DOI: 10.1214/EJP.v12-413

Electronic Journal of Probability

Vol. 12 (2007), Paper no. 23, pages 661–683.

Journal URL
http://www.math.washington.edu/~ejpecp/

Occupation laws for some time-nonhomogeneous Markov chains*

Zach Dietz
Department of Mathematics
Tulane University
6823 St. Charles Ave., New Orleans, LA 70118
zdietz@math.tulane.edu

Sunder Sethuraman
Department of Mathematics
Iowa State University
396 Carver Hall, Ames, IA 50011
sethuram@iastate.edu

Abstract

We consider finite-state time-nonhomogeneous Markov chains whose transition matrix at time n is $I + G/n^{\zeta}$ where G is a "generator" matrix, that is $G(i,j) > 0$ for i,j distinct, and $G(i,i) = -\sum_{k \neq i} G(i,k)$, and $\zeta > 0$ is a strength parameter. In these chains, as time grows, the positions are less and less likely to change, and so form simple models of age-dependent time-reinforcing schemes. These chains, however, exhibit a trichotomy of occupation behaviors depending on parameters.

We show that the average occupation or empirical distribution vector up to time n, when variously $0 < \zeta < 1$, $\zeta > 1$ or $\zeta = 1$, converges in probability to a unique "stationary" vector ν_G, converges in law to a nontrivial mixture of point measures, or converges in law to a distribution μ_G with no atoms and full support on a simplex respectively, as $n \uparrow \infty$. This last type of limit can be interpreted as a sort of "spreading" between the cases $0 < \zeta < 1$ and $\zeta > 1$.

*Research supported in part by NSA-H982300510041 and NSF-DMS-0504193

In particular, when G is appropriately chosen, μ_G is a Dirichlet distribution, reminiscent of results in Pólya urns.

Key words: laws of large numbers, nonhomogeneous, Markov, occupation, reinforcement, Dirichlet distribution.

AMS 2000 Subject Classification: Primary 60J10; Secondary: 60F10.

Submitted to EJP on September 29 2006, final version accepted April 30 2007.

1 Introduction and Results

In this article, we study asymptotic occupation laws for a class of finite space time-nonhomogeneous Markov chains where, as time increases, positions are less likely to change. Although these chains feature simple age-dependent time-reinforcing dynamics, some different, perhaps unexpected, asymptotic occupation behaviors emerge depending on parameters. A specific case, as in Example 1.1, was first introduced in Gantert [7] in connection with the analysis of certain simulated annealing laws of large numbers phenomena.

Example 1.1. Suppose there are only two states 1 and 2, and that the chain moves between the two locations in the following way: At large times n, the chain switches places with probability c/n, and stays put with complementary probability $1 - c/n$ for $c > 0$. The chain, as it ages, is less inclined to leave its spot, but nonetheless switches infinitely often. It can be seen that the probability of being in state 1 tends to 1/2 regardless of the initial distribution. One may ask, however, how the average occupation, or frequency up to time n of state 1, $n^{-1}\sum_{i=1}^{n} 1_1(X_i)$, behaves asymptotically as $n \uparrow \infty$. For this example, it was shown in [7] and Ex. 4.7.1 [27], surprisingly, that the frequency could not converge to a constant, or even more generally converge in probability to a random variable, without further investigation of the limit properties. However, a quick consequence of our results is that the frequency of state 1 converges in law to the Beta(c, c) distribution (Theorem 1.4).

More specifically, we consider a general version of this scheme with $m \geq 2$ possible locations, and moving and staying probabilities $G(i,j)/n^\zeta$ and $1 - \sum_{k\neq i} G(i,k)/n^\zeta$ from $i \to j \neq i$ and $i \to i$ respectively at time n where $G = \{G(i,j)\}$ is an $m \times m$ matrix and $\zeta > 0$ is a strength parameter. After observing the location probability vector, $\langle \mathbb{P}_\pi^{G,\zeta}(X_n = k) : 1 \leq k \leq m\rangle$, tends to a vector $\nu_{G,\pi,\zeta}$, as $n \uparrow \infty$, which depends on G, ζ, and initial probability π when $\zeta > 1$, but does not depend on ζ and π, $\nu_{G,\pi,\zeta} = \nu_G$, when $\zeta \leq 1$ (Theorem 1.1), the results on the limit of the average occupation or empirical distribution vector, $n^{-1}\sum_{i=1}^{n}\langle 1_1(X_i), \ldots, 1_m(X_i)\rangle$, as $n \uparrow \infty$, separate into three cases depending on whether $0 < \zeta < 1$, $\zeta = 1$, or $\zeta > 1$.

When $0 < \zeta < 1$, following [7], the empirical distribution vector is seen to converge to ν_G in probability; and when more specifically $0 < \zeta < 1/2$, this convergence is proved to be a.s. When $\zeta > 1$, as there are only a finite number of switches, the position eventually stabilizes and the empirical distribution vector converges in law to a mixture of point measures (Theorem 1.2).

Our main results are when $\zeta = 1$. In this case, we show the empirical distribution vector converges in law to a non-atomic distribution μ_G, with full support on a simplex, identified by its moments (Theorems 1.3 and 1.5). When, in particular, G takes form $G(i,j) = \theta_j$ for all $i \neq j$, that is when the transititions into a state j are constant, μ_G takes the form of a Dirichlet distribution with parameters $\{\theta_j\}$ (Theorem 1.4). The proofs of these statements follow by the method of moments, and some surgeries of the paths.

The heuristic, with respect to the asymptotic empirical distribution behavior, is that when $0 < \zeta < 1$ the chance of switching is strong and sufficient mixing gives a deterministic limit, but when $\zeta > 1$ there is little movement and the chain gets stuck in finite time. The boundary case $\zeta = 1$ is the intermediate "spreading" situation leading to non-atomic limits. For example, with respect to Ex. 1.1, when the switching probability at time n is c/n^ζ, the Beta(c, c) limit when $\zeta = 1$ interpolates, as c varies on $(0, \infty)$, between the point measure at 1/2, the weak frequency

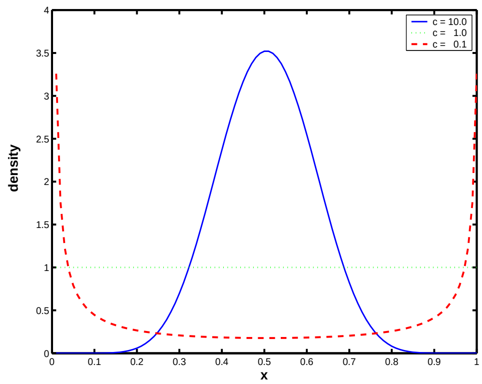

Figure 1: Beta(c, c) occupation law of state 1 in Ex. 1.1.

limit of state 1 when $0 < \zeta < 1$, and the fair mixture of point measures at 0 and 1, the limit when $\zeta > 1$ and starting at random (cf. Fig. 1).

In the literature, there are only a few results on laws of large numbers-type asymptotics for time-nonhomogeneous Markov chains, often related to simulated annealing and Metropolis algorithms which can be viewed in terms of a generalized model where $\zeta = \zeta(i, j)$ is a non-negative function. These results relate to the cases, "$\max \zeta(i, j) < 1$" or when the "landscape function" has a unique minimum, for which the average occupation or empirical distribution vector limits are constant [7], Ch. 7 [27], [10]. See also Ch. 1 [15], [18],[26]; and texts [5], [13],[14] for more on nonhomogeneous Markov chains. In this light, the non-degenerate limits μ_G found here seem to be novel objects. In terms of simulated annealing, these limits suggest a more complicated asymptotic empirical distribution picture at the "critical" cooling schedule when $\zeta(i, j) = 1$ for some pairs i, j in the state space with respect to general "landscapes."

The advent of Dirichlet limits, when G is chosen appropriately, seems of particular interest, given similar results for limit color-frequencies in Pólya urns [4], [8], as it hints at an even larger role for Dirichlet measures in related but different "reinforcement"-type models (see [16], [21], [20], and references therein, for more on urn and reinforcement schemes). In this context, the set of "spreading" limits μ_G in Theorem 1.3, in which Dirichlet measures are but a subset, appears intriguing as well (cf. Remarks 1.4, 1.5 and Fig. 2).

In another vein, although different, Ex. 1.1 seems not so far from the case of independent Bernoulli trials with success probability $1/n$ at the nth trial. For such trials much is known about the spacings between successes, and connections to GEM random allocation models and Poisson-Dirichlet measures [25], [1], [2], [3], [22], [23].

We also mention, in a different, neighbor setting, some interesting but distinct frequency limits have been shown for arrays of time-*homogeneous* Markov sequences where the transition matrix P_n for the nth row converges to a limit matrix P [6], [9], Section 5.3 [14]; see also [19] which comments on some "metastability" concerns.

We now develop some notation to state results. Let $\Sigma = \{1, 2, \ldots, m\}$ be a finite set of $m \geq 2$ points. We say a matrix $M = \{M(i, j) : 1 \leq i, j \leq m\}$ on Σ is a *generator* matrix if $M(i, j) \geq 0$ for all distinct $1 \leq i, j \leq m$, and $M(i, i) = -\sum_{j \neq i} M(i, j)$ for $1 \leq i \leq m$. In particular, M is a generator with *nonzero entries* if $M(i,j) > 0$ for $1 \le i,j \le m$ distinct, and $M(i,i) < 0$ for $1 \le i \le m$.

To avoid technicalities, e.g. with reducibility, we work with the following matrices,

$$\mathbb{G} \;=\; \Big\{ G \in \mathbb{R}^{m\times m} : G \text{ is a generator matrix with nonzero entries}\Big\},$$

although extensions should be possible for a larger class. For $G \in \mathbb{G}$, let $n(G,\zeta) = \lceil \max_{1\le i\le m} |G(i,i)|^{1/\zeta}\rceil$, and define for $\zeta > 0$

$$P_n^{G,\zeta} \;=\; \begin{cases} I & \text{for } 1 \le n \le n(G,\zeta) \\ I + G/n^{\zeta} & \text{for } n \ge n(G,\zeta)+1 \end{cases}$$

where I is the $m \times m$ identity matrix. Then, for all $n \ge 1$, $P_n^{G,\zeta}$ is ensured to be a stochastic matrix.

