CLUMPING, STICK-BREAKING, AND AN INHOMOGENEOUS MARKOV CHAIN

by

William Lippitt

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by ***William Lippitt***, titled ***Clumping, Stick-breaking, and an Inhomogeneous Markov Chain*** and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Sunder Sethuraman Date: 6/20/2019

David Glickenstein Date: 6/20/19

Hao Helen Zhang Date: 06/20/19

Edward Bedrick Date: 6/20/19

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Sunder Sethuraman Date: 6/20/2019
Dissertation Committee Chair
Department of Mathematics

Acknowledgments

My deepest thanks go to my advisor, Sunder Sethuraman. Without your unwavering support and patience through my struggles, your encouragement to pursue every idea, and the necessary pushes you gave to keep improving, I'd never have accomplished half so much. Thank you for helping me to do work I can take pride in.

To Dave Glickenstein, Helen Zhang, and Ed Bedrick: Thank you all for serving on my committee and for your support, time, and patience. In particular, Dave, I thank you for your efforts as Associate Head, for showing me a path to success when I was struggling in my early years to find one and helping me to set out upon it.

My thanks also for the support I received during my studies through the ARO grant W911NF-18-1-0311 and a Daniel Bartlett graduate fellowship.

I thank my partner, Hannah. Without you, I'd have lost my mind by the end. As it is, I still lost it a few times, but you always helped me to find it again. I thank my friends Ben, Georgia, and Angie for distracting me with good fun when I needed it, and commiserating with me when I couldn't be distracted.

For leading me down this road in the first place: Dad, thank you for your ridiculous (yet correct) idea that factoring polynomials with your son and teaching him about slopes at a point sounded like good fun. Jim Conklin, my undergraduate thesis advisor and professor of half of my courses, thank you for showing me that math, *real math*, isn't just fun, it's beautiful.

And finally, I thank my 6th grade math and English teacher, Mrs. Connie Patterson. You gave me my very first lesson in true mathematics when, upon seeing my tendency to compute before thinking, you said "A good mathematician is lazy. It's just that they put a phenomenal amount of effort into finding the most efficient way to be lazy." I've never forgotten.

Dedication

To my Dad

For looking forward to reading this.

Table of Contents

TABLE OF CONTENTS—*Continued*

List of Figures

Abstract

We study the connections among three types of objects: Residual Allocation Models (RAMs), a generalized class of stick-breaking processes which include Dirichlet processes, and the occupation laws of certain discrete space time-inhomogeneous Markov chains. These connections are established through a new clumping procedure on RAMs according to a clumping sequence. We introduce and analyze two methods of its application with respect to homogeneous Markov chains. Clumping according to the first method results in a class of stick-breaking measures we later identify as the limit of the empirical occupation measures for particular time-inhomogeneous Markov chains. Clumping according to the second method leads to a general study of self-similarity as a characterization of random probability measures, which we apply in context.

We further discuss the occupation law of an inhomogeneous Markov chain related to simulated annealing and its connections to stick-breaking measures. By introducing a reverse-chronological clumping procedure, we define and study the local occupations up to time n of the Markov chain. As n gets large, these local occupations converge jointly to the clumped RAM structure resulting from the first method above.

We then explore additional properties of the generalized stick-breaking measure uncovered by the connection to the occupation law, and consider potential applications of results.

Chapter 1
Introduction

We consider the connections among discrete space time-inhomogeneous Markov chains related to simulated annealing, clumped residual allocation models (RAMs), and a 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 generalized 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 or types of mRNA dynamics. 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. One purpose of this work is to develop what seems to be an unexpected connection between these a priori different subjects.

Before discussing the context of this work in depth, we give a brief overview of the objects of interest and how connections among them might be established. A formal treatment will follow in the main chapters.

We consider Markov chains on discrete spaces $\mathfrak{X} \subseteq \mathbb{N}$ consisting of a finite or a countably infinite number of elements. Let G be a generator matrix on $\mathfrak{X}$, that is $G_{i,j} \geq 0$ for $i \neq j \in \mathfrak{X}$ 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 for now that each K_n is irreducible and positive recurrent for sufficiently large n. Extensions will be formally considered in depth in later chapters.

We will be interested in the time-inhomogeneous Markov chain $\mathbf{M} = (M_n)_{n\geq 1}$ on $\mathfrak{X}$ associated to kernels $(K_n)_{n\geq 1}$. In such chains, every point in $\mathfrak{X}$ 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. Such chains can be considered as a simplified models of simulated annealing.

For finite spaces $\mathfrak{X} = \{1, 2, ..., k\}$, empirical measures of these chains do not converge a.s. or in probability, as would be the case for a homogeneous Markov chain. However, for generators G without zero entries, weak convergence to an empirical occupation law

$$\nu \stackrel{d}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \delta_{M_j} \tag{1.2}$$

was identified by computing its moments in [13]. When G is of the form $G = \theta(Q - I)$ for $\theta > 0$ and Q a stochastic matrix with constant rows μ, it was shown that ν is a Dirichlet distribution with parameters $\{\theta\mu_i\}_{i=1}^{k}$ by matching the moments.

In this context, one of our goals is to understand more constructively the limit (1.2) and its generalization to a countably infinite state space by considering so-called local occupations of the chain. These local occupations lead to identification of the limit ν as a stick-breaking process.

To understand this structure, we now consider RAMs and stick-breaking processes.

Consider the infinite-dimensional simplex Δ_∞ of all discrete probability distributions on $\mathbb{N} = \{1, 2, \ldots\}$. A Residual Allocation Model (RAM) is a class of distribution on Δ_∞ developed to address problems of apportionment: Let $\mathbf{X} = (X_n)_{n\geq 1}$ be independent $[0, 1]$-valued random variables, called residual proportions. Consider the

associated process $\mathbf{P} = (P_n)_{n\geq 1} \in [0,1]^{\mathbb{N}}$, given by $P_1 = X_1$ and

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

If $\sum_{n\geq 1} P_n \stackrel{a.s.}{=} 1$, the distribution $\mathbf{P}$ on Δ_∞ is called the associated RAM.

The RAM $\mathbf{P}$ when the proportions $\mathbf{X}$ are iid Beta$(1,\theta)$ random variables is the well-known Griffiths-Engen-McCloskey (GEM) distribution with parameter θ. The GEM distribution is a fundamental building block of Dirichlet processes, a class of random probability measures which often serve as priors in Bayesian nonparametric statistics.

The stick-breaking representation of the Dirichlet process with parameter measure $\theta\mu$, given in terms of a GEM(θ) distribution $(P_n)_{n\geq 1}$, and an independent sequence of iid random variables $(\omega_n)_{n\geq 1}$ with common distribution μ, is given by

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

In this work, we focus on a class of generalized stick-breaking processes on $\mathfrak{X} \subseteq \mathbb{N}$: Let $\mathbf{P}$ be a GEM(θ) distribution and let $\mathbf{T} = (T_n)_{n\geq 1}$ be an independent stationary Markov chain with irreducible transition kernel Q and stationary distribution μ on $\mathfrak{X}$. The random stick-breaking measures

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

are natural generalizations of the stick-breaking representation of the Dirichlet process, here with respect to a stationary Markov chain $\mathbf{T}$ instead of the iid sequence in (1.3). We note that other generalizations of Dirichlet processes have been considered, among them, Polya tree [44], Pitman-Yor [54], [58], and Beta processes [8]. This generalization ν appears new, and might find application in future in the Bayesian context.

This generalization of course includes the Dirichlet process as a case. For a large class of stationary chains $\mathbf{T}$, however, ν will not be a Dirichlet process. When ν is

not a Dirichlet process, ν may have marginal distributions which are nontrivial Beta Products, sometimes with complex parameters. In the finite dimensional case, the moments can be completely characterized. We will also show that ν satisfies a so-called self-similarity equation uniquely characterizing its distribution. This equation is reminiscent of the regenerative structure present in traditional stick-breaking [60], in integral constructions of the Dirichlet processs [43], [59], and in other related settings [27], [26].

Another characteristic of our generalization ν is that it is not 'permutation exchangeable' when the Markov chain $\mathbf{T}$ is not an iid sequence in the sense that the GEM distribution $\mathbf{P}$ may not be replaced by an arbitrary permutation of $\mathbf{P}$ without changing the distribution of ν. In contrast, when $\mathbf{T}$ is iid and ν is the Dirichlet process, such an exchangeability property holds. For example, the Poisson-Dirichlet order statistics $(\hat{P}_n)_{n\geq 1}$ of $\mathbf{P}$ may be used in place of $\mathbf{P}$ without changing the Dirichlet process.

Given this nonexchangeability, standard tools and representations for analysis of exchangeable constructions, such as those related to Dirichlet processes, the Poisson-Dirichlet distribution, and Gamma processes, are not readily applicable. We will rely on the nuts-and-bolts definition of the GEM distribution, the structure of the Markov chain $\mathbf{T}$, and the clumping procedure to analyze ν.

Returning to the time-inhomogeneous Markov chain $\mathbf{M} = (M_n)_{n\geq 1}$ with kernels $(K_n)_{n\geq 1}$ (1.1), we now describe the connection with the stick-breaking measure (1.4). Consider the random empirical occupation measure on $\mathfrak{X}$,

$$\nu^n = \frac{1}{n}\sum_{j=1}^{n} \delta_{M_i}.$$

Note the 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-chronological clumping procedure to gather adjacent observations by $\mathbf{M}$ of the same

state into 'clumped' or 'local' 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.

To be precise, let $1 = \tilde{V}_1 < \tilde{V}_2 < \cdots$ be the successive times when the Markov chain changes state. The differences $(\tilde{V}_{j+1} - \tilde{V}_j)_{j\geq 1}$ are the lengths of clumps of the chain $\mathbf{M}$. Going backwards from time n, let $\mathbf{P}_n = (P_{n,j})_{j\geq 1}$ be the sequence of local occupations up to time n, i.e. $P_{n,j} = \frac{\tilde{V}_{J_n+2-j} - \tilde{V}_{J_n+1-j}}{n}$ for $2 \leq j \leq n$ where J_n is the number of clumps of $\mathbf{M}$ up to time n, $P_{n,1}$ has an appropriately modified definition as a boundary occupation, and $P_{n,k} = 0$ for $k > J_n$. Let $\mathbf{Y}_n = (Y_{n,j})_{j\geq 1}$ be the sequence of states corresponding with the local occupations, taking by convention $Y_{n,k} = M_1$ for $k > J_n$, such that the empirical measure ν^n can be written as

$$\nu^n = \sum_{j=1}^{J_n} P_{n,j}\delta_{y_{n,j}} = \sum_{j=1}^{\infty} P_{n,j}\delta_{Y_{n,j}}.$$

We show 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 $\alpha_j G_{j,i}/[\alpha_i(-G_{i,i})]$, where α is the stochastic left eigenvector of G. Furthermore, the pair $(\mathbf{P}_n, \mathbf{Y}_n)$ converges jointly in distribution to the pair $(\mathbf{R}, \mathbf{Z})$ where conditional on $\mathbf{Z}$, we have $\mathbf{R} = (R_j)_{j\geq 1}$ is a RAM constructed from independent proportions distributed as $\text{Beta}(1, -G_{Z_j,Z_j})$. We say the pair $(\mathbf{R}, \mathbf{Z})$ has the MCcGEM(G') distribution with respect to α, where $G'_{i,j} = \frac{\alpha_j}{\alpha_i} G_{j,i}$ for $i, j \in \mathfrak{X}$.

To match the limit of ν^n to ν in (1.4), we will also need to clump the construction of ν and consider its intermediate structure and correspondence to the local occupations above with MCcGEM distribution. To this end, analogous to the switching times $\tilde{\mathbf{V}}$ of $\mathbf{M}$, suppose $1 = V_1 < V_2 < ...$ are the times when the stationary Markov chain $\mathbf{T}$ changes state. With respect to $\mathbf{P}$, consider the clumped probabilities $P_i^V = \sum_{j=V_i}^{V_{i+1}-1} P_j$ for $i \geq 1$, this time in chronological order. We show that the joint law of

$(\mathbf{P^V}, \mathbf{Y})$ is a MCcGEM($\theta(Q-I)$) distribution with respect to μ.

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

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

In particular, when $G' = \theta(Q-I)$ and $\alpha = \mu$, the constructions (1.2) and (1.5) from intermediate structures can be matched.

Much of this work was developed jointly with Zach Dietz and Sunder Sethuraman. In particular, Theorems 2.5, 3.2, 3.4, 3.7, 3.8, 3.9, 4.9, and a subcase of Theorem 2.8; Propositions 2.4, 2.12, 3.5, and 4.1; and Corollary 2.4.1 of this work may also be found in [12]. In some unpublished notes, Dietz established a connection between the traditional stick-breaking construction of the Dirichlet measure and the occupation law of an inhomogeneous Markov chain with kernels (1.13) where G is a generator which can be decomposed into a constant stochastic matrix Q and positive constant θ. The work published in [12] builds on these notes by refining and generalizing Dietz's methods to apply more broadly.

In the remainder of this chapter, we introduce more formally the objects discussed here and their context in previous works. We begin with some preliminary material.

1.1 Preliminaries

We start with notational conventions integral to understanding this dissertation. We then give a brief overview of random probability distributions in Section 1.1.2, mainly over discrete spaces, and a review of basic Markov chain theory in Section 1.1.3

We assume some knowledge of common probability distributions such as Beta, Beta Product, Gamma, and Geometric, precise definitions and properties of which may be found in Appendix A.1.

1.1.1 Notational Conventions

Empty sums are taken to equal 0, and empty products are taken to equal 1. The infimum of the empty set is understood to equal $+\infty$. The natural numbers are taken to start at 1: $\mathbb{N} = \{1, 2, 3, ...\}$.

We will say that a deterministic sequence $\mathbf{z}$ is a 'possible' sequence for a stochastic process $\mathbf{Z} = (Z_j)_{j \geq 1}$ on $\mathfrak{X}$ if the event $\{Z_i = z_i : 1 \leq i \leq n\}$ has positive probability for each $n \geq 1$.

For a number α and a positive number k, the Pochhammer symbol $(\alpha)_k$ is used to denote the rising factorial $(\alpha)_k = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}$.

1.1.2 Random Probability Distributions

Much of the work in this dissertation involves random probability distributions over a discrete space $\mathfrak{X}$. Such a random distribution takes values in the space of probability distributions over $\mathfrak{X}$, which is commonly denoted $\Delta_{\mathfrak{X}}$:

Definition 1.1 ($\Delta_{\mathfrak{X}}$)**.** Let $\mathfrak{X}$ be a discrete space. We define the space of probability distributions on $\mathfrak{X}$ as follows:

$$\Delta_{\mathfrak{X}} = \left\{ (p_x)_{x \in \mathfrak{X}} \in [0,1]^{\mathfrak{X}} : \sum_{x \in \mathfrak{X}} p_x = 1 \right\}$$

In the case that $\mathfrak{X} = \{1, 2, 3, ..., k\}$, we denote $\Delta_{\mathfrak{X}} = \Delta_k$. In the case that $\mathfrak{X} = \mathbb{N}$, we denote $\Delta_{\mathfrak{X}} = \Delta_\infty$.

For example, random variables with Dirichlet distribution are random probability distributions taking value in Δ_k for some $k \in \mathbb{N}$:

Definition 1.2 (Dirichlet Distribution)**.** Let $\alpha \in (\mathbb{R}^+)^k$. We say a random k-vector P taking values (a.s.) in Δ_k has Dirichlet distribution with parameters α, denoted

$\mathrm{D}(\alpha)$, if it has probability density function relative to $\prod_{j=1}^{k-1} dp_j$ given by:

$$f_\alpha(p) = \Gamma\left(\sum_{j=1}^{k} \alpha_j\right) \prod_{j=1}^{k} \frac{p_j^{\alpha_j - 1}}{\Gamma(\alpha_j)} \ .$$

A Dirichlet process is a generalization of the Dirichlet distribution which takes values in the space of probability measures over a measurable space $\mathfrak{Y}$. Dirichlet processes are particularly important examples of random probability distributions.

Definition 1.3 (Dirichlet Process)**.** Let $\mathfrak{Y}$ be a measurable space, and let ν be a random variable taking values (a.s.) in the space of probability measures on $\mathfrak{Y}$. Let μ be a deterministic probability measure on $\mathfrak{Y}$ and θ a positive constant. If for each finite, measurable partition $\{A_j\}_{j=1}^n$ of $\mathfrak{Y}$, we have

$$(\nu(A_1), \nu(A_2), ..., \nu(A_n)) \sim \mathrm{Dir}\left(\theta\mu(A_1), \theta\mu(A_2), ..., \theta\mu(A_n)\right)$$

we say ν is a Dirichlet process with parameter $\alpha = \theta\mu$, i.e. $\nu \sim \mathrm{DP}(\alpha)$.

1.1.3 Markov Chains

In this dissertation, we work with Markov chains in discrete space and time, a variety of stochastic process, i.e. random function, on $\mathbb{N}$ taking values in a countable space $\mathfrak{X}$. In the context of Markov chains, we call $\mathfrak{X}$ the state space, and will assume that $\mathfrak{X}$ contains at least two elements, potentially infinitely many.

Given such a space $\mathfrak{X}$, a Markov chain $\mathbf{M}$ on $\mathfrak{X}$ can then be defined:

Definition 1.4 (Markov Chain - Discrete Space, Time)**.** A Markov chain on $\mathfrak{X}$ is a sequence of random variables $\mathbf{M} = (M_j)_{j\geq 1}$ on a shared probability space taking values in $\mathfrak{X}$ which satisfies the Markov Property:

For each $n \in \mathbb{N}$ and each deterministic sequence $\mathbf{m} = (m_j)_{j\geq 1}$ taking values in $\mathfrak{X}$, whenever $\mathscr{P}(M_j = m_j : 1 \leq j \leq n) > 0$, it holds

$$\mathscr{P}(M_{n+1} = m_{n+1} | M_j = m_j : 1 \leq j \leq n) = \mathscr{P}(M_{n+1} = m_{n+1} | M_n = m_n).$$

The distribution of $\mathbf{T}$ is then characterized by its transition kernels $(K_n)_{n\geq 1}$, equivalently transition matrices, given by

$$K_n(x,y) = \mathscr{P}(M_{n+1} = y | M_n = x),$$

and its initial distribution μ, given by $\mu_x = \mathscr{P}(M_1 = x)$.

We see that μ and the kernels K_n, as examples of a probability distribution and a collection of conditional laws respectively, are necessarily stochastic. That is

Definition 1.5 (Stochastic)**.** Given a discrete space $\mathfrak{X}$, a stochastic vector on $\mathfrak{X}$ is a vector $\mu = (\mu_x)_{x\in\mathfrak{X}}$ such that

$$\forall x \in \mathfrak{X}, \quad \mu_x \geq 0 \qquad \text{and} \qquad \sum_{x\in\mathfrak{X}} \mu_x = 1.$$

A stochastic matrix on $\mathfrak{X}$ is a square matrix $K = (K_{x,y})_{x,y\in\mathfrak{X}}$ such that

$$\forall x, y \in \mathfrak{X}, \quad K_{x,y} \geq 0 \qquad \text{and} \qquad \forall x \in \mathfrak{X}, \ \sum_{y\in\mathfrak{X}} K_{x,y} = 1.$$

Often of particular interest is a Markov chain for which the transition kernels K_n do not depend on time, i.e. for some stochastic kernel K and for all n, we have $K_n = K$.

Definition 1.6 (Homogeneous Markov Chain)**.** Let $\mathbf{M}$ be a discrete time and space Markov chain with transition kernels $(K_n)_{n\geq 1}$. If there exists a stochastic matrix K such that for all n, $K_n = K$, then $\mathbf{M}$ is called time-homogeneous, or just homogeneous. In this case, the distribution of $\mathbf{M}$ is characterized by K and its initial distribution. If a Markov chain is not time-homogeneous, it is called time-inhomogeneous, or just inhomogeneous.

A homogeneous Markov chain can be understood as a random walk. Suppose $\mathfrak{X}$ is a set of locations which an individual can visit. The individual walks randomly from location to location in such a way that might be termed forgetful. The individual

randomly selects their next stop based on their current location without any concern for where they've been before. If we try to predict their location according to some information regarding where they're been, the only relevant information is their most recent location and how long ago they visited. This forgetfulness, i.e. the irrelevance of all but the most recent past to the future, is the Markov property.

In the context of this random walk, we might be interested in the locations the individual visits frequently, or the chances that the individual may always, or perhaps never, return to a particular location. This leads to the following definitions.

Definition 1.7 (Communication Classes)**.** Let $\mathbf{M}$ be a homogeneous Markov chain on $\mathfrak{X}$ with transition kernel Q, and let $x, y \in \mathfrak{X}$. With respect to $\mathbf{M}$, equivalently with respect to Q, we say that x leads to y if

$$\mathbb{1}(x = y) + \sum_{n=1}^{\infty} \mathscr{P}(T_{n+1} = y | T_1 = x) > 0$$

If x leads to y and y leads to x, we say x and y communicate. A communication class of $\mathbf{M}$, equivalently of Q, is a maximal subset C of $\mathfrak{X}$ such that all pairs of states in C communicate.

Definition 1.8 (Recurrent, Transient, and Absorbing States)**.** Let $\mathbf{M}$ be a homogeneous Markov chain on $\mathfrak{X}$ with transition kernel Q. Define for each $x \in \mathfrak{X}$ the first time $\mathbf{M}$ visits state x after starting: $\tau_x = \inf\{n > 1 : M_n = x\}$, allowing for the possibility of no return and $\tau_x = \infty$. Then, if

$$\mathscr{P}(\tau_x < \infty | M_1 = x) = 1,$$

we call x recurrent. Otherwise, x is called transient. If a recurrent state x satisfies

$$\mathbb{E}[\tau_x] < \infty,$$

then x is called positive recurrent. If x is recurrent but not positive recurrent, it is called null recurrent.

If all states in $\mathfrak{X}$ are transient, positive recurrent, or null recurrent, then $\mathbf{M}$, equivalently Q, is said to be transient, positive recurrent, or null recurrent respectively.

If for a state x, $Q_{x,x} = 1$, then x is called an absorbing state.

In the context of a random walk on locations $\mathfrak{X}$, we see that one location leads to another if an individual can potentially travel from the first location to the second with positive probability. If the individual has positive probability of making a return trip, the two locations are said to communicate. Note that 'leading to' is not symmetric. In a random walk, this lack of symmetry would happen if two locations are linked only by one way streets pointing North or East. However, communication is reflexive, symmetric, and associative, thus forming an equivalence relation, the classes of which are the communication classes.

Recurrence and transience then describe the individual's frequency of return to a location upon leaving it. If the individual on the random walk returns almost surely (a.s.), i.e. with probability 1, then the location is called recurrent. In fact, if the individual reaches a recurrent location, then they will a.s. return to it infinitely often. However, a transient location is one the individual may not ever return to. With probability one, transient locations are visited only finitely often, if at all. Recurrent locations then have the two sub-classifications of positive or null recurrent. We expect the individual to return to positive recurrent locations in finite time, but may wait indefinitely for the individual to return to null recurrent locations. An absorbing location is one which is inescapable.

Note that transience, positive recurrence, and null recurrence are class properties: if one state of a communication class is positive recurrent, all states are positive recurrent. An absorbing state is always in its own communication class as it communicates only with itself.

When, starting from any location, the individual can potentially travel to any other location, we call their walk irreducible:

Definition 1.9 (Irreducible)**.** If a homogeneous Markov chain $\mathbf{M}$ with transition kernel Q has a single communication class, both $\mathbf{M}$ and Q are called irreducible. Otherwise, $\mathbf{M}$ and Q are called reducible.

If the individual's travels are reducible, then there are at least two locations x and y such that, if the individual is observed at location x, they will from that point on never be observed at location y.

When a homogeneous Markov chain with kernel Q has at least one positive recurrent state, we can associate to $\mathbf{M}$ a stochastic vector μ such that, if $\mathbf{M}$ has initial distribution μ, the marginal distribution of each step M_j in the chain is also μ.

Definition 1.10 (Stationary)**.** Let $\mathbf{M}$ be a homogeneous Markov chain with transitional kernel Q. If μ is a stochastic vector such that for each state $y \in \mathfrak{X}$, it holds

$$\sum_{x \in \mathfrak{X}} \mu_x Q_{x,y} = \mu_y$$

then μ is called a stationary distribution of $\mathbf{M}$ and Q.

Every positive recurrent communication class of a stochastic kernel Q is uniquely associated with a stationary distribution. All other stationary distributions of Q can then be expressed as the weighted average of the stationary distributions associated with communication classes.

Proposition 1.11. *Let Q be an irreducible, positive recurrent stochastic matrix on $\mathfrak{X}$. Then Q has a unique stationary distribution. If additionally there exists a positive integer power of Q with all strictly positive entries, then for every stochastic vector π on $\mathfrak{X}$, it holds entry-wise that*

$$\pi^t Q^n \longrightarrow \mu^t \qquad \text{as } n \text{ approaches infinity,}$$

where π^t denotes the transpose of π, and $\pi^t Q^n$ denotes the usual vector and matrix multiplication.

For a proof, see, for example, Durrett [17].

A concrete example of a reducible homogeneous Markov chain in the context of a random walk may be found in Appendix A.2 with communication classes and stationary distributions computed. Consider also the following example:

Example 1.12. Let $\mathbf{M} = (M_j)_{j\geq 1}$ be an iid random sequence taking values in $\mathfrak{X}$ with common distribution μ. Then $\mathbf{M}$ is a stationary homogeneous Markov chain with initial distribution μ and stochastic kernel Q given by $Q_{x,y} = \mu_y$ for each pair of states $x, y \in \mathfrak{X}$.

Such a matrix Q with identical rows μ is called constant stochastic. Every distribution π then satisfies $\pi^t Q = \mu^t$. Thus, a Markov chain run with a constant stochastic matrix Q is an independent sequence, and Q has μ as its unique stationary distribution.

We can then consider a stationary homogeneous Markov chain to be a natural generalization of an iid sequence.

One aspect of a Markov chain we might consider is its empirical measure, or empirical occupation measure:

Definition 1.13. Let $\mathbf{M} = (M_j)_{j\geq 1}$ be a Markov chain taking values in $\mathfrak{X}$. Then the nth empirical measure ν^n of $\mathbf{M}$ is a probability measure on $\mathfrak{X}$ defined as follows:

$$\nu^n = \frac{1}{n}\sum_{j=1}^{n} \delta_{M_j} \qquad\qquad \forall x \in \mathfrak{X}: \quad \nu^n_x = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}(M_j = x)$$

We see that ν^n_x is then the observed proportion of time that the chain $\mathbf{M}$ spends at state x in the first n time steps. By limiting, we might study the global occupation measure of $\mathbf{M}$ over all of time, i.e.

$$\lim_{n\to\infty} \nu^n,$$

where this limit might be taken almost surely, in probability, or weakly (in distribution), depending on context.

Definition 1.14 (Occupation Law)**.** The occupation law of a Markov chain $\mathbf{M}$ on a state space $\mathfrak{X}$, if it exists, is a (potentially random) probability measure ν on $\mathfrak{X}$ which is the weak limit of the empirical measures $(\nu^n)_{n\geq 1}$ of $\mathbf{M}$. In other words,

$$\nu \stackrel{d}{=} \lim_{n\to\infty} \nu^n.$$

For homogeneous Markov chains, this limit converges almost surely, but not necessarily to a probability distribution. For example, if $\mathbf{M}$ is a homogeneous Markov chain consisting only of transient states, the global occupation measure is the zero measure. If, however, $\mathbf{M}$ is an irreducible, positive recurrent homogeneous Markov chain with strictly positive transition kernel Q, the limit of the empirical measures of $\mathbf{M}$ is almost surely the unique stationary distribution of $\mathbf{M}$.

For an inhomogeneous Markov chain, the situation is more complicated. The empirical measures may converge almost surely, in probability, in distribution, or not at all.

1.2 Background

1.2.1 The Dirichlet Process

The history of the Dirichlet process which follows aims to give a sense of the development of the Dirichlet process and the surrounding subjects at the surface level. The uses and impacts of Dirichlet processes and their cousins the Poisson-Dirichlet and GEM distributions are so wide and varied that a complete history would involve lengthy discussions in number theory, combinatorics, population genetics, Bayesian statistics, and more. In this recount, we focus more on developments related to Bayesian statistics and population genetics. Contributions from the study of population genetics were vital to the introduction of objects studied in this dissertation, and

we see certain aspects of this dissertation as motivated by application in Bayesian statistics.

In 1973, the Dirichlet process was introduced by Ferguson [22] as a practical example of a Bayesian prior that could be used for nonparametric problems. A prior distribution on a measurable space $\mathfrak{Y}$ is a distribution on probability distributions on $\mathfrak{Y}$. For example, in the case that $\mathfrak{Y} = \{1, 2, ..., k\}$, a prior distribution is a distribution on Δ_k, such as a Dirichlet distribution. If the prior distribution is supported over an infinite dimensional space of probability distributions, for example when $\mathfrak{Y} = [0, 1]$, then the context is deemed nonparametric.

Let us denote the prior distribution by π. If ν is a random variable with distribution π, and $\boldsymbol{X} = (X_1, ..., X_n)$ is a sequence of random variables such that $\mathscr{P}(X_j \in A_j : 1 \leq j \leq n|\nu) = \prod_{j=1}^{n} \nu(A_j)$ (equivalently, $\boldsymbol{X}|\nu$ is iid $\sim \nu$), then the conditional law $\pi(\,\cdot\,|\boldsymbol{X})$ of ν given $\boldsymbol{X}$ is called the posterior distribution.

Ferguson was particularly concerned in [22] with finding a prior distribution for nonparametric problems having both a suitably large support and a posterior distribution that could be reasonably handled analytically. He introduced the Dirichlet process as such a prior, having a support dense in the set of probability distributions absolutely continuous with respect to the parameter measure and being a conjugate prior. Specifically, a family of prior distributions (such as the family of Dirichlet processes) is considered conjugate when the posterior distribution is a member of the same family:

Theorem 1.15 (Ferguson, [22])**.** *Let $\mathfrak{Y}$ be a measurable space and α a measure on $\mathfrak{Y}$ such that $0 < \alpha(\mathfrak{Y}) < \infty$. If ν has distribution* $\mathrm{DP}(\alpha)$ *and $\boldsymbol{X} = (X_1, ..., X_n)$ is a sequence of random variables taking value in $\mathfrak{Y}$ such that $\boldsymbol{X}|\nu$ is iid $\sim \nu$, then*

$$\nu|\boldsymbol{X} \;\sim\; \mathrm{DP}(\tilde{\alpha})$$

where $\tilde{\alpha} = \alpha + \sum_{j=1}^{n} \delta_{X_j}$.

