y}$ and $G_{x,y} \geq 0$ for $x \neq y$), one obtains that $\alpha = (1-\mu_x - G_{x,x})^{-1}\left[G^2_{x,x} - G_{x,x} + \sum_{y\neq x} G_{x,y}G_{y,x}\frac{\mu_y}{\mu_x}\right] > -G_{x,x}$ for $x \in \mathscr{X}$. Hence, we may write $G = \alpha(U-I)$, where U is a stochastic kernel. Moreover, since μ is a stationary distribution of U, its norm on $\mathscr{H}$ is bounded by 1: For $f \in \mathscr{H}$, $\|Uf\|^2 = \sum_x \left|\sum_y U_{x,y}f(y)\right|^2 \mu_x \leq \sum_{x,y}\mu_x U_{x,y}|f(y)|^2 = \|f\|^2$.

From (3.20), one obtains

$$S = (\alpha+1)\left(\frac{U+U'}{2}\right) - \alpha UU', \tag{3.21}$$

where U' is the adjoint of U: $U'_{x,y} = (\mu_y/\mu_x)U_{y,x}$ for $x, y \in \mathscr{X}$. One also sees readily from (3.21) that $UU' = U'U$.

Hence, U is a bounded, normal operator. Moreover, on $\mathscr{K}$, since S vanishes on $\mathscr{K}$, we have

$$UU' = \frac{\alpha+1}{\alpha}\left(\frac{U'+U}{2}\right).$$

Then, $0 \leq \|Ug\|^2 = \langle Ug, Ug\rangle = [(\alpha+1)/\alpha]\langle g, [(U'+U)/2]g\rangle = [(\alpha+1)/\alpha]|\langle g, Ug\rangle|$ for $g \in \mathscr{K}$, and

$$\sup_{\substack{g\in\mathscr{K}\\ \|g\|=1}} \langle Ug, Ug\rangle = \frac{\alpha+1}{\alpha}\sup_{\substack{g\in\mathscr{K}\\ \|g\|=1}} |\langle g, Ug\rangle|.$$

The left-hand side equals $\|U\|^2_{\mathscr{K}}$, the square of the norm of the restriction of U to $\mathscr{K}$. However, since U is normal, the right-side is equal to $[(\alpha+1)/\alpha]\|U\|_{\mathscr{K}}$ (cf. [3], Proposition 3.2.25 in [42]). Therefore, if $\|U\|_{\mathscr{K}} \neq 0$, it must be equal to $(\alpha+1)/\alpha > 1$, which is impossible since $\|U\|_{\mathscr{K}} \leq \|U\| \leq 1$.

We conclude then that U restricted to $\mathscr{K}$ vanishes. Every function in $\mathscr{H}$ can be uniquely written as a sum of a constant function and a function in K, namely

$f = \langle 1, f\rangle \mathbf{1} + \big(f - \langle 1, f\rangle\big)\mathbf{1}\big)$. Noting $U\mathbf{1} = \mathbf{1}$, we conclude that U is the projection onto constant functions. That is, $U = S$ and $G = \alpha(S - I)$ as desired. □

3.7. **Proof of Theorem 2.21: Type of nonexchangeability.** One way follows immediately: That is, when $\mathbf{T}$ is an iid sequence, necessarily with common distribution μ, the measure ν is a Dirichlet (μ, θ) process from its known stick-breaking representation and hence its representation is permutation exchangeable (cf. Subsection 2.3).

Suppose now that the representation of ν is permutation exchangeable. We will derive several identities which will be helpful. Let π^k be the permutation where $\pi^k(1) = k$, $\pi^k(k) = 1$ and $\pi^k(i) = i$ for $i \neq 1, k$. Let also $\nu^{(k)}$ be the measure $\sum_{i\geq 1} P_{\pi^k(i)}\delta_{T_i}$. For any subset $A \subseteq \mathscr{X}$, we must have $E[\nu(A)^2] = E[\nu^{(k)}(A)^2]$ by exchangeability. Then, by abbreviating $\delta_{T_i}(A) = \delta_{T_i}$, we can write

$$\begin{aligned}
0 &= E\left[\big(\nu(A) - \nu^{(k)}(A)\big)\big(\nu(A) + \nu^{(k)}(A)\big)\right] \\
&= E\big[\big((P_1 - P_k)(\delta_{T_1} - \delta_{T_k}\big)\big((P_1 + P_k)(\delta_{T_1} + \delta_{T_k}) \\
&\qquad + 2P_2\delta_{T_2} + \cdots + 2P_{k-1}\delta_{T_{k-1}} + 2\sum_{i\geq k+1} P_i\delta_{T_i}\big)\big].
\end{aligned} \tag{3.22}$$