Let π be a distribution on Σ, and let $\mathbb{P}_\pi^{G,\zeta}$ be the (nonhomogeneous) Markov measure on the sequence space $\Sigma^{\mathbb{N}}$ with Borel sets $\mathcal{B}(\Sigma^{\mathbb{N}})$ corresponding to initial distribution π and transition kernels $\{P_n^{G,\zeta}\}$. That is, with respect to the coordinate process, $\mathbf{X} = \langle X_0, X_1, \ldots\rangle$, we have $\mathbb{P}_\pi^{G,\zeta}(X_0 = i) = \pi(i)$ and the Markov property

$$\mathbb{P}_\pi^{G,\zeta}(X_{n+1} = j | X_0, X_1, \ldots, X_{n-1}, X_n = i) = P_{n+1}^{G,\zeta}(i,j)$$

for all $i,j \in \Sigma$ and $n \ge 0$. Our convention then is that $P_{n+1}^{G,\zeta}$ controls "transitions" between times n and $n+1$. Let also $\mathbb{E}_\pi^{G,\zeta}$ be expectation with respect to $\mathbb{P}_\pi^{G,\zeta}$. More generally, E_μ denotes expectation with respect to measure μ.

Define the average occupation or empirical distribution vector, for $n \ge 1$,

$$\mathbf{Z}_n = \langle Z_{1,n}, \cdots, Z_{m,n}\rangle \quad \text{where} \quad Z_{k,n} \;=\; \frac{1}{n}\sum_{i=1}^{n} 1_k(X_i)$$

for $1 \le k \le m$. Then, $\mathbf{Z}_n$ is an element of the $m-1$-dimensional simplex,

$$\Delta_m \;=\; \Big\{ \mathbf{x} : \sum_{i=1}^{m} x_i = 1, 0 \le x_i \le 1 \text{ for } 1 \le i \le m \Big\}.$$

The first result is on convergence of the position of the process. For $G \in \mathbb{G}$, let ν_G be the stationary vector corresponding to G (of the associated continuous time homogeneous Markov chain), that is the unique left eigenvector, with positive entries, normalized to unit sum, of the eigenvalue 0.

Theorem 1.1. *For $G \in \mathbb{G}$, $\zeta > 0$, initial distribution π, and $k \in \Sigma$,*

$$\lim_{n\to\infty} \mathbb{P}_\pi^{G,\zeta}\big(X_n = k\big) \;=\; \nu_{G,\pi,\zeta}(k)$$

where $\nu_{G,\pi,\zeta}$ is a probability vector on Σ depending in general on ζ, G, and π. When $0 < \zeta \le 1$, $\nu_{G,\pi,\zeta}$ does not depend on π and ζ and reduces to $\nu_{G,\pi,\zeta} = \nu_G$.

Remark 1.1. For $\zeta > 1$, with only finitely many moves, actually X_n converges a.s. to a random variable with distribution $\nu_{G,\pi,\zeta}$. Also, $\nu_{G,\pi,\zeta}$ is explicit when $G = V_G D_G V_G^{-1}$ is diagonalizable with D_G diagonal and $D_G(i,i) = \lambda_i^G$, the ith eigenvalue of G, for $1 \leq i \leq m$. By calculation, $\nu_{G,\pi,\zeta} = \pi^t \prod_{n\geq 1} P_n^{G,\zeta} = \pi^t V_G D' V_G^{-1}$ with D' diagonal and $D'(i,i) = \prod_{n\geq n(G,\zeta)+1}(1+\lambda_i^G/n^\zeta)$.

We now consider the cases $\zeta \neq 1$ with respect to average occupation or empirical distribution vector limits. Let $\mathbf{i}$ be the basis vector $\mathbf{i} = \langle 0,\ldots,0,1,0,\ldots,0\rangle \in \Delta_m$ with a 1 in the ith component and $\delta_{\mathbf{i}}$ be the point measure at $\mathbf{i}$ for $1 \leq i \leq m$.

Theorem 1.2. *Let $G \in \mathbb{G}$, and π be an initial distribution. Under $\mathbb{P}_\pi^{G,\zeta}$, as $n \uparrow \infty$, we have that*

$$\mathbf{Z}_n \longrightarrow \nu_G$$

converges to the vector ν_G in probability when $0 < \zeta < 1$; when more specifically $0 < \zeta < 1/2$, this convergence is $\mathbb{P}_\pi^{G,\zeta}$-a.s.

However, when $\zeta > 1$, $\mathbf{Z}_n$ converges $\mathbb{P}_\pi^{G,\zeta}$ a.s. to a random variable, and

$$\lim_{n\to\infty} \mathbf{Z}_n \stackrel{d}{=} \sum_{i=1}^m \nu_{G,\pi,\zeta}(i)\delta_{\mathbf{i}} \ .$$

Remark 1.2. Simulations suggest that actually a.s. convergence might hold also on the range $1/2 \leq \zeta < 1$ (with worse convergence rates as $\zeta \uparrow 1$).

Let now $\gamma_1, \ldots, \gamma_m \geq 0$, be integers such that $\bar{\gamma} = \sum_{i=1}^m \gamma_i \geq 1$. Define the list $A = \{a_i : 1 \leq i \leq \bar{\gamma}\} = \{\underbrace{1,\ldots,1}_{\gamma_1}, \underbrace{2,\ldots,2}_{\gamma_2}, \ldots, \underbrace{m,\ldots,m}_{\gamma_m}\}$. Let $\mathbb{S}(\gamma_1,\ldots,\gamma_m)$ be the $\bar{\gamma}!$ permutations of A, although there are only $\binom{\bar{\gamma}}{\gamma_1,\gamma_2,\cdots,\gamma_m}$ distinct permutations; that is, each permutation appears $\prod_{k=1}^m \gamma_k!$ times.

Note also, for $G \in \mathbb{G}$, being a generator matrix, all eigenvalues of G have non-positive real parts (indeed, $I + G/k$ is a stochastic matrix for k large; then, by Perron-Frobenius, the real parts of its eigenvalues satisfy $-1 \leq 1 + \mathrm{Re}(\lambda_i^G)/k \leq 1$, yielding the non-positivity), and so the resolvent $(xI - G)^{-1}$ is well defined for $x \geq 1$.

We now come to our main results on the average occupation or empirical distribution vector limits when $\zeta = 1$

Theorem 1.3. *For $\zeta = 1$, $G \in \mathbb{G}$, and initial distribution π, we have under $\mathbb{P}_\pi^{G,\zeta}$ as $n \uparrow \infty$ that*

$$\mathbf{Z}_n \stackrel{d}{\longrightarrow} \mu_G$$

where μ_G is a measure on the simplex Δ_m characterized by its moments: For $1 \leq i \leq m$,

$$E_{\mu_G}\big(x_i\big) = \lim_{n\to\infty} \mathbb{E}_\pi^{G,\zeta}\big(Z_{i,n}\big) = \nu_G(i),$$

and for integers $\gamma_1, \ldots, \gamma_m \geq 0$ when $\bar{\gamma} \geq 2$,

$$\begin{aligned} E_{\mu_G}\Big(x_1^{\gamma_1}\cdots x_m^{\gamma_m}\Big) &= \lim_{n\to\infty} \mathbb{E}_\pi^{G,\zeta}\Big(Z_{1,n}^{\gamma_1}\cdots Z_{m,n}^{\gamma_m}\Big) \\ &= \frac{1}{\bar{\gamma}} \sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)} \nu_G(\sigma_1) \prod_{i=1}^{\bar{\gamma}-1} \Big(iI - G\Big)^{-1}(\sigma_i,\sigma_{i+1}). \end{aligned}$$

Remark 1.3. However, as in Ex. 1.1 and [7], when $\zeta = 1$ as above, $\mathbf{Z}_n$ cannot converge in probability to a random variable (as the tail field $\cap_n \sigma\{X_n, X_{n+1}, \ldots\}$ is trivial by Theorem 1.2.13 and Proposition 1.2.4 [15] and (2.3), but the limit distribution μ_G is not a point measure by say Theorem 1.5 below). This is in contrast to Pólya urns where the color frequencies converge a.s.

We now consider a particular matrix under which μ_G is a Dirichlet distribution. For $\theta_1, \ldots, \theta_m > 0$, define

$$\Theta = \begin{bmatrix} \theta_1 - \bar{\theta} & \theta_2 & \theta_3 & \cdots & \theta_m \\ \theta_1 & \theta_2 - \bar{\theta} & \theta_3 & \cdots & \theta_m \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_m - \bar{\theta} \end{bmatrix}$$

where $\bar{\theta} = \sum_{l=1}^m \theta_l$. It is clear $\Theta \in \mathbb{G}$. Recall identification of the Dirichlet distribution by its density and moments; see [17], [24] for more on these distributions. Namely, the Dirichlet distribution on the simplex Δ_m with parameters $\theta_1, \ldots, \theta_m$ (abbreviated as $\mathrm{Dir}(\theta_1, \ldots, \theta_m)$) has density

$$\frac{\Gamma(\bar{\theta})}{\Gamma(\theta_1)\cdots\Gamma(\theta_m)} x_1^{\theta_1 - 1} \cdots x_m^{\theta_m - 1}.$$

The moments with respect to integers $\gamma_1, \ldots, \gamma_m \geq 0$ with $\bar{\gamma} \geq 1$ are

$$E\left(x_1^{\gamma_1} \cdots x_m^{\gamma_m}\right) = \frac{\prod_{i=1}^m \theta_i(\theta_i + 1)\cdots(\theta_i + \gamma_i - 1)}{\prod_{i=0}^{\bar{\gamma}-1}(\bar{\theta} + i)}, \tag{1.1}$$

where we take $\theta_i(\theta_i + 1)\cdots(\theta_i + \gamma_i - 1) = 1$ when $\gamma_i = 0$.

Theorem 1.4. *We have* $\mu_\Theta = \mathrm{Dir}(\theta_1, \ldots, \theta_m)$.

Remark 1.4. Moreover, by comparing the first few moments in Theorem 1.3 with (1.1), one can check μ_G is not a Dirichlet measure for many G's with $m \geq 3$. However, when $m = 2$, then any G takes the form of Θ with $\theta_1 = G(2,1)$ and $\theta_2 = G(1,2)$, and so $\mu_G = \mathrm{Dir}(G(2,1), G(1,2))$.