Ferguson's constructions of the Dirichlet process involved Gamma processes, which are fundamentally connected to the study of the Dirichlet process through its construction via Poisson point processes.

Blackwell and MacQueen provide an alternative realization of the Dirichlet process through a modified Pólya urn scheme allowing a continuum of colors, also in 1973 [6]. This realization gives the Dirichlet process as the limit of the empirical measures of an exchangeable sequence. Blackwell [5] also expanded upon Ferguson's work to show that with probability 1, an observation of a Dirichlet process on a separable space $\mathfrak{Y}$ is a discrete probability distribution. That is, Dirichlet processes are examples of random discrete probability distributions. While Ferguson's proof in [22] made use of the construction of a Dirichlet process through a Gamma process, Blackwell's did not appeal to a particular construction, but rather to assumptions regarding the underlying space $\mathfrak{Y}$.

A distribution closely related to the Dirichlet process is the Poisson-Dirichlet distribution, which takes values in Δ_∞. Kingman defined the Poisson-Dirichlet distribution in 1975 [38] and connected it to the Dirichlet process through Ferguson's use of a Gamma process, ultimately introducing the machinery of Poisson point processes to the study of the Dirichlet process. Specifically, to produce a random measure with distribution $\mathrm{DP}(\alpha)$ on $\mathfrak{Y}$, one observes a Poisson point process Π on $\mathfrak{Y} \times (0, \infty)$ with underlying intensity measure $\alpha \times \beta$ with $d\beta = x^{-1}e^{-x}dx$, and defines $m(A) = \sum_{(y,x)\in\Pi;\ y\in A} x$. Then $m(A)$ has Gamma distribution for each measurable A. On the one hand, $\nu(\cdot) = m(\cdot)/m(\mathfrak{Y})$ is a Dirichlet process with parameter measure α. On the other, the order statistics of the set $\{x : (y, x) \in \Pi\}$ normalized by their total sum are said to have Poisson-Dirichlet distribution with parameter $\alpha(\mathfrak{Y})$ as a matter of definition. Kingman gave a review of this connection in his book [41].

At this time, Kingman and others were considering models for studying allele frequency within a population in which the Poisson-Dirichlet distribution naturally arose. In fact, much of the development of the theory surrounding Poisson-Dirichlet

distributions began in the context of these models. With this in mind, we now provide a brief description of the development of these models. Much of this exposition follows Ewens recount of the history of allele frequency models in [21] as it ultimately relates to the Dirichlet process. Ewens' work with such models has been fundamental, and his account of the history is comprehensive.

Classic allele frequency models such as Wright-Fisher [23, 63], its exchangeable generalization by Cannings [9], and the Moran model [48] describe the way the frequency of an allele in a population of fixed size changes over time under the assumption of finitely many alleles. In particular, the Wright-Fisher model tracks the number of instances of allele type 1 in a population of fixed size M over generations of genes. If one generation of genes has i instances of allele type 1, then the probability of j instances in the next generation is given by

$$p_{ij} = \binom{M}{j} \left(\frac{i}{M}\right)^j \left(\frac{M-i}{M}\right)^{M-j}.$$

Wright and Fisher considered a diffusion approximation to their model derived by sending the population size M to infinity and rescaling the number of generations represented by unit time appropriately. This diffusion approximation, with minor modification, holds also for the Cannings and Moran models. These models were then modified to simultaneously track frequencies of multiple alleles. In the case of a fixed number k of alleles with a fixed common rate of mutation from one allele type to another, the stationary distribution of allele frequency was found to be a Dirichlet distribution on the k alleles with equal parameters related to choices of scaling and the mutation rate.

A new model type emerged, referred to as infinitely many alleles models, in which it is assumed that when a mutation occurs, the mutation produces an allele type never seen before rather than mutating to a different previously seen allele type. Under such assumptions, all allele types eventually disappear over time. However, in a population of fixed size, one can study the relative frequences of allele types

independent of labeling by considering the distribution of the random partition of the population defined by allele type. Exact results for the modified Wright-Fisher and Cannings models under this infinitely many alleles assumption are difficult to come by, but the diffusion approximations provide some analytic results.

For purposes of application to actual data, the distribution of the partition of a sample taken from a population governed by the diffusion model was of great interest, where the partition is defined by the allele types of the sample. Specifically, the distribution of the number of alleles and the joint distribution of the numbers of each allele (unlabeled) were desired. This distribution is given by the Ewens sampling formula, conjectured by Ewens [20] and proved by Karlin and McGregor [35] in 1972.

Kingman then introduced the Kingman coalescent in 1982 [37, 40], a model by which we understand a sample of genes through their ancestral tree. At each time in the past, the state of the ancestral tree gives a partition of the sample, where two genes are in the same class of an equivalence relation at that time if they share a common ancestor by that time. If this process has mutations introduced to it such that ancestral trees are capped at mutations and older common ancestors discarded, a partition on the sample is also induced. Under natural assumptions regarding the timing of splits in the ancestral tree and timing of mutations, Kingman showed Ewens sampling formula can be directly derived. Because the formula may be derived from basic assumptions regarding genetics, it appears as an approximation to many models proposed (e.g. Wright-Fisher) as well as an exact formula in some cases (e.g. Moran).

The study of distributions of such sample partitions motivates the study of partition structures. A consistent partition structure is a family $(P_n)_{n\geq 1}$ of distributions on random partitions of n objects such that deletion of an object from a random partition governed by P_{n+1} results in a random partition governed by P_n. This corresponds with the idea that a random sample of size n taken from the population should have the same distribution of a subsample of size n taken from a sample of larger size. Kingman gave a representation theorem of a large class of consistent

partition structures in 1978 [36, 39]. He then showed that the consistent partition structure given by Ewens sampling formula was represented by the Poisson-Dirichlet distribution, i.e. that the joint distribution of the (normalized) counts of classes of a sample ordered by decreasing frequency limits to Poisson-Dirichlet as the sample size gets large.

The allele frequency models discussed so far primarily concern the study of partitions labeled according to size. At the time, a chronological labeling of classes of a partition according to when the defining mutation of the class occurred was also of some interest. This corresponds to the abstract notion of a size-biased permutation $\mathbf{P}^{\pi} = (P_{\pi(1)}, P_{\pi(2)}, ..)$ of a distribution $\mathbf{P} = (P_1, P_2, ..)$ on $\mathbb{N}$, where if $(X_1, X_2, X_3, ...)|\mathbf{P}$ is conditionally an iid sequence with common distribution $\mathbf{P}$, we define $\pi_j = X_{k_j}$ for $k_1 = 1$ and $k_j = \inf\{k : X_k \notin \{X_{k_l}\}_{l=1}^{j-1}\}$.

Hoppe introduced a modified Pólya's urn model in 1987, later called Hoppe's urn, that he found to be connected to Ewens sampling formula and the Poisson-Dirichlet distribution [33]. His model established a context in which the size-biased permutation of the Poisson-Dirichlet distribution, i.e. the limiting distribution of relative frequencies according to age-based labeling, could be identified. In Hoppe's urn model, at time 0, a black ball of mass $\theta > 0$ is the only ball in the urn. At time $n \in \mathbb{N}$, an individual removes a ball from the urn at random with probabilities proportional to mass. If the removed ball is black, the individual returns the black ball with a unit mass ball of a new color, and if the removed ball is not black, the individual returns the ball with a unit mass ball of the same color. If the colors are enumerated at each time according to descending relative frequency of the color, then the limiting distribution of colors in the urn over time is Poisson-Dirichlet. If, however, the colors are labeled in chronological order of appearance, the limiting distribution of colors has GEM distribution:

Definition 1.16 (GEM(θ))**.** 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 random distribution $\mathbf{P}$ taking values in Δ_∞ given by

$$P_j = X_j \prod_{i=1}^{j-1} (1 - X_i)$$

is said to have GEM(θ) distribution.

McCloskey and then Engen separately introduced the GEM distribution to the context of population genetics modeling in 1965 [46] and 1975 [18]. In particular, they did so in the context of the chronological labeling of alleles, with Engen identifying the GEM distribution as an appropriate limit of a Dirichlet distribution and McCloskey studying its analytic properties, especially those related to size-biased permutations. After studying the Poisson-Dirichlet distribution in 1979 [29], Griffiths more firmly established in unpublished work the relevance of the GEM distribution to population genetics models. For their collective contributions to the subject, the distribution was then named the Griffiths-Engen-McCloskey (GEM) distribution by Ewens in [21].

Patil and Taillie recognized Hoppe's urn model as establishing the GEM distribution as the size-biased permutation of the Poisson-Dirichlet distribution, and hence the Poisson-Dirichlet distribution as the order statistics of the GEM distribution in 1977 [51]. A formal proof was given by Donnelly and Joyce in 1989 [15].

In 1982 [61], J. Sethuraman and Tiwari introduced and made use of J. Sethuraman's construction of the Dirichlet process valid on any measurable space using a stick-breaking construction and the GEM distribution. Specifically, to construct a random measure with distribution DP(α) on a measurable space $\mathfrak{Y}$, define $\theta = \alpha(\mathfrak{Y})$ and $\mu = \theta^{-1}\alpha$. If $\mathbf{T}$ is an iid sequence with common distribution μ, and $\mathbf{X}$ is an iid sequence independent of $\mathbf{T}$ with common distribution Beta$(1, \theta)$, then

$$\sum_{j=1}^{\infty} \left(X_j \prod_{i=1}^{j-1} (1 - X_i) \right) \delta_{T_j} \sim \mathrm{DP}(\alpha) \tag{1.6}$$

where δ_T denotes the dirac measure on $\mathfrak{Y}$ giving measure 1 to every set containing the point T and 0 measure to every other set. J. Sethuraman gave a standalone analysis of the Dirichlet process in 1994 [60] using only this construction, verifying standard properties such as the posterior distribution and the fact observations of Dirichlet processes are almost surely discrete probability measures. This construction provided a highly computable representation of the Dirichlet process, both analytically and in application in Bayesian nonparametric statistics.

This dissertation is concerned in part with the study of a generalization of the Dirichlet process as constructed by J. Sethuraman above, for which the iid sequence $\mathbf{T}$ is permitted to be a stationary homogeneous Markov chain on a state space $\mathfrak{X} \subseteq \mathbb{N}$. This generalization is not only new, but a notable departure from from the current literature as it is nonexchangeable. By nonexchangeable, we mean the following: Let $\mathbf{P}$ have GEM distribution, and let $\mathbf{T}$ be a stationary homogeneous Markov chain independent of $\mathbf{P}$ which is not iid. Then there exists a finite permutation π of $\mathbb{N}$ such that

$$\sum_{j=1}^{\infty} P_j \delta_{T_j} \stackrel{d}{\neq} \sum_{j=1}^{\infty} P_{\pi(j)} \delta_{T_j}.$$

See Section 3.2.1 for a precise statement. In contrast, the constructions of the Dirichlet process given by J. Sethuraman using the GEM distribution and by Kingman using a Poisson point process are exchangeable. The constructions of the Dirichlet process using either GEM distributions or Poisson-Dirichlet distributions may in fact be identified because the GEM distribution is the size-biased permutation of the Poisson-Dirichlet distribution and, for the Dirichlet process, $\mathbf{T}$ is an iid sequence.

It should be noted, however, that concepts of exchangeability permeate the allele frequency models we discussed here and have deeply motivated much of the modern work related to GEM distributions, Poisson-Dirichlet distributions, and stick-breaking constructions, the last of which we will define and discuss shortly. This loss of exchangeability for our generalization impedes analysis via the standard tools developed

in the context of allele frequency models and that have since been expanded upon.

Kingman's introduction and study of random partition structures has since motivated development of a rich theory. Overviews of exchangeability and combinatorial structures and processes related to random partition structures may be found in such books as by Arratia et al. [3], Aldous et al. [1], and Pitman [56], with the inescapable connections to the Poisson-Dirichlet distribution, GEM distribution, and Ewens sampling formula.

Pitman in particular has made significant contributions to the studies of exchangeable random partitions in [52], regenerative random partitions [26, 27] with Gnedin, and size-biasing in [53] and with Yakubovich in [57]. Pitman [54] also extended the urn scheme of Blackwell and MacQueen [6] to a more general family of random discrete distributions involving a specific variety of exchangeable sequences. Additionally, along with Dubins, Pitman developed the Chinese Restaurant Process, a reformulation of Hoppe's urn that more readily connects to the study of partition structures first appearing in [1].

In 1997, Pitman and Yor famously introduced two-parameter generalizations of the Poisson-Dirichlet and GEM distributions [58]. They established a theory surrounding these generalizations with the depth and breadth of the theory surrounding their one-parameter origins, and their work led also to consideration of the Pitman-Yor process, the two-parameter analogue of the Dirichlet process.

New characterizations of the Poisson-Dirichlet distribution, GEM distribution, and Dirichlet process are still being found, as in but not limited to Arratia et al. in 1999 [2], Gnedin and Kerov in 2001 [25], Pitman in 2002 [55], Diaconis et al. in 2004 [11], Schiavo and Lytvynov in 2017 [59], and Last in 2018 [43].

1.2.2 RAMs and Stick-breaking

A subset of the modern developments related to Dirichlet processes focus on stick-breaking, of which the GEM distribution and J. Sethuraman's construction (1.6) of the Dirichlet process may be considered examples. A primary building block for stick-breaking measures is the Residual Allocation Model (RAM) introduced by Halmos in 1944 [30]. RAMs are a classic example of distributions on Δ_∞, i.e. of random probability distributions on $\mathbb{N}$.

Halmos developed RAMs to address the modernized question of how to randomly distribute one full pint of beer among (countably) infinitely many graduate students. We begin by giving a random proportion X_1 of the pint to the first student in line. To the second student, we give a random proportion X_2 of the remaining beer where X_2 is independent of X_1. We then give to the third student a random proportion X_3 of what remains after the first two students, where X_3 is independent of X_1 and X_2. We continue in this manner for all students. As long as we aren't stingy by letting the proportions $\mathbf{X} = (X_j)_{j\geq 1}$ get too small too quickly, the entire pint of beer should be given out eventually.

If we denote by P_j the proportion of the whole pint that student j received, we come to the following definition:

Definition 1.17 (Residual Allocation Model - RAM)**.** Let $\mathbf{X} = (X_j)_{j\geq 1}$ be a sequence of random proportions, i.e. independent $[0,1]$-valued random variables. Define

$$P_1 = X_1 \quad \text{and} \quad P_j = X_j \left(1 - \sum_{i=1}^{j-1} P_i\right) \quad \text{for} \quad j \geq 2. \tag{1.7}$$

Then, if $\sum_{j=1}^{\infty} P_j \overset{a.s.}{=} 1$, that is, if $\mathbf{P} = (P_1, P_2, ...)$ is almost surely a probability measure on $\mathbb{N}$, we say $\mathbf{P}$ is a RAM. Furthermore, if $\mathbf{X}$ consists of iid proportions, we say $\mathbf{P}$ is self-similar.

If one considers distributing portions of a stick to graduate students rather than beer, then X_j is the proportion of the stick remaining which is broken off and handed

to the jth graduate student. While this context of breaking sticks may justify the name, a great variety of contexts can and have been used, and we prefer that of beer in this matter.

The recursive definition (1.7) is intuitive in context but unwieldy in practice. A RAM $\mathbf{P}$ can be given a non-recursive definition:

$$P_j = X_j \prod_{i=1}^{j-1}(1 - X_i). \tag{1.8}$$

To see that the recursive formula (1.7) is equivalent to the direct formula (1.8), we need the following algebraic identity given by Halmos [30]:

Lemma 1.18 (Halmos; [30])**.** *For any sequence of numbers a_j and integer $k \geq 1$, we have*

$$\prod_{j=1}^{k}(1 - a_j) + \sum_{j=1}^{k} a_j \prod_{i=1}^{j-1}(1 - a_i) = 1. \tag{1.9}$$

In the context of a RAM, $\prod_{j=1}^{k}(1 - X_j)$ is the proportion of the pint of beer remaining after the first k students, and $\sum_{j=1}^{k} X_j \prod_{i=1}^{j-1}(1 - X_i) = \sum_{j=1}^{k} P_j$ is the sum of proportions of the pint of beer given to the first k students. The sum of the proportions already given out and the proportion of beer remaining should indeed equal 1 at any given time. Since manipulations using this identity and related identities are the basis of many proofs in this dissertation, the proof is included in Appendix A.3.

To complete the identification of definition (1.7) with the direct formula (1.8), we give a simple extension of Lemma 1.18. This extension also plays a role in the proof of later results in this work, and a proof may be found in Appendix A.3.

Lemma 1.19. *Consider a probability distribution $\mathbf{P} = (P_j)_{j\geq 1}$ on $\mathbb{N}$ and sequence $\mathbf{X} = (X_j)_{j\geq 1}$ satisfying*

$$X_j = P_j \left(1 - \sum_{i=1}^{j-1} P_i\right)^{-1}$$

for every $j \in \mathbb{N}$ *such that* $\sum_{i=1}^{j-1} P_i < 1$. *Then for every* $j \geq 1$,

$$P_j = X_j \prod_{i=1}^{j-1}(1 - X_i). \tag{1.10}$$

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

$$\sum_{j=1}^{r} P_j = 1 - \prod_{j=1}^{r}(1 - X_j)$$

$$\textit{and} \tag{1.11}$$

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

Returning to the definition of a RAM, we note that $\mathbf{P} = (P_n)_{n\geq 1}$ need not sum to 1 for every choice of distribution on $\mathbf{X}$. In light of Lemmas 1.18 and 1.19, a simple condition equivalent to $\sum_{n\geq 1} P_n \overset{a.s.}{=} 1$ is that $\prod_{j=1}^{\infty}(1 - X_j) \overset{a.s.}{=} 0$ as

$$1 - \sum_{j=1}^{n} P_j = \prod_{j=1}^{n}(1 - X_j).$$

For example, this is the case for an iid sequence $\mathbf{X} = (X_j)_{j\geq 1}$ of random proportions with $\mathscr{P}(X_1 > 0) = 1$.

The GEM distribution and two of its variants, the disordered GEM and the two-parameter GEM, are examples of RAMs constructed from proportions $\mathbf{X}$ with Beta distributions. We recall the definition of GEM in the language of RAMs:

Definition 1.20 (GEM(θ))**.** 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 have GEM(θ) distribution.

Note that a GEM(θ) distribution is self-similar as it is constructed from iid proportions. The two-parameter GEM of Pitman and Yor [58] and the disordered GEM considered in this work are non-self-similar variants:

Definition 1.21 (Two-parameter GEM)**.** Let $0 \leq \alpha < 1$ and $\theta > -\alpha$, and let $\mathbf{X} = (X_j)_{j\geq 1}$ be a sequence of independent variables with $X_j \sim \text{Beta}(1-\alpha, \theta + j\alpha)$. Then the RAM $\mathbf{P}$ constructed from $\mathbf{X}$ is said to have GEM(α, θ) distribution.

Definition 1.22 (Disordered GEM)**.** Consider a sequence $(\theta_j)_{j\geq 1}$ of positive constants, and let $\mathbf{X}$ be a sequence of independent random variables where X_j has Beta$(1, \theta_j)$ distribution for $j \geq 1$. When the measure $\mathbf{P}$, constructed from $\mathbf{X}$, is a RAM, we will say it has a disordered GEM distribution with parameters $(\theta_j)_{j\geq 1}$.

It is possible that the measure $\mathbf{P}$ constructed in Definition 1.22 is not a RAM if the probabilities $\mathbf{P}$ do not sum to 1 almost surely. This may occur if the parameter sequence $(\theta_j)_{j\geq 1}$ diverges quickly. However, we note that a sufficient (but not necessary) condition on the parameter sequence is that it be bounded. If, for example, the parameter sequence is bounded above by a positive constant θ^m, then for each j we have

$$\mathscr{P}(X_j \geq 1 - 2^{-1/\theta^m}) \geq \mathscr{P}(X_j \geq 1 - 2^{-1/\theta_j}) = \frac{1}{2}.$$

As every proportion X_j has probability at least $\frac{1}{2}$ of being greater than $1 - 2^{-1/\theta^m}$ and the proportions are independent, we have by the second Borel Cantelli lemma that this occurs for infinitely many values of j almost surely. Thus,

$$0 \leq 1 - \sum_{j=1}^{\infty} P_j = \prod_{j=1}^{\infty}(1 - X_j) \overset{a.s.}{=} 0$$

and $\mathbf{P}$ is a RAM. This boundedness condition is sufficient for our use of the disordered GEM distribution in this work.

A stick-breaking measure is then typically understood to be a random measure $\tilde{\nu}$ on a measurable space $\mathfrak{Y}$ constructed from an iid sequence $\tilde{\mathbf{T}}$ taking values in $\mathfrak{Y}$ and a RAM $\tilde{\mathbf{P}}$:

$$\tilde{\nu} = \sum_{j=1}^{\infty} \tilde{P}_j \delta_{\tilde{T}_j}.$$

By the work of J. Sethuraman [60], we see the Dirichlet process to be an example. The Pitman-Yor process, for which $\tilde{\mathbf{P}} \sim \text{Gem}(\alpha, \theta)$ in the above definition of $\tilde{\nu}$, is another standard example. Other closely related constructions involving the stick-breaking procedure without the complete RAM structure have been considered, such as the stick-breaking representation of the Beta process given by Broderick et al. in 2012 [8].

In the context of Bayesian statistics, a stick-breaking construction is generally understood as a random discrete measure that can be represented as

$$\sum_{j=1}^{\infty} p_k \delta_{Z_k},$$

for an iid sequence $(Z_k)_{k\geq 1}$ and random probability distribution $(p_k)_{k\geq 1}$ on $\mathbb{N}$. Ishwaran and James [34] did foundational work in 2001 regarding sampling stick-breaking priors, their approximations, and methods of simulating them. Stick-breaking representations of the Dirichlet process and other priors contributed immensely to the development of nonparametric Bayesian statistics by providing practical nonparametric priors for computation and simulation; see Müller et al. [49] for an introduction to stick-breaking generalizations of the Dirichlet process, such as Dirichlet process mixtures and Pitman's species sampling models from [54], and their applications in nonparametric Bayesian statistics.

In this dissertation, we consider a very broad notion of stick-breaking measures in Chapter 2 appealing to RAM structure but forgoing the restriction to an iid sequence of dirac measures and instead permitting a sequence of random probability measures. This very broad notion of stick-breaking will provide a context in which we can study our generalized stick-breaking measure

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j} \tag{1.12}$$

in which $\mathbf{P}$ has GEM(θ) distribution and $\mathbf{T}$ is a stationary homogeneous Markov chain through a clumping procedure on terms in the sum. We introduce these clumping procedures in Chapter 2 as a method of studying ν in absence of traditional methodologies pertaining to exchangeability.

Beyond their potential application in Bayesian statistics, our study of these generalized stick-breaking measures is motivated by their appearance as the occupation laws of a particular class of inhomogeneous Markov chain, which we introduce next.

1.2.3 An Inhomogeneous Markov Chain

We now consider a particular inhomogeneous Markov chain of interest whose empirical measures converge in distribution only. The occupation law of this inhomogeneous Markov chain is a primary object of study in this work, and represented by the generalized stick-breaking measure (1.12) of the previous section.

This particular class of inhomogeneous Markov chain has a discrete state space $\mathfrak{X} \subseteq \mathbb{N}$ and transition matrices built out of generator matrices, which we first define:

Definition 1.23 (Generator matrix)**.** Let $G = \{G_{i,j} : i, j \in \mathfrak{X}\}$ be a square matrix on $\mathfrak{X} \subseteq \mathbb{N}$. We say that G is a generator matrix if it satisfies

(a) $G_{i,j} \geq 0$ for each $i \neq j \in \mathfrak{X}$,

(b) $G_{i,i} = -\sum_{j \neq i} G_{i,j}$ for each $i \in \mathfrak{X}$,

(c) and $\sup_i |G_{i,i}| < \infty$.

We say G has a particular trait, e.g. is irreducible or has a positive recurrent state, if the stochastic kernel $I + G/\sup_i |G_{i,i}|$ has that trait.

A generator matrix is always a matrix of the form $\theta(Q - I)$ where $\theta > 0$ and Q is a stochastic matrix. This decomposition of G into the pair (Q, θ) is non-unique, but extremely useful in establishing connections with homogeneous Markov chains.

Specifically, we note that for any $\theta \geq \sup_i |G_{i,i}|$, the matrix $Q = I + G/\theta$ is necessarily stochastic. Also note that if (Q, θ) is a decomposition of G, then Q has an absorbing state x if and only if $G_{x,x} = 0$.

In this dissertation, we consider an inhomogeneous Markov chain $\mathbf{M} = (M_j)_{j \geq 1}$ with transition kernels of the form

$$K_n = I + \frac{G}{n} \mathbb{1}(n > c). \tag{1.13}$$

where c is chosen to be sufficiently large that K_n is stochastic, i.e. $c \geq \sup_i |G_{i,i}| - 1$. Note that if (Q, θ) is a decomposition of G, then we may rewrite

$$K_n = I + \frac{\theta(Q - I)}{n} \mathbb{1}(n > c) = \left(1 - \frac{\theta}{n} \mathbb{1}(n > c)\right) I + \frac{\theta}{n} \mathbb{1}(n > c) Q.$$

We observe that K_n is a Markov chain which gets 'sticky' over time, since the probability of $M_n = M_{n+1}$ approaches 1 as n approaches infinity. However, as the probability of transitioning from one state to a different state scales to 0 like $\frac{1}{n}$, the chain $\mathbf{M}$ changes states infinitely often as long as G is without absorbing states. Otherwise, $\mathbf{M}$ would have positive probability of becoming eventually constant at some state $x \in \mathfrak{X}$:

$$\mathscr{P}(M_j = x \text{ for } j \geq n | M_n = x) = \prod_{j=n}^{\infty} \left(1 - \frac{G_{x,x}}{n}\right) = 0 \quad \text{when} \quad G_{x,x} \neq 0.$$

Such chains can be considered a simplified model of simulated annealing. In this context, every point in $\mathfrak{X}$ represents a valley from which the chain rarely but almost surely exits to enter another point valley. Thus, a certain 'landscape' $\mathfrak{X}$ is explored.

This particular model has been studied before. Gantert considered laws of large numbers (i.e. occupation laws) for inhomogeneous Markov chains modeling simulated annealing in 1990, specifically Gibbs schemes and Metropolis schemes [24]. As a

particular example, which she credited to conversations with H.-R. Künsch, Gantert considered an inhomogeneous Markov chain with kernels

$$K_n = \begin{bmatrix} 1 - \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & 1 - \frac{1}{n} \end{bmatrix} \tag{1.14}$$

and noted that the associated empirical measures cannot converge in probability, as did Winkler [62] in 1995.

Dietz and S. Sethuraman generalized this example in 2007 to inhomogeneous Markov chains with kernels of the form

$$K_n^{\zeta} = I + \frac{G}{n^{\zeta}} \mathbb{1}(n > c)$$

for finite states spaces and characterized convergence of the empirical measures according to the cases $\zeta < 1$, $\zeta = 1$, and $\zeta > 1$. They found convergence at least in probability for $\zeta < 1$ to a common stationary distribution with almost sure convergence established for $\zeta < \frac{1}{2}$, convergence almost surely to a random point mass when $\zeta > 1$, and convergence in distribution only for $\zeta = 1$. Through a method of moments, they identified the distribution of the occupation law when $\zeta = 1$. Generally, the marginal distribution may be seen from their formulas to be of Beta Products having potentially complex parameters, and hence not corresponding with a Dirichlet process. However, when G could be decomposed as (Q, θ) where Q was constant stochastic, Dietz and S. Sethuraman found the occupation law to be a Dirichlet distribution.

They noted the apparent connection between their Dirichlet result and Dirichlet results for other 'reinforcement'-type models such as Pólya urn schemes in which changes occur more slowly with time, as well as Poisson-Dirichlet and GEM results for independent sequences of Bernoulli trials with probabilities that scale like $\frac{1}{n}$; see Arratia et. al. [3] and Pitman [54, 56]. The desire to firmly establish and understand these connections was an impetus for the work of this dissertation.

A specific generalization of the example (1.14) was considered by Engländer and Volkov in 2018 in which $\frac{1}{n}$ is replaced with a sequence $(a_n)_{n \geq 1}$ [19]. They studied

the nature of the convergence of an occupation law for these arbitrary scaling sequences $(a_n)_{n\geq 1}$, as well as a central limit theorem, extending the results of Dietz and S. Sethuraman in the two-state case and building upon the direct implications of Dobrushin's Central Limit Theorem [14] in their context.

Another generalization of the Markov chains of Dietz and S. Sethuraman to freezing Markov chains was studied by Bouguet and Cloez in 2018 [7]. Freezing Markov chains are characterized by transition kernels on finite state spaces of the form

$$R_n = (1 - a_n)I + a_n(G + Q_n) \tag{1.15}$$

for a generator matrix G, a sequence of stochastic kernels $(Q_n)_{n\geq 1}$ converging entrywise to the identity suitably quickly, and a positive decreasing sequence $(a_n)_{n\geq 1}$ such that the kernels R_n are stochastic. Engländer and Volkov's Markov chains are a subcase. Using stochastic approximation techniques, Bouguet and Cloez were able to extend some results of Dietz and S. Sethuraman regarding occupation laws. They then characterized these laws as stationary distributions of piecewise-deterministic Markov processes in continuous time.

Bouguet and Cloez's work establishes a connection between the 2018 model of Herbach [31] for mRNA production and the inhomogeneous Markov chains studied by Dietz and S. Sethuraman. In particular, Herbach considers mRNA production in the context of a multistate promoter, where the (categorical) state of a gene's promoter in continuous time is Markovian and different chemicals affecting mRNA production are continuously produced depending on that state. The relative distribution of these chemicals associated to each state is a piecewise-deterministic Markov process in Herbach's model. We can identify Herbach's piecewise-deterministic Markov process as corresponding with the inhomogeneous Markov chains studied by Dietz and S. Sethuraman through the work of Bouguet and Cloez.