In the above expression, from independence of $\mathbf{P}$ and $\mathbf{T}$, the term

$$\begin{aligned}
&E\left[\big((P_1 - P_k)(\delta_{T_1} - \delta_{T_k}\big)\big((P_1 + P_k)(\delta_{T_1} + \delta_{T_k})\big)\right] \\
&= E[\big(P_1^2 - P_k^2]E[\delta_{T_1} - \delta_{T_k}]
\end{aligned} \tag{3.23}$$

vanishes since $E[\delta_{T_j}] = \mu(A)$ for all $j \geq 1$ by stationarity of $\mathbf{T}$.

Also, by reapportioning, we have that $\langle F^{(k)}_{i-k} = P_i/(1 - P_1 - \cdots - P_k) : i \geq k+1\rangle$ has GEM(θ) distribution and is independent of $\{P_1, \ldots, P_k\}$, and as before $\mathbf{T}$. Then, the term

$$\begin{aligned}
&2E\Big[\big((P_1 - P_k)(\delta_{T_1} - \delta_{T_k}\big)\sum_{i\geq k+1} P_i\delta_{T_i}\Big] \\
&= 2E\left[(P_1 - P_k)(1 - P_1 - \cdots - P_k)\right] E\Big[\sum_{i\geq k+1} F^{(k)}_{i-k}(\delta_{T_1} - \delta_{T_k})\delta_{T_i}\Big].
\end{aligned} \tag{3.24}$$

Let X be a Beta$(1, \theta)$ random variable, and note that $E[F^{(k)}_{i-k}] = b(1 - b)^{i-k-1}$ where $b = (1 + \theta)^{-1} = E[X]$. Denote also for $1 \leq i < j$ that

$$r^{j-i} = r^{j-i}(A) := E[\delta_{T_i}(A)\delta_{T_j}(A)].$$

We compute, as $\sum_{i\geq 1} P_i = 1$ a.s. and X is nontrivial, that the factor

$$E\left[(P_1 - P_k)(1 - P_1 - \cdots - P_k)\right] = E\big[(P_1 - P_k)\sum_{j\geq k+1} P_k\big] \tag{3.25}$$

$$= E[X(1 - X)]\sum_{j\geq k+1} b(1 - b)^{j-1-k}\big(E[(1 - X)]^{k-1} - E[(1 - X)^2]^{k-1}\big) > 0.$$

To set up an induction, consider the cases $k = 2, 3$. Recall that $\mathbf{P}$ and $\mathbf{T}$ are independent. In (3.22), the 'cross-term'

$$E\left[\big(P_1 - P_k\big)\big(\delta_{T_1} - \delta_{T_k}\big)\big(2P_2\delta_{T_2} + \cdots + 2P_{k-1}\delta_{T_{k-1}}\big)\right], \tag{3.26}$$

when $k = 2$ does not appear and, when $k = 3$, it vanishes as it reduces to $2E[(P_1 - P_2)P_2](r^1 - r^1) = 0$.

When $k=2$, noting (3.23), (3.24) and (3.25), we have that the equation (3.22) reduces to

$$0=\sum_{i\geq 3} E[F^{(2)}_{i-2}]E[\delta_{T_1}\delta_{T_i}-\delta_{T_2}\delta_{T_i}]=\sum_{i\geq 3} b(1-b)^{i-3}\big(r^{i-1}-r^{i-2}\big)$$

which after summing by parts may be rewritten as $r^1=\sum_{j\geq 1} b(1-b)^{j-1}r^j$.

When $k=3$, correspondingly, we have that (3.22) reduces to $0=\sum_{i\geq 4} b(1-b)^{i-4}\big(r^{i-1}-r^{i-3}\big)$ which is equivalent after simple manipulation to $r^1=r^2$.