We now characterize the measures $\{\mu_G : G \in \mathbb{G}\}$ as "spreading" measures different from the limits when $0 < \zeta < 1$ and $\zeta > 1$.

Theorem 1.5. *Let $G \in \mathbb{G}$. Then, (1) $\mu_G(U) > 0$ for any non-empty open set $U \subset \Delta_m$. Also, (2) μ_G has no atoms.*

Remark 1.5. We suspect better estimates in the proof of Theorem 1.5 will show μ_G is in fact mutually absolutely continuous with respect to Lebesgue measure on Δ_m. Of course, in this case, it would be of interest to find the density of μ_G. Meanwhile, we give two histograms, found by calculating 1000 sample averages, each on a run of time-length 10000 starting at random on Σ at time $n(G,1)$ (= 3, 1 respectively), in Figure 2 of the empirical density when $m = 3$ and G takes forms

$$G_{\text{left}} = \begin{bmatrix} -3 & 1 & 2 \\ 2 & -3 & 1 \\ 1 & 2 & -3 \end{bmatrix}, \quad \text{and} \quad G_{\text{right}} = \begin{bmatrix} -.4 & .2 & .2 \\ .3 & -.6 & .3 \\ .5 & .5 & -1 \end{bmatrix}.$$

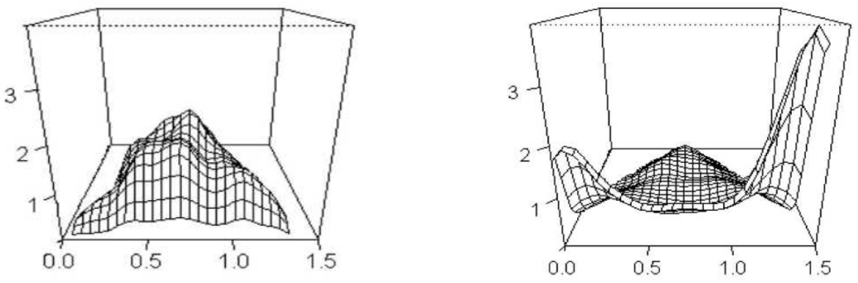

Figure 2: Empirical μ_G densities under G_{left} and G_{right} respectively.

To help visualize plots, Δ_3 is mapped to the plane by linear transformation $f(\mathbf{x}) = x_1 f(\langle 1,0,0\rangle) + x_2 f(\langle 0,1,0\rangle) + x_3 f(\langle 0,0,1\rangle)$ where $f(\langle 1,0,0\rangle) = \langle\sqrt{2},0\rangle$, $f(\langle 0,1,0\rangle) = \langle 0,0\rangle$ and $f(0,0,1) = \sqrt{2}\langle 1/2, \sqrt{3}/2\rangle$. The map maintains a distance $\sqrt{2}$ between the transformed vertices.

We now comment on the plan of the paper. The proofs of Theorems 1.1 and 1.2, 1.3, 1.4, and 1.5 (1) and (2) are in sections 2,3,4, 5, and 6 respectively. These sections do not depend structurally on each other.

2 Proofs of Theorems 1.1 and 1.2

We first recall some results for nonhomogeneous Markov chains in the literature. For a stochastic matrix P on Σ, define the "contraction coefficient"

$$\begin{aligned} c(P) &= \max_{x,y} \frac{1}{2} \sum_z \Big| P(x,z) - P(y,z) \Big| \\ &= 1 - \min_{x,y} \sum_z \min \Big\{ P(x,z), P(y,z) \Big\} \end{aligned} \tag{2.1}$$

The following is implied by Theorem 4.5.1 [27].

Proposition 2.1. *Let X_n be a time-nonhomogeneous Markov chain on Σ connected by transition matrices $\{P_n\}$ with corresponding stationary distributions $\{\nu_n\}$. Suppose, for some $n_0 \geq 1$, that*

$$\prod_{n=k}^{\infty} c(P_n) = 0 \text{ for all } k \geq n_0, \quad \textit{and} \quad \sum_{n=n_0}^{\infty} \|\nu_n - \nu_{n+1}\|_{\text{Var}} < \infty. \tag{2.2}$$

Then, $\nu = \lim_{n\to\infty} \nu_n$ exists, and, starting from any initial distribution π, we have for each $k \in \Sigma$ that

$$\lim_{n\to\infty} P(X_n = k) \ = \ \nu(k).$$

A version of the following is stated in Section 2 [7] as a consequence of results (1.2.22) and Theorem 1.2.23 in [15].

Proposition 2.2. *Given the setting of Proposition 2.1, suppose (2.2) is satisfied for some* $n_0 \geq 1$*. Define* $c_n = \max_{n_0 \leq i \leq n} c(P_i)$ *for* $n \geq n_0$*. Let also* π *be any initial distribution, and* f *be any function* $f : \Sigma \to \mathbb{R}$*. Then, we have convergence, as* $n \to \infty$*,*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \ \to \ E_\nu[f]$$

in the following senses:

(i) In probability, when $\lim_{n\to\infty} n(1-c_n) = \infty$*.*

(ii) a.s. when $\sum_{n\geq n_0} 2^{-n}(1-c_{2^n})^{-2} < \infty$ *(with convention the sum diverges if* $c_{2^n} = 1$ *for an* $n \geq n_0$*).*

Proof of Theorem 1.1. We first consider when $\zeta > 1$. In this case there are only a finite number of movements by Borel-Cantelli since $\sum_{n\geq 1} \mathbb{P}_\pi^{G,\zeta}(X_n \neq X_{n+1}) \leq C\sum_{n\geq 1} n^{-\zeta} < \infty$. Hence there is a time of last movement $N < \infty$ a.s. Then, $\lim_{n\to\infty} X_n = X_N$ a.s., and, for $k \in \Sigma$, the limit distribution $\nu_{G,\pi,\zeta}$ is defined and given by $\mathbb{P}_\pi^{G,\zeta}(X_N = k) = \nu_{G,\pi,\zeta}(k)$.

When $0 < \zeta \leq 1$, as $G \in \mathbb{G}$, by calculation with (2.1), $c(P_n^{G,\zeta}) = 1 - C_G/n^\zeta$, with respect to a constant $C_G > 0$, for all $n \geq n_0(G,\zeta)$ for an index $n_0(G,\zeta) > n(G,\zeta)$ large enough. Then, for $k \geq n_0(G,\zeta)$,

$$\prod_{n\geq k}\left(1 - \frac{C_G}{n^\zeta}\right) \ = \ 0. \tag{2.3}$$

Since for $n \geq n_0(G,\zeta)$, $\nu_G^t P_n^{G,\zeta} = \nu_G^t(I - G/n^\zeta) = \nu_G^t$, the last condition of Proposition 2.1 is trivially satisfied, and hence the result follows. □

Proof of Theorem 1.2. When $\zeta > 1$, as mentioned in the proof of Theorem 1.1, there are only a finite number of moves a.s., and so a.s. $\lim_{n\to\infty} \mathbf{Z}_n = \sum_{k=1}^m 1_{[X_N=k]}\mathbf{k}$ concentrates on basis vectors $\{\mathbf{k}\}$. Hence, as defined in proof of Theorem 1.1, $\mathbb{P}_\pi^{G,\zeta}(X_N = k) = \nu_{G,\pi,\zeta}(k)$, and the result follows.

When $0 < \zeta < 1$, we apply Proposition 2.2 and follow the method in [7]. First, recalling the proof of Theorem 1.1, (2.2) holds, and $c_n = \max_{n_0(G,\zeta)\leq i\leq n} c(P_i^{G,\zeta}) = 1 - C_G/n^\zeta$ for a constant $C_G > 0$ and $n \geq n_0(G,\zeta)$. Then, $n(1-c_n) = C_G n^{1-\zeta} \uparrow \infty$ to give the probability convergence in part (i). For a.s. convergence in part (ii) when $0 < \zeta < 1/2$, note

$$\sum_{n\geq n_0(G,\zeta)} \frac{1}{2^n(1-c_{2^n})^2} \ = \sum_{n\geq n_0(G,\zeta)} \frac{1}{2^n(C_G/(2^n)^\zeta)^2} \ = \sum_{n\geq n_0(G,\zeta)} \frac{1}{C_G^2(2^{1-2\zeta})^n} \ < \ \infty. \qquad \square$$

3 Proof of Theorem 1.3.

In this section, as $\zeta = 1$ is fixed, we suppress notational dependence on ζ. Also, as Z_n takes values on the compact set Δ_m, the weak convergence in Theorem 1.3 follows by convergence of the moments.

The next lemma establishes convergence of the first moments.

Lemma 3.1. *For $G \in \mathbb{G}$, $1 \le k \le m$, and initial distribution π,*

$$\lim_{n\to\infty} \mathbb{E}_\pi^G\Big(Z_{k,n}\Big) \;=\; \nu_G(k)$$

Proof. From Theorem 1.1, and Cesaro convergence,

$$\lim_n \mathbb{E}_\pi^G\Big(Z_{k,n}\Big) \;=\; \lim_n \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\pi^G\Big(1_k(X_i)\Big) \;=\; \lim_n \frac{1}{n}\sum_{i=1}^n \mathbb{P}_\pi^G(X_i = k) \;=\; \nu_G(k). \qquad \square$$

We now turn to the joint moment limits in several steps, and will assume in the following that $\gamma_1, \ldots, \gamma_m \ge 0$ with $\bar\gamma \ge 2$. The first step is an "ordering of terms."