Returning to the use of more general Markov chains for simulated annealing, there has been some interest in studying the 'valleys' or 'cycles' of the chain in which

the chain is likely to get stuck for an extended period of time before exiting. In particular, Catoni and Cerf [10] and Miclo [47] studied the distribution of the amount of time spent in a cycle before exiting, and the path by which the Markov chain is likely to exit the cycle, as the chain becomes more stagnant (sticky). Of particular interest were contexts of metastability, in which a law of large numbers type result may be associated with a cycle of the Markov chain. This suggests a variety of local convergence of the chain. We study here some local occupation laws which are reminiscent of but distinct from the results of [10] and [47]. In particular, for Markov chains with kernels (1.13), the entire landscape is reduced to point valleys, which removes consideration of cycles and exit paths. On the other other hand, we study all local occupations jointly in large time.

In this dissertation, we build upon Dietz and Sethuraman's work in [13] for $\zeta = 1$ by allowing generator matrices to have zero entries, be reducible, and be defined on countable state spaces, whether finite or infinite. Furthermore, we analyze the occupation laws through intermediate structures, termed local occupations, constructed through clumping. These intermediate structures may be matched with clumped versions of the stick-breaking measures (1.12), providing a construction for the occupation law and a more explicit connection to Dirichlet processes as constructed by J. Sethuraman [60].

1.3 Motivation and Organization

One of our goals is to understand more constructively the limiting occupation law ν of the inhomogeneous Markov chain $\mathbf{M}$ with kernels (1.13) discussed in Section 1.2.3 by considering an intermediate clumped structure. By connecting the occupation law of the inhomogeneous Markov chain $\mathbf{M}$ to a concrete generalized stick-breaking measure through the intermediate structure, the law ν may be more directly studied.

Additionally, identification of the intermediate structure is itself a description of $\mathbf{M}$.

On the other hand, by connecting generalized stick-breaking measures constructed from a pair $(\mathbf{P}, \mathbf{T})$ of a GEM distribution and a stationary Markov chain to the occupation law of $\mathbf{M}$, we provide a description of the distribution of the stick-breaking measures via their moments, which may be useful in application with, for example, Bayesian statistics.

The clumping methodology may be of its own interest as a way of studying so-called local occupations of an inhomogeneous Markov chain or finding intermediate structure in a stick-breaking measure. In fact, an alternate application of the clumping procedure to a stick-breaking pair $(\mathbf{P}, \mathbf{T})$ leads to a unique characterization of ν in terms of self-similarity equations (Section 2.2). Satisfaction of these self-similarity equations is reminiscent of the regenerative structure present in Dirichlet stick-breaking measures [60], in integral constructions of the Dirichlet processs [43], [59], and in other related settings [26, 27]. Such a characterization has been used to identify Dirichlet processes [32] and, in the context of Bayesian statistics, to identify their posterior distribution [60].

In Chapter 2, we consider an extremely broad notion of stick-breaking which encompasses all particular instances of stick-breaking relevant to this work. In Section 2.1, we introduce a clumping procedure on RAMs, and analyze two methods of its application with respect to homogeneous Markov chains. The first method leads to a class of stick-breaking measures encountered in Chapter 3. The second method motivates further study of self-similarity in Section 2.2. In Section 2.2, we introduce generalized RAMs and explore the correspondence between self-similarity of a generalized RAM and self-similarity properties of its push-forward measure. We then consider the self-similarity results in the context of the second method of applying the clumping procedure.

In Chapter 3, we discuss the occupation law of an inhomogeneous Markov chain

$\mathbf{M}$ and its connections to a general class of stick-breaking measure. In Section 3.1, we introduce a reverse-clumping procedure which we use to define and study the local occupations of $\mathbf{M}$. We then consider RAMs and homogeneous Markov chains which combine via clumping to mimic the structure of the local occupations in the limit. In Section 3.2, we identify the occupation law of $\mathbf{M}$ as a stick-breaking measure and consider the results of Section 2.2 in this context.

In Chapter 4, we consider future progressions of this work, including use of the general stick-breaking measure of Chapter 3 in a Bayesian context and a potential application to inhomogeneous generalizations of teleporting random walks which may also be considered as a partial generalization of freezing Markov chains.

Chapter 2

Residual Allocation Models and Stick-breaking Constructions

Stick-breaking is often used to describe two special objects: Residual Allocation Models (RAMs) and what we'll refer to as traditional stick-breaking measures. RAMs are a class of random probability measures on $\mathbb{N}$, and traditional stick-breaking measures are random probability measures on a measurable space $\mathfrak{X}$ formed from a RAM and an iid sequence taking values in $\mathfrak{X}$.

A generalization of these objects arises naturally when studying the limiting occupation law of a particular class of inhomogeneous Markov chain, introduced in Section 1.2.3. In this chapter, we will study a much broader generalization of these objects in order to highlight certain properties and establish a more complete theory.

Let $\mathfrak{X}$ be a measurable space. At its most general, we will think of a stick-breaking construction on $\mathfrak{X}$ as a random probability measure ν on $\mathfrak{X}$ constructed as follows:

Definition 2.1 (Stick-breaking construction)**.** Let $\mathbf{X} = (X_1, X_2, ...)$ be a sequence of $[0,1]$-valued random variables. Define the sequence $\mathbf{P} = (P_1, P_2, ...)$ from $\mathbf{X}$ which for $j \geq 1$ is given by

$$P_j = X_j \prod_{i=1}^{j-1} (1 - X_i). \tag{2.1}$$

Let $\mathscr{M} = (\eta_1, \eta_2, ...)$ be a sequence of random probability measures on a measurable space $\mathfrak{X}$. If $\sum_{j=1}^{\infty} P_j = 1$ almost surely, then the stick-breaking measure ν constructed from $\mathbf{P}$ and $\mathscr{M}$ is defined as follows for measurable sets $A \subseteq \mathfrak{X}$:

$$\nu(A) = \sum_{j=1}^{\infty} P_j \eta_j(A).$$

We may equivalently refer to ν as being constructed from $\mathbf{X}$ and $\mathscr{M}$.

This definition will encompass every instance of stick-breaking we discuss in this work, including RAMs, traditional stick-breaking constructions, and their respective generalizations. For the purposes of this work, if ν is a stick-breaking measure constructed from $\mathbf{X}$ and $\mathscr{M}$, it will typically be the case that $\mathbf{X}$ is either an independent sequence or a conditionally independent sequence given $\mathscr{M}$, and that $\mathscr{M}$ is a deterministic, independent, or Markovian sequence.

Recall the definition of a RAM:

Definition 2.2 (Residual Allocation Model - RAM)**.** Let $\mathbf{X} = (X_j)_{j\geq 1}$ be a sequence of random proportions, i.e. independent $[0,1]$-valued random variables. Define

$$P_1 = X_1 \quad \text{and} \quad P_j = X_j\left(1 - \sum_{i=1}^{j-1} P_i\right) \quad \text{for} \quad j \geq 2. \tag{2.2}$$

Then, if $\sum_{j=1}^{\infty} P_j \overset{a.s.}{=} 1$, that is, if $\mathbf{P} = (P_1, P_2, ...)$ is almost surely a probability measure on $\mathbb{N}$, we say $\mathbf{P}$ is a RAM. Furthermore, if $\mathbf{X}$ consists of iid proportions, we say $\mathbf{P}$ is self-similar.

In other words, a RAM is a stick-breaking construction for which $\mathfrak{X} = \mathbb{N}$, $\eta_j = \delta_j$ deterministically, and $\mathbf{X}$ is an independent sequence. In this context, $\mathbf{P}$ and ν are two representations of the same random measure on $\mathbb{N}$ where $\nu(\{j\}) = P_j$.

The other object stick-breaking is often used to describe is a traditional stick-breaking measure. Suppose $\mathbf{P}$ is a RAM, $\mathscr{M}$ is a random sequence of iid Dirac measures on a measurable space $\mathfrak{X}$, and $\mathbf{P}$ is independent of $\mathscr{M}$. If ν is the stick-breaking measure constructed from $\mathbf{P}$ and $\mathscr{M}$, then ν is a traditional stick-breaking measure. When discussing traditional stick-breaking measures, we often reformulate $\mathscr{M}$ as a sequence $\mathbf{T} = (T_1, T_2, ...)$ where T_j is the random point of $\mathfrak{X}$ to which the Dirac measure η_j assigns non-zero probability. We then represent ν as

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j}.$$

In this chapter, we explore a generalization of the traditional stick-breaking measure. Instead of working with a RAM $\mathbf{P}$ and an iid sequence $\mathbf{T}$ independent of $\mathbf{P}$, we allow $\mathbf{T}$ to be a homogeneous Markov chain in a countable space $\mathfrak{X} \subseteq \mathbb{N}$. When $\mathbf{P}$ is self-similar, we establish multiple stick-breaking constructions associated with the same measure ν using a clumping procedure and two distinct methods of applying the procedure. The first method arises naturally in the context of studying local occupations of the inhomogeneous Markov chain in Section 3.1. The second method helps us establish uniquely identifying characteristics of ν.

After formally defining a clumping procedure and the two methods of application, we establish various structural properties that the clumping procedure and application methods respect. Then, we further explore the idea of self-similarity, an integral notion in the analysis of clumping and one which may aid identification of a posterior distribution in the context of Bayesian statistics. A generalized RAM is defined for spaces of the form $\mathfrak{X}\times\mathbb{N}$ for a measurable space $\mathfrak{X}$. Through push-forward measures of these generalized RAMs onto the space $\mathfrak{X}$, we understand self-similarity properties of random probability measures ν on the space $\mathfrak{X}$. Certain self-similarity properties will be shown to uniquely characterize the distribution of a random probability measure ν, which might be used in practice to identify the law of a random measure.

Some of this work was developed jointly with Zach Dietz and Sunder Sethuraman. In particular, Propositions 2.4 and 2.12 and Corollary 2.4.1, Theorem 2.5, and a subcase of Theorem 2.8 may also be found in [12].

2.1 Clumping Procedure

We begin by defining how to clump a probability measure $\mathbf{P}$ on $\mathbb{N}$ with respect to a clumping sequence $\mathbf{u}$ to form a clumped probability measure $\mathbf{P^u}$ on $\mathbb{N}$:

Definition 2.3 (Clumped measure)**.** Let $\mathbf{u} = (u_j)_{j\geq 1}$ be a non-decreasing sequence in $\mathbb{N} \cup \{\infty\}$ with $u_1 = 1$ and $\lim_{j\to\infty} u_j = \infty$, and let $\mathbf{P} = (P_j)_{j\geq 1}$ be a probability

measure on $\mathbb{N}$. We clump $\mathbf{P}$ according to $\mathbf{u}$ to construct a new probability measure $\mathbf{P^u} = (P_j^u)_{j\geq 1}$ on $\mathbb{N}$ where, for $j \geq 1$,

$$P_j^u = \begin{cases} \sum_{i=u_j}^{u_{j+1}-1} P_i & \text{if } u_j < \infty \\ \\ 0 & \text{if } u_j = \infty. \end{cases}$$

Intuitively, the clumping sequence $\mathbf{u}$ will mark the boundaries of clumps of probabilities. We remark, when $\mathbf{u}$ takes the value infinity at an entry u_{j+1} in the sequence, necessarily $\mathbf{P^u}$ is a distribution supported on $\{1, 2, \ldots, j\}$.

In the context that $\mathbf{P}$ is a RAM, we naturally consider when $\mathbf{P^u}$ is also a RAM. In fact, so long as $\mathbf{u}$ is deterministic, $\mathbf{P^u}$ is always a RAM.

Proposition 2.4 (Clumped RAMs)**.** *Let* $\mathbf{P}$ *be a RAM constructed from proportions* $\mathbf{X}$*. Fix a deterministic, non-decreasing sequence* $\mathbf{u} = (u_j)_{j\geq 1}$ *in* $\mathbb{N} \cup \{\infty\}$ *with* $u_1 = 1$ *and* $\lim_{j\to\infty} u_j = \infty$*. Then,* $\mathbf{P^u}$ *is a RAM with respect to proportions* $\mathbf{X^u} = (X_j^u)_{j\geq 1}$ *where:*

$$X_j^u = \begin{cases} \sum_{i=u_j}^{u_{j+1}-1} X_i \prod_{l=u_j}^{i-1} (1 - X_l) & \text{if } u_j < \infty \\ \\ 1 & \text{if } u_j = \infty. \end{cases}$$

However, the situation is more involved when a random clumping sequence is used. We are interested in two types of random clumping sequences constructed from a Markov chain $\mathbf{T} = (T_j)_{j\geq 1}$ on the discrete space $\mathfrak{X}$. The first sequence $\mathbf{V}$ comes from considering clumps of repeated values in $\mathbf{T}$; that is, $\mathbf{V}$ will keep track of the times when $\mathbf{T}$ switches states. The second sequence $\mathbf{W}$ arises in considering the times when $\mathbf{T}$ returns to its initial value T_1.

For example, if $\mathbf{T} = (6, 6, 2, 2, 2, 2, 4, 6, 6, 5, \ldots)$ is observed, then:

$$\mathbf{V} = (1, 3, 7, 8, 10, \ldots) \quad \text{and} \quad \mathbf{W} = (1, 2, 8, 9, \ldots).$$

More formally, let $V_1 = W_1 = 1$ and, for $j \geq 1$, set

$$V_{j+1} = \inf\{v > V_j : T_v \neq T_{v-1}\} \tag{2.3}$$
$$W_{j+1} = \inf\{w > W_j : T_w = T_1\}.$$

We understand the infimum of the empty set to be $+\infty$. In the case that $\mathbf{T}$ reaches an absorbing state, denoted T_∞, the chain $\mathbf{T}$ is eventually constant and $\mathbf{V}$ eventually takes the value infinity. In the case that T_1 is a transient state, the chain $\mathbf{T}$ returns to the first state finitely many times and $\mathbf{W}$ eventually takes the value infinity.

Motivated by the study of stick-breaking constructions of the form

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j}, \tag{2.4}$$

we also clump the chain $\mathbf{T}$ in two different ways corresponding to $\mathbf{P^V}$ and $\mathbf{P^W}$ in order to preserve ν.

Define $\mathbf{Y} = (Y_j)_{j\geq 1}$ by

$$Y_j = T_{V_j} \tag{2.5}$$

for $j \geq 1$. Intuitively, we think of $\mathbf{Y}$ as the sequence of values taken by $\mathbf{T}$ without repetition. When $\mathbf{T}$ does not reach an absorbing state, this is precisely the case. However, if $\mathbf{T}$ meets an absorbing state T_∞, $\mathbf{Y}$ will be non-repeating until becoming constant at the value T_∞. By construction, we then have

$$\nu = \sum_{i=1}^{\infty} P_i \delta_{T_i} = \sum_{j=1}^{\infty} \sum_{i=V_j}^{V_{j+1}-1} P_i \delta_{T_i} = \sum_{j=1}^{\infty} P_j^V \delta_{Y_j}.$$

Given $\mathbf{T}$ and a distribution $\mathbf{P} = (P_j)_{j\geq 1}$ on $\mathbb{N}$, define $\mathscr{M} = (\eta_j)_{j\geq 1}$ by

$$\eta_j = \begin{cases} \dfrac{\sum_{i=W_j}^{W_{j+1}-1} P_i \delta_{T_i}}{\sum_{i=W_j}^{W_{j+1}-1} P_i} & \text{if } P_j^W > 0 \\ \delta_{T_1} & \text{if } P_j^W = 0. \end{cases} \tag{2.6}$$

We intuitively understand $\mathscr{M}$ to be the clumped version of the sequence $(\delta_{T_i})_{i\geq 1}$ according to cycles, where η_j is a weighted average of the Dirac measures on the jth cycle of $\mathbf{T}$. By construction, we then have

$$\nu = \sum_{i=1}^{\infty} P_i \delta_{T_i} = \sum_{j=1}^{\infty} \sum_{i=W_j}^{W_{j+1}-1} P_i \delta_{T_i} = \sum_{j=1}^{\infty} P_j^W \eta_j.$$

With these two methods of applying the clumping procedure in hand, we can now consider when RAM structure is still present after clumping. Simple application of Proposition 2.4 gives the following corollary:

Corollary 2.4.1. *Let* $\mathbf{P}$ *be a RAM constructed from proportions* $\mathbf{X}$*, and let* $\mathbf{T} = (T_j)_{j\geq 1}$ *be a Markov chain independent of* $\mathbf{P}$*. Define* $\mathbf{V}$ *and* $\mathbf{W}$ *as in* (2.3) *with respect to* $\mathbf{T}$*. Then* $\mathbf{P}^{\mathbf{V}}\big|\mathbf{T} = \mathbf{t}$ *and* $\mathbf{P}^{\mathbf{W}}\big|\mathbf{T} = \mathbf{t}$ *are conditionally RAMs.*

With the additional assumption that $\mathbf{P}$ is self-similar (i.e. constructed from iid proportions $\mathbf{X}$), we obtain a finer result which is more convenient for the stick-breaking construction ν of interest (see (2.4)). We first consider the clumped pair $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y})$:

Theorem 2.5. *Let* $\mathbf{P}$ *be a self-similar RAM constructed from proportions* $\mathbf{X}$*, and let* $\mathbf{T} = (T_j)_{j\geq 1}$ *be homogeneous a Markov chain independent of* $\mathbf{P}$ *having transition kernel* Q*. Define* $\mathbf{V}$ *and* $\mathbf{Y}$ *as in* (2.3) *and* (2.5) *with respect to* $\mathbf{T}$*. Let* $\mathbf{t}$ *be a possible sequence in* $\mathfrak{X}$ *with respect to* $\mathbf{T}$*, and let* $\mathbf{y}$ *be a possible sequence in* $\mathfrak{X}$ *with respect to* $\mathbf{Y}$*.*

(1) $\mathbf{Y}$ *is a homogeneous Markov chain with transition kernel* K *given by*

$$K(z,w) = \begin{cases} \dfrac{Q_{z,w}}{1-Q_{z,z}} & \text{for } z \neq w;\ Q_{z,z} \neq 1 \\ 1 & \text{for } z = w;\ Q_{z,z} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

If Q is without absorbing states, the above formula reduces to

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

(2) $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y} = \mathbf{y}$ *is conditionally a RAM with respect to proportions* $\mathbf{X}^{\mathbf{V}}$.

This theorem establishes that, when $\mathbf{P}$ is self-similar, the clumped pair $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y})$ has a structure that might be described as a Markov Chain conditional RAM. In the next two examples, we explore the precise nature of the conditional structure of $\mathbf{P}^{\mathbf{V}}|\mathbf{Y}$ and then the integral role of self-similarity of $\mathbf{P}$ in deducing this structure.

Example 2.6. Suppose that $\mathbf{T}$ is a homogeneous Markov chain independent of $\mathbf{P}$ with stochastic kernel Q where Q has constant diagonal entries $Q_{x,x} = q$ for $x \in \mathfrak{X}$. Note that even if $\mathbf{P}$ is not a RAM, proportions $\mathbf{X}$ can be defined such that $\mathbf{P}$ is constructed from $\mathbf{X}$ like a RAM via Lemma 1.19. Define $\mathbf{V}$ and $\mathbf{Y}$ with respect to $\mathbf{T}$. According to the computations in the proof of Theorem 2.5 Part (2), the proportions $\mathbf{X}^{\mathbf{V}}$ depend on $\mathbf{Y} = (Y_1, Y_2, ...)$ only through the corresponding diagonal entries $(Q_{Y_1,Y_1}, Q_{Y_2,Y_2}, ...)$ of Q. Specifically, given an observed chain $\mathbf{Y} = \mathbf{y}$, we see that conditionally $(V_{j+1} - V_j)_{j\geq 1}$ is a collection of geometric variables with parameters $(Q_{y_j,y_j})_{j\geq 1}$. However, since the diagonal entries of Q are all the same constant value q, $\mathbf{V}$ and $\mathbf{Y}$ are independent, and so $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y} = \mathbf{P}^{\mathbf{V}}$.

Thus, if $\mathbf{P}$ is a self-similar RAM and $\mathbf{T}$ is a homogeneous Markov chain independent of $\mathbf{P}$ with stochastic kernel Q where Q has constant diagonal entries, then $\mathbf{P}^{\mathbf{V}}$ is independent of $\mathbf{Y}$ and a RAM unconditionally.

We now consider the importance of self-similarity of $\mathbf{P}$ in deducing in full generality that $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y}$ is a RAM. For this example, we use a well-known RAM $\mathbf{P}$ called a two-parameter GEM, and a Markov chain $\mathbf{T}$ consisting of iid Bernoulli variables. As we will show, this pair constitutes a RAM $\mathbf{P}$ which is not self-similar and for which $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y}$ is not conditionally a RAM.

Example 2.7. We now consider a RAM $\mathbf{P}$ constructed from independent proportions $X_j \sim$ 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 independent of $\mathbf{P}$. Thought of as a Markov chain on the 2-state space $\mathfrak{X} = \{1, 2\}$, every entry of the stochastic kernel Q of $\mathbf{T}$ equals 1/2. By the discussion in Example 2.6, as the diagonal entries of Q are the constant $q = 1/2$, we have $\mathbf{P^V}\big|\mathbf{Y} = \mathbf{P^V}$.

We now observe that $\mathbf{P^V}$ is not a RAM by considering the associated non-atomic proportions $\mathbf{X^V}$. Compute

$$\mathbb{E}\left[1 - X_1^V | V_2 - V_1 = m, V_3 - V_2 = n\right] = \mathbb{E}\left[\prod_{j=1}^{m}(1 - X_j)\right] = \prod_{j=1}^{m}\frac{2+j}{3+j} = \frac{3}{3+m}$$

$$\begin{aligned}\mathbb{E}\left[1 - X_2^V | V_2 - V_1 = m, V_3 - V_2 = n\right] &= \frac{3+m}{3+m+n} \\ \mathbb{E}\left[(1 - X_1^V)(1 - X_2^V) | V_2 - V_1 = m, V_3 - V_2 = n\right] &= \frac{3}{3+m+n}.\end{aligned}$$

Then,

$$\begin{aligned}\mathbb{E}\left[1 - X_1^V\right] &= \sum_{m\geq 1}\frac{3}{3+m}(.5)^m \\ \mathbb{E}\left[1 - X_2^V\right] &= \sum_{n,m\geq 1}\frac{3+m}{3+m+n}(.5)^{m+n} \\ \mathbb{E}\left[(1 - X_1^V)(1 - X_2^V)\right] &= \sum_{m,n\geq 1}\frac{3}{3+m+n}(.5)^{m+n}.\end{aligned}$$

Hence, $\mathrm{Cov}[1 - X_1^V, 1 - X_2^V] \approx -.005391$, and so the non-atomic proportions are not independent, and $\mathbf{P^V}$ cannot be a RAM.

Notably, Example 2.7 is not a pathological example, but rather a standard pairing of a non-self-similar RAM and iid sequence to which we would want to extend the result. The stick-breaking measure constructed from a two parameter GEM and an

iid sequence is known as the Pitman-Yor process, and is a well-known generalization of the Dirichlet process (see [34] and [58]).

If $\mathbf{P}$ is not self-similar, then X_j^V will generically depend not only on $V_{j+1} - V_j$, but on V_j, introducing dependence among clumped proportions even after conditioning on $\mathbf{Y}$. Self-similarity plays an integral role in the result by removing this dependence on the length of previous clumps.

That said, self-similarity is not strictly necessary. If, for example, $\mathbf{X}$ is a sequence of independent variables such that $(X_2, X_3, ...)$ is an iid sequence and X_1 has a different distribution from X_2, we know surely that the odd proportion out, X_1, contributes to X_1^V. Thus, conditional independence of the factors $\mathbf{X^V}$ given $\mathbf{Y}$ is preserved.

We next consider the clumped pair $(\mathbf{P^W}, \mathscr{M})$:

Theorem 2.8. *Let* $\mathbf{P}$ *be a self-similar RAM and* $\mathbf{T}$ *be a homogeneous Markov chain independent of* $\mathbf{P}$*. Let* $t_1 \in \mathfrak{X}$*. Define* $\mathbf{W}$ *and* $\mathscr{M}$ *as in* (2.3) *and* (2.6) *with respect to* $\mathbf{T}$ *and* $\mathbf{P}$*. Then, given* $T_1 = t_1$*:*

$$\left(P_j^W \eta_j\right)_{j\geq 1} \stackrel{d}{=} \left(\tilde{P}_j \tilde{\eta}_j\right)_{j\geq 1}$$

where $(\tilde{X}_j, \tilde{\eta}_j)_{j\geq 1} | T_1 = t_1$ *is an iid sequence of copies of* $(X_1^W, \eta_1) | T_1 = t_1$ *and* $\tilde{\mathbf{P}}$ *is constructed from* $\tilde{\mathbf{X}}$*. In particular,* $\mathbf{P^W} | T_1 = t_1$ *is conditionally distributed as a self-similar RAM.*

If additionally t_1 *is a recurrent state, then* $(X_j^W, \eta_j)_{j\geq 1} | T_1 = t_1$ *is conditionally an iid sequence and* $\mathbf{P^W} | T_1 = t_1$ *is conditionally a self-similar RAM with respect to proportions* $\mathbf{X^W}$*.*

Notably, clumping according to $\mathbf{W}$ preserves conditionally self-similar structure of $\mathbf{P}$ when forming $\mathbf{P}^W$. Additionally, the stick-breaking construction

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j} = \sum_{j=1}^{\infty} P_j^W \eta_j$$

is, in some sense, given more structure through clumping as we see the emergence of a conditionally iid sequence $(\tilde{X}_j, \tilde{\eta}_j)_{j\geq 1}$ reminiscent of the traditional stick-breaking measure.

To understand why self-similarity is preserved by this clumping sequence, we consider cycles of a homogeneous Markov chain $\mathbf{T}$ starting at a specified state t_1. In other words, we consider the collection of states, in order, which $\mathbf{T}$ visits between one observation of t_1 and the next. In the context of Theorem 2.8, the first such cycle would be $C_1 = (T_1, T_2, ..., T_{W_2-1})$ and, if $W_2 < \infty$, the second such cycle would be $C_2 = (T_{W_2}, T_{W_2+1}, T_{W_2+2}, ..., T_{W_3-1})$, and so on. If we assume first for simplicity that t_1 is recurrent, almost surely $\mathbf{T}$ will have infinitely many cycles starting at an observation of t_1. By the Markov property, the distribution of a cycle of a homogeneous Markov chain is completely determined by the starting state t_1 and the transition kernel of the chain, and all such cycles are independent of each other. In other words, if a homogeneous Markov chain is broken up into its cycles, an iid sequence is created.

If t_1 is transient, almost surely $\mathbf{T}$ has only finitely many cycles starting at an observation of t_1, the last cycle being of infinite length. On the one hand, this clearly prevents construction of an infinite iid sequence. On the other hand, these cycles still exhibit some iid behavior conditioned on the number of cycles. We ultimately compare the collection of the finitely many cycles that arise from $\mathbf{T}$ to a list of infinitely many iid copies of the first cycle of $\mathbf{T}$ which we cut off after the first infinite copy, and find that the two finite collections have the same distribution. In other words, when t_1 is transient, the picture becomes more complex but the intuition from the recurrent case holds.

To be more concrete, we briefly discuss the technical role recurrence versus transience of state t_1 plays in the self-similarity of $\mathbf{P}^W$. If t_1 is transient, then almost surely an observation of $\mathbf{T}$ with $T_1 = t_1$ will have only finitely many cycles starting at t_1. If J is the number of cycles, then typically $X_j^W < 1$ for $j < J$ and always $X_j^W = 1$ for $j \geq J$. Since J is not a function of T_1 only, this clearly suggests that $\mathbf{X^W}|T_1 = t_1$

is not an independent sequence, let alone the iid sequence necessary for a self-similar RAM.

At first, this may seem contradictory to Theorem 2.8. However, it is possible for a random probability measure to be distributed as a RAM but still be constructed from non-independent proportions. Specifically, this is possible when the distribution in question is of a RAM $\mathbf{P}$ constructed from $\mathbf{X}$ where some proportion X_j has non-zero probability of taking the value 1. In this situation, the proportions $\mathbf{X}$ we associate with $\mathbf{P}$ are not uniquely determined by $\mathbf{P}$.

For example, consider an iid sequence $\mathbf{X}'$ of nontrivial Bernoulli variables. Construct from $\mathbf{X}'$ a sequence $\mathbf{X}$ given by $X_j = \max_{1\leq i\leq j} X'_i$, which is certainly not an iid sequence. If $\mathbf{P}$ is the RAM constructed from $\mathbf{X}'$, then $\mathbf{P}$ can be equivalently constructed from $\mathbf{X}$. This is because if the sequence of proportions $\mathbf{X}'$ takes the value 1 at some finite minimal index J, then all probability is allocated by time J, i.e. $\sum_{j=1}^{J} P_j = 1$. Thus, the proportions X'_j for $j > J$ are independent of $\mathbf{P}$.

When t_1 is transient, the sequence of proportions $\mathbf{X^W}\big|T_1 = t_1$ almost surely takes the value 1 at some finite minimal time J. Specifically, in defining the proportions $\mathbf{X^W}$, we imposed the condition that $X_j^W = 1$ for $j > J$ where J is the random number of cycles of $\mathbf{T}$, and so we imposed dependence on the proportions. In fact, picking another convention in which for $j > J$, X_j^W is taken to be an iid copy of X_1^W, provides the self-similar RAM structure seen in Theorem 2.5. However, since this second convention introduces new random variables, potentially requiring a larger probability space, we make conclusions about the distribution only.

In the context that $\mathbf{P}$ is a self-similar RAM and $\mathbf{T}$ is a homogeneous Markov chain independent of $\mathbf{P}$, Theorem 2.5 says that much of the RAM and Markovian structure is retained after clumping together probabilities P_j according to adjacent repeated values of the chain $\mathbf{T}$.

On the other hand, if $\mathbf{P}$ is a self-similar RAM constructed from $\mathbf{X}$, $\mathbf{T}$ is a homoge-

neous Markov chain independent of $\mathbf{P}$ with probability one of T_1 being recurrent, and $\mathbf{W}$ and $\mathscr{M}$ are constructed from $\mathbf{T}$, then the sequence of pairs (η_j, X_j^W) is conditionally an iid sequence by Theorem 2.8. This stick-breaking construction of ν from an iid sequence (η_j, X_j^W) is reminiscent of both traditional stick-breaking measures and self-similarity of a RAM, and will lead to a characterization of ν in terms of so-called self-similarity equations in the next section.