We now claim that $r^1=r^{k-1}$ for $k\geq 2$. To show the induction step, suppose that $r^1=\cdots=r^{k-2}$. Then, the cross-term (3.26) vanishes as $E[\delta_{T_1}\delta_{T_j}]=r^{j-1}=r^{k-j}=E[\delta_{T_k}\delta_{T_j}]$ for $2\leq j\leq k-1$. Hence, noting (3.23), (3.24) and (3.25), we have that (3.22) is equivalent to $0=\sum_{i\geq k+1} b(1-b)^{i-k-1}\big(r^{i-1}-r^{i-k}\big)$ or after simple calculation, using the induction assumption, that $r^{k-1}=r^1$.

We now show that the Markov chain $\mathbf{T}$ is aperiodic. Let Q be its transition matrix. Taking $A=\{c\}$ for $c\in\mathscr{X}$, from the proved claim, we have that $\mu(c)Q_{c,c}=r^1=r^k=\mu(c)Q^k_{c,c}$ for $k\geq 1$. Since the chain is irreducible, $\mu(c)>0$, and also there is a time $\tilde{k}$ where $Q^{\tilde{k}}_{c,c}>0$. Hence, $Q_{c,c}=Q^k_{c,c}>0$ for each $k\in\mathbb{N}$ and in particular, the chain is aperiodic.

As a consequence, we have for $c,d\in\mathbb{N}$ that $Q^k_{c,d}\to\mu(d)$ as $k\to\infty$. Hence, for any $A\subseteq\mathscr{X}$, we have that $r^1(A)=r^k(A)=\lim_{j\to\infty} r^j(A)=\mu(A)^2$.

Define stochastic matrices S and Q' by

$$S_{a,b}=\mu(b)\quad\text{ and }\quad Q'_{a,b}=\frac{\mu(b)}{\mu(a)}Q_{b,a} \tag{3.27}$$

for all $a,b\in\mathscr{X}$. Note Q' is the transition probability of the reversed chain. Since $r^k=r^1=\mu(A)^2$, when we take $A=\{a,b\}$ for $a\neq b$, we have

$$\mu(a)Q^k_{a,a}+\mu(a)Q^k_{a,b}+\mu(b)Q^k_{b,a}+\mu(b)Q^k_{b,b}=\big(\mu(a)+\mu(b)\big)^2.$$

It follows that $S=\frac{1}{2}\left(Q^k+(Q')^k\right)$ for $k\geq 1$.

Consider now Hilbert space $\mathscr{H}=L^2(\mathscr{X},\mu)$ with Hermitian inner-product

$$\langle f,g\rangle=\sum \bar{f}(x)g(x)\mu(x)$$

where $\bar{f}$ is the complex conjugate, as in the proof of Theorem 2.15. As before, we consider a kernel M on $\mathscr{X}$ as on operator from $\mathscr{H}$ to $\mathscr{H}$, where $(Mf)(x)=\sum_{y\in\mathscr{X}} M_{x,y}f(y)$. Note that Q' is the adjoint of Q when considered as operators. From (3.27), the operator S is a projection onto constant functions. We now claim that $Q=S$.

Indeed, consider the closed subspace $K=\{g\in\mathscr{H}:\langle 1,g\rangle=\sum_{x\in\mathscr{X}} g(x)\mu_x=0\}\subset\mathscr{H}$. On K, since $Sg=\mathbf{0}$, where $\mathbf{0}$ is the constant function with value 0, we have that $Q^kg=-(Q')^kg$ for $k\geq 1$. By computation, $\sum(Qg)(x)\mu(x)=\sum_{x,y}\mu(x)Q_{x,y}g(y)=\sum_y \mu(y)g(y)$, and so $QK\subseteq K$. Then, we have

$$Q^2g=Q(Qg)=-Q'(-Q'g)=(Q')^2g=-Q^2g$$

for each $g\in K$. In other words, Q^2 annihilates K. Thus, for $g\in K$, we have

$$0=\langle g,Q^2g\rangle=\langle Q'g,Qg\rangle=-\langle Qg,Qg\rangle=-\|Qg\|^2.$$

That is, Q also annihilates K. Let $\mathbf{1}$ be the constant function with value 1. Then, every function in $\mathscr{H}$ can be uniquely written as a sum of a constant function and a function in K, namely $f = \langle 1, f\rangle \mathbf{1} + \big(f - \langle 1, f\rangle\big)\mathbf{1}\big)$. Noting $Q\mathbf{1} = \mathbf{1}$, we conclude that Q is the projection onto constant functions. That is, $Q = S$.