Lemma 3.2. *For $G \in \mathbb{G}$, and initial distribution π, we have*

$$\lim_{n\to\infty} \Bigg| \mathbb{E}_\pi^G\Big(Z_{1,n}^{\gamma_1}\cdots Z_{m,n}^{\gamma_m}\Big) - \sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)} \frac{1}{n^{\bar\gamma}} \sum_{i_1=1}^{n-\bar\gamma+1}\sum_{i_2>i_1}^{n-\bar\gamma+2}\cdots\sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n} \mathbb{E}_\pi^G\Big(\prod_{l=1}^{\bar\gamma} 1_{\sigma_l}(X_{i_l})\Big)\Bigg| \;=\; 0.$$

Proof. By definition of $\mathbb{S}(\gamma_1, \ldots, \gamma_m)$,

$$\mathbb{E}_\pi^G\Big(Z_{1,n}^{\gamma_1}\cdots Z_{m,n}^{\gamma_m}\Big) \;=\; \frac{1}{\bar\gamma!}\frac{1}{n^{\bar\gamma}} \sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ 1\le i_1,\ldots,i_{\bar\gamma}\le n}} \mathbb{E}_\pi^G\Big(1_{\sigma_1}(X_{i_1})1_{\sigma_2}(X_{i_2})\cdots 1_{\sigma_{\bar\gamma}}(X_{i_{\bar\gamma}})\Big).$$

Note now

$$\sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ 1\le i_1,\ldots,i_{\bar\gamma}\le n}} 1 = \bar\gamma! n^{\bar\gamma}, \text{ and } \sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ 1\le i_1,\ldots,i_{\bar\gamma}\le n,\ \text{distinct}}} 1 = \bar\gamma!\bar\gamma!\binom{n}{\bar\gamma}.$$

Let $\mathcal{K}$ be those indices $\langle i_1, \ldots, i_{\bar\gamma}\rangle$, $1 \le i_1, \ldots, i_{\bar\gamma} \le n$ which are not distinct, that is $i_j = i_k$ for some $j \ne k$. Then,

$$\begin{aligned}
&\frac{1}{\bar\gamma!}\frac{1}{n^{\bar\gamma}} \Bigg| \sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ 1\le i_1,\ldots,i_{\bar\gamma}\le n}} \mathbb{E}_\pi^G\Big(\prod_{l=1}^{\bar\gamma} 1_{\sigma_l}(X_{i_l})\Big) - \sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ 1\le i_1,\ldots,i_{\bar\gamma}\le n,\ \text{distinct}}} \mathbb{E}_\pi^G\Big(\prod_{l=1}^{\bar\gamma} 1_{\sigma_l}(X_{i_l})\Big)\Bigg| \\
&= \frac{1}{\bar\gamma!}\frac{1}{n^{\bar\gamma}} \sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ \langle i_1,\ldots,i_{\bar\gamma}\rangle\in\mathcal{K}}} \mathbb{E}_\pi^G\Big(1_{\sigma_1}(X_{i_1})\cdots 1_{\sigma_{\bar\gamma}}(X_{i_{\bar\gamma}})\Big) \\
&\le \frac{1}{\bar\gamma!}\frac{1}{n^{\bar\gamma}}\Big(\bar\gamma! n^{\bar\gamma} - \bar\gamma!\bar\gamma!\binom{n}{\bar\gamma}\Big) \;=\; o(1).
\end{aligned}$$

But,

$$\sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ 1\le i_1,\ldots,i_{\bar\gamma}\le n,\ \text{distinct}}} \mathbb{E}_\pi^G\Big(\prod_{l=1}^{\bar\gamma} 1_{\sigma_l}(X_{i_l})\Big) \;=\; \bar\gamma! \sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ 1\le i_1<\cdots<i_{\bar\gamma}\le n}} \mathbb{E}_\pi^G\Big(\prod_{l=1}^{\bar\gamma} 1_{\sigma_l}(X_{i_l})\Big). \qquad \square$$

The next lemma replaces the initial measure with ν_G. Let $P_{i,j}^G = \prod_{l=i}^j P_l^G$ for $1 \le i \le j$.

Lemma 3.3. *For $G \in \mathbb{G}$ and initial distribution π, we have*

$$\lim_{n\to\infty}\Bigg|\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)}\frac{1}{n^{\bar\gamma}}\sum_{i_1=1}^{n-\bar\gamma+1}\sum_{i_2>i_1}^{n-\bar\gamma+2}\cdots\sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n}\mathbb{E}_\pi^G\Big(\prod_{l=1}^{\bar\gamma}1_{\sigma_l}(X_{i_l})\Big) \tag{3.1}$$
$$-\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)}\frac{\nu_G(\sigma_1)}{n^{\bar\gamma}}\sum_{i_1=1}^{n-\bar\gamma+1}\sum_{i_2>i_1}^{n-\bar\gamma+2}\cdots\sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n}\prod_{l=1}^{\bar\gamma-1}P^G_{i_l+1,i_{l+1}}(\sigma_l,\sigma_{l+1})\Bigg| \;=\; 0.$$

Proof. As $\mathbb{P}^G_\pi(X_j = t|X_i = s) = P^G_{i+1,j}(s,t)$ for $1 \le i < j$ and $s,t \in \Sigma$, we have

$$\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)}\frac{1}{n^{\bar\gamma}}\sum_{i_1=1}^{n-\bar\gamma+1}\sum_{i_2>i_1}^{n-\bar\gamma+2}\cdots\sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n}\mathbb{E}_\pi^G\Big(\prod_{l=1}^{\bar\gamma}1_{\sigma_l}(X_{i_l})\Big)$$
$$=\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)}\frac{1}{n^{\bar\gamma}}\sum_{i_1=1}^{n-\bar\gamma+1}\sum_{i_2>i_1}^{n-\bar\gamma+2}\cdots\sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n}\mathbb{P}_\pi^G(X_{i_1}=\sigma_1)\prod_{l=1}^{\bar\gamma-1}P^G_{i_l+1,i_{l+1}}(\sigma_l,\sigma_{l+1})$$

which differs from the second expression in (3.1) by at most

$$\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)}\frac{1}{n}\sum_{i_1=1}^{n-\bar\gamma+1}\Big|\mathbb{P}_\pi^G(X_{i_1}=\sigma_1)-\nu_G(\sigma_1)\Big|,$$

which vanishes by Theorem 1.1. □

We now focus on a useful class of diagonalizable matrices

$$\mathbb{G}^* \;=\; \Big\{G \in \mathbb{R}^m\times\mathbb{R}^m:\ \mathrm{Re}(\lambda_l^G) < 1 \text{ for } 1 \le l \le m, \text{ and } G \text{ is diagonalizable}\Big\}$$

where $\{\lambda_l^G\}$ are the eigenvalues of G. As $\mathrm{Re}(\lambda_l^G) \le 0$ for $1 \le l \le m$ when $G \in \mathbb{G}$, certainly all diagonalizable $G \in \mathbb{G}$ belong to $\mathbb{G}^*$. The relevance of this class, in the subsequent arguments, is that for $G \in \mathbb{G}^*$ the resolvent $(xI - G)^{-1}$ exists for $x \ge 1$.

For $G \in \mathbb{G}^*$, let V_G be the matrix of eigenvectors and D_G be a diagonal matrix with corresponding eigenvalue entries $D_G(i,i) = \lambda_i^G$ so that $G = V_G D_G V_G^{-1}$. Define also for $1 \le s,t,k \le m$,

$$g(k;s,t) \;=\; V_G(s,k)V_G^{-1}(k,t).$$

We also denote for $a_1,\ldots,a_m \in \mathbb{C}$, the diagonal matrix $\mathrm{Diag}(a_\cdot)$ with ith diagonal entry a_i for $1 \le i \le m$. We also extend the definitions of P_n^G and $P_{i,j}^G$ to $G \in \mathbb{G}^*$ with the same formulas. In the following, we use the principal value of the complex logarithm, and the usual convention $a^{b+ic} = e^{(b+ic)\log(a)}$ for $a,b,c \in \mathbb{R}$ with $a > 0$.

Lemma 3.4. *For $G \in \mathbb{G}^*$, $s,t \in \Sigma$, and $C \le i \le j$ where $C = C(G)$ is a large enough constant,*

$$P^G_{i,j}(s,t) \;=\; \sum_{k=1}^m \nu(k;i,j)g(k;s,t)\Big(\frac{j}{i-1}\Big)^{\lambda_k^G};$$

moreover, $\nu(k;i,j) \to 1$ as $i \uparrow \infty$ uniformly over k and j.

Proof. Straightforwardly,

$$P_{i,j}^G \; = \; V_G \prod_{k=i}^{j} \Big(I + \frac{1}{k} D_G \Big) V_G^{-1} \; = \; V_G \mathrm{Diag}\Big(\prod_{k=i}^{j} \Big(1 + \frac{\lambda_\cdot^G}{k} \Big) \Big) V_G^{-1}.$$

To expand further, we note for $z \in \mathbb{C}$ such that $|z-1| < 1$, we have

$$\log(z) \; = \; (z-1) + (z-1)^2 \sum_{n=0}^{\infty} (-1)^{n+1} \frac{1}{n+2} (z-1)^n.$$

and estimate

$$\Big| \sum_{n=0}^{\infty} (-1)^{n+1} \frac{1}{n+2} (z-1)^n \Big| \; \leq \; \sum_{n=0}^{\infty} |z-1|^n \; = \; \Big(1 - |z-1| \Big)^{-1}.$$

Let now L be so large such that $\max_{1 \leq u \leq m} |\lambda_u^G|/L < 1/2$. Then, for $1 \leq s \leq m$ and $k \geq L$,

$$\log \left(1 + \frac{\lambda_s^G}{k} \right) \; = \; \frac{\lambda_s^G}{k} + \left(\frac{\lambda_s^G}{k} \right)^2 C_{s,k}$$

for some $C_{s,k} \in \mathbb{C}$ with $|C_{s,k}| \leq (1 - \max_{1 \leq u \leq m} |\lambda_u^G|/L)^{-1} \leq 2$. Then, for $i \geq L$,

$$\prod_{k=i}^{j} \left(1 + \frac{\lambda_s^G}{k} \right) \; = \; \exp\left(\sum_{k=i}^{j} \log \left(1 + \frac{\lambda_s^G}{k} \right) \right) \; = \; \exp\left(\sum_{k=i}^{j} \frac{\lambda_s^G}{k} + c(s;i,j) \right)$$

where $c(s;i,j) = \sum_{k=i}^{j} (\lambda_s^G/k)^2 C_{s,k}$ satisfies

$$|c(s;i,j)| \; \leq \; 2 \max_{1 \leq u \leq m} |\lambda_u^G|^2 \sum_{k=i}^{\infty} \frac{1}{k^2} \; \to \; 0 \;\; \text{uniformly over } s \text{ and } j \text{ as } i \uparrow \infty.$$

Let now

$$d(s;i,j) \; = \; \lambda_s^G \Big(\sum_{k=i}^{j} \frac{1}{k} - \int_{i-1}^{j} \frac{dx}{x} \Big)$$

and note by the simple estimate

$$\sum_{k=i}^{j} \frac{1}{k} \; < \; \int_{i-1}^{j} \frac{dx}{x} \; < \; \sum_{k=i-1}^{j-1} \frac{1}{k}$$

that

$$|d(s;i,j)| \; \leq \; \max_{1 \leq u \leq m} |\lambda_u^G| \Big(\frac{1}{j} + \frac{1}{i-1} \Big) \; \leq \; \max_{1 \leq u \leq m} |\lambda_u^G| \Big(\frac{1}{i} + \frac{1}{i-1} \Big) \; \to \; 0$$

uniformly over j and s as $i \uparrow \infty$. This allows us to write

$$\prod_{k=i}^{j} \left(1 + \frac{\lambda_s^G}{k} \right) \; = \; \exp\Big(c(s;i,j) + d(s;i,j) \Big) \Big(\frac{j}{i-1} \Big)^{\lambda_s^G}.$$

Defining $\nu(s;i,j) = \exp(c(s;i,j) + d(s;i,j))$ gives after multiplying out that

$$\begin{aligned} P^G_{i,j} &= V_G \text{Diag}\left(\nu(\cdot;i,j)\left(\frac{j}{i-1}\right)^{\lambda^G_\cdot}\right)V_G^{-1} \\ &= \left[\sum_{k=1}^m \nu(k;i,j)g(k;s,t)\left(\frac{j}{i-1}\right)^{\lambda^G_k}\right]_{s,t\in\Sigma} \end{aligned}$$

completing the proof. □

The next lemma estimates a "boundary" contribution.