2.2 Self-similarity

In this section, for the purpose of further exploring notions of self-similarity, we consider a generalized notion of a RAM. Like RAMs, generalized RAMs are examples of stick-breaking constructions built from an independent sequence of random variables. However, they are random probability measures on spaces of the form $\mathfrak{X} \times \mathbb{N}$ rather than strictly on $\mathbb{N}$:

Definition 2.9 (Generalized RAM)**.** Let $\mathfrak{X}$ be a measurable space. Let $\mathscr{M} = (\eta_j)_{j\geq 1}$ be a sequence of random probability measures on $\mathfrak{X}$ and $\mathbf{X} = (X_j)_{j\geq 1}$ a sequence of $[0,1]$-valued random variables such that the sequence of pairs (η_j, X_j) is an independent sequence. For each measurable $A \subseteq \mathfrak{X} \times \mathbb{N}$, let

$$\alpha(A) = \sum_{j=1}^{\infty} P_j(\eta_j \times \delta_j)(A)$$

where $\mathbf{P}$ is the RAM constructed from $\mathbf{X}$. Then if $\alpha(\mathfrak{X} \times \mathbb{N}) \overset{a.s.}{=} 1$, we call α a generalized RAM on $\mathfrak{X} \times \mathbb{N}$ constructed from $\mathbf{P}$ and $\mathscr{M}$. We may equivalently refer to α as a generalized RAM constructed from $\mathbf{X}$ and $\mathscr{M}$.

Furthermore, if $\{(\eta_j, X_j)\}_{j\geq 1}$ is an iid sequence, we call α self-similar.

Clearly, α is a stick-breaking construction on $\mathfrak{X} \times \mathbb{N}$. Note that if $\mathfrak{X} = \{x\}$ is a space with a single element, then necessarily $\eta_j = \delta_x$ is degenerate. In this case, the

generalized RAM effectively reduces to a RAM as identification of $\{x\} \times \mathbb{N}$ with $\mathbb{N}$ identifies the measures α and $\mathbf{P}$.

When introducing a RAM in Section 1.2.2, we described the structure as a method of allocating a pint of beer to infinitely many graduate students by assigning a proportion X_j of what beer remains after the first $j-1$ students to the jth student. Here, rather than infinitely many individuals, we think of infinitely many copies of the space $\mathfrak{X}$. Random variables $\mathbf{X}$ are used to allocate a proportion of the probability remaining to the jth copy of $\mathfrak{X}$ in $\mathfrak{X} \times \mathbb{N}$, and the random measure η_j allocates that probability across that copy.

From a generalized RAM on $\mathfrak{X} \times \mathbb{N}$, we can construct a stick-breaking construction on $\mathfrak{X}$ via pushforward:

Definition 2.10 (Pushforward Measure)**.** Suppose α is a generalized RAM on $\mathfrak{X} \times \mathbb{N}$ constructed from $\mathbf{P}$ and $\mathscr{M}$. Let $\pi : \mathfrak{X} \times \mathbb{N} \to \mathfrak{X}$ be the projection map. We define the pushforward measure $\pi_*\alpha$ of α on $\mathfrak{X}$ as follows: for measurable $A \subseteq \mathfrak{X}$

$$\pi_*\alpha(A) = \alpha(A \times \mathbb{N}) = \sum_{j=1}^{\infty} P_j \eta_j(A).$$

We now explore what self-similarity of a generalized RAM α implies about its pushforward $\pi_*\alpha$. For motivation, recall the stick-breaking measure ν constructed from a self-similar RAM $\mathbf{P}$ and a homogeneous Markov chain $\mathbf{T}$ independent of $\mathbf{P}$ as in equation (2.4). We have that ν is conditionally distributed as the pushforward of a self-similar generalized RAM given T_1 by Theorem 2.8.

Since self-similar generalized RAMs are characterized by an iid sequence of pairs (η_j, X_j), we can equivalently characterize a self-similar generalized RAM as having distribution invariant under a left shift. Specifically, suppose α is a generalized RAM constructed from proportions $\mathbf{X}$ and measures $\mathscr{M}$, and denote the left shifts of $\mathbf{X}$ and $\mathbf{M}$ by $\mathbf{X}' = (X_2, X_3, ...)$ and $\mathscr{M}' = (\eta_2, \eta_3, ...)$. If α' is constructed from $\mathbf{X}'$ and $\mathscr{M}'$, and L acts on subsets of $\mathfrak{X} \times \mathbb{N}$ as a left shift by sending elements of the form

$(x, j+1)$ to (x, j) and annihilating elements of the form $(x, 1)$, then for all $A \subseteq \mathfrak{X} \times \mathbb{N}$

$$\alpha(A) = X_1(\eta_1 \times \delta_1)(A) + (1 - X_1)\alpha'(L(A))$$

where α' is independent of the pair (η_1, X_1). Furthermore, α is self-similar if and only if α and α' have the same law. In this sense, α is self-similar if and only if its law is invariant under a left shift.

Self-similarity of α can be equivalently expressed through the pushforward $\pi_*\alpha$ using:

$$\pi_*\alpha = X_1\eta_1 + (1 - X_1)\pi_*\alpha'$$

noting $\pi_*\alpha'$ is independent of (η_1, X_1). Here, $\pi_*\alpha$ and $\pi_*\alpha'$ have the same distribution if α is self-similar. This motivates the following abstraction of self-similarity for random probability measures on $\mathfrak{X}$:

Definition 2.11 (Self-similarity Equation)**.** Let $\mathfrak{X}$ be a measurable space and let ν and η be two random probability measures on $\mathfrak{X}$. Let X be a random variable taking values in $[0, 1]$. Suppose there exists a triple (η', X', ν') where (η', X') has the same law as (η, X), ν' has the same law as ν, and ν' is independent of the pair (η', X'), such that

$$\nu \stackrel{d}{=} X'\eta' + (1 - X')\nu'.$$

Then we say ν satisfies a self-similarity equation with respect to (η, X).

Note that self-similar generalized RAMs are not the objects meant to satisfy self-similarity equations. Rather, if α is a generalized RAM, self-similarity of α as a generalized RAM corresponds with its pushforward $\pi_*\alpha$ satisfying a self-similarity equation. In fact:

Proposition 2.12. *Let $\mathfrak{X}$ be a measurable space, X be a $[0, 1]$-valued random variable with $\mathscr{P}(X = 0) < 1$, and η be a random probability measure on $\mathfrak{X}$, not necessarily independent. Then*

(a) if $(\eta_j, X_j)_{j\geq 1}$ is a sequence of iid copies of (η, X) and α is the generalized RAM constructed from this sequence, then $\pi_*\alpha$ satisfies a self-similarity equation with respect to (η, X), and

(b) random measures ν on $\mathfrak{X}$ satisfying a self-similarity equation with respect to (η, X) are unique in law.

We note that this proposition has been proven more broadly before and previously applied to the Dirichlet process. This particular statement is a reformulation of a lemma in [12] in the language of generalized RAMs. See also [60] for an application of self-similarity equations to the Dirichlet process, [43] for related integral characterizations of the Dirichlet process, and [28] for a study of equations like self-similarity equations.

This proposition states that all distributions of measures satisfying self-similarity equations arise from pushforwards of self-similar generalized RAMs, and that all such pushforwards produce measures satisfying self-similarity equations. Additionally, this proposition says that satisfaction of a self-similarity equation uniquely characterizes the law of a random measure, and so may be used to identify the law of that random measure.

Theorem 2.13. *Let $\mathfrak{X} \subseteq \mathbb{N}$ and let $\mathbf{P}$ be a self-similar RAM constructed from iid proportions $\mathbf{X}$. Let $\mathbf{T}$ be a homogeneous Markov chain on the state-space $\mathfrak{X}$ independent of $\mathbf{X}$ with initial distribution μ, and let $\mathbf{W}$ and $\mathscr{M}$ be defined as in* (2.3) *and* (2.6) *with respect to $\mathbf{T}$ and $\mathbf{P}$. Define the stick-breaking construction*

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j}.$$

Then the joint distribution of (ν, T_1) is uniquely characterized by μ and the conditional joint distributions of $\{(\eta_1, X_1^W)|T_1 = x\}_{x\in\mathfrak{X}}$ in the following sense:

1. $T_1 \sim \mu$

2. *For each* $x \in \mathfrak{X}$, $\nu\big|T_1 = x$ *satisfies a self-similarity equation with respect to* $(\eta_1, X_1^W)|T_1 = x$.

For clarity, the theorem above characterizes T_1 and ν in terms of related variables (η_1, X_1^W). Note, however, that variables independent of T_1 and ν may be used:

Corollary 2.13.1. *Consider the setting of Theorem 2.13. Let* $\tilde{\mathbf{X}}$ *be identically distributed to* $\mathbf{X}$ *and for each* $x \in \mathfrak{X}$, *let* $\mathbf{T}(x)$ *be a homogenous Markov chain with the same transition kernel as* $\mathbf{T}$ *and* $T_1(x) = x$. *For each* $x \in \mathfrak{X}$, *define* $W(x) = \inf\{j > 1 : T_j(x) = x\}$ *and*

$$\tilde{X}(x) = \sum_{j=1}^{W(x)-1} \tilde{X}_j \prod_{i=1}^{j-1}(1 - \tilde{X}_i)$$

$$\eta(x) = \begin{cases} \tilde{X}(x)^{-1} \displaystyle\sum_{j=1}^{W(x)-1} \delta_{T_j(x)} \tilde{X}_j \prod_{i=1}^{j-1}(1 - \tilde{X}_i) & \text{if } \tilde{X}(x) > 0 \\ \delta_x & \text{if } \tilde{X}(x) = 0. \end{cases}$$

Then the joint distribution of (ν, T_1) *is uniquely characterized by* μ *and the joint distributions of* $\{(\eta(x), \tilde{X}(x))\}_{x\in\mathfrak{X}}$ *in the following sense:*

1. $T_1 \sim \mu$

2. *For each* $x \in \mathfrak{X}$, $\nu\big|T_1 = x$ *satisfies a self-similarity equation with respect to* $(\eta(x), \tilde{X}(x))$.

2.3 Proofs

2.3.1 Proofs for Section 2.1

Proof of Proposition 2.4. The new proportions $\mathbf{X^u} = (X_j^u)_{j\geq 1}$ defined from $\mathbf{X}$ can be reformulated using Lemma 1.19: For $j \geq 1$ such that $u_j < \infty$,

$$X_j^u = \sum_{i=u_j}^{u_{j+1}-1} X_i \prod_{l=u_j}^{i-1}(1 - X_l) = 1 - \prod_{i=u_j}^{u_{j+1}-1}(1 - X_i). \tag{2.7}$$

We need to show that $\mathbf{P^u}$ is constructed from $\mathbf{X^u}$, that $\mathbf{X^u}$ is a sequence of independent variables, and that $\mathbf{P^u}$ is almost surely a probability measure. Let $B = \sup\{j : u_j < \infty\}$, i.e. the index of the last nontrivial clump. Noting (2.7), write for finite $1 \leq j \leq B$ that

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

For $j > B$, note $P_j^u = 0$ and

$$\prod_{i=1}^{B}(1-X_i^u) = 1 - \sum_{i=1}^{B} P_i^u = 1 - \sum_{i=1}^{\infty} P_i = 0.$$

Then, $X_j^u \prod_{i=1}^{j-1}(1-X_i^u) = 0 = P_j^u$.

Since $\mathbf{X}$ is composed of independent variables, so is $\mathbf{X^u}$. Furthermore, we have $\sum_{j\geq 1} P_j^u = \sum_{j\geq 1} P_j \overset{a.s.}{=} 1$, so $\mathbf{P^u}$ is a RAM constructed from $\mathbf{X^u}$. ■

Proof of Corollary 2.4.1. The sequences $\mathbf{V}$ and $\mathbf{W}$ are functions of $\mathbf{T}$. Therefore, it follows immediately from the Proposition 2.4 that $\mathbf{P^V}\big|\mathbf{T} = \mathbf{t}$ and $\mathbf{P^W}\big|\mathbf{T} = \mathbf{t}$ are conditionally RAMs. ■

Proof of Theorem 2.5.

Part (1):

We begin by establishing the structure of a possible sequence $\mathbf{y}$ of $\mathbf{Y}$ before directly computing the distribution of $\mathbf{Y}$.

Let $\mathbf{y} = (y_i)_{i\geq 1}$ be a possible sequence for $\mathbf{Y}$. Recall $Y_j = T_{V_j}$ and $V_{j+1} = \inf\{v > V_j : T_v \neq T_{v-1}\}$ for all $j \geq 1$, where T_∞ is understood as the absorbing state of $\mathbf{T}$

should the sequence $\mathbf{V}$ be finite for only finitely many values. We see that for each $j \geq 1$, the event $\{Y_j = Y_{j+1}\}$ is then contained in, and hence equivalent to, the event $\{Y_j = Y_{j+1} = Y_{j+2} = ...\}$. Thus, a possible sequence $\mathbf{y}$ either is non-repeating, or is non-repeating until reaching a finite time J, after which $\mathbf{y}$ is constant. Accordingly, define $J = \inf\{j \geq 1 : y_j = y_{j+1}\}$.

We understand J as the time $\mathbf{y}$ initially reaches an absorbing state, with $J = \infty$ if $\mathbf{y}$ never reaches an absorbing state. Thus, for $i < J$, necessarily y_i is not an absorbing state of Q since $\mathbf{Y}$, and hence $\mathbf{T}$, has positive probability of changing states from y_i. Similarly, if $J < \infty$ and $J \leq k$, y_k is necessarily an absorbing state of Q. We conclude $Q_{y_i,y_i} \neq 1 = Q_{y_k,y_k}$ for $i < J \leq k$.

For $1 \leq n < J$, the event $Y_i = y_i$ for $1 \leq i \leq n$ is equivalently the event where $\mathbf{T}$ starts at y_1, staying there until time V_2 when it switches to y_2, remaining until time V_3, and so on up to time V_n when $\mathbf{T}$ switches to y_n. For $n < J$, we have

$$\begin{aligned}
& \mathscr{P}\left(Y_i = y_i : 1 \leq i \leq n\right) \\
= & \sum_{l_1=1}^{\infty} \cdots \sum_{l_{n-1}=1}^{\infty} \mathscr{P}\left(Y_i = y_i : 1 \leq i \leq n;\ V_{i+1} - V_i = l_i : 1 \leq i \leq n-1\right) \\
= & \sum_{l_1=1}^{\infty} \cdots \sum_{l_{n-1}=1}^{\infty} \mathscr{P}(T_1 = y_1) \prod_{i=1}^{n-1} Q_{y_i,y_i}^{l_i-1} Q_{y_i,y_{i+1}} \\
= & \mathscr{P}(T_1 = y_1) \prod_{i=1}^{n-1} \left(Q_{y_i,y_{i+1}} \sum_{l=1}^{\infty} Q_{y_i,y_i}^{l-1} \right) \\
= & \mathscr{P}(T_1 = y_1) \prod_{i=1}^{n-1} \frac{Q_{y_i,y_{i+1}}}{1 - Q_{y_i,y_i}} \\
= & \mathscr{P}(T_1 = y_1) \prod_{i=1}^{n-1} K(y_i, y_{i+1}). \qquad (2.8)
\end{aligned}$$

Suppose $J < \infty$. Then, y_J is an absorbing state of Q and, for $i \geq J$, we have

$Q_{y_i,y_i} = K(y_i, y_i) = K(y_i, y_{i+1}) = 1$. Define $l_J = \infty$ and write for $n \geq J$ that

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

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

The reduction of the kernel K in the case that Q is without absorbing states follows immediately from the definition of K and the characterization of an absorbing state z of Q by the property $Q_{z,z} = 1$.

Part (2):

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 only need to show that the associated proportions $\mathbf{X^V}$ 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 $\mathbf{Y}$, and associate to $\mathbf{y}$ the time J as in the proof of part (1). We begin by considering the conditional joint law of $(V_{i+1} - V_i)_{i=1}^n$ given $\mathbf{Y}$. With respect to a possible sequence $\mathbf{y}$ of $\mathbf{Y}$ and a sequence of numbers $l_j \in \mathbb{N}$ for $1 \leq j < J$ and $l_J = \infty$, denote by A_n and B_n the events

$$A_n = \{Y_j = y_j : 1 \leq j \leq n\} \qquad \text{and} \qquad B_n = \{V_{j+1} - V_j = l_j : 1 \leq j \leq \min(n, J)\}.$$

Then we have for $m \leq n < J$ that

$$\begin{aligned}
& \mathscr{P}\left(A_{n+1} \cap B_m\right) \\
= \; & \mathscr{P}\left(T_1 = y_1\right) \left(\prod_{i=1}^{m} Q_{y_i,y_i}^{l_i-1} Q_{y_i,y_{i+1}} \right) \left(\prod_{i=m+1}^{n} \sum_{l=1}^{\infty} Q_{y_i,y_i}^{l-1} Q_{y_i,y_{i+1}} \right) \\
= \; & \mathscr{P}\left(T_1 = y_1\right) \left(\prod_{i=1}^{n} \frac{Q_{y_i,y_{i+1}}}{1-Q_{y_i,y_i}} \right) \left(\prod_{i=1}^{m} Q_{y_i,y_i}^{l_i-1} (1-Q_{y_i,y_i}) \right) \\
= \; & \mathscr{P}\left(A_{n+1}\right) \prod_{i=1}^{m} Q_{y_i,y_i}^{l_i-1} (1-Q_{y_i,y_i}). \qquad (2.9)
\end{aligned}$$

Suppose $J < \infty$. For $n \geq J$ we have that:

$$\begin{aligned}
& \mathscr{P}\left(A_{n+1} \cap B_J\right) \\
= \; & \mathscr{P}\left(A_J \cap B_J\right) \\
= \; & \mathscr{P}\left(T_1 = y_1\right) \prod_{i=1}^{J-1} Q_{y_i,y_i}^{l_i-1} Q_{y_i,y_{i+1}} \\
= \; & \mathscr{P}\left(T_1 = y_1\right) \left(\prod_{i=1}^{J-1} \frac{Q_{y_i,y_{i+1}}}{1-Q_{y_i,y_i}} \right) \left(\prod_{i=1}^{J-1} Q_{y_i,y_i}^{l_i-1} (1-Q_{y_i,y_i}) \right) \\
= \; & \mathscr{P}\left(A_J\right) \prod_{i=1}^{J-1} Q_{y_i,y_i}^{l_i-1} (1-Q_{y_i,y_i}) \\
= \; & \mathscr{P}\left(A_{n+1}\right) \prod_{i=1}^{J-1} Q_{y_i,y_i}^{l_i-1} (1-Q_{y_i,y_i}).
\end{aligned}$$

When $\mathbf{X}$ is composed of iid variables, i.e. when $\mathbf{P}$ is a self-similar RAM, we will argue now that the proportions $\mathbf{X}^{\mathbf{V}} \big| \mathbf{Y} = \mathbf{y}$ form a conditionally independent sequence, and therefore $\mathbf{P}^{\mathbf{V}} \big| \mathbf{Y} = \mathbf{y}$ is RAM. We split into subcases, $J = \infty$ versus $J < \infty$.

When $J = \infty$, let $r \geq n \geq 1$, and $(a_i)_{i=1}^{n} \in (0,1)^n$. Then

$$\begin{aligned}
& \mathscr{P}\left(1 - X_j^V \leq a_j : 1 \leq j \leq n \middle| A_{r+1} \right) \\
= & \sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \mathscr{P}\left(1 - X_j^V \leq a_j : 1 \leq j \leq n \middle| B_n \cap A_{r+1} \right) \mathscr{P}\left(B_n \middle| A_{r+1} \right). \qquad (2.10)
\end{aligned}$$

Relative to $(l_j)_{j=1}^n$, define the sequence $\mathbf{v} = (v_j)_{j=1}^{n+1}$ where $v_j = 1 + \sum_{k=1}^{j-1} l_k$ for $1 \leq j \leq n+1$, which marks the first n times when $\mathbf{T}$ changes states. In particular, on the event B_n, we have $V_j = v_j$ for $1 \leq j \leq n+1$. Given B_n, we then have from (2.7) that the proportions $(X_j^V)_{j=1}^n$ satisfy $1 - X_j^V = \prod_{k=v_j}^{v_{j+1}-1}(1 - X_k)$ for $1 \leq j \leq n$ and are independent of each other and $\mathbf{Y}$. Thus, we have that (2.10) equals

$$\begin{aligned}
&\sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \mathscr{P}\left(\prod_{k=v_j}^{v_{j+1}-1} (1 - X_k) \leq a_j : 1 \leq j \leq n \right) \left[\prod_{j=1}^{n} Q_{y_j,y_j}^{l_j-1}(1 - Q_{y_j,y_j}) \right] \\
=&\sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \prod_{j=1}^{n} \mathscr{P}\left(\prod_{k=v_j}^{v_{j+1}-1} (1 - X_k) \leq a_j \right) Q_{y_j,y_j}^{l_j-1}(1 - Q_{y_j,y_j}) \\
=&\sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \prod_{j=1}^{n} \mathscr{P}\left(\prod_{k=1}^{l_j} (1 - X_k) \leq a_j \right) Q_{y_j,y_j}^{l_j-1}(1 - Q_{y_j,y_j}) \\
=&\prod_{j=1}^{n} \sum_{l=1}^{\infty} \mathscr{P}\left(\prod_{k=1}^{l} (1 - X_k) \leq a_j \right) Q_{y_j,y_j}^{l-1}(1 - Q_{y_j,y_j}) \qquad (2.11)
\end{aligned}$$

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

When $J < \infty$, note that the collection $(X_j^V|\mathbf{Y} = \mathbf{y})_{j \geq J}$ is a deterministic sequence of 1s. Thus, we need only show the proportions $(X_j^V|\mathbf{Y} = \mathbf{y})_{j=1}^{J-1}$ are independent. Define $l_J = \infty$. For $n \geq J$, we have

$$\begin{aligned}
&\mathscr{P}\left(1 - X_j^V \leq a_j : 1 \leq j < J \middle| A_{n+1} \right) \\
=&\sum_{l_1=1}^{\infty} \cdots \sum_{l_{J-1}=1}^{\infty} \mathscr{P}\left(1 - X_j^V \leq a_j : 1 \leq j < J \middle| B_J \cap A_{n+1} \right) \mathscr{P}\left(B_J \middle| A_{n+1} \right). \qquad (2.12)
\end{aligned}$$

Define for $j \leq J+1$ the sequence v_j as before, and note $v_{J+1} = \infty$. One derives as before that (2.12) equals

$$\prod_{j=1}^{J-1} \sum_{l=1}^{\infty} \mathscr{P}\left(\prod_{k=1}^{l} (1 - X_k) \leq a_j \right) Q_{y_j,y_j}^{l-1}(1 - Q_{y_j,y_j}), \qquad (2.13)$$

in factored form. Therefore, the proportions $\mathbf{X}^{\mathbf{V}}$ are conditionally independent as desired and $\mathbf{P}^{\mathbf{V}}\big|\mathbf{Y} = \mathbf{y}$ is a RAM also in the case $J < \infty$. ■

Proof of Theorem 2.8. Without loss of generality, we assume T_1 is deterministically the fixed state t_1 and prove the result without conditioning upon the event $T_1 = t_1$. Let $\mathbf{X}$ be the iid sequence from which $\mathbf{P}$ is constructed.

The main idea of this proof is that cycles of a homogeneous Markov chain, conditioned upon the initial state t_1 of the cycle, are nearly iid as as result of the Markov property. If t_1 is recurrent, cycles are exactly iid, and if t_1 is transient, cycles are iid after conditioning on cycles having finite length. To establish Theorem 2.8, we consider directly the distribution of the cycles $\mathbf{C} = \{C_j\}$ and the corresponding partition $\mathbf{S} = \{S_j\}$ of the proportions $\mathbf{X}$. To then connect to an iid sequence $\{(\tilde{X}_j, \tilde{\eta}_j)\}_{j\geq 1}$, we choose to construct $\{(\tilde{X}_j, \tilde{\eta}_j)\}_{j\geq 1}$ from an iid sequence $\{(S'_j, C'_j)\}$ of copies of (S_1, C_1). We then appropriately match the distributions of the true cycles and partitioned proportions $\{(S_j, C_j)\}$ with the iid copies $\{(S'_j, C'_j)\}$.

We begin by formalizing these objects (Step 1), and then consider separately the cases that t_1 is recurrent (Step 2) or transient (Step 3).

Step 1: Define $B = \sup\{j : W_j < \infty\}$. We understand B to be the number of cycles of $\mathbf{T}$ starting at t_1. If t_1 is a recurrent state, then B is almost surely infinite. However, if t_1 is a transient state, then B will be almost surely finite, as the chain will visit the state t_1 only finitely many times.

For $j \geq 1$, define:

$$C_j = \begin{cases} (T_{W_j}, T_{W_j+1}, T_{W_j+2}, ..., T_{W_{j+1}-1}) & \text{if } j < B \\ (T_{W_B}, T_{W_B+1}, T_{W_B+2}, ...) & \text{if } j = B \\ \emptyset & \text{if } j > B, \end{cases}$$

$$S_j = \begin{cases} (X_{W_j}, X_{W_j+1}, X_{W_j+2}, ..., X_{W_{j+1}-1}) & \text{if } j < B \\ (X_{W_B}, X_{W_B+1}, X_{W_B+2}, ...) & \text{if } j = B \\ \emptyset & \text{if } j > B. \end{cases}$$

The collection $\mathbf{C} = \{C_j\}_{j\geq 1}$ is of the cycles of $\mathbf{T}$, which are of finite length for $j < B$,

of infinite length for the last cycle $j = B$, and non-existent for $j > B$. We define the partition $\mathbf{S} = \{S_j\}_{j\geq 1}$ of the proportions $\mathbf{X}$ to correspond to the cycles.

If for a list C we denote its cardinality by $\#C$ and its entries by $C = (C(i))_{i=1}^{\#C}$, then we have

$$X_j^W = \begin{cases} 1 - \prod_{i=1}^{\#S_j}(1 - S_j(i)) & \#S_j > 0 \\ \\ 1 & \#S_j = 0, \end{cases}$$

$$\eta_j = \begin{cases} (X_j^W)^{-1} \sum_{i=1}^{\#S_j} \delta_{C_j(i)} S_j(i) \prod_{l=1}^{i-1}(1 - S_j(l)) & X_j^W > 0,\ \#S_j > 0 \\ \\ \delta_{t_1} & X_j^W = 0 \text{ or } \#S_j = 0. \end{cases}$$

We build an iid sequence $\{(\tilde{X}_j, \tilde{\eta}_j)\}_{j\geq 1}$ of copies of (X_1^W, η_1) through a similar construction: Let $(\mathbf{S}', \mathbf{C}')$ be an iid sequence of copies of (S_1, C_1). Define $\mathbf{W}' = (W_1', W_2', ...)$ by $W_1' = 1$ and for $j \geq 1$:

$$W_{j+1}' = 1 + \sum_{i=1}^{j} \#C_i' \qquad \text{and} \qquad B' = \sup\{j : W_j' < \infty\}.$$

Note that $\mathbf{W}'$ is the same function of $\mathbf{C}'$ that $\mathbf{W}$ is of $\mathbf{C}$, just as B' is the same function of $\mathbf{W}'$ that B is of $\mathbf{W}$. Define for each $j \geq 1$

$$\tilde{X}_j = 1 - \prod_{i=1}^{\#S_j'}(1 - S_j'(i)),$$

$$\tilde{\eta}_j = \begin{cases} (\tilde{X}_j)^{-1} \sum_{i=1}^{\#S_j'} \delta_{C_j'(i)} S_j'(i) \prod_{l=1}^{i-1}(1 - S_j'(l)) & \tilde{X}_j^W > 0 \\ \\ \delta_{t_1} & \tilde{X}_j^W = 0, \end{cases}$$

and note that as $\#S_j' > 0$ almost surely for each j, $\{(\tilde{X}_j, \tilde{\eta}_j)\}_{j\geq 0}$ is the same function of $(\mathbf{S}', \mathbf{C}')$ that $\{(X_j^W, \eta_j)\}_{j\geq 0}$ is of $(\mathbf{S}, \mathbf{C})$.

If we construct $\mathbf{P}^{\mathbf{W}}$ from $\mathbf{X}^{\mathbf{W}}$ and $\tilde{\mathbf{P}}$ from $\tilde{\mathbf{X}}$, then to show the desired result

$$\left(P_j^W \eta_j\right)_{j\geq 1} \overset{d}{=} \left(\tilde{P}_j \tilde{\eta}_j\right)_{j\geq 1}, \tag{2.14}$$

it is sufficient to compare the distributions of $(\mathbf{S}, \mathbf{C})$ and $(\mathbf{S}', \mathbf{C}')$, where the precise nature of that comparison is determined by the transience or recurrence of t_1.