Therefore, Q is composed of identical rows μ, and as a consequence $\mathbf{T}$ is an iid sequence with common distribution μ. This finishes the proof of the theorem. □

Acknowledgements. Research was partially supported by ARO W911NF-18-1-0311, a Simons Foundation Sabbatical grant, and a Daniel Bartlett graduate fellowship.

References

[1] Arratia, R., Barbour, A.D., Tavaré, S.: The Poisson-Dirichlet distribution and the scale-invariant Poisson process. *Combin. Probab. Comput.* **8**(5), 407–416 (1999)

[2] Arratia, R., Barbour, A.D., Tavaré, S.: *Logarithmic combinatorial structures: a probabilistic approach*, vol. 1. European Mathematical Society, Zürich (2003)

[3] Bernau, S.J., Smithies, F.: A note on normal operators. *Proc. Comb. Phil. Soc.* **59** 727–729 (1963)

[4] Blackwell, D., MacQueen, J.B.: Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**(2), 353–355 (1973)

[5] Bouguet, F., Cloez, B.: Fluctuations of the empirical measure of freezing Markov chains. *Elec. J. Probab.* **23**, 1–31 (2018)

[6] Bovier, A., Den Hollander, F.: *Metastability: a Potential-Theoretic Approach*, *Grundlehren der mathematischen Wissenschaften*, vol. 351. Springer, Berlin (2016)

[7] Broderick, T., Jordan, M.I., Pitman, J.: Beta processes, stick-breaking and power laws. *Bayesian Anal.* **7**(2), 439–476 (2012)

[8] Catoni, O., Cerf, R.: The exit path of a Markov chain with rare transitions. *ESAIM: P&S* **1** 95–144.

[9] Crane, H.: Rejoinder: The ubiquitous Ewens sampling formula. *Statist. Sci.* **31**(1), 37–39 (2016)

[10] Crane, H.: The ubiquitous Ewens sampling formula. *Statist. Sci.* **31**(1), 1–19 (2016)

[11] Diaconis, P., Mayer-Wolf, E., Zeitouni, O., Zerner, M.P.W.: The Poisson-Dirichlet law is the unique invariant distribution for uniform split-merge transformations. *Ann. Probab.* **32**(1B), 915–938 (2004)

[12] Dietz, Z., Sethuraman, S.: Occupation laws for some time-nonhomogeneous markov chains. *Elec. J. Probab.* **12**, 661–683 (2007)

[13] Dietz, Z., Sethuraman, S., Lippitt, W.: Stick-breaking processes, clumping, and Markov chain occupation laws. arXiv: 1901.08135v1

[14] Donnelly, P., Tavaré, S.: The population genealogy of the infinitely-many neutral alleles model. *J. Math. Biol.* **25**(4), 381–391 (1987)

[15] Dufresne, D.: The beta product distribution with complex parameters. *Comm. Statist. Theory Methods* **39**(5), 837–854 (2010)

[16] Engen, S.: A note on the geometric series as a species frequency model. *Biometrika* **62**(3), 697–699 (1975)

[17] Engländer, J., Volkov, S.: Turning a coin over instead of tossing it. *J. Theoret. Probab.* **31**(2), 1097–1118 (2018)

[18] Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. *Ann. Stat.* pp. 209–230 (1973)

[19] Gantert, N.: Laws of large numbers for the annealing algorithm. *Stoch. Proc. Appl.* **35**(2), 309–313 (1990)

[20] Ghosal, S., Van der Vaart, A.: *Fundamentals of nonparametric Bayesian inference*, vol. 44. Cambridge University Press, Cambridge (2017)

[21] Ghosh, J., Ramamoorthi, R.: *Bayesian Nonparametrics.* Springer, New York (2003)