Lemma 3.5. *For $G \in \mathbb{G}$,*

$$\lim_{\epsilon\downarrow 0}\lim_{n\uparrow\infty} \sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)} \frac{\nu_G(\sigma_1)}{n^{\bar\gamma}} \sum_{i_1=1}^{\lfloor n\epsilon\rfloor} \sum_{i_2>i_1}^{n-\bar\gamma+2} \cdots \sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n} \prod_{l=1}^{\bar\gamma-1} P^G_{i_l+1,i_{l+1}}(\sigma_l,\sigma_{l+1}) \;=\; 0.$$

Proof. For any $\sigma \in \mathbb{S}(\gamma_1,\ldots,\gamma_m)$,

$$\begin{aligned} 0 \;&\le\; \lim_\epsilon \lim_n \frac{1}{n^{\bar\gamma}} \sum_{i_1=1}^{\lfloor n\epsilon\rfloor} \sum_{i_2>i_1}^{n-\bar\gamma+2} \cdots \sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n} \prod_{l=1}^{\bar\gamma-1} P^G_{i_l+1,i_{l+1}}(\sigma_l,\sigma_{l+1}) \\ &\le\; \lim_\epsilon \lim_n \frac{1}{n^{\bar\gamma}}(n\epsilon)n^{\bar\gamma-1} \;=\; 0. \end{aligned}$$

□

To continue, define for $G \in \mathbb{G}^*$ the function $T^G_{x,y}(s,t) : (0,1]^2 \times \Sigma^2 \to \mathbb{C}$ by

$$T^G_{x,y}(s,t) \;=\; \sum_{k=1}^m g(k;s,t)\left(\frac{x}{y}\right)^{-\lambda^G_k}.$$

Lemma 3.6. *For $G \in \mathbb{G}^*$, $\sigma \in \mathbb{S}(\gamma_1,\ldots,\gamma_m)$, and $\epsilon > 0$,*

$$\begin{aligned} &\lim_{n\uparrow\infty} \frac{1}{n^{\bar\gamma}} \sum_{i_1=\lfloor n\epsilon\rfloor+1}^{n-\bar\gamma+1} \sum_{i_2>i_1}^{n-\bar\gamma+2} \cdots \sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n} \prod_{l=1}^{\bar\gamma-1} P^G_{i_l+1,i_{l+1}}(\sigma_l,\sigma_{l+1}) \\ &= \int_{\epsilon\le x_1\le x_2\le\cdots\le x_{\bar\gamma}\le 1} \prod_{l=1}^{\bar\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l,\sigma_{l+1})\; dx_1dx_2\cdots dx_{\bar\gamma} \end{aligned}$$

Proof. From Lemma 3.4, as $\nu(s;i,j) \to 1$ as $i \uparrow \infty$ uniformly over j and s, $T_{x,y}(s,t)$ is bounded, continuous on $[\epsilon,1]^2$ for fixed s,t, and Riemann convergence, we have

$$\begin{aligned} &\lim_n \frac{1}{n^{\bar\gamma}} \sum_{i_1=\lfloor n\epsilon\rfloor+1}^{n-\bar\gamma+1} \sum_{i_2>i_1}^{n-\bar\gamma+2} \cdots \sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n} \prod_{l=1}^{\bar\gamma-1} P^G_{i_l+1,i_{l+1}}(\sigma_l,\sigma_{l+1}) \\ &= \lim_n \frac{1}{n^{\bar\gamma}} \sum_{i_1=\lfloor n\epsilon\rfloor+1}^{n-\bar\gamma+1} \cdots \sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n} \prod_{l=1}^{\bar\gamma-1} \sum_{k=1}^m \nu(k;i_l+1,i_{l+1})g(k;\sigma_l,\sigma_{l+1})\left(\frac{i_l/n}{i_{l+1}/n}\right)^{-\lambda^G_k} \\ &= \int_{\epsilon\le x_1\le x_2\le\cdots\le x_{\bar\gamma}\le 1} \prod_{l=1}^{\bar\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l,\sigma_{l+1})dx_1dx_2\cdots dx_{\bar\gamma}. \end{aligned}$$

□

Lemma 3.7. *For $G \in \mathbb{G}^*$ and $\sigma \in \mathbb{S}(\gamma_1, \ldots, \gamma_m)$,*

$$\lim_{\epsilon\downarrow 0} \int_{\epsilon\le x_1\le x_2\le\cdots\le x_{\bar\gamma}\le 1} \prod_{l=1}^{\bar\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l,\sigma_{l+1})\ dx_1dx_2\cdots dx_{\bar\gamma}$$
$$= \int_0^1\int_0^{x_{\bar\gamma}}\cdots\int_0^{x_2} T^G_{x_{\bar\gamma-1},x_{\bar\gamma}}(\sigma_{\bar\gamma-1},\sigma_{\bar\gamma})\cdots T^G_{x_1,x_2}(\sigma_1,\sigma_2)dx_1dx_2\cdots dx_{\bar\gamma}.$$

Proof. Let

$$f_\epsilon \ = \ 1_{\{\epsilon\le x_1\le x_2\le\cdots\le x_{\bar\gamma}\le 1\}} \prod_{l=1}^{\bar\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l,\sigma_{l+1}).$$

Then,

$$\lim_{\epsilon} f_\epsilon \ = \ 1_{\{0< x_1\le x_2\le\cdots\le x_{\bar\gamma}\le 1\}} \prod_{l=1}^{\bar\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l,\sigma_{l+1}),$$

and f_ϵ is uniformly bounded over ϵ as

$$|f_\epsilon| \ \le \ \bar f \ = \ 1_{\{0< x_1\le x_2\le\cdots\le x_{\bar\gamma}\le 1\}} \prod_{l=1}^{\bar\gamma-1}\sum_{k=1}^m |g(k;\sigma_l,\sigma_{l+1})|\Big(\frac{x_l}{x_{l+1}}\Big)^{-\mathrm{Re}(\lambda^G_k)}.$$

The right-hand bound is integrable: Indeed, by Tonelli's Lemma and induction, we have

$$\int \bar f dx_1\cdots dx_{\bar\gamma} \ = \ \int_0^1\int_0^{x_{\bar\gamma}}\cdots\int_0^{x_2} \prod_{l=1}^{\bar\gamma-1}\sum_{k=1}^m |g(k;\sigma_l,\sigma_{l+1})|\Big(\frac{x_l}{x_{l+1}}\Big)^{-\mathrm{Re}(\lambda^G_k)} dx_1\cdots dx_{\bar\gamma}$$
$$= \ \frac{1}{\bar\gamma}\prod_{l=1}^{\bar\gamma-1}\Big(\sum_{k=1}^m \frac{|g(k;\sigma_l,\sigma_{l+1})|}{l-\mathrm{Re}(\lambda^G_k)}\Big).$$

Hence, the lemma follows by dominated convergence and Fubini's Theorem. □

Lemma 3.8. *For $G \in \mathbb{G}^*$ and $\sigma \in \mathbb{S}(\gamma_1, \ldots, \gamma_m)$,*

$$\int_0^1\int_0^{x_{\bar\gamma}}\cdots\int_0^{x_2}\prod_{l=1}^{\bar\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l,\sigma_{l+1})dx_1\cdots dx_{\bar\gamma} \ = \ \frac{1}{\bar\gamma}\prod_{l=1}^{\bar\gamma-1}\Big(lI-G\Big)^{-1}(\sigma_l,\sigma_{l+1}).$$

Proof. By induction, the integral equals

$$\int_0^1\int_0^{x_{\bar\gamma}}\cdots\int_0^{x_2} T^G_{x_{\bar\gamma-1},x_{\bar\gamma}}(\sigma_{\bar\gamma-1},\sigma_{\bar\gamma})\cdots T^G_{x_1,x_2}(\sigma_1,\sigma_2)dx_1\cdots dx_{\bar\gamma}$$
$$= \ \frac{1}{\bar\gamma}\prod_{l=1}^{\bar\gamma-1}\Big(\sum_{k=1}^m \frac{g(k;\sigma_l,\sigma_{l+1})}{l-\lambda^G_k}\Big).$$

However, for $x \ge 1$, we have

$$\Big(xI-G\Big)^{-1}(s,t) \ = \ V_G\Big(xI-D_G\Big)^{-1}V_G^{-1}(s,t) \ = \ \sum_{k=1}^m \frac{g(k;s,t)}{x-\lambda^G_k}$$

to finish the identification. □

At this point, by straightforwardly combining the previous lemmas, we have proved Theorem 1.3 for $G \in \mathbb{G}$ diagonalizable. The method in extending to non-diagonalizable generators is accomplished by approximating with suitable "lower" and "upper" diagonal matrices.