Step 2: Suppose first that t_1 is transient. As argued earlier, B is finite almost surely. Furthermore, $\mathscr{P}(\#C_1 = \infty) > 1$, and so B' is finite almost surely. Thus for $i > B$ and $j > B'$, $P_i^W = \tilde{P}_j = 0$, and to conclude (2.14) it is sufficient to determine

$$\{(S_j, C_j)\}_{j=1}^{B} \stackrel{d}{=} \left\{(S'_j, C'_j)\right\}_{j=1}^{B'}. \tag{2.15}$$

Let $k \in \mathbb{N}$ and for $1 \leq j \leq k$, let A_j be a measurable subset in the space that the elements (S_j, C_j) take values. A base set for measurable sets may be understood as having one of two forms: Let $n \in \mathbb{N}$, let $B_j \subset [0,1]$ be a measurable subset for each j, and let $x_j \in \mathfrak{X}$. A base set A may be given as

$$A = \{(S, C) : \#S = \#C = n, \text{ for } 1 \leq j \leq n : \ S(j) \in B_j, \ C(j) = x_j\}$$

or

$$A = \{(S, C) : \#S = \#C = \infty, \text{ for } 1 \leq j \leq n : \ S(j) \in B_j, \ C(j) = x_j\}.$$

Note that the element $(\emptyset, \emptyset)$ of that space has probability zero of being observed by (S_1, C_1) since the chain $\mathbf{T}$ must have a first cycle. On the one hand, since $(\mathbf{S}', \mathbf{C}')$ is a sequence of iid copies of (S_1, C_1), we have

$$\begin{aligned}
&\mathscr{P}\left((S'_j, C'_j) \in A_j : 1 \leq j \leq k; \ B' = k\right) \\
&= \left[\prod_{j=1}^{k-1} \mathscr{P}\left((S'_j, C'_j) \in A_j, \#C'_j \neq \infty\right)\right] \mathscr{P}\left((S'_k, C'_k) \in A_k, \#C'_1 = \infty\right) \\
&= \left[\prod_{j=1}^{k-1} \mathscr{P}\left((S_1, C_1) \in A_j, \#C_1 \neq \infty\right)\right] \mathscr{P}\left((S_1, C_1) \in A_k, \#C_1 = \infty\right).
\end{aligned}$$

On the other hand, by the Markov property we have

$$\begin{aligned}
&\mathscr{P}\left((S_j,C_j)\in A_j : 1\le j\le k;\ B=k\right)\\
=&\mathscr{P}\left((S_j,C_j)\in A_j,\ \#C_j\neq 0,\infty : 1\le j\le k-1;\ (S_k,C_k)\in A_k,\ \#C_k=\infty\right)\\
=&\mathscr{P}\left((S_1,C_1)\in A_1,\#C_1\neq 0,\infty\right)\\
&\qquad\times\left[\prod_{j=2}^{k-1}\mathscr{P}\left((S_j,C_j)\in A_j,\ \#C_j\neq 0,\infty|\#C_{j-1}\neq 0,\infty\right)\right]\\
&\qquad\times\mathscr{P}\left((S_k,C_k)\in A_k,\ \#C_k=\infty|\#C_{k-1}\neq 0,\infty\right)\\
=&\mathscr{P}\left((S_1,C_1)\in A_1,\#C_1\neq 0,\infty\right)\left[\prod_{j=2}^{k-1}\mathscr{P}\left((S_1,C_1)\in A_j,\ \#C_1\neq 0,\infty\right)\right]\\
&\qquad\times\mathscr{P}\left((S_1,C_1)\in A_k,\ \#C_1=\infty\right)\\
=&\left[\prod_{j=1}^{k-1}\mathscr{P}\left((S_1,C_1)\in A_j,\#C_1\neq\infty\right)\right]\mathscr{P}\left((S_1,C_1)\in A_k,\#C_1=\infty\right).
\end{aligned}$$

Note that if any of the above conditional probabilities conditions on an event of measure 0, then the original probability is 0, as is the final expression. Thus, in the case that t_1 is transient, we have (2.15), and from it the desired result (2.14).

Step 3: If instead t_1 is recurrent, then B is almost surely infinity. Furthermore, $\mathscr{P}(\#C_1=\infty)=0$, and so B' will be almost surely infinity. Thus, we aim to show

$$\{(S_j,C_j)\}_{j=1}^{\infty}\ \stackrel{d}{=}\ \left\{(S'_j,C'_j)\right\}_{j=1}^{\infty},$$

which implies not only (2.14) but also that the sequence $\{(X_j^W,\eta)\}_{j\ge 1}$ is iid. We do so by showing $\{(S_j,C_j)\}_{j=1}^{k}\stackrel{d}{=}\left\{(S'_j,C'_j)\right\}_{j=1}^{k}$ for each $k\in\mathbb{N}$. Let k and the sets $\{A_j\}_{j=1}^{k}$ be as in Step 2. Then by computations analogous to those in the case t_1 is transient, we have

$$\begin{aligned}
\mathscr{P}\left((S'_j,C'_j)\in A_j : 1\le j\le k\right)&=\prod_{j=1}^{k}\mathscr{P}\left((S_1,C_1)\in A_j\right)\\
&=\mathscr{P}\left((S_j,C_j)\in A_j : 1\le j\le k\right).
\end{aligned}$$

The result follows. ■

2.3.2 Proofs for Section 2.2

Proof of Proposition 2.12.

(a) Let $(\eta_j, X_j)_{j\geq 1}$ be an iid sequence of copies of (η, X). Additionally, define for $j \geq 1$ the pairs $(\eta'_j, X'_j) = (\eta_{j+1}, X_{j+1})$. Let α and α' be the generalized RAMs constructed from the sequences of pairs $(\eta_j, X_j)_{j\geq 1}$ and $(\eta'_j, X'_j)_{j\geq 1}$ respectively. Since $(\eta_j, X_j)_{j\geq 1}$ is an iid sequence, α' and $\pi_*\alpha'$ are independent of (η_1, X_1), and α' and $\pi_*\alpha'$ are identically distributed to α and $\pi_*\alpha$ respectively. We compute:

$$\begin{aligned}
\pi_*\alpha &= \sum_{j=1}^{\infty} \eta_j X_j \prod_{i=1}^{j-1}(1 - X_i) \\
&= \eta_1 X_1 + (1 - X_1) \sum_{j=2}^{\infty} \eta_j X_j \prod_{i=2}^{j-1}(1 - X_i) \\
&= \eta X + (1 - X) \sum_{j=1}^{\infty} \eta'_j X'_j \prod_{i=1}^{j-1}(1 - X'_i) \\
&= \eta_1 X_1 + (1 - X_1)\pi_*\alpha'.
\end{aligned}$$

Thus, $\pi_*\alpha$ satisfies a self-similarity equation with respect to (η, X).

(b) Let ν and ν' be two random measures satisfying a self-similarity equation with respect to (η, X). Let $(\eta_j, X_j)_{j\geq 1}$ be an iid sequence of copies of (η, X) independent of ν and ν'. Define $\nu_1 = \nu$ and $\nu'_1 = \nu'$, and for $j > 1$ define

$$\nu_{j+1} = X_j\eta_j + (1 - X_j)\nu_j \qquad \text{and} \qquad \nu'_{j+1} = X_j\eta_j + (1 - X_j)\nu'_j.$$

By construction, $\{\nu_j\}_{j\geq 1}$ and $\{\nu'_j\}_{j\geq 1}$ are two iid sequences of copies of ν and ν' respectively. Thus, in terms of the variational norm $||\cdot||$,

$$\begin{aligned}
\left|\left|\nu_{j+1} - \nu'_{j+1}\right|\right| &= |1 - X_j| \left|\left|\nu_j - \nu'_j\right|\right| \\
&= \left(\prod_{i=1}^{j} |1 - X_i|\right) ||\nu_1 - \nu'_1|| \\
&\leq \prod_{i=1}^{j} |1 - X_i|.
\end{aligned}$$

Since $\{X_j\}_{j\geq 1}$ is an iid sequence of copies of X and $P(X = 0) < 1$, the final product converges to 0 almost surely. Thus, $\nu_j - \nu'_j$ converges almost surely to the 0 measure. Since $\{\nu_j\}_{j\geq 1}$ and $\{\nu'_j\}_{j\geq 1}$ are two iid sequences, this implies that $\nu = \nu_1$ and $\nu' = \nu'_1$ have the same law.

■

Proof of Theorem 2.13. This follows immediately from Theorem 2.8 and Proposition 2.12. ■

Proof of Corollary 2.13.1. Note that Definition 2.11 uses the distribution of (η, X) and of ν, but not the joint distribution of (η, X, ν). As such, Corollary 2.13.1 is simply a restatement of Theorem 2.13. ■

Chapter 3

An Inhomogenous Markov Chain and Stick-breaking Constructions

In this chapter, we study the occupation law of a particular class of inhomogeneous Markov chain and connect that law to a family of generalized stick-breaking constructions built from a homogeneous Markov chain and a GEM distribution. We do so by identifying an intermediate clumped object, the Markov Chain conditional GEM (MCcGEM) distribution.

Recall the definition of a generator matrix:

Definition 3.1 (Generator matrix)**.** Let $G = \{G_{i,j} : i, j \in \mathfrak{X}\}$ be a square matrix on $\mathfrak{X} \subseteq \mathbb{N}$. We say that G is a generator matrix if it satisfies

(a) $G_{i,j} \geq 0$ for each $i \neq j \in \mathfrak{X}$,

(b) $G_{i,i} = -\sum_{j \neq i} G_{i,j}$ for each $i \in \mathfrak{X}$,

(c) and $\sup_i |G_{i,i}| < \infty$.

We say G has a particular trait, e.g. is irreducible or has a positive recurrent state, if the stochastic kernel $I + G/\sup_i |G_{i,i}|$ has that trait.

Recall that a generator matrix is always a matrix of the form $\theta(Q - I)$ where $\theta > 0$ and Q is a stochastic matrix.

We consider here a class of inhomogeneous Markov chain with transition kernels of the following form: Let G be a generator matrix on $\mathfrak{X} \subseteq \mathbb{N}$ without zero rows and define stochastic kernels

$$K_n = I + \frac{G}{n}\mathbb{1}(n > c) \tag{3.1}$$

FIGURE 3.1. Connecting Diagram

$$
\begin{array}{lll}
(a) & \displaystyle\sum_{j=1}^{\infty} P_j \delta_{T_j} & (\mathbf{P}, \mathbf{T}) \\
& = & \downarrow \textit{ via clumping} \\
(b) & \displaystyle\sum_{j=1}^{\infty} P_j^V \delta_{Y_j} & (\mathbf{P^V}, \mathbf{Y}) \\
& \stackrel{d}{=} & \textit{as } (\mathbf{P^V}, \mathbf{Y}) \stackrel{d}{=} (\mathbf{R}, \mathbf{Z}) \\
(c) & \displaystyle\sum_{j=1}^{\infty} R_j \delta_{Z_j} & (\mathbf{R}, \mathbf{Z}) \\
& \uparrow & \textit{convergence in distribution} \\
(d) & \displaystyle\sum_{j=1}^{N_n - 1} P_{n,j} \delta_{Y_{n,j}} & (\mathbf{P}_n, \mathbf{Y}_n) \\
& = & \uparrow \textit{ via reverse clumping} \\
(e) & \displaystyle\sum_{j=1}^{n} \frac{1}{n} \delta_M & \mathbf{M}
\end{array}
$$

for some fixed integer $c \geq \sup_i |G_{i,i}| - 1$. Let $\mathbf{M} = (M_n)_{n\geq 1}$ be a time-inhomogeneous Markov chain on $\mathfrak{X}$ associated to transition kernels $\{K_n\}_{n\geq 1}$.

In the first section, we will consider the so-called local occupations of a Markov chain $\mathbf{M}$ with kernels $(K_n)_{n\geq 1}$. As time passes, the chain $\mathbf{M}$ typically spends more and more time on a single state before switching states. We will refer colloquially to a single stretch of time that $\mathbf{M}$ spends on a state as a clump of $\mathbf{M}$ on that state. Thus, later clumps of the chain on a state are typically larger than earlier clumps. Once normalized, we think of these clump sizes up to time n as the collection of local occupations by $\mathbf{M}$ of the states visited up to time n. We study these local occupations as n gets large as a method of studying the (global) occupation law of $\mathbf{M}$. To keep the local occupations from tending to zero as n gets large as a result of normalization by n, we study the local occupations in reverse chronological order. The process of forming the local occupations of M is then a reverse-clumping procedure akin to the forward-in-time procedure of Chapter 2.

By forming these local occupations of the Markov chain and limiting, we recover a Markov Chain conditional RAM structure, termed the MCcGEM distribution. Then, we show that stick-breaking constructions built from a paired GEM distribution $\mathbf{P}$ and independent homogeneous Markov chain $\mathbf{T}$ are a concrete example of measures with the same law as the occupation law of the inhomogeneous Markov chain.

We include for the reader's reference Figure 3 depicting the relations we will establish in this chapter. In the right column are the objects studied in Section 3.1, and in the left column are the related measures studied in Section 3.2, with the manner of connecting each set of objects noted in the middle.

Some of this work was developed jointly with Zach Dietz and Sunder Sethuraman, and can also be found in [12]. In particular, Theorems 3.2, 3.4, 3.7, 3.8, and 3.9, and Proposition 3.5.

3.1 Local Occupations

We begin by defining the 'local occupations' of $\mathbf{M}$. Let $\tilde{\mathbf{V}}$ be the sequence of times $(\tilde{V}_1, \tilde{V}_2, \tilde{V}_3, ...)$ at which $\mathbf{M}$ changes states. As in the previous chapter, we specify that $\mathbf{M}$ 'switches states' upon observing the first state, i.e. $\tilde{V}_1 = 1$, and then define

$$\tilde{V}_{j+1} = \inf\{v > \tilde{V}_j : M_v \neq M_{v-1}\}.$$

Then, we define for each $n \geq 1$ the first time index N_n that $\mathbf{M}$ switches states after time n, $N_n = \inf\{j : \tilde{V}_j > n\}$. Note that as G is without zero rows, every diagonal entry of G is strictly negative and almost surely N_n diverges to infinity:

$$\begin{aligned}
\mathscr{P}\left(\lim_{n\to\infty} N_n < \infty\right) &= \sum_{x\in\mathfrak{X}} \sum_{j=1}^{\infty} \mathscr{P}(M_j \neq x = M_{j+1} = M_{j+2} = M_{j+3} = \cdots) \\
&= \sum_{x\in\mathfrak{X}} \sum_{j=1}^{\infty} \mathscr{P}(M_j \neq x = M_{j+1}) \prod_{n=1+\max(c,j)}^{\infty} \left(1 + \frac{G_{x,x}}{n}\right) \\
&= 0,
\end{aligned}$$

where the last equality follows from every product $\prod_{n=k}^{\infty}\left(1 + \frac{G_{x,x}}{n}\right)$ being 0. Thus, the chain $\mathbf{M}$ switches states infinitely often almost surely, and $\tilde{\mathbf{V}}$ is a strictly increasing sequence of integers with 0 probability of taking on an infinite value.

We consider the two collections

$$\left\{\frac{\tilde{V}_{j+1} - \tilde{V}_j}{n}\right\}_{1\leq j<N_n-1} \cup \left\{\frac{n+1-\tilde{V}_{N_n-1}}{n}\right\} \qquad \text{and} \qquad \left\{M_{\tilde{V}_j}\right\}_{1\leq j\leq N_n-1}$$

to be the local occupations of $\mathbf{M}$ up to time n. In particular, for $j < N_n - 1$, the jth local occupation of $\mathbf{M}$ is given at time n by the proportion $\frac{\tilde{V}_{j+1}-\tilde{V}_j}{n}$ on state $M_{\tilde{V}_j}$.

As a concrete example, consider an observation

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

Then at time $n = 4$, we have $N_4 = 3$ and the set of local occupations is given by

$$\left\{\frac{3}{4}, \frac{1}{4}\right\} \qquad \text{and} \qquad \{1, 6\}.$$

Similarly, at time $n = 7$, we have $N_4 = 5$ and the set of local occupations is given by

$$\left\{\frac{3}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{7}\right\} \qquad \text{and} \qquad \{1, 6, 1, 3\}.$$

These local occupations can be used to construct the empirical occupation measure:

$$\nu^n := \frac{1}{n}\sum_{j=1}^{n} \delta_{M_j} = \sum_{j=1}^{N_n-2} \frac{\tilde{V}_{j+1} - \tilde{V}_j}{n} \delta_{M_{\tilde{V}_j}} + \frac{n+1-\tilde{V}_{N_n-1}}{n} \delta_{M_{\tilde{V}_{N_n-1}}} \tag{3.2}$$

By studying the collection of local occupations as n gets large, we aim to better understand the global occupation of $\mathbf{M}$, i.e. the limiting distribution of the empirical occupations ν^n of $\mathbf{M}$. However, we must be careful to study local occupations in reverse chronological order as n gets large. The jth local occupation of $\mathbf{M}$, given at time n, will converge almost surely to 0 as n gets large. However, we will show that the $(n+1-j)$th local occupation at time n will converge in distribution non-trivially as n gets large. In fact, indexed in reverse chronological order, the collection of local occupations will converge jointly in distribution and, very importantly, without loss of mass.

Formally: for $1 < k < N_n$, $j \geq N_n$, and all $l \in \mathbb{N}$, define

$$\begin{aligned} \tau_{n,1} &= n + 1 - \tilde{V}_{N_n-1} & Y_{n,1} &= M_{\tilde{V}_{N_n-1}} \\ \tau_{n,k} &= \tilde{V}_{N_n-(k-1)} - \tilde{V}_{N_n-k} & Y_{n,k} &= M_{\tilde{V}_{N_n-k}} \\ \tau_{n,j} &= 0 & Y_{n,j} &= T_1 \end{aligned} \tag{3.3}$$

$$P_{n,l} = \frac{\tau_{n,l}}{n}$$

Then the pair $(\mathbf{P}_n, \mathbf{Y}_n)$ are the local occupations of $\mathbf{M}$ up to time n.

We think of $\tau_{n,k}$ as the length of time $\mathbf{M}$ visited state $Y_{n,k}$, where the sequence $\mathbf{Y}_n = (Y_{n,1}, Y_{n,2}, Y_{n,3}, ...)$ is of states $\mathbf{M}$ visited in reverse chronological order starting at time n. We specify that $\mathbf{Y}_n$ is eventually constant at M_1, though other conventions may be chosen. The sequence $\mathbf{P}_n = (P_{n,1}, P_{n,2}, P_{n,3}, ...)$ is then the sequence of

proportions of time $\mathbf{M}$ visited the states $\mathbf{Y}_n$, with $P_{n,j} = 0$ after all local occupations have been recorded.

Returning to the example above, with observation

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

we have for $n = 4$ that 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 $\mathbf{Y}_7 = (3, 1, 6, 1, 1, 1, 1, \ldots)$.

The nth empirical measure ν^n of $\mathbf{M}$ as in (3.2) can now be rewritten as

$$\nu^n = \sum_{j=1}^{n} \frac{1}{n} \delta_{M_j} = \sum_{j=1}^{N_n - 1} P_{n,j} \delta_{Y_{n,j}} = \sum_{j=1}^{\infty} P_{n,j} \delta_{Y_{n,j}}.$$

We now consider the limiting distribution of the pair $(\mathbf{P}_n, \mathbf{Y}_n)$ as n gets large. Note, we consider a weak limit of the pair, where weak convergence is in the sense of finite-dimensional distributions, the natural sense associated to the product space $[0, 1]^{\mathbb{N}} \times \mathfrak{X}^{\mathbb{N}}$ endowed with the product topology.

Theorem 3.2 (Local Occupations of a Time-inhomogenous Markov Chain)**.** *Let G be a generator matrix on $\mathfrak{X}$ without zero rows and having at least one positive recurrent state. Define $\theta^G = \inf\{r \in \mathbb{R}^+ : I + G/r$ is non-negative$\}$, and let $c, \theta \geq \theta^G$ such that $c \in \mathbb{N}$. Define $Q = I + G/\theta$. Let also π be a distribution and μ be a stationary distribution of Q so that entry-wise,*

$$\pi^t Q^n \to \mu^t \quad \text{as} \quad n \to \infty. \tag{3.4}$$

Define stochastic kernels $\{K_n\}_{n \geq 1}$ by $K_n = I + \frac{G}{n} \mathbb{1}(n > c)$, and also define the generator matrix G' by

$$G'_{ij} = \frac{\mu_j}{\mu_i} G_{ji} \mathbb{1}(\mu_i \neq 0). \tag{3.5}$$

Let $\mathbf{M}$ be an inhomogeneous Markov chain with initial distribution π and transition kernels $\{K_n\}_{n \geq 1}$, and define $(\mathbf{P}_n, \mathbf{Y}_n)$ as in (3.3) with respect to $\mathbf{M}$. Then the pairs

$(\mathbf{P}_n, \mathbf{Y}_n)$ converge weakly to $(\mathbf{R}, \mathbf{Z})$ where $\mathbf{Z}$ is a stationary, homogeneous Markov chain with stationary distribution μ and stochastic kernel

$$K_{G'}(w,z) = \begin{cases} \dfrac{G'_{w,z}}{-G'_{w,w}} & \text{if } w \neq z;\ G'_{w,w} \neq 0 \\ 1 & \text{if } w = z;\ G'_{w,w} = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.6}$$

and $\mathbf{R}\big|\mathbf{Z}$ is conditionally distributed as a RAM constructed from proportions $\tilde{\mathbf{X}}$ with $\tilde{X}_j\big|\mathbf{Z} \sim \text{Beta}(1, -G'_{Z_j,Z_j})$.

Corollary 3.2.1. *Let G be an irreducible, positive recurrent generator matrix on $\mathfrak{X}$. Let $c \in \mathbb{N}$ be such that $c \geq \inf\{r \in \mathbb{R}^+ : I + r^{-1}G$ is non-negative$\}$. Let μ be the stationary distribution of G.*

Define stochastic kernels $\{K_n\}_{n\geq 1}$ by $K_n = I + \frac{G}{n}\mathbb{1}(n > c)$, and also define the generator matrix G' by

$$G'_{ij} = \frac{\mu_j}{\mu_i} G_{ji}. \tag{3.7}$$

Let $\mathbf{M}$ be an inhomogeneous Markov chain with transition kernels $\{K_n\}_{n\geq 1}$, and define $(\mathbf{P}_n, \mathbf{Y}_n)$ as in (3.3) with respect to $\mathbf{M}$. Then the pairs $(\mathbf{P}_n, \mathbf{Y}_n)$ converge weakly to $(\mathbf{R}, \mathbf{Z})$ where $\mathbf{Z}$ is a stationary, homogeneous Markov chain with stationary distribution μ and stochastic kernel

$$K_{G'}(w,z) = \frac{G'_{w,z}}{-G'_{w,w}}\mathbb{1}(w \neq z)$$

and $\mathbf{R}\big|\mathbf{Z}$ is conditionally distributed as a RAM constructed from proportions $\tilde{\mathbf{X}}$ with $\tilde{X}_j\big|\mathbf{Z} \sim \text{Beta}(1, -G'_{Z_j,Z_j})$.

Note that the conditional law of $\mathbf{R}|\mathbf{Z}$ described in Theorem 3.2 is that of a disordered GEM with parameters $\theta_j = -G'_{Z_j,Z_j}$. Recall from discussion below Definition 1.22 that as the entries of a generator matrix G are bounded, $\mathbf{R}$ is a.s. a probability distribution.

We discuss now the technical condition of the convergence (3.4). This is a condition on the structure of positive recurrent states of a homogeneous Markov chain run with kernel $Q = I + G/\theta$. Since the limit μ is a stationary distribution with respect to Q, the chain must have a positive recurrent state, and μ is positive only on such states. The initial distribution π must be such that observation of a positive recurrent state occurs almost surely.

In general, μ depends on π when there is more than one irreducible class of positive recurrent states. In the case that Q has a single class of positive recurrent states, then μ will be the unique stationary distribution associated with Q and will not depend on π. Corollary 3.2.1 gives the associated simplification when Q is irreducible, positive recurrent, and aperiodic.

We note, along with positive recurrent states, there may also be null recurrent and transient states associated with Q. It could be that Q has an infinite number of null recurrent or transient states, in addition to positive recurrent states. However, the requirement that μ be stationary means that a chain run with kernel Q starting from π cannot visit a null recurrent state or remain indefinitely on transient states a.s. This reflects that the limit of $(\mathbf{P}_n, \mathbf{Y}_n)$ corresponds to the long time average occupations of $\mathbf{M}$, a Markov chain with kernels

$$K_n = I + \frac{G}{n}\mathbb{1}(n > c) = I + \frac{\theta}{n}(Q - I)\mathbb{1}(n > c).$$

Theorem 3.2 identifies the weak limit $(\mathbf{R}, \mathbf{Z})$ of $(\mathbf{P}_n, \mathbf{Y}_n)$ as an example of a Markov Chain conditional RAM, a structure we first encountered in Theorem 2.5. In fact, upon inspection, the kernel $K_{G'}$ defined in (3.6) and the kernel K of Theorem 2.5 are

identical with respect to a decomposition $G' = \theta(Q' - I)$:

$$\begin{cases} \dfrac{G'_{w,z}}{-G'_{w,w}} & \text{if } w \neq z;\ G'_{w,w} \neq 0 \\ 1 & \text{if } w = z;\ G'_{w,w} = 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \dfrac{Q'_{w,z}}{1 - Q'_{w,w}} & \text{if } w \neq z;\ Q'_{w,w} \neq 1 \\ 1 & \text{if } w = z;\ Q'_{w,w} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Here, we should understand G' as a reverse-chronological version of G taken with respect to a specified stationary distribution. Specifically, consider a decomposition $G = \theta(Q - I)$ of an irreducible, positive recurrent generator G with stationary distribution μ, the context of Corollary 3.2.1. If A is a random variable with distribution μ and B is a random variable such that the conditional law of B given A is Q, then $Q' = I + G'/\theta$ is the conditional law of A given B, where $G'_{ij} = (\mu_j/\mu_i)G_{ji}$ as in (3.7).

To further connect the work of Theorem 3.2 with the work of Theorem 2.5, we then seek a self-similar RAM $\mathbf{P}$ such that when clumped as in Theorem 2.5 according to a stationary Markov chain with kernel Q', the resulting Markov Chain conditional RAM is the one described in Theorem 3.2. That RAM, as we will see, is is the GEM distribution.

Recall the definitions of GEM and its variant the disordered GEM distribution on page 34. We now formally name the Markov Chain conditional GEM (MCcGEM) structure of Theorem 3.2:

Definition 3.3 (MCcGEM distribution)**.** Let G be a generator matrix and μ be a distribution on $\mathfrak{X}$. Let $\mathbf{Z}$ be a homogeneous Markov chain with initial distribution μ and transition kernel K_G on $\mathfrak{X}$ given by

$$K_G(w,z) = \begin{cases} \frac{G_{w,z}}{-G_{w,w}} & \text{if } w \neq z;\ G_{w,w} \neq 0 \\ 1 & \text{if } w = z;\ G_{w,w} = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Consider variables $\tilde{\mathbf{X}} = (\tilde{X}_j)_{j \geq 1}$, on the same probability space as $\mathbf{Z}$, such that $\tilde{\mathbf{X}} \big| (\mathbf{Z} = \mathbf{z})$ is an independent sequence and $\tilde{X}_j \big| (\mathbf{Z} = \mathbf{z}) \sim \text{Beta}(1, -G_{z_j,z_j})$. De-

fine $\mathbf{R}$ where $R_j = \tilde{X}_j \prod_{i=1}^{j-1} \left(1 - \tilde{X}_i\right)$ for $j \geq 1$. We say that the pair $(\mathbf{R}, \mathbf{Z})$ has MCcGEM(G) distribution with respect to μ.

We now confirm that by clumping a GEM distribution according to a stationary homogeneous Markov chain, we obtain a MCcGEM distribution:

Theorem 3.4 (GEM to MCcGEM)**.** *Let $\theta > 0$ and $\mathbf{P}$ have* GEM(θ) *distribution. Let also $\mathbf{T} = (T_j)_{j \geq 1}$ be an independent homogeneous Markov chain with kernel Q and initial distribution μ. Recall the associated switch times $\mathbf{V}$ and the Markov chain $\mathbf{Y}$ as defined in* (2.3) *and* (2.5)*, as well as the clumped distribution $\mathbf{P}^{\mathbf{V}}$.*

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

Here, it is worth revisiting the nature of the decomposition of a generator matrix into a pair (Q, θ) of a stochastic matrix and positive constant. There is a degree of freedom in choosing such a decomposition and, by convention, we consider θ to be the free parameter. Specifically, for all $\theta \geq \sup_{i \in \mathfrak{X}} |G_{i,i}|$, the matrix $Q = I + G/\theta$ is stochastic and so (Q, θ) constitutes a valid decomposition of G (see Section 4.1.2). This gives a family of distributions on a pair $(\mathbf{P}, \mathbf{T})$ of GEM distribution and homogeneous Markov chain that can be associated to a single MCcGEM(G) distribution and, consequently, to a single inhomogeneous Markov chain $\mathbf{M}$. Specifically, the following proposition restates Theorem 3.2 to identify the limit of the local occupations as MCcGEM(G'):

Proposition 3.5. *Let G be a generator matrix on $\mathfrak{X}$ without zero rows and having at least one positive recurrent state, and let π, μ, $\mathbf{M}$, $(\mathbf{P}_n, \mathbf{Y}_n)$, and G' be as in Theorem 3.2. Then the local occupations $(\mathbf{P}_n, \mathbf{Y}_n)$ of $\mathbf{M}$ converge weakly in distribution to a pair $(\mathbf{R}, \mathbf{Z})$ having* MCcGEM(G') *distribution with respect to μ.*

Before progressing to (global) occupation laws from the local occupations of this section, we extend Example 2.6 in light of the connection between clumped stick-breaking measures and local occupations of $\mathbf{M}$ establish by Theorems 3.2 and 3.4:

Example 3.6. Suppose G has constant diagonal entries g. If $(\mathbf{R}, \mathbf{Z})$ is distributed as MCcGEM(G'), then $\mathbf{R}$ and $\mathbf{Z}$ are independent. In the context of the inhomogeneous Markov chain $\mathbf{M}$, we then have $\mathbf{P}_n$ and $\mathbf{Y}_n$ are asymptotically independent and $\mathbf{P}_n$ will converge to a GEM($-g$) distribution.

3.2 Global Occupation

We now characterize the limit occupation law ν of the empirical measures ν^n 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 induced topology on $\Delta_{\mathfrak{X}} \subset [0,1]^{\mathfrak{X}}$, the space of probability measures on $\mathfrak{X}$, where $[0,1]^{\mathfrak{X}}$ is given the product topology.

Theorem 3.7 (Occupation laws to MCcGEM and stick-breaking measures)**.** *Consider the setting of Theorem 3.2. Observe that μ is a stationary distribution of $Q' = I + G'/\theta$, and let $\mathbf{T}$ be the homogeneous and stationary Markov chain with kernel Q' and initial distribution μ. Let $\mathbf{P}$ have* GEM(θ) *distribution independent of $\mathbf{T}$.*

Then, $\nu^n = \sum_{j=1}^n \frac{1}{n}\delta_{M_j} \xrightarrow{d} \nu$, where

$$\nu \overset{d}{=} \sum_{j=1}^{\infty} R_j \delta_{Z_j} \overset{d}{=} \sum_{j=1}^{\infty} P_j \delta_{T_j}. \tag{3.8}$$

We may also reverse the connection. 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, almost a corollary of Theorem 3.7.