[22] Gnedin, A., Kerov, S.: A characterization of GEM distributions. *Combin. Probab. Comput.* **10**(3), 213–217 (2001)

[23] Gnedin, A., Pitman, J.: Regenerative composition structures. *Ann. Probab.* **33**(2), 445–479 (2005)

[24] Gnedin, A., Pitman, J.: Regenerative partition structures. *Elec. J. Combin.* **11**(2), Research Paper 12 (2005)

[25] Goldie, C.M.: Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**(1) 126–166 (1991)
[26] Halmos, P.R.: Random alms. *Ann. Math. Stat.* **15**(2), 182–189 (1944)
[27] Herbach, U.: Stochastic gene expression with a multistate promoter: breaking down exact distributions. arXiv:1804.05740 (2018)
[28] Hirth, U.M.: From GEM back to Dirichlet via Hoppe's urn. *Combin. Probab. Comput.* **6**(2), 185–195 (1997)
[29] Hjort, N.L., Holmes, C., Müller, P., Walker, S.G. (eds.): *Bayesian nonparametrics*, vol. 28. Cambridge University Press, Cambridge (2010)
[30] Hoppe, F.M.: The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.* **25**(2), 123–159 (1987)
[31] Kingman, J.F.C.: Random discrete distributions. *J. Royal Stat. Soc. Ser. B. Stat. Methodol.* pp. 1–22 (1975)
[32] Kingman, J.F.C.: *Poisson Processes*. Clarendon Press, Oxford (1992)
[33] Landim, C.: Metastable Markov chains. arXiv:1807.04144 (2018)
[34] Last, G.: An integral characterization of the Dirichlet process. arXiv:1803.11465 (2018)
[35] Lavine, M.: Some aspects of Polya tree distributions for statistical modeling. *Ann. Stat.* **20**(3), 1222–1235 (1992)
[36] Mathai, A.M., Saxena, R.K.: *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, vol. 348. Springer-Verlag (1973)
[37] McCloskey, J.W.: A model for the distribution of individuals by species in an environment. Ph.D. thesis, Michigan State University, Department of Statistics (1965)
[38] Miclo, L.: Sur let problémes de sortie discrets inhomogénes. *Ann. Appl. Probab.* **6** (4) 1112–1156.
[39] Müller, P., Quintana, F.A., Jara, A., Hanson, T.: *Bayesian Nonparametric Data Analysis.* Springer, New York (2015)
[40] Olivieri, E., Vares, M.E.: *Large Deviations and Metastability*, vol. 100. Cambridge University Press, Cambridge (2005)
[41] Patil, G., Taillie, C.: Diversity as a concept and its implications for random communities. *Bull. Int. Stat. Inst.* **47**, 497–515 (1977)
[42] Pederson, G.T..: *Analysis Now*, Revised printing, GTM **118**, Springer-Verlag, New York, (1989)
[43] Pitman, J.: Random discrete distributions invariant under size-biased permutation. *Adv. in Appl. Probab.* **28**(2), 525–539 (1996)
[44] Pitman, J.: Some developments of the Blackwell-MacQueen urn scheme. Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell **30**, 245–267 (1996)
[45] Pitman, J.: *Combinatorial Stochastic Processes*: Ecole d'Eté de Probabilités de Saint-Flour XXXII-2002. Springer-Verlag, Berlin (2006)
[46] Pitman, J., Yakubovich, Y., et al.: Ordered and size-biased frequencies in GEM and Gibbs models for species sampling. *Ann. Appl. Probab.* **28**(3), 1793–1820 (2018)
[47] Pitman, J., Yor, M.: The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900 (1997)
[48] Schiavo, L.D., Lytvynov, E.W.: A Mecke-type characterization of the Dirichlet-Ferguson measure. arXiv:1706.07602 (2017)
[49] Sethuraman, J.: A constructive definition of Dirichlet priors. *Statist. Sinica* pp. 639–650 (1994)
[50] Winkler, G.: *Image Analysis, Random Fields and Dynamic Monte Carlo Methods.* Springer-Verlag, Berlin (1995)

E-mail: zedietz@yahoo.com, wlippitt@email.arizona.edu, sethuram@math.arizona.edu

W.L. & S.S.: Department of Mathematics, University of Arizona, Tucson, AZ 85721