Lemma 3.9. *For* $G \in \mathbb{G}$*,*

$$\lim_{n\to\infty} \sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)} \frac{\nu_G(\sigma_1)}{n^{\bar\gamma}} \sum_{i_1=1}^{n-\bar\gamma+1} \sum_{i_2>i_1}^{n-\bar\gamma+2} \cdots \sum_{i_{\bar\gamma}>i_{\bar\gamma-1}}^{n} \prod_{l=1}^{\bar\gamma-1} P^G_{i_l+1,i_{l+1}}(\sigma_l,\sigma_{l+1})$$
$$= \frac{1}{\bar\gamma} \sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)} \nu_G(\sigma_1) \prod_{l=1}^{\bar\gamma-1} \Big(lI - G\Big)^{-1}(\sigma_l,\sigma_{l+1}). \tag{3.2}$$

Proof. For an $m\times m$ matrix A, let $G[A] = G + A$. Let $\|\cdot\|_{\mathrm{M}}$ be the matrix norm $\|A\|_{\mathrm{M}} = \max\{|A(s,t)| : 1 \le s,t \le m\}$. Now, for small $\epsilon > 0$, choose matrices A_1 and A_2 with non-negative entries so that $\|A_1\|_{\mathrm{M}}, \|A_2\|_{\mathrm{M}} < \epsilon$, $I + G[-A_1]/l, I + G[A_2]/l$ have positive entries for all l large enough, and $G[-A_1], G[A_2] \in \mathbb{G}^*$: This last condition can be met as (1) the spectrum varies continuously with respect to the matrix norm $\|\cdot\|_{\mathrm{M}}$ (cf. Appendix D [12]), and (2) diagonalizable real matrices are dense (cf. Theorem 1 [11]).

Then, for $s,t \in \Sigma$, and l large enough, we have $0 < (I + G[-A_1]/l)(s,t) \le (I + G/l)(s,t) \le (I + G[A_2]/l)(s,t)$. Hence, for $i \le j$ with i large enough,

$$P^{G[-A_1]}_{i,j}(s,t) \;\le\; P^G_{i,j}(s,t) \;\le\; P^{G[A_2]}_{i,j}(s,t).$$

By Lemmas 3.5, 3.6, 3.7 and 3.8, the left-side of (3.2), that is in terms of liminf and limsup, is bounded below and above by

$$\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)} \frac{1}{\bar\gamma}\nu_G(\sigma_1) \prod_{l=1}^{\bar\gamma-1} \Big(lI - G[-A_1]\Big)^{-1}(\sigma_l,\sigma_{l+1}),$$

and

$$\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)} \frac{1}{\bar\gamma}\nu_G(\sigma_1) \prod_{l=1}^{\bar\gamma-1} \Big(lI - G[A_2]\Big)^{-1}(\sigma_l,\sigma_{l+1})$$

respectively. On the other hand, for $\sigma \in \mathbb{S}(\gamma_1,\ldots,\gamma_m)$, both

$$\prod_{l=1}^{\bar\gamma-1} \big(lI - G[-A_1]\big)^{-1}(\sigma_l,\sigma_{l+1}), \prod_{l=1}^{\bar\gamma-1} \big(lI - G[A_2]\big)^{-1}(\sigma_l,\sigma_{l+1}) \;\to\; \prod_{l=1}^{\bar\gamma-1} \big(lI - G\big)^{-1}(\sigma_l,\sigma_{l+1})$$

as $\epsilon \to 0$, completing the proof. □

4 Proof of Theorem 1.4

The proof follows by evaluating the moment expressions in Theorem 1.3 when $G = \Theta$ as those corresponding to the Dirichlet distribution with parameters $\theta_1,\ldots,\theta_m$ (1.1).

Lemma 4.1. *The stationary distribution ν_Θ is given by $\nu_\Theta(l) = \theta_l/\bar{\theta}$ for $l \in \Sigma$.*

Also, for $2 \leq l \leq \bar{\gamma}$, let F_l be the $m \times m$ matrix with entries

$$F_l(j,k) = \begin{cases} \theta_k & \text{for } k \neq j \\ \theta_j + l - 1 & \text{for } k = j. \end{cases}$$

Then,

$$\Big(lI - \Theta\Big)^{-1} = \frac{1}{l(l+\bar{\theta})} F_{l+1}.$$

Proof. The form of ν_Θ follows by inspection. For the second statement, write $F_{l+1} = lI + \hat{\Theta}$ where the matrix $\hat{\Theta}$ has ith column equal to $\theta_i(1,\ldots,1)^t$. Then, also $\Theta = \hat{\Theta} - \bar{\theta}I$. As $(1,\ldots,1)^t$ is an eigenvector of Θ with eigenvalue 0, we see $(lI - \Theta)(lI + \hat{\Theta}) = (l^2 + l\bar{\theta})I$ finishing the proof. □

The next statement is an immediate corollary of Theorem 1.3 and Lemma 4.1.

Lemma 4.2. *The μ_Θ-moments satisfy $E_{\mu_\Theta}[x_i] = \theta_i/\bar{\theta}$ for $1 \leq i \leq m$ and, when $\bar{\gamma} \geq 2$,*

$$\begin{aligned} E_{\mu_\Theta}\Big[\prod_{i=1}^m x_i^{\gamma_i}\Big] &= \sum_{\sigma \in \mathbb{S}(\gamma_1,\ldots,\gamma_m)} \nu_\Theta(\sigma_1)\frac{1}{\bar{\gamma}} \prod_{l=1}^{\bar{\gamma}-1} \Big(lI - \Theta\Big)^{-1}(\sigma_l, \sigma_{l+1}) \\ &= \sum_{\sigma \in \mathbb{S}(\gamma_1,\ldots,\gamma_m)} \frac{\theta_{\sigma_1}\prod_{l=2}^{\bar{\gamma}} F_l(\sigma_{l-1},\sigma_l)}{\bar{\gamma}!\prod_{l=0}^{\bar{\gamma}-1}(\bar{\theta}+l)}. \end{aligned}$$

We now evaluate the last expression of Lemma 4.2 by first specifying of the value of $\sigma_{\bar{\gamma}}$. Recall, by convention $\theta_l \cdots (\theta_l + \gamma_l - 1) = 1$ when $\gamma_l = 0$ for $1 \leq l \leq m$.

Lemma 4.3. *For $\bar{\gamma} \geq 2$ and $1 \leq k \leq m$,*

$$\sum_{\substack{\sigma \in \mathbb{S}(\gamma_1,\ldots,\gamma_m) \\ \sigma_{\bar{\gamma}} = k}} \theta_{\sigma_1} \prod_{l=2}^{\bar{\gamma}} F_l(\sigma_{l-1},\sigma_l) = \gamma_k(\bar{\gamma}-1)! \prod_{l=1}^m \theta_l \cdots (\theta_l + \gamma_l - 1). \tag{4.1}$$

Proof. The proof will be by induction on $\bar{\gamma}$.

Base Step: $\bar{\gamma} = 2$. If $\gamma_k = 1$ and $\gamma_i = 1$ for $i \neq k$, the left and right-sides of (4.1) both equal $\theta_i F_2(i,k) = \theta_i\theta_k$. If $\gamma_k = 2$, then the left and right-sides of (4.1) equal $2\theta_k F_2(k,k) = 2\theta_k(\theta_k + 1)$.

Induction Step. Without loss of generality and to ease notation, let $k = 1$. Then, by specifying the next-to-last element $\sigma_{\bar{\gamma}-1}$, and simple counting, we have

$$\begin{aligned} \sum_{\substack{\sigma \in \mathbb{S}(\gamma_1,\ldots,\gamma_m) \\ \sigma_{\bar{\gamma}} = 1}} \theta_{\sigma_1} \prod_{l=2}^{\bar{\gamma}} F_l(\sigma_l,\sigma_{l-1}) &= \gamma_1(\theta_1 + \bar{\gamma} - 1) \sum_{\substack{\sigma \in \mathbb{S}(\gamma_1-1,\ldots,\gamma_m) \\ \sigma_{\bar{\gamma}-1} = 1}} \theta_{\sigma_1} \prod_{l=2}^{\bar{\gamma}-1} F_l(\sigma_l,\sigma_{l-1}) \\ &\quad + \sum_{j=2}^m \gamma_1\theta_1 \sum_{\substack{\sigma \in \mathbb{S}(\gamma_1-1,\ldots,\gamma_m) \\ \sigma_{\bar{\gamma}-1} = j}} \theta_{\sigma_1} \prod_{l=2}^{\bar{\gamma}-1} F_l(\sigma_l,\sigma_{l-1}). \end{aligned}$$

We now use induction to evaluate the right-side above as

$$\begin{aligned}
&\theta_1\cdots(\theta_1+\gamma_1-2)\prod_{i=2}^m\theta_i\cdots(\theta_i+\gamma_i-1)\\
&\qquad\times\left\{\gamma_1(\theta_1+\bar\gamma-1)(\gamma_1-1)(\bar\gamma-2)!+\sum_{j=2}^m\gamma_1\theta_1\gamma_j(\bar\gamma-2)!\right\}\\
&= \theta_1\cdots(\theta_1+\gamma_1-2)\prod_{i=2}^m\theta_i\cdots(\theta_i+\gamma_i-1)\\
&\qquad\times\left\{\gamma_1(\theta_1+\bar\gamma-1)(\gamma_1-1)(\bar\gamma-2)!+\gamma_1\theta_1(\bar\gamma-\gamma_1)(\bar\gamma-2)!\right\}\\
&= \theta_1\cdots(\theta_1+\gamma_1-2)\prod_{i=2}^m\theta_i\cdots(\theta_i+\gamma_i-1)\\
&\qquad\times\gamma_1(\bar\gamma-2)!\left\{(\theta_1+\gamma_1-1)(\bar\gamma-1)\right\}\\
&= \gamma_1(\bar\gamma-1)!\prod_{l=1}^m\theta_l\cdots(\theta_l+\gamma_l-1).
\end{aligned}$$

□

By now adding over $1\le k\le m$ in the previous lemma, we finish the proof of Theorem 1.4.