Theorem 3.8 (Stick-breaking measures to Occupation laws)**.** *Let $\theta > 0$ and* $\mathbf{P}$ *have* GEM(θ) *distribution. Let Q be a stochastic matrix without absorbing states and with stationary distribution μ. Suppose* $\mathbf{T}$ *is a homogeneous Markov chain independent of* $\mathbf{P}$ *with kernel Q starting from μ. Let* $\mathbf{M}$ *be an inhomogeneous Markov chain, as in the setting of Theorem 3.2, with respect to generator matrix G' and starting from any distribution π satisfying $\pi^t (Q')^n \to \mu^t$ entry-wise.*

Then,

$$\sum_{j=1}^{\infty} P_j \delta_{T_j} \stackrel{d}{=} \nu,$$

where $\nu \stackrel{d}{=} \lim_{n\to\infty} \nu^n$ is the occupation law defined with respect to $\mathbf{M}$*.*

Here, we associate to (Q, θ, μ) the generator matrices G and G' and the stochastic matrix Q', where $G = \theta(Q - I)$, G' is given by $G'_{i,j} = (\mu_j/\mu_i)\, G_{j,i} \mathbb{1}(\mu_i \neq 0)$, and $Q' = I + G'/\theta$.

We discuss some technical remarks regarding zero rows of G and absorbing states before discussing the implications of Theorem 3.7

Although we specified that G has no zero rows in Theorems 3.2 and 3.7, i.e. $\mathbf{M}$ has no absorbing states, one can extend some of the statements trivially to the case when there are absorbing states. In particular, when $\mu = \delta_z$ where z is an absorbing state of Q, we have $G'_{z,z} = G_{z,z} = 0$ and $K_n(z, z) = K_{G'}(z, z) = 1$. Then, the state z is also an absorbing state for the inhomogeneous Markov chain $\mathbf{M}$, reached in finite time a.s. starting from π. Also, the chain $\mathbf{T}$, starting from μ, is the constant sequence of z's. In addition, $P_{n,1}$ tends to 1 a.s.

We conclude that $\mathbf{P}_n$ converges weakly to $\mathbf{R} = (1, 0, \ldots)$, a GEM with constant proportions 1=Beta$(1, 0)$, and $\mathbf{Y}_n$ converges weakly to the constant sequence $\mathbf{Z} = (z, z, z, ...)$. Moreover, the empirical distribution ν^n of the chain $\mathbf{M}$ converges weakly to δ_z. We also observe that $\sum_{j\geq 1} R_j \delta_{Z_j}$ and $\sum_{j\geq 1} P_j \delta_{T_j}$ both equal δ_z in distribution, i.e. almost surely.

Then, the full case when G may or may not have zero rows may be obtained as a weighted average of the Dirac measures on absorbing states and the distribution obtained in Theorem 3.7, with weights according to μ.

Whether or not G has zero rows, zero rows may be introduced in G' when reversing time. Any null recurrent or transient state of the Markov chain run with Q corresponds to a zero row of G', i.e. an absorbing state for the chains $\mathbf{T}$ and $\mathbf{Z}$. However, such absorbing states are never visited by $\mathbf{T}$ or $\mathbf{Z}$. The initial distribution μ is a stationary distribution of Q, which vanishes on these states. Moreover, as μ is also a stationary distribution of Q', the chain $\mathbf{T}$ can only move on the positive recurrent states of Q to which μ assigns positive weight, the states $\{i \in \mathfrak{X} : \mu_i > 0\}$.

Similarly, starting from μ, the chain $\mathbf{Z}$ moves only on states $\{i \in \mathfrak{X} : \mu_i > 0\}$, given that $G'_{w,z} = K_{G'}(w, z) = 0$ when $w \neq z$ and either $\mu_z = 0$ or $\mu_w = 0$.

Finally, we can also consider the implications of Theorem 2.13 in this context of occupation laws and the MCcGEM distribution, and give a self-similarity equation that uniquely characterizes the joint distribution of (ν, T_1). Let $\mathbf{P}$ and $\mathbf{T}$ be as in Theorem 3.4. Define the associated stick-breaking measure

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j}.$$

Let $\tilde{\mathbf{X}}$ be an iid sequence of Beta$(1, \theta)$ variables and for each $x \in \mathfrak{X}$, let $\mathbf{T}(x)$ be a homogeneous Markov chain with transition kernel Q and $T_1(x) = x$. For each $x \in \mathfrak{X}$, define $W(x) = \inf\{j > 1 : T_j(x) = x\}$ and

$$\tilde{X}(x) = \sum_{j=1}^{W(x)-1} \tilde{X}_j \prod_{i=1}^{j-1} (1 - \tilde{X}_i)$$

$$\eta(x) = \begin{cases} \tilde{X}(x)^{-1} \displaystyle\sum_{j=1}^{W(x)-1} \delta_{T_j(x)} \tilde{X}_j \prod_{i=1}^{j-1} (1 - \tilde{X}_i) & : \ \tilde{X}(x) > 0 \\ \delta_x & : \ \tilde{X}(x) = 0. \end{cases}$$

Theorem 3.9 (Self-similarity for GEM)**.** *The joint law of* (ν, T_1) *is uniquely characterized by* μ *and the joint distributions of* $\{(\eta(x), \tilde{X}(x))\}_{x\in\mathfrak{X}}$ *in the following sense:*

1. $T_1 \sim \mu$
2. *For each* $x \in \mathfrak{X}$, $\nu\big|T_1 = x$ *satisfies a self-similarity equation with respect to* $(\eta(x), \tilde{X}(x))$.

3.2.1 On Nonexchangeability

Let $\pi : \mathbb{N} \to \mathbb{N}$ be a permutation of $\mathbb{N}$, that is π is one-to-one and onto, and let $\mathbf{P}$ have GEM(θ) distribution. Suppose that $\mathbf{T}$ is a positive-recurrent, irreducible Markov chain, starting from stationary distribution μ.

Definition 3.10. We say that the stick-breaking measure $\nu = \sum_{i\geq 1} P_i\delta_{T_i}$ is permutation exchangeable when, for each permutation π fixing all but finitely many integers,

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

Theorem 3.11. *The measure* ν *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 parameter measure* $\theta\mu$.

A proof of this will be included in a future iteration of [12]. By computing and matching second moments of the measure ν and its permuted measure $\pi\nu$ for arbitrary transpositions π, we can deduce that the stochastic kernel of $\mathbf{T}$ is constant stochastic when ν is exchangeable, and so $\mathbf{T}$ is iid.

3.2.2 New Constructions of the Dirichlet Process

Let $\mathfrak{X}$ be a countable space. In light of Theorems 3.7 and 3.8, we now describe a family of stick-breaking constructions for Dirichlet processes on $\mathfrak{X}$.

Let α be a finite measure on $\mathfrak{X}$ such that for each $x \in \mathfrak{X}$, $\alpha(x) > 0$. It is well-known that a Dirichlet process with parameter measure α may be constructed on $\mathfrak{X}$ through stick-breaking as follows. Define $\theta = \alpha(\mathfrak{X})$ and $\mu = \theta^{-1}\alpha$. Let $\mathbf{P}$ have GEM(θ) distribution, and let $\mathbf{T}$ be an iid sequence independent of $\mathbf{P}$ with common measure μ. Then

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j}$$

has DP(α) distribution. This construction can be equivalently understood as a paired GEM distribution $\mathbf{P}$ and stationary homogeneous Markov chain $\mathbf{T}$ independent of $\mathbf{P}$ having transition kernel Q, where Q is the constant stochastic matrix with rows μ. This associates to DP(α) a generator matrix G where

$$G_{i,j} = \theta(Q - I)_{i,j} = \alpha(\{j\}) - \alpha(\mathfrak{X})\mathbb{1}(i = j).$$

Note here that if G' is the generator matrix given by $G'_{ij} = (\mu_j/\mu_i)G_{ji}$, then $G' = G$. Furthermore, the matrix $\tilde{Q} = I + G/\tilde{\theta}$ is stochastic if and only if we have $\tilde{\theta} \geq \sup_{x\in\mathfrak{X}} \theta(1 - \mu_x) = \sup_{x\in\mathfrak{X}}[\alpha(\mathfrak{X}) - \alpha(\{x\})]$.

For each $\tilde{\theta} \geq \sup_{x\in\mathfrak{X}} \theta(1 - \mu_x)$, let $\mathbf{P}^{\tilde{\theta}}$ have GEM($\tilde{\theta}$) distribution and let $\mathbf{T}^{\tilde{\theta}}$ be a stationary homogeneous Markov chain independent of $\mathbf{P}^{\tilde{\theta}}$ with transition kernel $\tilde{Q} = I + G/\tilde{\theta}$. Then the stick-breaking construction

$$\nu^{\tilde{\theta}} = \sum_{j=1} P_j^{\tilde{\theta}} \delta_{T_j^{\tilde{\theta}}}$$

also has DP(α) distribution by Theorem 3.7. Furthermore, if (R, Z) has MCcGEM(G) distribution, then the stick-breaking construction

$$\tilde{\nu} = \sum_{j=1} R_j \delta_{Z_j}$$

also has DP(α) distribution by Theorem 3.7.

These constructions of the Dirichlet process appear new. Furthermore, in light of Section 3.2.1, we see the Dirichlet process has both exchangeable and nonexchangeable stick-breaking representations.

3.3 Proofs

3.3.1 Proofs for Section 3.1

We first specify certain asymptotics which aid in the proof of Theorem 3.2.

Lemma 3.12. *For $\gamma > 0$ and integers $1 \le m \le n$, let*

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

Then, for $0 < a < b$ and integers $M \ge 0$, we have

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

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 3.13. *Let $r \ge 1$ 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 $1 \le j \le r$. Then,*

$$\lim_{s_0\to\infty} \sum_{s_1=\lceil a_1 s_0 \rceil}^{\lceil b_1 s_0 \rceil - 1} \cdots \sum_{s_r=\lceil a_r s_{r-1} \rceil}^{\lceil b_r s_{r-1} \rceil - 1} \left[\prod_{j=1}^{r} {s_j}^{-1} f_{s_j}^{\lfloor b_j s_{j-1} \rfloor - 1}(\gamma_j) \right] = \prod_{j=1}^{r} \gamma_j^{-1} \Big(1 - \Big(\frac{a_j}{b_j}\Big)^{\gamma_j}\Big).$$

Proof. The argument follows by inputting the asymptotics in Lemma 3.12. 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

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

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

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

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 compute

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

Then, noting (3.9), the proposition follows for $r = 1$. ■

We now show a form of 'weak ergodicity' for the chain $\mathbf{T}$ of Theorem 3.2.

Lemma 3.14. *Let $\mathfrak{X} \subseteq \mathbb{N}$. For a generator matrix G with indices in $\mathfrak{X}$, define $\theta^G = \inf\{r \in \mathbb{R}^+ : I + G/r \text{ is non-negative}\}$, and let $c, \theta \geq \theta^G$ such that $c \in \mathbb{N}$. Define $Q = I + G/\theta$. Recall that $K_n = I + \frac{G}{n}\mathbb{1}(n > c)$ for $n \geq 1$. Let π be a stochastic vector and μ be a stationary distribution for Q such that $\pi^t Q^n \to \mu^t$ entry-wise. Then, as $n \to \infty$:*

(a) $\mu^n := \pi^t \prod_{i=1}^n K_i \to \mu^t$ entrywise, and

(b) $\left(\mu^n\right)^t Q \to \mu^t$ entry-wise.

Proof. We separate into three steps.

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

matrix:

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

as a polynomial in Q with positive coefficients. 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 3.12, $[Q^0]_n = f_m^n(\theta) \to 0$ as $n \to \infty$. Also, as $f_m^n(\theta) \sim n^{-\theta}$ by Lemma 3.12, 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 2. For each $x \in \mathfrak{X}$, let e_x denote the vector in $\mathbb{R}^{\mathfrak{X}}$ with a 1 in the entry corresponding to state x and 0's elsewhere. Since Q is a stochastic kernel, observe that for each $x \in \mathfrak{X}$ and $n \geq m$:

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

Also, as μ is a stationary eigenvector of Q, note that μ is also a stationary eigenvector of K_n for each $n \geq 1$. Recall that $(\pi - \mu)^t Q^n \to 0$ as $n \to \infty$ entry-wise, and $\mu^m = \pi^t \prod_{i=1}^{m} K_i$. Hence, $(\mu^m - \mu)^t Q^n \to 0$ as $\prod_{i=1}^{m} K_i$ is a polynomial in Q.

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

we may bound:

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

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

Step 3. Finally, by Fatou's lemma, the proved first limit (a), and that μ is a stationary vector of Q, we have for each $j \in \mathfrak{X}$ that

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

Now, suppose for a particular $k \in \mathfrak{X}$ 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

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

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

Proof of Theorem 3.2.

We will argue in steps.

Step 1. Recall the definition of kernel G' (cf. (3.5)). 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 $G'_{i,j} = (\mu_j/\mu_i) G_{j,i} 1(\mu_i \neq$

$0) \geq 0$ for $i \neq j$, and $\sum_j G'_{i,j} = \frac{1(\mu_i \neq 0)}{\mu_i} \sum_j \mu_j G_{j,i} = 0$. Moreover,

$$\sup_i |G'_{i,i}| \;=\; \sup_i |G_{i,i} 1(\mu_i \neq 0)| \;\leq\; \sup_i |G_{i,i}| \;<\; \infty.$$

Thus, G' is a generator matrix.

Step 2. Recall the Markov chain $\mathbf{M}$, with transition kernels $\{K_n\}_{n\geq 1}$ (cf. (3.1)), starting from π. Recall the associated variable N_n and sequence $\mathbf{P}_n$.

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

$$X_{n,j} = P_{n,j} / \left(1 - \sum_{i=1}^{j-1} P_{n,i}\right) \qquad \text{and} \qquad X_{n,i} = 1. \tag{3.10}$$

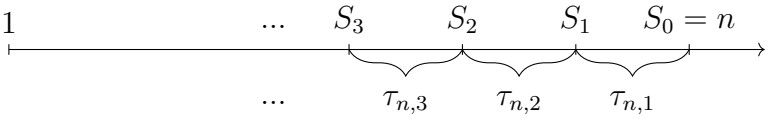
The variables $\mathbf{X}_n = (X_{n,i})_{i\geq 1}$ are the associated proportions to the distribution $\mathbf{P}_n$ on $\mathbb{N}$ and, by Lemma 1.19, for $j \geq 1$,

$$P_{n,j} = X_{n,j} \prod_{i=1}^{j-1} (1 - X_{n,i}) \;\text{ and }\; 1 - \sum_{i=1}^{j-1} P_{n,j} = \prod_{i=1}^{j-1} (1 - X_{n,i})\,. \tag{3.11}$$

For $j \geq 0$, also define

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

In terms of the switching times $\tilde{\mathbf{V}}$, and the first time N_n that the chain $\mathbf{M}$ switches after time n, we have $S_0 = n$, $S_j = \tilde{V}_{N_n - j} - 1$ 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 $\mathbf{M}$ switches states when considered in reverse order, and $\{\tau_{n,j}\}$ are the lengths of the associated local occupations in the depiction below.



Step 3. Recall the sequence $\mathbf{Y}_n$ where $Y_{n,j} = M_{\tilde{V}_{N_n - j}}$ for $1 \leq j \leq N_n - 1$ and $Y_{n,i} = M_1$ for $i \geq N_n$. 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 $r \geq 1$, 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 $1 \leq j \leq r$, are all integers. Set also $s_0 = n$ and recall $S_0 = n$.

Note from (3.11) and (3.12) that

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

Then, with respect to a possible sequence y, we have

$$\begin{aligned} &\mathscr{P}\left(X_{n,j} = \beta_j, Y_{n,j} = y_j : 1 \leq j \leq r\right) \\ =&\mathscr{P}\left(\tau_{n,j} = s_{j-1} - s_j,\ Y_{n,j} = y_j : 1 \leq j \leq r\right) \\ =&\sum_{\substack{y_{r+1} \in \mathfrak{X} \\ y_{r+1} \neq y_r}} \mathscr{P}(M_{s_r} = y_{r+1}) \prod_{j=1}^{r} \mathscr{P}\left(M_{s_{j-1}} = \cdots = M_{s_j+1} = y_j \middle| M_{s_j} = y_{j+1}\right). \end{aligned} \tag{3.13}$$

Note the computation for $c \leq l < n$ and $z \neq y$,

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

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

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

Then, (3.13) equals

$$\sum_{\substack{y_{r+1}\in\mathfrak{X}\\ y_{r+1}\neq y_r}} \mu_{y_{r+1}}^{s_r} \prod_{j=1}^{r} \frac{G_{y_{j+1},y_j}}{s_j} f_{s_j}^{s_{j-1}-1}\left(-G_{y_j,y_j}\right)$$

$$= \left[(\mu^{s_r})^t Q e_{y_r} - \mu_{y_r}^{s_r} Q_{y_r,y_r}\right] \frac{\theta}{s_r} f_{s_r}^{s_{r-1}-1}\left(-G_{y_r,y_r}\right) \prod_{j=1}^{r-1} \frac{G_{y_{j+1},y_j}}{s_j} f_{s_j}^{s_{j-1}-1}\left(-G_{y_j,y_j}\right) \quad (3.14)$$

Step 4. We now sum (3.14) over all appropriate values of $\{s_j\}_{j=1}^r$ such that $0 < X_{n,j} \leq \beta_j$ for $1 \leq j \leq r < N_n$, where we recall N_n is the time the chain switches after time n. Then, we have from (3.12) that

$$1 \leq \tau_{n,j} = nP_{n,j} = S_{j-1} - S_j = X_{n,j} S_{j-1}. \quad (3.15)$$

Moreover, we have from (3.12) that $s_r \geq n\prod_{j=1}^r (1-\beta_j)$ diverges to infinity as $n \to \infty$.

Recall $s_0 = n$ and $\lim_{n\to\infty} N_n = \infty$ a.s. Then, with equation (3.15) in hand,

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

Step 5. From (3.12), the sum index $s_r \geq n\prod_{j=1}^r (1-\beta_j)$ diverges to infinity as $n \to \infty$. Also, by Lemma 3.14, we have $\lim_{s\to\infty} \mu_y^s = \mu_y$ and $\lim_{s\to\infty} (\mu^s)^t Q e_y = \mu_y$ for each $y \in \mathfrak{X}$. Therefore, as $n \to \infty$, we have

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

Note that $-G_{i,i} > 0$ for each $i \in \mathfrak{X}$ since by assumption G has no zero rows. Thus,

by Proposition 3.13, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathscr{P}\left(0 < X_{n,j} \le \beta_j,\ Y_{n,j} = y_j : 1 \le j \le r\right) \\
=&\mu_{y_r}(-G_{y_r,y_r}) \prod_{j=1}^{r-1} G_{y_{j+1},y_j} \\
&\qquad \times \lim_{n\to\infty} \sum_{s_1=\lceil s_0(1-\beta_1)\rceil}^{s_0-1} \cdots \sum_{s_r=\lceil s_{r-1}(1-\beta_r)\rceil}^{s_{r-1}-1} \prod_{j=1}^{r} {s_j}^{-1} f_{s_j}^{s_{j-1}-1}\left(-G_{y_j,y_j}\right) \\
=&\mu_{y_r}(-G_{y_r,y_r}) \left[\prod_{j=1}^{r-1} G_{y_{j+1},y_j}\right] \left[\prod_{j=1}^{r} \left(-G_{y_j,y_j}\right)^{-1} \left(1-(1-\beta_j)^{-G_{y_j,y_j}}\right)\right]. \qquad (3.16)
\end{aligned}$$

Hence, if $\mu_{y_k} = 0$ for some $1 \le k \le r$, the limit (3.16) vanishes by bounding

$$\mathscr{P}(0 < X_{n,j} \le \beta_j, Y_{n,j} = y_j : 1 \le j \le r) \le \mathscr{P}(0 < X_{n,j} \le \beta_j, Y_{n,j} = y_j : 1 \le j \le k).$$

Now, suppose that $\{y_j\}_{j=1}^r$ is such that $\mu_{y_j} > 0$ for each $1 \le j \le r$. We may write the limit (3.16) as

$$\begin{aligned}
&\mu_{y_1} \left[\prod_{j=1}^{r-1} \frac{\mu_{y_{j+1}}}{\mu_{y_j}} \frac{G_{y_{j+1},y_j}}{-G_{y_j,y_j}}\right] \left[\prod_{j=1}^{r} \left(1-(1-\beta_j)^{-G_{y_j,y_j}}\right)\right] \\
=&\mu_{y_1} \left[\prod_{j=1}^{r-1} \frac{G'_{y_j,y_{j+1}}}{-G'_{y_j,y_j}}\right] \left[\prod_{j=1}^{r} \left(1-(1-\beta_j)^{-G'_{y_j,y_j}}\right)\right],
\end{aligned}$$

decomposed as a product of i) the transition probability of the chain $\mathbf{M}$, with kernel $K_{G'}$ (cf. (3.6)) 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. ■

Proof of Corollary 3.2.1. We show that the conditions of Theorem 3.2 are satisfied for every choice of θ and π:

Recall that G is said to be irreducible and positive recurrent if for a decomposition $G = \theta(Q - I)$, the stochastic matrix Q is irreducible and positive recurrent. First, note

that this definition does not depend on Q: if (Q, θ) and $(\tilde{Q}, \tilde{\theta})$ are both decompositions of G, then

$$\tilde{Q} = \left(1 - \frac{\theta}{\tilde{\theta}}\right) I + \frac{\theta}{\tilde{\theta}} Q$$

and so Q and $\tilde{Q}$ have identical class structure and share stationary distributions.

Second, if G has a zero row, then $Q = I + G/\theta$ has a row with all zeros except a 1 in the diagonal entry. Thus, Q has an absorbing state and isn't irreducible. Since we assume G is irreducible, G has no zero rows. The simplification of $K_{G'}$ follows.

Third, if G is irreducible and positive recurrent, it has a unique stationary distribution μ such that for every $i \in \mathfrak{X}$, $\mu_i > 0$. The simplification of G' follows.

Finally, for every distribution π on $\mathfrak{X}$ and every decomposition (Q, θ) of G, $\pi^t Q^n \to \mu^t$ as Q is irreducible and positive recurrent. ■

Proof of Theorem 3.4. In Part (1) of Theorem 2.5, we showed that the associated sequence $\mathbf{Y}$ is a Markov chain with transition kernel K on $\mathfrak{X}$ such that

$$K(z, w) = \begin{cases} \frac{Q_{z,w}}{1 - Q_{z,z}} & \text{if } z \neq w,\ Q_{z,z} \neq 1 \\ 1 & \text{if } z = w,\ Q_{z,z} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

By inspection, the kernel $K = K_G$ where $G = \theta(Q - I)$.

Recall now the switch times $\mathbf{V}$ with respect to the chain $\mathbf{T}$ (cf. (2.3)). In Part (2) of Theorem 2.5, as $\mathbf{P}$ is a self-similar RAM, we proved that $\mathbf{P}^{\mathbf{V}}$ is conditionally a RAM given $\mathbf{Y}$. In particular, we showed that the associated proportions $\mathbf{X}^{\mathbf{V}} = (X_j^V)_{j \geq 1}$ are conditionally independent variables given $\mathbf{Y}$. Hence, to identify the joint distribution of $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y})$, we need only find the conditional distribution of each proportion $X_j^V \big| \mathbf{Y}$, for $j \geq 1$.

To this end, recall (2.11) and (2.13), in which for $j < J$ and $a \in (0, 1)$ we determine

$$\mathscr{P}\left(1 - X_j^V \leq a | Y = y\right) = \sum_{l=1}^{\infty} \mathscr{P}\left(\prod_{k=1}^{l} (1 - X_k) \leq a\right) Q_{y_j, y_j}^{l-1} (1 - Q_{y_j, y_j}).$$

This is equivalent to the probability that $\prod_{j=1}^{N} (1 - X_j) \leq a$ where N has distribution Geometric$(1 - Q_{y_j, y_j})$. The variable $\prod_{j=1}^{N} (1 - X_j) \leq a$ is a Beta$(\theta(1 - Q_{y_j, y_j}), 1)$

random variable (see Lemma A.6), and so $X_j^V|\mathbf{Y} = \mathbf{y} \sim \text{Beta}(1, \theta(1 - Q_{y_j,y_j}))$ for $j < J$.

When $J \leq j < \infty$, recall that y_j is an absorbing state, and so $Q_{y_j,y_j} = 1$ and $X_j^V = 1$. Thus $X_j^V|(\mathbf{Y} = \mathbf{y}) \sim \text{Beta}(1, 0) = \text{Beta}(1, \theta(1 - Q_{y_j,y_j}))$.

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

Proof of Proposition 3.5. This proposition is a restatement of Theorem 3.2 in light of the newly introduced definition of MCcGEM. Refer to the proof of Theorem 3.2 earlier in this proofs section. ■

3.3.2 Proofs for Section 3.2

In the setting of Theorems 3.2 and 3.7, consider the pairs $\{(\mathbf{P}_n, \mathbf{Y}_n)\}_{n\geq 1}$, $(\mathbf{R}, \mathbf{Z})$, and $(\mathbf{P}, \mathbf{T})$. These objects belong to $[0,1]^{\mathbb{N}} \times \mathfrak{X}^{\mathbb{N}}$. We now discuss the topology on this space and its relatives, before going to the proofs of the theorems.

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 $[0,1]^{\mathbb{N}} \times \mathfrak{X}^{\mathbb{N}}$ defined as follows:
For $(p, y), (p', y') \in [0,1]^{\mathbb{N}} \times \mathfrak{X}^{\mathbb{N}}$,

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

The corresponding σ-field on $[0,1]^{\mathbb{N}} \times \mathfrak{X}^{\mathbb{N}}$, generated by ρ, is the usual product σ-field formed from the Borel σ-fields on copies of $[0,1]$ and $\mathfrak{X}$. Importantly, weak

convergence of probability measures on $[0,1]^{\mathbb{N}} \times \mathfrak{X}^{\mathbb{N}}$ translates to finite dimensional convergence of these laws. Moreover, $([0,1]^{\mathbb{N}} \times \mathfrak{X}^{\mathbb{N}}, \rho)$ 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

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

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

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

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

is measurable, and also continuous on the subset $\Delta_\infty \times \mathfrak{X}^{\mathbb{N}}$.

Now, denote the collection of probabilities on $\mathfrak{X}$,

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

and endow it with the metric $\rho^2(p,p') = \sum_{n \geq 1} 2^{-n} |p_n - p'_n|$, and the associated Borel σ-field. Define $g : \Delta_\infty \times \mathfrak{X}^{\mathbb{N}} \to \Delta_{\mathfrak{X}}$ by

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

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

$$\sum_{j \geq A} p_j^n \mathbb{1}_l(y_j^n) \leq \sum_{j \geq A} p_j^n = 1 - \sum_{j < A} p_j^n \xrightarrow{n \to \infty} 1 - \sum_{j < A} p_j.$$

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) $\sum_{j\geq A} p_j \xrightarrow{A\to\infty} 1$.

Proof of Theorem 3.7. 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 $\Delta_\infty \times \mathfrak{X}^{\mathbb{N}}$. Clearly, $\{(\mathbf{P}_n, \mathbf{Y}_n)\}_{n\geq 1}$ surely lives in $\Delta_\infty \times \mathfrak{X}^{\mathbb{N}}$ by construction. Also, $(\mathbf{R}, \mathbf{Z})$ and $(\mathbf{P}, \mathbf{T})$ lie almost surely in $\Delta_\infty \times \mathfrak{X}^{\mathbb{N}}$ since we have that $\mathbf{P}$ and $\mathbf{R}$ are, respectively, a RAM and conditionally a RAM. Thus $\sum_{j=1}^\infty R_j \stackrel{d}{=} \sum_{j=1}^\infty P_j \stackrel{a.s.}{=} 1$.

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 3.2, we have $\nu_n = g((\mathbf{P}_n, \mathbf{Y}_n)) = g \circ f((\mathbf{P}_n, \mathbf{Y}_n))$ converges weakly to $\nu = g \circ f((\mathbf{R}, \mathbf{Z}))$ by the continuous mapping theorem, and so the left equality in (3.8) 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 3.4. Recall that $\mathbf{T}$ is a Markov chain with kernel $Q' = I + G'/\theta$ and initial stationary distribution μ. Then, noting that $G' = \theta(Q' - I)$, we have by Theorem 3.4 that $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y})$ has a MCcGEM(G') distribution. Hence, $(\mathbf{P}^{\mathbf{V}}, \mathbf{Y}) \stackrel{d}{=} (\mathbf{R}, \mathbf{Z})$. Since almost surely,

$$g \circ f\left((\mathbf{P}^{\mathbf{V}}, \mathbf{Y})\right) = g \circ f((\mathbf{P}, \mathbf{T})) = \left(\sum_{j\geq 1} P_j \mathbb{1}(T_j = x)\right)_{x\in\mathfrak{X}},$$

we have $g \circ f((\mathbf{R}, \mathbf{Z})) \stackrel{d}{=} g \circ f((\mathbf{P}, \mathbf{T}))$, and the right equality of (3.8) holds. ■

Proof of Theorem 3.8. The claim follows from Theorem 3.7 once we verify that a homogeneous Markov chain with kernel Q and a homogeneous Markov chain with kernel $(Q')' = I + (\tilde{G}')'/\theta$, each with initial distribution μ, are equivalent in distribution.

To this end, for any generator matrix $G = \theta(Q - I)$ and associated stationary

distribution μ, we observe that $(G')'_{ij} = G_{ij}$ when μ_i and μ_j are both positive:

$$\begin{aligned}(G')'_{ij} &= \frac{\mu_j}{\mu_i} G'_{ji} \mathbb{1}(\mu_i \neq 0) \\ &= \frac{\mu_j}{\mu_i} \frac{\mu_i}{\mu_j} G_{ij} \mathbb{1}(\mu_i \neq 0) \mathbb{1}(\mu_j \neq 0) \\ &= G_{ij} \mathbb{1}(\mu_i \neq 0) \mathbb{1}(\mu_j \neq 0).\end{aligned}$$

Since $Q = I + G/\theta$ and $(Q')' = I + (G')'/\theta$, we conclude that $Q_{ij} = (Q')'_{ij}$ when μ_i and μ_j are both positive.

Finally, as μ is a stationary distribution, μ is only positive on positive recurrent states and for each recurrence class of Q, μ either assigns 0 weight to each state in that class or strictly positive weights to each state in that class. Hence, homogeneous Markov chains with kernels Q and $(Q')'$, starting from μ, are equal in distribution. ■

Proof of Theorem 3.9. The theorem follows immediately from Theorem 2.13. ■

Chapter 4
Future Directions

4.1 Bayesian Statistics

The Dirichlet process is an example of a Bayesian prior. Abstractly, this means that a Dirichlet process is a random variable ν taking value almost surely in a space of probability measures over a measurable space $\mathfrak{X}$. In the context of statistics, we may view the observation of such a random variable ν as giving the underlying distribution of a simple random sample $\boldsymbol{X} = (X_1, ..., X_n)$ we collect, specifically with $\boldsymbol{X}$ being a collection of n many random variables taking values in $\mathfrak{X}$ satisfying

$$\mathscr{P}(X_j \in A_j : 1 \leq j \leq n | \nu) \; = \; \prod_{j=1}^{n} \nu(A_j)$$

for each collection of measurable subsets $A_j \subseteq \mathfrak{X}$. The prior distribution in this context then refers to the (marginal) distribution of ν, and the posterior distribution refers to the conditional distribution of ν given the data, i.e. the conditional law of ν given $\boldsymbol{X}$.

The Dirichlet process in particular is an important example of a Bayesian prior. A constructive example of a Dirichlet process may be given on any measurable space $\mathfrak{X}$, as shown by J. Sethuraman in [60]. Specifically, if $\mathbf{T}$ is an iid sequence of random variables taking value in $\mathfrak{X}$ with distribution μ and $\mathbf{P}$ has GEM(θ) distribution and is independent of $\mathbf{T}$, then

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j}$$

is a Dirichlet process on $\mathfrak{X}$ with parameter measure $\alpha = \theta\mu$. Furthermore, in the context above where $\boldsymbol{X}|\nu$ is a simple random sample with common distribution ν, the posterior distribution of ν may be succinctly described as another Dirichlet process, this time with updated parameter measure $\tilde{\alpha} = \alpha + \sum_{j=1}^{n} \delta_{X_j}$.

In Bayesian nonparametric statistics, the space of probability distributions a prior takes values in is taken to be infinite dimensional. The Dirichlet process is then an attractive building block for Bayesian nonparametrics as a result of the above properties and properties relevant to computational efficiency.

One goal of this work is to begin laying the foundation for use of our generalization of a Dirichlet process in the context of Bayesian statistics. Doing so necessitates the computation of or the ability to efficiently simulate a posterior distribution. Doing so well involves generalizing some of the work of Chapter 3 from countable state spaces to measurable ones, and clearly identifying what flexibility or advantages our generalization of Dirichlet processes provides over the Dirichlet process.

In the rest of this section, we discuss some preliminary work for such a foundation, upon which we hope to build in future.

4.1.1 Recasting Moments

Recall the generalization of the Dirichlet process discussed in Chapter 3:

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j} \tag{4.1}$$

where $\mathbf{P}$ and $\mathbf{T}$ are independent random sequences with $\mathbf{P} \sim \text{GEM}(\theta)$ and $\mathbf{T}$ a stationary Markov chain on state space $\mathfrak{X}$ with transition matrix Q and stationary measure μ.

In this section, we discuss and expand upon some previous work by Dietz and Sethuraman in [13] regarding the moments of the measure ν, where we understand ν as a construction of the occupation law of the inhomogeneous Markov chain discuss in Chapter 3. In keeping with their context, we restrict our discussion to the following setting: The state space $\mathfrak{X}$ is given by $\{1, 2, ..., k\}$ for some $k \in \mathbb{N}$. All stochastic kernels Q and generator matrices G are $k \times k$ matrices without 0 entries. While this is a significantly restricted selection of generator matrices G, a description of

ν through its moments begins to shed light on what flexibility ν may have over its Dirichlet sub-case. For example, it may allow the assertion of greater or lesser dependence among the measures of certain sets using the more general prior.

In this restricted context, Dietz and Sethuraman [13] showed the following: Let $G = \theta(Q - I)$ be a generator without 0 entries. If we represent the random measure ν of (4.1) by a vector $(\nu_1, ..., \nu_k)$ taking value almost surely in Δ_k, the law of ν is identified through its moments. Let $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ be the whole numbers. For $\mathbf{m} = (m_1, \ldots, m_k) \in \mathbb{N}_0^k$ and $N = \sum_{j=1}^k m_j > 0$, we have

$$\mathbb{E}\left(\prod_{j=1}^k \nu_j^{m_j}\right) = \binom{N}{m_1, ..., m_k}^{-1} \sum_{\sigma \in \mathbb{S}(\mathbf{m})} \mu_{\sigma_1} \prod_{j=1}^{N-1} \left[\left(I - \frac{G}{j}\right)^{-1}\right]_{\sigma_j, \sigma_{j+1}} \tag{4.2}$$

where μ is the unique stochastic eigenvector of G, and $\mathbb{S}(\mathbf{m})$ is the set of $\binom{N}{m_1,...,m_k}$ distinct permutations of the list of N integers consisting of m_1 many 1's, m_2 many 2's, up through m_k many k's.

We now recast the moment result (4.2) in an algebraic form where it can be more easily exploited. Let $p_{min}(\lambda)$ be the minimal polynomial of G and $q(\lambda) = \sum_{k=0}^n a_k \lambda^k$ be the polynomial such that $p_{min}(\lambda) = \lambda q(\lambda)$. Define, for $j \geq 0$,

$$p_j(\lambda) := \frac{p_{min}(\lambda) - p_{min}(j)}{\lambda - j} = \sum_{k=0}^n \lambda^k \sum_{l=k}^n a_l j^{l-k}. \tag{4.3}$$

Proposition 4.1 (Recasting moments)**.** *In the above context, we have $p_0(G)/q(0)$ is the matrix with constant rows μ, and $p_j(G)/q(j) = (I - G/j)^{-1}$ for $j \geq 1$. As a consequence, for $\mathbf{m} \in \mathbb{N}_0^k$ with $\sum_{i=1}^k m_i = N > 0$ and fixed constant $\sigma_0 \in \mathfrak{X}$,*

$$\mathbb{E}\left[\prod_{i=1}^k \nu_i^{m_i}\right] = \binom{N}{m_1, ..., m_k}^{-1} \sum_{\sigma \in \mathbb{S}(\mathbf{m})} \prod_{j=0}^{N-1} \left[\frac{p_j(G)}{q(j)}\right]_{\sigma_j \sigma_{j+1}}. \tag{4.4}$$

This proposition may also be found in [12].

One can now recover the moments of the marginals from the joint moments. We note that these marginals were first identified by Herbach in his work on mRNA

modeling [31]. However, the identification of the connection between Herbach's model and ν established through a piecewise deterministic Markov process as in Bouguet and Cloez [7] was not previously noticed.

Corollary 4.1.1 (Marginals; Herbach [31]). *Let $\{\lambda_l\}_{l=1}^n$ be the non-zero roots of q, i.e. the non-zero eigenvalues of G. Let also $\{\gamma_{i,l}\}_{l=1}^n$ be the zeros of $[p_j(G)]_{ii}$ considered as a polynomial in j. Then,*

$$\mathbb{E}[\nu_i^N] = \prod_{l=1}^{n} \frac{(-\gamma_{i,l})_N}{(-\lambda_l)_N} \tag{4.5}$$

We recognize the above marginal moments as those of a Beta product (see Definition A.2). Furthermore, note that if λ is an eigenvalue of Q and $G = \theta(Q - I)$, then $\theta(\lambda - 1)$ is an eigenvalue of G. We see that the marginal distribution of ν_i may well be a Beta product with complex parameters, and hence not represented by an actual product of independent Beta variables.

Herbach showed that the zeros $\{\gamma_{i,l}\}$ of $[p_j(G)]_{ii}$ may be realized as the eigenvalues of the matrix $G(i)$ obtained by deleting the ith row and column of G. This identifies $[p_j(G)]_{ii}$ as the minimal polynomial of $G(i)$. Further work in this direction may uncover an efficient method for constructing generator matrices whose corresponding occupation laws ν have pre-specified marginal parameters.

4.1.2 Understanding θ as a Free Parameter

In Theorems 3.7 and 3.8, we related a family of stick-breaking constructions built out of a paired GEM distribution and homogeneous Markov chain to a single clumped construction built out of the MCcGEM distribution. We then associated these constructions to the occupation laws initially described by Dietz and Sethuraman in [13]. Either construction of the occupation law involves an infinite sum: an observation ν of the random measure of Theorem 3.8, parametrized by generator matrix G, may be

constructed as

$$\nu = \sum_{j=1}^{\infty} R_j \delta_{Z_j}$$

where $(\mathbf{R}, \mathbf{Z}) \sim \text{MCcGEM}(G)$ or, equivalently as

$$\nu = \sum_{j=1}^{\infty} P_j \delta_{T_j}$$

where $P \sim \text{GEM}(\theta)$, T is a stationary Markov chain independent of P with stochastic kernel Q, and $G = \theta(Q - I)$.

Since each of these sums converge almost surely, one way of simulating a random sample of measures with a distribution approximating that of ν would be to truncate the sum at a sufficiently large index N such that the truncated measure is close to the infinite sum. In the context of Dirichlet measures, typically the remaining probability not yet allocated at the time of truncation is attributed to state T_N to keep the truncated measure a probability measure. This truncated measure is said to be a truncated Dirichlet process, and is useful for simulating posterior distributions for Dirichlet process mixtures; see [34] and [49].

Theorem 3.7 provides two classes of stick-breaking construction we might wish to truncate. We note that while the MCcGEM distribution is parametrized by a generator matrix G and an initial distribution, the paired GEM distribution and Markov chain are parametrized by a decomposition (Q, θ) of $G = \theta(Q - I)$ and an initial distribution. This association is many-to-one, and θ is a natural parameter for the family of GEM-Markov chain pairs associated to a single MCcGEM distribution. To be precise, if we fix G, we can index by θ all pairs (Q, θ) associated to G via:

$$Q(\theta) = I + G/\theta.$$

If we choose to use a GEM-Markov chain pair when simulating the measure ν, we have freedom to choose a value of θ. Our interest is then what role θ plays in the stick-breaking construction. From this point forward, we take G to be a fixed generator matrix.

First, we consider what values of θ correspond with a valid decomposition of G. Clearly, θ must be positive since it acts as a GEM parameter. Additionally, the matrix $Q(\theta)$ must be a stochastic matrix. Since G is a generator matrix, $Q(\theta)$ will have row sums of 1, but we need to ensure that $Q(\theta)$ has non-negative entries:

$$\begin{aligned}
\inf Q(\theta)_{ij} \geq 0 &\iff \inf[I + G/\theta]_{ij} \geq 0 \\
&\iff \inf 1 + G_{ii}/\theta \geq 0 \\
&\iff \theta \geq -\inf G_{ii} \\
&\iff \theta \geq \sup |G_{ii}|
\end{aligned}$$

Recall that $\sup |G_{ii}| < \infty$ by definition. Thus, we can associate to G a minimal value of θ, denoted $\theta^G = \sup |G_{ii}|$, such that when $\theta \geq \theta^G$, we have that $Q(\theta)$ is stochastic and $(Q(\theta), \theta)$ is a valid decomposition of G. A simple computation gives

$$Q(\theta) = \left(1 - \frac{\theta^G}{\theta}\right) I + \frac{\theta^G}{\theta} Q(\theta^G).$$

On the one hand, this suggests that smaller values of θ result in a stochastic kernel with less repetition. On the other, θ controls the size of GEM proportions, each having Beta$(1, \theta)$ distribution. This family of Beta distributions is stochastically ordered, meaning here that if $X^\theta \sim$ Beta$(1, \theta)$ and $x \in (0, 1)$, then:

$$\frac{\partial}{\partial \theta} P(X^\theta > x) = \frac{\partial}{\partial \theta} (1 - x)^\theta = \ln(1 - x)(1 - x)^\theta < 0.$$

We can interpret this to mean that a random proportion with distribution Beta$(1, \theta)$ tends to be larger if θ is smaller.

Thus, smaller values of θ correspond with less repetitive Markov chains and typically larger random proportions used to allocate probability. This suggests that θ acts as a concentration parameter in the following sense:

Proposition 4.2. *Let $P^\theta \sim GEM(\theta)$ and T^θ be a Markov chain independent of P^θ having stochastic kernel $Q(\theta)$ and initial distribution μ on a state space $\mathfrak{X} \subseteq \mathbb{N}$. Then*

define the measures ν *and* ν^k *on* $\mathfrak{X}$ *for each* $k \in \mathbb{N}$ *by:*

$$\nu^k = \sum_{j=1}^{k} P_j^\theta \delta_{T_j^\theta} \qquad\qquad \nu = \sum_{j=1}^{\infty} P_j^\theta \delta_{T_j^\theta}$$

For each $k \in \mathbb{N}$ *and* $d \in (0,1)$*, we then have*

$$\frac{\partial}{\partial \theta} P\left(||\nu - \nu^k|| \geq d\right) > 0$$

where

$$||\nu - \nu^k|| = \sum_{x \in \mathscr{X}} \left|\nu(\{x\}) - \nu^k(\{x\})\right|.$$

In the context of simulation, one interpretation of this is that if we use a truncated measure with fixed cutoff k to approximate an exact observation ν of the stick-breaking measure associated to G, then increasing θ typically increases the maximal error in the approximation. Thus, barring any other consideration, simulations using fixed truncations will generally have improved approximation quality with smaller θ.

Furthermore, it strongly suggests that, in the same sense as the above proposition, a construction of ν from $(\mathbf{R}, \mathbf{Z}) \sim \text{MCcGEM}(G)$,

$$\sum_{j=1}^{\infty} R_j \delta_{Z_j},$$

will be more concentrated than any of the constructions

$$\sum_{j=1}^{\infty} P_j^\theta \delta_{T_j^\theta}.$$

4.1.3 Posterior Distribution

Establishing a posterior distribution of ν distributed as in Theorem 3.7 has proven difficult. On the one hand, ν is a member of a family of generalizations of Dirichlet processes. Dirichlet processes are known to be conjugate priors, i.e. to have posterior distributions which are also Dirichlet, and so we remain hopeful that this generalization is also conjugate.

If so, one might consider the Dirichlet process as a sub-case in an attempt to uncover the posterior. Specifically, consider a finite measure α on a countable space $\mathfrak{X}$. If we define a generator matrix G such that $G_{i,j} = \alpha(\{j\})$ for $i \neq j$, then the stick-breaking measure ν associated to G is a Dirichlet process with parameter α (see Section 3.2.2).

We know that the posterior distribution for a Dirichlet process given observed data $(X_1, X_2, ..., X_n)$ is given by another Dirichlet process with parameter measure $\tilde{\alpha} = \alpha + \sum_{j=1}^{n} \delta_{X_j}$, and so we may write down explicitly a posterior generator matrix

$$\tilde{G} = G + \left(\sum_{j=1}^{n} E_{X_j} \right) - nI$$

where E_x denotes the square matrix with ones in the x-column and zeros elsewhere. Whether this expression of the posterior in terms of generator matrices extends beyond the Dirichlet process is yet to be determined.

Establishing a posterior distribution may involve the self-similarity equations discussed in Chapter 2. Notably, in his discussion of the Dirichlet process as constructed through stick-breaking [60], J. Sethuraman used self-similarity equations satisfied by Dirichlet processes both to identify the distribution of the stick-breaking prior as a Dirichlet process and establish the posterior distribution. Should the generalization in this work also be conjugate, the posterior might similarly be confirmed.

4.2 Wizard's Walks

In Chapter 3, we considered the occupation law of an inhomogeneous Markov chain $\mathbf{M}$ with transition kernels

$$K_n = I + \frac{G}{n}\mathbb{1}(n > c) = I\left(1 - \frac{\theta}{n}\mathbb{1}(n > c)\right) + Q\frac{\theta}{n}\mathbb{1}(n > c)$$

where G is a generator matrix and (Q, θ) is a decomposition of G.

The collections $(R_n)_{n\geq 1}$ and $(P_n)_{n\geq 1}$ defined below constitute two examples of generalizations of $(K_n)_{n\geq 1}$ to which we might want to extend the work of Chapter 3. Define for $n \in \mathbb{N}$

$$R_n = P + a_n G \qquad \text{and} \qquad P_n = (1 - b_n)P + b_n Q \tag{4.6}$$

where P and Q are stochastic matrices, G is a generator matrix, and $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ are sequences of real numbers such that R_n and P_n are stochastic for each $n \geq 1$.

When the stochastic matrix P has diagonal entries with positive lower bound, we find that the kernels $(R_n)_{n\geq 1}$ can always be represented in the form of kernels $(P_n)_{n\geq 1}$:

Example 4.3. Let P be a stochastic matrix with diagonal entries bounded below by $p > 0$, G be a generator matrix, and $(a_n)_{n\geq 1}$ be a sequence of non-negative constants such that R_n, as defined in (4.6), is a stochastic matrix for each n.

Define $g = \sup_{x\in\mathfrak{X}} -G_{x,x}$, $Q = P + \frac{p}{g}G$, and $b_n = \frac{a_n g}{p}$, and consider P_n as defined in (4.6). Then for all $n \in \mathbb{N}$, $R_n = P_n$.

On the other hand, if Q and P are stochastic matrices such that $Q - P$ is not a generator matrix, then no choice of generator matrix will make $(R_n)_{n\geq 1}$ and $(P_n)_{n\geq 1}$ equivalent unless $(b_n)_{n\geq 1}$ is a constant sequence. This is argued in the example below.

Technically, we see that neither generalization of $(K_n)_{n\geq 1}$ in (4.6) is strictly a subclass of the other, but an example illustrates that we may reasonably consider the generalization $(P_n)_{n\geq 1}$ to be broader than the generalization $(R_n)_{n\geq 1}$.

Example 4.4. Let $\mathfrak{X} = \{1, 2, ..., k\}$ be a finite set, and endow the space $(\Delta_k)^k$ of stochastic matrices on $\mathfrak{X}$ with a $k(k-1)$-dimensional Lebesgue measure.

On one hand, almost every stochastic matrix P on $\mathfrak{X}$ has strictly positive entries, and so has diagonal entries with a strictly positive lower bound. By Example 4.3, we

then have that for almost every P and for every G and $(a_n)_{n\geq 1}$, a sequence of kernels of the form $(R_n)_{n\geq 1}$ has a representation as a sequence of kernels of the form $(P_n)_{n\geq 1}$.

On the other hand, consider a pair of stochastic matrices Q and P such that $Q-P$ is not a generator matrix. If a sequence of kernels $(P_n)_{n\geq 1}$ built from Q, P, and a sequence $(b_n)_{n\geq 1}$ can be represented as a sequence of kernels $(R_n)_{n\geq 1}$ when $Q-P$ is not a generator matrix, we can deduce $(b_n)_{n\geq 1}$ to be a constant sequence for any countable state space $\mathfrak{X}$:

Suppose the sequence $(P_n)_{n\geq 1}$ built from Q, P, and a sequence $(b_n)_{n\geq 1}$ can be represented as a sequence of kernels $(R_n)_{n\geq 1}$ built from $\tilde{P}$, G, and $(a_n)_{n\geq 1}$. Setting $R_n = P_n$, we note that as $P-\tilde{P}$ is free of n, so too must be $b_n(Q-P)-a_nG$. As $Q-P$ has row sums of zero but is not a generator matrix, $Q-P$ and G must differ in at least two entries. This relation that these two entries impose on a_n and b_n for all n then implies that $(b_n)_{n\geq 1}$ is constant or $(a_n)_{n\geq 1}$ is constant, which in turn implies that $(b_n)_{n\geq 1}$ is constant.

Thus, if we exclude constant sequences $(b_n)_{n\geq 1}$ (and hence homogeneity), we have that $(P_n)_{n\geq 1}$ can be represented as a sequence of kernels $(R_n)_{n\geq 1}$ precisely when $Q-P$ is a generator matrix. Returning to the restricted setting $\mathfrak{X} = \{1, 2, ..., k\}$, the set of pairs (Q, P) for which $Q-P$ is not a generator matrix has positive measure as such pairs are characterized by $Q_{i,j} \geq P_{i,j}$ for all $i \neq j \in \mathfrak{X}$. Thus, the set of pairs (Q, P) for which every non-constant sequence of kernels of the form $(P_n)_{n\geq 1}$ cannot be represented as a sequence of kernels of the form $(R_n)_{n\geq 1}$ has positive measure.

We now focus on inhomogeneous Markov chains with transition kernels of the form $(P_n)_{n\geq 1}$ as in (4.6), termed wizard's walks.

Definition 4.5 (Wizard's random walk)**.** Consider a countable state space $\mathfrak{X}$. Let P and Q be stochastic matrices on $\mathfrak{X}$, and let $(a_n)_{n\geq 1}$ be a deterministic sequence of positive numbers such that the matrices $(P_n)_{n\geq 1}$, as defined in (4.6), are stochastic. An inhomogeneous Markov chain with transition kernels $(P_n)_{n\geq 1}$ is called a wizard's

random walk, or simply wizard's walk, in P, Q, and $(a_n)_{n\geq 1}$.

In the case that $a = a_n$ is a constant sequence, a wizard's walk with transition kernels $P_n = P_1$ is a homogeneous Markov chain which is represented as a teleporting random walk. In this context, we think of P as representing a standard random walk and Q as representing teleportation probabilities between states, with a denoting the probability of teleportation occurring at any given time. Such constructions can be useful in network theory, where one wants to execute a random walk on the network via P but doesn't want to get stuck in any one particular communication class of P without observing the rest of the network. Teleportation across the network via Q is introduced in essence to force irreducibility of the walk. Teleporting random walks are sometimes referred to as random web surfer models, and were famously considered by Page et al. [50] in the Google PageRank algorithm. See Langville and Meyer [42] and Bianchini [4] for more mathematical discussions.

The kernels P_n then constitute an inhomogeneous generalization of the teleporting random walk, in which the odds of teleportation a_n change over time. In particular, if a_n converges to zero slowly, such as when $a_n = \frac{\theta}{n}$ for large n, then an inhomogeneous Markov chain with transition kernels P_n will teleport infinitely often, but less and less frequently over time.

Our main interest in this section will involve the sequence $(a_n)_{n\geq 1}$ going to zero, with $a_n = \frac{\theta}{n}$ a prototypical example.

We now consider how wizard's walks behave with respect to the communication classes of the main underlying walk, P. When possible, it is convenient to study this behavior by lumping the wizard's walk according to the communication classes of P.

Definition 4.6 (Lumped stochastic matrix)**.** Let P be a stochastic matrix with state space $\mathfrak{X}$. If $\mathscr{A} = \{A_\alpha\}_{\alpha\in D}$ is a partition of $\mathfrak{X}$ such that for each $\alpha, \beta \in D$, the submatrix $P(A_\alpha, A_\beta)$ of P consisting of transition probabilities from A_α to A_β has

constant row sums, denoted $s(\alpha, \beta)$, we call P lumpable with respect to $\mathscr{A}$. Let $pr_{\mathscr{A}} : \mathfrak{X} \to \mathscr{A} \cong D$ be given by projection. When P is lumpable with respect to $\mathscr{A}$, we define the lumped matrix $P^{\mathscr{A}}$ of P by $\mathscr{A}$, by

$$P^{\mathscr{A}}_{\alpha\beta} = s(\alpha, \beta)$$

for $\alpha, \beta \in D$.

Recall that the communication classes of a stochastic matrix P form a partition $\mathscr{A}$ of the state space.

Proposition 4.7. *If* $\mathbf{M}$ *is a wizard's walk on* $\mathfrak{X}$ *in* P, Q, *and* $(a_n)_{n\geq 1}$, *then when* P *and* Q *are both lumpable with respect to a partition* $\mathscr{A} = \{A_\alpha\}_{\alpha\in D}$ *of* $\mathfrak{X}$, *the process* $\mathbf{M}^{\mathscr{A}}$ *given by* $M_j^{\mathscr{A}} := pr_{\mathscr{A}}(M_j)$ *is a wizard's walk on* $\mathscr{A} \cong D$ *in* $P^{\mathscr{A}}$, $Q^{\mathscr{A}}$, *and* $(a_n)_{n\geq 1}$.

In the case that P and Q are both lumpable with respect to the communication classes of P, this returns us to an inhomogeneous Markov chain related to those studied in Chapter 3 and, more broadly, to the classes of inhomogeneous Markov chains studied by Dietz and S. Sethuraman [13] and Bouguet and Cloez [7]:

Proposition 4.8. *Let* P *be a recurrent, reducible stochastic matrix on countable state space* $\mathfrak{X}$. *Let* $\mathscr{A} = \{A_\alpha\}_{\alpha\in D}$ *be the partition of* $\mathfrak{X}$ *given by the communication classes of* P, *and note* $P^{\mathscr{A}} = I$. *Let* Q *be a stochastic matrix on* $\mathfrak{X}$ *lumpable with respect to* $\mathscr{A}$. *If* $\mathbf{M}$ *is wizard's walk in* P, Q, *and* $(a_n)_{n\geq 1}$, *and* $\tilde{\mathbf{M}}$ *is an inhomogeneous Markov chain on* D *with transition kernels* $K_n = I + a_n G$ *where* $G = Q^{\mathscr{A}} - I$, *then*

$$\tilde{\mathbf{M}} \stackrel{d}{=} \mathbf{M}^{\mathscr{A}}.$$

We then see that when lumping is possible, a wizard's walk can be lumped to a freezing Markov chain. Corollary 4.8.1 below is a direct application of [7], [13], and Theorem 3.7 of this work, to Proposition 4.8. In the case of lumpability with $a_n = \frac{\theta}{n}$, it establishes the occupation law on communication classes as mimicking the occupation laws studied in Chapter 3.

Corollary 4.8.1. *Consider the context of Proposition 4.8. Suppose G is irreducible and $(a_n)_{n\geq 1}$ is a decreasing sequence such that $\sum_{n=1}^{\infty} a_n = \infty$.*

1. *If $0 < \lim_{n\to\infty} na_n < \infty$, the empirical measures of $M^{\mathscr{A}}$ converge in distribution to the stationary distribution of a piecewise-deterministic Markov process (the exponential zig-zag process, detailed in [7]).*

2. *If $\lim_{n\to\infty}(na_n)^{-1} = 0$, the empirical measures of $M^{\mathscr{A}}$ converge in probability to the stationary distribution of G.*

3. *If $\sum_{n=0}^{\infty}(n^2 a_n)^{-1} < \infty$, the empirical measures of $M^{\mathscr{A}}$ converge almost surely to the stationary distribution of G.*

4. *If $a_n = \frac{\theta}{n}$ for large n, then the empirical measures of $\mathbf{M}^{\mathscr{A}}$ converge in distribution only, and*
$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \delta_{M_j^{\mathscr{A}}} \stackrel{d}{=} \sum_{j=1}^{\infty} R_j \delta_{Z_j}$$
where $(\mathbf{R}, \mathbf{Z}) \sim MCcGEM(\theta G')$.

What remains in the case of lumpability is then to establish the (normalized) occupation law for an individual communication class A_α, i.e. the measure on A_α given by
$$\lim_{n\to\infty} \frac{\sum_{j=1}^{n} \delta_{M_j}}{\sum_{j=1}^{n} \delta_{M_j}(A_\alpha)}$$
As one might expect in the context that a_n approaches 0 slowly and $P(A_\alpha, A_\alpha)$ is positive recurrent, preliminary computations suggest that this limit converges at least in probability to the unique stationary distribution of $P(A_\alpha, A_\alpha)$.

In future work, we intend to establish this limit and extend results to the case that P has transient states. From there, we hope to slowly relax the lumpability condition in order to understand more completely the occupation law of a wizard's walk. Also of interest would be to transport some of the results and methods of Bouguet and Cloez [7] to wizard's walks.

4.3 Another Inhomogeneous Chain

Building on the work of S. Sethuraman and Dietz in [13], discussed in Section 4.1, we can provide another alternative description of the moments of our generalization of the Dirichlet process, ν.

Let G be a $k \times k$ generator matrix without zero entries. First, we observe that the matrices $\tilde{K}_n := \left(I - G/n\right)^{-1}$ for $n \geq 1$ are each stochastic. For the reader familiar with continuous time Markov chains, we recognize $\tilde{K}_n$ as a resolvent operator with respect to the transition function $\{\mathscr{P}_s^G : s \geq 0\}$ of a continuous time Markov chain with generator G. In particular, it is standard to write

$$\left(I - \frac{G}{n}\right)^{-1}_{l,m} = \int_0^\infty n e^{-ns} \mathscr{P}_s^G(l,m) ds.$$

As a consequence, $\tilde{K}_n$ is a stochastic kernel on $\mathfrak{X}$.

Let $\mathbf{Z} = (Z_n)_{n \geq 1}$ be the inhomogeneous Markov chain with initial distribution μ and transition kernels $\tilde{K}_n$ for $n \geq 1$. We observe a type of 'duality' relation between the moments of ν and $\mathbf{Z}$. As usual, let $\mathbf{M}$ be an inhomogeneous Markov chain with kernels K_n.

Theorem 4.9 (Recasting moments II)**.** *In the setting of [13] given above: The stochastic kernel $\tilde{K}_n = K_n + O(n^{-2})$ and the occupation law ν with respect to $\mathbf{M}$ is also the occupation law with respect to $\mathbf{Z}$,*

$$\nu \stackrel{d}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \delta_{Z_j} \stackrel{d}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \delta_{M_j} \tag{4.7}$$

Moreover, the moments of ν may be expressed in terms of $\mathbf{Z}$,

$$\mathbb{E}\left[\prod_{i=1}^{k} \nu_i^{m_i}\right] = \binom{N}{m_1, ..., m_k}^{-1} \mathscr{P}\left(\sum_{j=1}^{N} \mathbb{1}(Z_j = i) = m_i : 1 \leq i \leq k\right), \tag{4.8}$$

and, in particular, $\mathbb{E}\left[\nu_i^N\right] = \mathscr{P}(Z_1 = ... = Z_N = i)$.

This theorem may also be found in [12].

The simplicity of the relation between $\mathbf{Z}$ and the moments of ν suggests that the chain $\mathbf{Z}$ is a natural object of study, and may be of equal interest as the chain $\mathbf{M}$ moving forward.

4.4 Proofs

4.4.1 Proofs for Section 4.1

Proof of Proposition 4.1. We break the argument into steps.