Lemma 4.4. *When* $\bar\gamma\ge 2$,

$$\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)}\frac{\theta_{\sigma_1}\prod_{l=2}^{\bar\gamma}F_l(\sigma_{l-1},\sigma_l)}{\bar\gamma!\prod_{l=0}^{\bar\gamma-1}(\bar\theta+l)} = \frac{\prod_{l=1}^m\theta_l\cdots(\theta_l+\gamma_l-1)}{\prod_{l=0}^{\bar\gamma-1}(\bar\theta+l)}.$$

Proof.

$$\begin{aligned}
\sum_{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)}\frac{\theta_{\sigma_1}\prod_{l=2}^{\bar\gamma}F_l(\sigma_{l-1},\sigma_l)}{\bar\gamma!\prod_{l=0}^{\bar\gamma-1}(\bar\theta+l)} &= \sum_{k=1}^m\sum_{\substack{\sigma\in\mathbb{S}(\gamma_1,\ldots,\gamma_m)\\ \sigma_{\bar\gamma}=k}}\frac{\theta_{\sigma_1}\prod_{l=2}^{\bar\gamma}F_l(\sigma_{l-1},\sigma_l)}{\bar\gamma!\prod_{l=0}^{\bar\gamma-1}(\bar\theta+l)}\\
&= \frac{\sum_{k=1}^m\gamma_k(\bar\gamma-1)!}{\bar\gamma!}\frac{\prod_{l=1}^m\theta_l\cdots(\theta_l+\gamma_l-1)}{\prod_{l=0}^{\bar\gamma-1}(\bar\theta+l)}\\
&= \frac{\prod_{l=1}^m\theta_l\cdots(\theta_l+\gamma_l-1)}{\prod_{l=0}^{\bar\gamma-1}(\bar\theta+l)}.
\end{aligned}$$

□

5 Proof of Theorem 1.5 (1)

Let $\mathbf{p}=\langle p_1,\ldots,p_m\rangle\in\mathrm{Int}\Delta_m$ be a point in the simplex with $p_i>0$ for $1\le i\le m$. For $\epsilon>0$ small, let $B(\mathbf{p},\epsilon)\subset\mathrm{Int}\Delta_m$ be a ball with radius ϵ and center $\mathbf{p}$. To prove Theorem 1.5 (1), it is enough to show for all large n the lower bound

$$\mathbb{P}^G_\pi\Big(\mathbf{Z}_n\in B(\mathbf{p},\epsilon)\Big) > C(\mathbf{p},\epsilon) > 0.$$

To this end, let $\bar{p}_0 = 0$ and $\bar{p}_i = \sum_{l=1}^{i} p_l$ for $1 \le i \le m$. Also, define, for $1 \le k \le l$, $\mathbf{X}_k^l = \langle X_k, \ldots, X_l \rangle$. Then, there exist small $\delta, \beta > 0$ such that

$$\begin{aligned}
&\big\{\mathbf{Z}_n \in B(\mathbf{p}, \epsilon)\big\} \qquad (5.1)\\
&\supset \cup_{0 \le k_1, \ldots, k_m \le \lfloor n\beta \rfloor} \left\{ \left\{ \mathbf{X}_{\lfloor n\delta \rfloor}^{\lfloor n\bar{p}_1 \rfloor - k_1} = \vec{1} \right\} \cap \left(\cap_{j=2}^{m} \left\{ \mathbf{X}_{\lfloor n\bar{p}_{j-1} \rfloor - \bar{k}_{j-1}+1}^{\lfloor n\bar{p}_j \rfloor - \bar{k}_j} = \vec{j} \right\} \right) \right\}
\end{aligned}$$

where $\bar{k}_a = \sum_{l=1}^{a} k_l$, and $\vec{i}$ is a vector with all coordinates equal to i of the appropriate length. The last event represents the process being in the fixed location j for times $\lfloor n\bar{p}_{j-1} \rfloor - \bar{k}_{j-1} + 1$ to $\lfloor n\bar{p}_j \rfloor - \bar{k}_j$ for $1 \le j \le m$ where we take $1 - \bar{k}_0 = \lfloor n\delta \rfloor$.

Now, as G has strictly negative diagonal entries, $C_1 = \max_s |G(s,s)| > 0$, and so for all large n,

$$\mathbb{P}_\pi^G \left(\mathbf{X}_{\lfloor n\bar{p}_{i-1} \rfloor - \bar{k}_{i-1}+1}^{\lfloor n\bar{p}_i \rfloor - \bar{k}_i} = \vec{i} \big| \mathbf{X}_{\lfloor n\bar{p}_{i-1} \rfloor - \bar{k}_{i-1}+1} = i \right) \ \ge \ \prod_{j=\lfloor n\delta \rfloor}^{n} 1 - \frac{C_1}{j} \ \ge \ \frac{\delta^{C_1}}{2}.$$

Also, as G has positive nondiagonal entries, $C_2 = \min_s G(s, s+1) > 0$. Then,

$$\mathbb{P}_\pi^G \left(X_{\lfloor n\bar{p}_{i-1} \rfloor - \bar{k}_{i-1}+1} = i \big| X_{\lfloor n\bar{p}_{i-1} \rfloor - \bar{k}_{i-1}} = i - 1 \right) \ \ge \ \frac{C_2}{\lfloor n\bar{p}_{i-1} \rfloor - \bar{k}_{i-i} + 1}.$$

Hence, for all large n, as $\mathbb{P}_\pi^G(X_{\lfloor n\delta \rfloor} = 1) \ge \nu_G(1)/2$ (Theorem 1.1),

$$\begin{aligned}
&\mathbb{P}_\pi^G \Big(\mathbf{Z}_n \in B(\mathbf{p}, \epsilon) \Big) \\
&\ge \sum_{0 \le k_1, \ldots, k_m \le \lfloor n\beta \rfloor} \mathbb{P}_\pi^G \left(\left\{ \mathbf{X}_{\lfloor n\delta \rfloor}^{\lfloor np_1 \rfloor - k_1} = \vec{1} \right\} \cap \left(\cap_{j=2}^{m} \left\{ \mathbf{X}_{\lfloor n\bar{p}_{j-1} \rfloor - \bar{k}_{j-1}+1}^{\lfloor n\bar{p}_j \rfloor - \bar{k}_j} = \vec{j} \right\} \right) \right) \\
&\ge \left[\frac{\delta^{C_1}}{2} \right]^m \sum_{0 \le k_1, \ldots, k_m \le \lfloor n\beta \rfloor} \frac{\nu_G(1)}{2} \prod_{j=2}^{m} \frac{C_2}{\lfloor n\bar{p}_{j-1} \rfloor - \bar{k}_{j-1} + 1} \\
&\ge \left[\frac{\delta^{C_1}}{2} \right]^m \sum_{0 \le k_1, \ldots, k_m \le \lfloor n\beta \rfloor} \frac{\nu_G(1)}{2} \prod_{j=2}^{m} \frac{C_2}{\lfloor n\bar{p}_{j-1} \rfloor - k_{j-1} + 1} \\
&\ge \frac{\nu_G(1)}{4} \left[\frac{C_2 \delta^{C_1}}{2} \right]^m \prod_{j=2}^{m} \log \left(\frac{\lfloor n\bar{p}_{j-1} \rfloor}{\lfloor n\bar{p}_{j-1} \rfloor - \lfloor n\beta \rfloor} \right) \\
&\ge \frac{\nu_G(1)}{8} \left[\frac{C_2 \delta^{C_1}}{2} \right]^m \prod_{j=2}^{m} \log \left(\frac{\bar{p}_{j-1}}{\bar{p}_{j-1} - \beta} \right).
\end{aligned}$$

□

6 Proof of Theorem 1.5 (2)

The proof of Theorem 1.5 (2) follows from the next two propositions.

Proposition 6.1. *For $G \in \mathbb{G}$, the m vertices of Δ_m, $\mathbf{1}, \ldots, \mathbf{m}$, are not atoms.*

Proof. From Theorem 1.3, moments $\alpha_{l,k} = E_{\mu_G}[(x_l)^k]$ satisfy $\alpha_{l,k+1} = (I - G/k)^{-1}(l,l)\alpha_{l,k}$ for $1 \leq l \leq m$ and $k \geq 1$. By the inverse adjoint formula, for large k,

$$\left(I - G/k\right)^{-1}(l,l) \;=\; \frac{1 - \frac{1}{k}(\mathrm{Tr}(G) - G(l,l))}{1 - \mathrm{Tr}(G)/k} + O(k^{-2}) \;=\; 1 + \frac{G(l,l)}{k} + O(k^{-2}).$$

As $G \in \mathbb{G}$, $G(l,l) < 0$. Hence, $\alpha_{l,k}$ vanishes at polynomial rate $\alpha_{l,k} \sim k^{G(l,l)}$. In particular, as $\mu_G(\{\mathbf{l}\}) \leq E_{\mu_G}[(x_l)^k] \to 0$ as $k \to \infty$, the point $\mathbf{l}$ cannot be an atom of the limit distribution. $\square$

Fix for the remainder $\mathbf{p} \in \Delta_m \setminus \{\mathbf{1}, \ldots, \mathbf{m}\}$, and define $\check{p} = \min\{p_i : p_i > 0, 1 \leq i \leq m\} > 0$. Let also $0 < \delta < \check{p}/2$, and consider $B(\mathbf{p}, \delta) = \{\mathbf{x} \in \Delta_m : |\mathbf{p} - \mathbf{x}| < \delta\}$.

Proposition 6.2. *For $G \in \mathbb{G}$, there is a constant $C = C(G, \mathbf{p}, m)$ such that*

$$\mu_G\Big(B(\mathbf{p}, \delta)\Big) \;\leq\; C \;\log\left(\frac{\check{p} + 2\delta}{\check{p} - \delta}\right).$$

Before proving Proposition 6.2, we will need some notation and lemmas. We will say a "switch" occurs at time $1 < k \leq n$ in the sequence $\omega^n = \langle \omega_1, \ldots, \omega_n \rangle \in \Sigma^n$ if $\omega_{k-1} \neq \omega_k$. For $0 \leq j \leq n-1$, let

$$T(j) \;=\; \Big\{\omega^n : \omega^n \text{ has exactly } j \text{ switches}\Big\}.$$

Note as $\mathbf{p} \in \Delta_m \setminus \{\mathbf{1}, \ldots, \mathbf{m}\}$ at least two coordinates of $\mathbf{p}$ are positive. Then, as $\delta < \check{p}/2$, when $(1/n)\sum_{i=1}^n \langle 1_1(\omega_i), \ldots, 1_m(\omega_i)\rangle \in B(\mathbf{p}, \delta)$, at least one switch is in ω^n.