Step 1. First, we show that $p_j(\lambda)$, $q(j)$, and their quotients are all well-defined. A generator matrix G can always be written as $G = \theta(Q - I)$ for some $\theta > 0$ and a stochastic matrix Q. The eigenvalues λ of Q correspond with the eigenvalues $\theta(\lambda - 1)$ of G. Additionally, since G has no zero entries, Q is irreducible. Therefore, the algebraic multiplicity of the eigenvalue 0 of G is 1. Thus, with respect to the minimal polynomial of G, $p_{min}(\lambda)$, there exists a polynomial q such that $p_{min}(\lambda) = \lambda q(\lambda)$ and $q(0) \neq 0$. Define

$$\theta^G = \min\left\{\theta \in \mathbb{R}^+ : I + G\theta^{-1} \text{ is non-negative}\right\}.$$

Since the eigenvalues of the stochastic matrix $I + G/\theta^G$ are bounded by 1, the (complex) eigenvalues $\tilde{\lambda}$ of G satisfy $\left|1 + \tilde{\lambda}/\theta^G\right| \leq 1$. Hence, the eigenvalues of G have non-positive real part. Since $p_{min}(\lambda) = \lambda q(\lambda)$ and $q(0) \neq 0$, we obtain that $j \in \mathbb{N}$ is not an eigenvalue of G and so $q(j) \neq 0$ for $j \geq 0$. Thus, $p_j(\lambda)/q(j)$ is well-defined for $j \geq 0$.

Step 2. We now verify for $j > 0$ that $\tilde{K}_j = (I - G/j)^{-1} = \frac{p_j(G)}{q(j)}$. Write

$$\begin{aligned}(j - \lambda)p_j(\lambda) = &\sum_{i=0}^{n} \lambda^i \sum_{l=i}^{n} a_l j^{l+1-i} - \sum_{i=0}^{n} \lambda^{i+1} \sum_{l=i}^{n} a_l j^{l-i} \\ &\sum_{i=0}^{n} \lambda^i \sum_{l=i}^{n} a_l j^{l+1-i} - \sum_{r=1}^{n+1} \lambda^r \sum_{l=r-1}^{n} a_l j^{l+1-r},\end{aligned}$$

which, after cancelling terms, equals

$$\sum_{l=0}^{n} a_l j^{l+1} - \sum_{r=1}^{n+1} \lambda^r a_{r-1} \;=\; jq(j) - \lambda q(\lambda).$$

In particular, as $Gq(G) = p_{min}(G) = 0$, we have

$$I = \frac{jq(j)I - Gq(G)}{jq(j)} = \frac{(jI - G)p_j(G)}{jq(j)} = \left(I - \frac{G}{j}\right)\frac{p_j(G)}{q(j)},$$

from which the desired identity follows.

Step 3: We now show that $p_0(G)/q(0)$ is the constant matrix with rows μ. Note that $p_0(\lambda)/q(0) = q(\lambda)/q(0)$ is well-defined in (4.3). Since row sums of G^k vanish for $k \geq 1$, we see that $p_j(G)/q(j)$ has constant row sums of $p_j(0)/q(j) = 1$. Now, necessarily, $p_0(G)G = 0$ as $q(\lambda)\lambda = p_{min}(\lambda)$ is the minimal polynomial of G. Since G is irreducible, we can conclude that $p_0(G)$ is a matrix with rows given by multiples of the unique stochastic eigenvector μ associated to G and eigenvalue 0. However, since $p_0(G)/q(0)$ has row sums of 1, the claim follows.

Moreover, noting $[p_0(G)/q(0)]_{i,j} = \mu_j$ for any $i, j \in \{1, 2, ..., k\}$, the moment identity (4.4) is a direct consequence of these calculations.

■

Proof of Proposition 4.2. First, we reformulate the lemma: Let $\{X_j^\theta\}_{j\geq 1}$ be the iid Beta$(1, \theta)$ factors associated to P^θ. Then for each $k \in \mathbb{N}$:

$$\begin{aligned}
\sum_{x\in\mathfrak{X}} \left|\nu(\{x\}) - \nu^k(\{x\})\right| &= \sum_{x\in\mathfrak{X}} \nu(\{x\}) - \nu^k(\{x\}) \\
&= \nu(\mathfrak{X}) - \nu^k(\mathfrak{X}) \\
&= 1 - \sum_{j=1}^{k} P_j^\theta \\
&= \prod_{j=1}^{k} (1 - X_j^\theta)
\end{aligned}$$

Thus, it is equivalent to prove for each $k \in \mathbb{N}$ and $d \in (0,1)$:

$$\frac{\partial}{\partial\theta} P\left(\prod_{j=1}^{k}(1-X_j^\theta) \geq d\right) > 0$$

The derivative may be easily computed after noting that $\sum_{j=1}^{k} -\ln(1-X_j^\theta)$ has distribution Gamma(k,θ) (see Lemma A.4):

$$\begin{aligned}
\frac{\partial}{\partial\theta} P\left(\prod_{j=1}^{k}(1-X_j^\theta) \geq d\right) &= \frac{\partial}{\partial\theta} P\left(\sum_{j=1}^{k} -\ln(1-X_j^\theta) \leq -\ln d\right) \\
&= \frac{\partial}{\partial\theta} \int_0^{-\ln(d)} \theta^k \frac{y^{k-1}e^{-y\theta}}{\Gamma(k)} dy \\
&= \frac{\partial}{\partial\theta} \int_0^{-\theta\ln(d)} \frac{z^{k-1}e^{-z}}{\Gamma(k)} dz \\
&= \frac{(-\theta\ln(d))^{k-1}\, e^{\theta\ln(d)}}{\Gamma(k)} (-\ln(d)) > 0
\end{aligned}$$

■

4.4.2 Proofs for Section 4.2

Proof of Proposition 4.7. Let α_j be a sequence in D, and define $P_j^{\mathscr{A}} = (1-a_j)P^{\mathscr{A}} + a_j Q^{\mathscr{A}}$. Then for each $1 \leq m \leq n$, we have

$$\begin{aligned}
&\mathscr{P}(M_j^{\mathscr{A}} = \alpha_j : 1 \leq j \leq n) = \mathscr{P}(M_j \in A_{\alpha_j} : 1 \leq j \leq n) \\
=& \sum_{b_1 \in A_{\alpha_1}} \sum_{b_2 \in A_{\alpha_2}} \cdots \sum_{b_n \in A_{\alpha n}} \mathscr{P}(M_1 = b_1) \prod_{j=1}^{n-1} \left[(1-a_j)P + a_j Q\right]_{b_j, b_{j+1}} \\
=& \sum_{b_1 \in A_{\alpha_1}} \sum_{b_2 \in A_{\alpha_2}} \cdots \sum_{b_m \in A_{\alpha m}} \mathscr{P}(M_1 = b_1) \cdot \\
&\qquad \prod_{j=1}^{m-1} \left[(1-a_j)P + a_j Q\right]_{b_j, b_{j+1}} \prod_{j=m}^{n-1} \left[(1-a_j)P^{\mathscr{A}} + a_j Q^{\mathscr{A}}\right]_{\alpha_j, \alpha_{j+1}} \\
=& \mathscr{P}(M_1 \in A_{\alpha_1}) \prod_{j=1}^{n-1} \left[P_j^{\mathscr{A}}\right]_{\alpha_j, \alpha_{j+1}} \\
=& \mathscr{P}(M_1^{\mathscr{A}} = \alpha_1) \prod_{j=1}^{n-1} \left[P_j^{\mathscr{A}}\right]_{\alpha_j, \alpha_{j+1}}
\end{aligned}$$

Thus, $\mathbf{M}^{\mathscr{A}}$ is a wizard's walk in $P^{\mathscr{A}}$, $Q^{\mathscr{A}}$, and $\{a_n\}$. ■

Proof of Proposition 4.8. The proposition follows immediately after noting that

$$K_n = (1 - a_n)P^{\mathscr{A}} + a_n Q^{\mathscr{A}}.$$

■

4.4.3 Proofs for Section 4.3

Proof of Theorem 4.9. First, since G is a $k \times k$ generator matrix with bounded entries and for large enough $j \in \mathbb{N}$

$$\tilde{K}_j = \left(I - \frac{G}{j}\right)^{-1} = \sum_{n=0}^{\infty} \frac{G^n}{j^n},$$

we verify that $\tilde{K}_j = K_j + O(j^{-2})$

To show (4.7), we next relate the occupation laws of the Markov chain $\mathbf{Z}$ having transition kernels $\{\tilde{K}_n\}$ and the Markov chain $\mathbf{M}$ having kernels $\{K_n\}$ through a Borel-Cantelli argument.

Define $A_j := \tilde{K}_j - K_j$, for $j \geq 1$, and note $A_j = O(j^{-2})$ has constant row sums of 0. Since G does not have 0 entries and $K_j = I + \frac{G}{j}\mathbf{1}(j > c)$, there exists an a such that $R_j := K_j + \frac{j^2}{a}A_j$ is a non-negative matrix, and hence stochastic. Note

$$\tilde{K}_j = \left(1 - \frac{a}{j^2}\right) K_j + \frac{a}{j^2} R_j. \tag{4.9}$$

Consider now an auxilliary sequence $\mathbf{B} = (B_j)_{j \geq 1}$ of independent Bernoulli(aj^{-2}) variables, possibly by enlargening the probability space. Define a process $\mathbf{Z}'\big|\mathbf{B}$ with $Z_1'\big|\mathbf{B} \sim \mu$ and

$$\begin{aligned} &\mathscr{P}\left(Z'_{j+1} = z_{j+1}\middle| Z'_l = z_l : 1 \leq l \leq j, \mathbf{B}\right) \\ =&\mathscr{P}\left(Z'_{j+1} = z_{j+1}\middle| Z'_j = z_j, B_j\right) = (1 - B_j)K_j(z_j, z_{j+1}) + B_j R_j(z_j, z_{j+1}). \end{aligned}$$

Noting (4.9), $\mathbf{Z}'$ is then marginally a Markov chain with initial distribution μ and transition kernels given by $\tilde{K}_j$:

$$\begin{aligned}
&\mathscr{P}\left(Z'_{j+1} = z_{j+1} \middle| Z'_j = z_j\right) \\
=&K_j(z_j, z_{j+1})\mathscr{P}(B_j = 0) + R_j(z_j, z_{j+1})\mathscr{P}(B_j = 1) \\
=&\tilde{K}_j(z_j, z_{j+1}).
\end{aligned}$$

Now, by the Borel Cantelli lemma, $\mathscr{P}(B_j = 1 \text{ i.o.}) = 0$ and so $L := \max\{j : B_j = 1\} < \infty$ a.s. Conditional on the event that $\{L = r\}$, the chain $(Z'_j)_{j>r}$ is a Markov chain with transition kernels $\{K_j\}_{j>r}$. Also, since G is irreducible in the setting of [13], the initial distribution does not matter in the calculation of the occupation law ν (see discussion below Corollary 3.2.1). Hence, the occupation law with respect to $\mathbf{Z}$ is also ν and (4.7) holds: Indeed, if $A \subseteq \Delta_k$ is measurable, then

$$\begin{aligned}
&\lim_{n\to\infty} \mathscr{P}\left(\frac{1}{n}\sum_{j=1}^{n} \delta_{Z_j} \in A\right) \\
=&\lim_{n\to\infty} \mathscr{P}\left(\frac{1}{n}\sum_{j=1}^{n} \delta_{Z'_j} \in A\right) \\
=&\lim_{R\to\infty}\lim_{n\to\infty} \mathscr{P}\left(\frac{1}{n}\sum_{j=1}^{n} \delta_{Z'_j} \in A\right) \\
=&\lim_{R\to\infty}\lim_{n\to\infty} \mathscr{P}\left(\frac{1}{n}\sum_{j=1}^{n} \delta_{Z'_j} \in A \text{ and } L < R\right) + o(1)_R \\
=&\lim_{n\to\infty} \mathscr{P}\left(\frac{1}{n}\sum_{j=1}^{n} \delta_{T_j} \in A\right),
\end{aligned}$$

where $o(1)_R$ is an expression which vanishes uniformly in n as $R \to \infty$.

Finally, (4.8) follows straightforwardly by gathering together terms.

■

Appendix A

A.1 Common Probability Distributions

Definition A.1 (Beta Distribution)**.** A real-valued random variable X is said to have Beta(α, β) distribution for positive constants α and β if for a measurable set A of real numbers,

$$\mathscr{P}(X \in A) = \int_A \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}(0<x<1)dx,$$

i.e. when X can be associated to probability density function (pdf)

$$f_{\alpha,\beta}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}(0<x<1)$$

One may then compute directly the raw moments of a random variable X with distribution Beta(α, β):

$$\mathbb{E}[X^N] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+N)}{\Gamma(\alpha+N+\beta)} = \frac{(\alpha)_N}{(\alpha+\beta)_N}$$

where $(\alpha)_N$ denotes the Pochhammer symbol, i.e. the rising factorial $(\alpha)_N = \frac{\Gamma(\alpha+N)}{\Gamma(\alpha)}$. A reader might note that the notation $(\alpha)_N$ does not have standard definition across the literature, and may also be used to denote the falling factorial. In this dissertation, only the rising factorial is encountered, and the Pochhammer symbol is used for this meaning only.

Definition A.2 (Beta Product Distribution)**.** A random variable X taking values in [0,1] is said to have Beta Product$(\boldsymbol{\alpha}, \boldsymbol{\beta})$ distribution for $\alpha, \beta \in \mathbb{C}^k$ if the moments of X are given by

$$\mathbb{E}[X^N] = \prod_{j=1}^{k} \frac{(\alpha_j)_N}{(\alpha_j+\beta_j)_N}.$$

The naming convention for the Beta Product distribution is justified by noting that if $(X_j)_{j=1}^k$ is a collection of independent variables with Beta distributions, then their product has a Beta Product distribution. This can be seen by matching the moments of the product to the moments of the Beta Product. However, one should note that not all Beta Product distributions arise in this way.

In [45], Mathai and Saxena gave the pdf of the Beta Product$(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in terms of a contour integral:

$$f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(x) = \frac{1}{2\pi i} \int_C \prod_{j=1}^{k} \frac{(\alpha_j)_s}{(\alpha_j + \beta_j)_s} x^{-s-1} ds \mathbb{1}(0 < x < 1).$$

This expression is derived from their study of Meijer's G functions, and more information, including the particulars of the contour and valid choices of complex parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for Meijer's G functions, may be found in their book [45].

In particular, the pdf for the Beta Product is still valid, i.e. real, non-negative, and integrating to 1, for certain choices of complex parameters. In the case that $k = 2$, Dufresne completely characterized in [16] the set of parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ for which the density is valid, a set which included some complex parameters.

We encounter Beta Product distributions in this work while investigating occupation laws of Markov chains and stick-breaking measures, including some Beta Products with complex parameters.

Definition A.3 (Gamma Distribution)**.** A real-valued random variable X is said to have Gamma(α, β) distribution for positive constants α and β if for a measurable set A of real numbers,

$$\mathscr{P}(X \in A) = \int_A \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1(x > 0) dx$$

The following lemma establishes a useful and well-known connection between Beta variables and Gamma variables:

Lemma A.4. *Let $(X_j)_{j=1}^n$ be a collection of independent variables with common distribution Beta$(1,\theta)$. Then*

$$Y := -\sum_{j=1}^{n} \ln(1 - X_j) \sim \text{Gamma}(n, \theta)$$

The lemma is proved by induction on n and direct computation of the cdf of Y.

Proof. Let $f_n(y)$ denote the pdf of a Gamma(n,θ) distribution and $G(x) = 1-(1-x)^\theta$ denote the cdf of a Beta$(1,\theta)$ variable. We proceed by induction.

Consider the case $n = 1$.

$$\frac{d}{dc}\mathscr{P}(-\ln(1-X) \le c) = \frac{d}{dc}\mathscr{P}(X \le 1 - e^{-c}) = \frac{d}{dc}\left[1 - e^{-\theta c}\right] = \theta e^{-\theta c} = f_1(c)$$

Suppose the result holds for some $n \in \mathbb{N}$. Let Y be a Gamma(n,θ) variable independent of a Beta$(1,\theta)$ variable X. Then

$$\begin{aligned}
\frac{d}{dc}\mathscr{P}(-\ln(1-X) + Y \le c) &= \frac{d}{dc}\mathscr{P}(X \le 1 - e^{Y-c}) \\
&= \frac{d}{dc}\int_0^c f_n(y) G\left(1 - e^{y-c}\right) dy \\
&= \frac{d}{dc}\int_0^c \left(1 - e^{\theta(y-c)}\right) \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y} dy \\
&= \frac{d}{dc}\left[\mathscr{P}(Y \le c) - e^{-\theta c}\frac{\theta^n}{\Gamma(n+1)}c^n\right] = f_{n+1}(c)
\end{aligned}$$

■

Definition A.5 (Geometric Distribution)**.** A random variable X taking values in $\mathbb{N} = \{1, 2, 3, ...\}$ is said to have Geometric(p) distribution if for each $n \in \mathbb{N}$:

$$\mathscr{P}(X = n) = p(1-p)^{n-1}$$

The following lemma establishes a relation between Geometric distributions and Beta distributions useful in a later section of this work.

Lemma A.6. *Let $(X_j)_{j\geq 1}$ be an iid sequence of random variables with common distribution Beta$(1,\theta)$. Let N have Geometric p distribution. Then,*

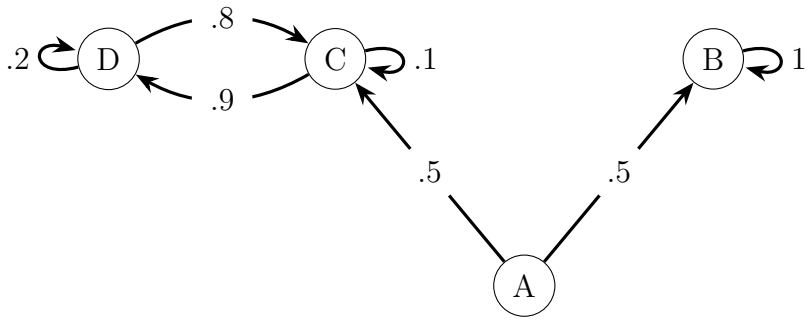
$$Y = \prod_{j=1}^{N}(1 - X_j) \sim \text{Beta}(p\theta, 1)$$

The lemma is proved by directly computing the moments of Y and matching the moments of Y to the moments of a Beta$(p\theta, 1)$ variable. Since the moments of Y are uniformly bounded between 0 and 1, these moments are sufficient to identify the distribution of Y as Beta$(p\theta, 1)$.

A.2 Markov Chain Example

Example A.7. Suppose an individual walks randomly among four locations, numbered $\{A, B, C, D\}$. When at location A, the individual travels at the next time step either to location B or location C with probability .5 each. When at location B, the individual becomes stuck and stays put at each time step for eternity with probability one. When at location C, at the next time step the individual either stays in that location with probability .1 or travels to location D with probability .9. When at location D, at the next time step the individual stays in that location with probability .2 or travels to locations C with probability .8. A map with associated probabilities of traveling between locations is in Figure A.1.

FIGURE A.1. A map for the random walk of Example A.7



Let $\mathbf{M} = (M_j)_{j\geq 0}$ be the associated stochastic process where at time j, the individual is at location $M_j \in \{A, B, C, D\}$. If μ is the stochastic vector on $\mathfrak{X} = \{A, B, C, D\}$ such that $\mu_x = \mathscr{P}(M_1 = x)$, then $\mathbf{M}$ is a Markov chain on $\mathfrak{X}$ with initial distribution μ and stochastic kernel

$$Q = \begin{bmatrix} 0 & .5 & .5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & .1 & .9 \\ 0 & 0 & .8 & .2 \end{bmatrix}.$$

Note that Q, as a block upper triangular matrix, is reducible in the standard (non-probability) sense of a matrix, with a singleton block each for locations A and B, and a block containing both locations C and D.

Upon inspection, we see that location A leads to itself vacuously, to B and C, and through C to D, but no other locations lead to A. Thus, A is a transient location as well as in its own communication class. Location B leads only to itself, and so is an absorbing state. Thus, B is positive recurrent and in its own communication class. Location C leads to itself and to location D, while location D leads to itself and to location C. Thus, C and D are in the same communication class. We compute

$$\begin{aligned} &\mathbb{E}[\tau_C | M_1 = C] \\ =&1\mathscr{P}(M_2 = C | M_1 = C) + \sum_{j=2}^{\infty} j\mathscr{P}(M_2 = ... = M_j = D, M_{j+1} = C | M_1 = C) \\ =&Q_{C,C} + \sum_{j=2}^{\infty} jQ_{C,D}(Q_{D,D})^{j-2}Q_{D,C} \\ =&.1 + \sum_{j=2}^{\infty} j * .9 * (.2)^{j-2} * .8 = 2.125 < \infty \end{aligned}$$

So C is positive recurrent, as is D.

Thus, $\mathbf{M}$ has three communication classes, one with a single recurrent state A, one with a positive recurrent absorbing state B, and one with two positive recurrent states C and D. $\mathbf{M}$ is then reducible, and has class structure corresponding with the matrix structure of the stochastic kernel Q of $\mathbf{M}$.

We can associate stationary distributions to $\mathbf{M}$ by finding left eigenvectors μ of Q having eigenvalue 1. Two such stochastic vectors are

$$\mu = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \tilde{\mu} = \begin{bmatrix} 0 \\ 0 \\ 8/17 \\ 9/17 \end{bmatrix}.$$

Note that μ assigns positive probability only to the positive recurrent class consisting only of B, and $\tilde{\mu}$ assigns probability only to the positive recurrent class consisting of states C and D. Since the multiplicity of the eigenvalue 1 of Q is 2, all stationary distributions can be obtained as an average of μ and $\tilde{\mu}$.

A.3 RAM identities

Proof of Lemma 1.18.

Recall that an empty product is taken to equal 1. We prove the identity for $k = 1$ and proceed by induction. When $k = 1$:

$$\prod_{j=1}^{1}(1-a_j) + \sum_{j=1}^{1} a_j \prod_{i=1}^{j-1}(1-a_i = (1-a_1) + a_1 \prod_{i=1}^{0}(1-a_i)$$
$$= (1-a_1) + a_1 = 1$$

Suppose the identity holds for k. Then the identity also holds for $k+1$:

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

By induction, the identity holds for all $k \in \mathbb{N}$. ■

Proof of Lemma 1.19. Note that empty sums are taken to equal 0 and empty products are taken to equal 1. For (1.10), we prove the identity for $j = 1$ and proceed by induction:

When $j = 1$, $\sum_{i=1}^{j-1} P_i = \sum_{i=1}^{0} P_i = 0 < 1$. Thus

$$X_1 = P_1 \left(1 - \sum_{i=1}^{0} P_i\right)^{-1} = P_1$$

and

$$P_1 = X_1 = X_1 \prod_{i=1}^{0} (1 - X_i).$$

Suppose (1.10) holds for j. Then

$$\sum_{i=1}^{j} P_i = \sum_{i=1}^{j} X_i \prod_{l=1}^{i-1} (1 - X_l) = 1 - \prod_{i=1}^{j} (1 - X_i)$$

where the last inequality follows from Lemma 1.18. If $\sum_{i=1}^{j} P_i = 1$, then $0 = \prod_{i=1}^{j} (1 - X_i)$ and also $0 = P_{j+1}$ as $\mathbf{P}$ is a probability distribution. Thus

$$0 = P_{j+1} = X_{j+1} \prod_{i=1}^{j} (1 - X_i)$$

If instead $\sum_{i=1}^{j} P_i < 1$, then

$$P_{j+1} = X_{j+1} \left(1 - \sum_{i=1}^{j} P_i\right) = X_{j+1} \prod_{i=1}^{j} (1 - X_i)$$

Either way, if (1.10) holds for j, it holds also for $j + 1$. By induction, (1.10) holds for all $j \in \mathbb{N}$. Then, (1.11) follows immediately from (1.10) and Lemma 1.18. ■

References

[1] Aldous, D. J., Ibragimov, I. A., and Jacod, J. *Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983*, vol. 1117. Springer, 2006.

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

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

[4] Bianchini, M., Gori, M., and Scarselli, F. Inside pagerank. *ACM Transactions on Internet Technology (TOIT) 5*, 1 (2005), 92–128.

[5] Blackwell, D. Discreteness of ferguson selections. *Ann. Statist. 1*, 2 (1973), 356–358.

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

[7] Bouguet, F., and Cloez, B. Fluctuations of the empirical measure of freezing markov chains. *Electron. J. Probab. 23* (2018), 1–31.

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

[9] Cannings, C. The latent roots of certain markov chains arising in genetics: a new approach, i. haploid models. *Adv. in Appl. Probab. 6*, 2 (1974), 260–290.

[10] Catoni, O., and Cerf, R. The exit path of a markov chain with rare transitions. *ESAIM Probab. Stat. 1* (1997), 95–144.

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

[12] Dietz, Z., Lippitt, W., and Sethuraman, S. Stick-breaking processes, clumping, and markov chain occupation laws. *arXiv preprint arXiv:1901.08135* (2019).

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

[14] Dobrushin, R. L. Central limit theorem for nonstationary markov chains. i,ii. *Theory Probab. Appl. 1*, 1 (1956), 65–80.

[15] Donnelly, P., and Joyce, P. Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Process. Appl. 31*, 1 (1989), 89–103.

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

[17] Durrett, R. *Probability: theory and examples*, vol. 49. Cambridge university press, 2019.

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

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

[20] Ewens, W. J. The sampling theory of selectively neutral alleles. *Theoretical population biology 3*, 1 (1972), 87–112.

[21] Ewens, W. J. Population genetics theory-the past and the future. In *Mathematical and statistical developments of evolutionary theory*. Springer, 1990, pp. 177–227.

[22] Ferguson, T. S. A bayesian analysis of some nonparametric problems. *Ann. Statist.* (1973), 209–230.

[23] Fisher, R. *The genetical theory of natural selection.* Clarendon Press, 1930.

[24] Gantert, N. Laws of large numbers for the annealing algorithm. *Stochastic Process. Appl. 35*, 2 (1990), 309–313.

[25] Gnedin, A., and Kerov, S. A characterization of gem distributions. *Combin. Probab. Comput. 10*, 3 (2001), 213–217.

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

[27] Gnedin, A., and Pitman, J. Regenerative partition structures. *Electron. J. Combin. 11*, 2 (2005), Research Paper 12.

[28] Goldie, C. M. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* (1991), 126–166.

[29] Griffiths, R. On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology 15*, 1 (1979), 140–158.

[30] Halmos, P. R. Random alms. *Ann. Math. Stat. 15*, 2 (1944), 182–189.

[31] Herbach, U. Stochastic gene expression with a multistate promoter: breaking down exact distributions. *arXiv preprint arXiv:1804.05740* (2018).

[32] Hirth, U. M. From gem back to dirichlet via hoppe's urn. *Combin. Probab. Comput. 6*, 2 (1997), 185–195.

[33] Hoppe, F. M. The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol. 25*, 2 (1987), 123–159.

[34] Ishwaran, H., and James, L. F. Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc 96*, 453 (2001), 161–173.

[35] Karlin, S., and McGregor, J. Addendum to a paper of ewen. *Theor. Pop. Biol* (1972).

[36] Kingman, J. F. The representation of partition structures. *J. Lond. Math. Soc. 2*, 2 (1978), 374–380.

[37] Kingman, J. F. On the genealogy of large populations. *J. Appl. Probab. 19*, A (1982), 27–43.

[38] Kingman, J. F. C. Random discrete distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* (1975), 1–22.

[39] Kingman, J. F. C. Random partitions in population genetics. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 361*, 1704 (1978), 1–20.

[40] Kingman, J. F. C. The coalescent. *Stochastic Process. Appl. 13*, 3 (1982), 235–248.

[41] Kingman, J. F. C. *Poisson processes*. Clarendon Press, Oxford, 1992.

[42] Langville, A. N., and Meyer, C. D. Deeper inside pagerank. *Internet Math. 1*, 3 (2004), 335–380.

[43] Last, G. An integral characterization of the dirichlet process. *arXiv preprint arXiv:1803.11465* (2018).

[44] Lavine, M. Some aspects of polya tree distributions for statistical modelling. *Ann. Statist. 20*, 3 (1992), 1222–1235.

[45] Mathai, A. M., and Saxena, R. K. *Generalized hypergeometric functions with applications in statistics and physical sciences*, vol. 348. Springer-Verlag, 1973.

[46] McCloskey, J. W. *A model for the distribution of individuals by species in an environment*. PhD thesis, Michigan State University, Department of Statistics, 1965.

[47] Miclo, L. Sur les temps d'occupations des processus de markov finis inhomogènes à basse température. *Stochastics 63*, 1-2 (1998), 65–137.

[48] Moran, P. A. P. Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society* (1958), vol. 54, Cambridge University Press, pp. 60–71.

[49] Müller, P., Quintana, F. A., Jara, A., and Hanson, T. *Bayesian nonparametric data analysis*. Springer, New York, 2015.

[50] Page, L., Brin, S., Motwani, R., and Winograd, T. The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab, 1999.

[51] Patil, G., and Taillie, C. Diversity as a concept and its implications for random communities. *Bull. Int. Stat. Inst 47* (1977), 497–515.

[52] Pitman, J. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields 102*, 2 (1995), 145–158.

[53] Pitman, J. Random discrete distributions invariant under size-biased permutation. *Adv. in Appl. Probab. 28*, 2 (1996), 525–539.

[54] Pitman, J. Some developments of the blackwell-macqueen urn scheme. *Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell 30* (1996), 245–267.

[55] Pitman, J. Poisson–dirichlet and gem invariant distributions for split-and-merge transformations of an interval partition. *Combin. Probab. Comput. 11*, 5 (2002), 501–514.

[56] Pitman, J. *Combinatorial Stochastic Processes: Ecole d'Eté de Probabilités de Saint-Flour XXXII-2002*. Springer-Verlag, Berlin, 2006.

[57] Pitman, J., Yakubovich, Y., et al. Ordered and size-biased frequencies in gem and gibbs' models for species sampling. *Ann. Appl. Probab. 28*, 3 (2018), 1793–1820.

[58] Pitman, J., and Yor, M. The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *Ann. Probab. 25* (1997), 855–900.

[59] Schiavo, L. D., and Lytvynov, E. W. A mecke-type characterization of the dirichlet-ferguson measure. *arXiv preprint arXiv:1706.07602* (2017).

[60] Sethuraman, J. A constructive definition of dirichlet priors. *Statist. Sinica* (1994), 639–650.

[61] Sethuraman, J., and Tiwari, R. C. Convergence of dirichlet measures and the interpretation of their parameter. In *Statistical decision theory and related topics III*. Elsevier, 1982, pp. 305–315.

[62] Winkler, G. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer-Verlag, Berlin, 1995.

[63] Wright, S. Evolution in mendelian populations. *Genetics 16*, 2 (1931), 97.