For $j \geq 1$ and a path in $T(j)$, let $\alpha_1, \ldots, \alpha_j$ denote the j switch times in the sequence; let also $\theta_1, \ldots, \theta_{j+1}$ be the $j+1$ locations visited by the sequence. We now partition $\{\omega^n : (1/n)\sum_{i=1}^n \langle 1_1(\omega_i), \ldots, 1_m(\omega_i)\rangle \in B(\mathbf{p}, \delta)\} \cap T(j)$ into non-empty sets $A_j(\mathbf{U}, \mathbf{V})$ where $\mathbf{U} = \langle U_1, \ldots, U_{j-1}\rangle$ and $\mathbf{V} = \langle V_1, \ldots, V_{j+1}\rangle$ denote possible switch times (up to the $j-1$st switch time) and visit locations respectively:

$$\begin{aligned} A_j(\mathbf{U}, \mathbf{V}) \;=\; \Big\{\omega^n &: \omega^n \in T(j), \frac{1}{n}\sum_{i=1}^n \langle 1_1(\omega_i), \ldots, 1_m(\omega_i)\rangle \in B(\mathbf{p}, \delta), \\ &\alpha_i = U_i, \theta_k = V_k \text{ for } 1 \leq i \leq j-1, 1 \leq k \leq j+1\Big\}. \end{aligned}$$

In this decomposition, paths in $A_j(\mathbf{U}, \mathbf{V})$ are in $1:1$ correspondence with jth switch times α_j–the only feature allowed to vary.

Now, for each set $A_j(\mathbf{U}, \mathbf{V})$, we define a path $\eta(j, \mathbf{U}, \mathbf{V}) = \langle \eta_1, \ldots, \eta_n\rangle$ where the last jth switch is "removed,"

$$\eta_l = \begin{cases} V_1 & \text{for } 1 \leq l < U_1 \\ V_k & \text{for } U_{k-1} \leq l < U_k, 2 \leq k \leq j-1 \\ V_j & \text{for } U_{j-1} \leq l \leq n. \end{cases}$$

Note that the sequence $\eta(j, \mathbf{U}, \mathbf{V})$ belongs to $T(j-1)$, can be obtained no matter the location V_{j+1} (which could range on the m values in the state space), and is in $1:1$ correspondence with pair $\langle U_1, \ldots, U_{j-1}\rangle$ and $\langle V_1, \ldots, V_j\rangle$. In particular, recalling $\mathbf{X}_1^n = \langle X_1, \ldots, X_n\rangle$ denotes the coordinate sequence up to time n, we have

$$\sum_{\mathbf{U},\mathbf{V}} \mathbb{P}_\pi^G\Big(\mathbf{X}_1^n = \eta(j, \mathbf{U}, \mathbf{V})\Big) \ \le\ m\ \mathbb{P}_\pi^G\Big(\mathbf{X}_1^n \in T(j-1)\Big) \tag{6.1}$$

where the sum is over all $\mathbf{U}, \mathbf{V}$ corresponding to the decomposition into sets $A_j(\mathbf{U}, \mathbf{V})$ of $\{\omega^n : (1/n)\sum_{i=1}^n \langle 1_1(\omega_i), \ldots, 1_m(\omega_i)\rangle \in B(\mathbf{p}, \delta)\} \cap T(j)$.

The next lemma estimates the location of the last switch time α_j, and the size of the set $A_j(\mathbf{U}, \mathbf{V})$. The proof is deferred to the end.

Lemma 6.1. *On $A_j(\mathbf{U}, \mathbf{V})$, we have $\lceil n(\check{p} - \delta) + 1\rceil \le \alpha_j$. Also, $|A_j(\mathbf{U}, \mathbf{V})| \le \lfloor 2n\delta + 1\rfloor$.*

A consequence of these bounds on the position and cardinality of α_j's associated to a fixed set $A_j(\mathbf{U}, \mathbf{V})$, is that

$$\sum{}' \frac{1}{U_j} \ \le \sum_{k=\lceil n(\check{p}-\delta)+1\rceil}^{\lceil n(\check{p}+\delta)+2\rceil} \frac{1}{k} \ \le\ \log\left(\frac{\check{p} + \delta + 3/n}{\check{p} - \delta}\right) \tag{6.2}$$

where $\sum'$ refers to adding over all last switch times U_j associated to paths in $A_j(\mathbf{U}, \mathbf{V})$.

Let now $\hat{G} = \max\{|G(i,j)| : 1 \le i, j \le m\}$.

Lemma 6.2. *For $\omega^n \in A_j(\mathbf{U}, \mathbf{V})$ such that $\alpha_j = U_j$, and all large n, we have*

$$\mathbb{P}_\pi^G\Big(\mathbf{X}_1^n = \omega^n\Big) \ \le\ \frac{\hat{G}(\check{p}/2)^{-2\hat{G}}}{U_j}\ \mathbb{P}_\pi^G\Big(\mathbf{X}^n = \eta(j, \mathbf{U}, \mathbf{V})\Big). \tag{6.3}$$

Proof. The path $\eta(j, \mathbf{U}, \mathbf{V})$ differs from ω^n only in that there is no switch at time U_j. Hence,

$$\frac{\mathbb{P}_\pi^G\big(\mathbf{X}^n = \omega^n\big)}{\mathbb{P}_\pi^G\big(\mathbf{X}^n = \eta(j, \mathbf{U}, \mathbf{V})\big)} \ =\ \frac{G(V_j, V_{j+1})}{U_j(1 + G(V_j, V_j)/U_j)} \prod_{l=U_j+1}^{n} \left(\frac{1 + G(V_{j+1}, V_{j+1})/l}{1 + G(V_j, V_j)/l}\right).$$

Now bounding $G(V_j, V_{j+1}) \le \hat{G}$, $1 + G(V_{j+1}, V_{j+1})/l \le 1$, $1 + G(V_j, V_j)/l \ge 1 - \hat{G}/l$, and noting $U_j \ge n(\check{p} - \delta) + 1$ (by Lemma 6.1), $-\ln(1-x) \le 2x$ for $x > 0$ small, and $\delta < \check{p}/2$, give for large n,

$$\frac{G(V_j, V_{j+1})}{1 + G(V_j, V_j)/U_j} \prod_{l=U_j+1}^{n} \left(\frac{1 + G(V_{j+1}, V_{j+1})/l}{1 + G(V_j, V_j)/l}\right) \le \hat{G}\left(\frac{n}{n(\check{p} - \delta)}\right)^{2\hat{G}} \le \hat{G}(\check{p}/2)^{-2\hat{G}}. \quad \square$$

Proof of Proposition 6.2. By decomposing over number of switches j and on the structure of the paths with j switches, estimates (6.3), (6.2), comment (6.1), and $\sum_j \mathbb{P}_\pi^G\big(\mathbf{X}^n \in T(j-1)\big) \le 1$, we have for all large n,

$$\begin{aligned}
\mathbb{P}^G_\pi\Big(\mathbf{Z}_n \in B(\mathbf{p},\delta)\Big) &= \sum_{j=1}^{n-1}\mathbb{P}^G_\pi\Big(\mathbf{Z}_n \in B(\mathbf{p},\delta), \mathbf{X}^n \in T(j)\Big)\\
&= \sum_{j=1}^{n-1}\sum_{\mathbf{U},\mathbf{V}}\mathbb{P}^G_\pi\Big(A_j(\mathbf{U},\mathbf{V})\Big)\\
&\le \sum_{j=1}^{n-1}\sum_{\mathbf{U},\mathbf{V}}\sum{}' \ \frac{C(G,\mathbf{p})}{U_j}\ \mathbb{P}^G_\pi\Big(\mathbf{X}^n = \eta(j,\mathbf{U},\mathbf{V})\Big)\\
&\le C(G,\mathbf{p})\log\left(\frac{\check{p}+2\delta}{\check{p}-\delta}\right)\sum_{j=1}^{n-1}\sum_{\mathbf{U},\mathbf{V}}\ \mathbb{P}^G_\pi\Big(\mathbf{X}^n = \eta(j,\mathbf{U},\mathbf{V})\Big)\\
&\le mC(G,\mathbf{p})\log\left(\frac{\check{p}+2\delta}{\check{p}-\delta}\right)\sum_{j=1}^{n-1}\ \mathbb{P}^G_\pi\Big(\mathbf{X}^n \in T(j-1)\Big)\\
&\le C(G,\mathbf{p},m)\log\left(\frac{\check{p}+2\delta}{\check{p}-\delta}\right).
\end{aligned}$$

The proposition follows by taking limit on n, and weak convergence. □

Proof of Lemma 6.1. For a path $\omega^n \in A_j(\mathbf{U},\mathbf{V})$ and $1 \le k \le j+1$, let τ_k be the number of visits to state V_k (some τ_k's may be the same if V_k is repeated). For $1 \le i \le \tau_k$, let $\underline{n}^k_i$ and $\overline{n}^k_i$ be the start and end of the ith visit to V_k. Certainly, $\sum_{i=1}^n 1_{V_k}(\omega_i) = \sum_{i=1}^{\tau_k}(\overline{n}^k_i - \underline{n}^k_i + 1)$. Moreover, as $(1/n)\sum_{i=1}^n\langle 1_1(\omega_i),\ldots,1_m(\omega_i)\rangle \in B(\mathbf{p},\delta)$, we have $|(1/n)\sum_{i=1}^n 1_{V_k}(\omega_i) - p_{V_k}| \le \delta$, and so

$$n(p_{V_k}-\delta) \ \le\ \sum_{i=1}^{\tau_k}\left(\overline{n}^k_i - \underline{n}^k_i + 1\right) \ \le\ n(p_{V_k}+\delta). \tag{6.4}$$

Hence, as the disjoint sojourns $\{[\underline{n}^k_i, \overline{n}^k_i] : 1 \le i \le \tau_k\}$ occur between times 1 and $\overline{n}^k_{\tau_k}$, their total sum length is less than $\overline{n}^k_{\tau_k}$, and we deduce $n(p_{V_k}-\delta) \le \overline{n}^k_{\tau_k}$.

Now, for $\mathbf{p} \in \Delta_m \setminus \{\mathbf{1},\ldots,\mathbf{m}\}$, at least one of the $\{p_{V_i} : V_i \ne V_{j+1}, 1 \le i \le j\}$ is positive: Indeed, there are two coordinates of $\mathbf{p}$, say p_s and p_t, which are positive. Say $V_{j+1} \ne s$; then, as $(1/n)\sum_{i=1}^n 1_s(\omega_i) = (1/n)\sum_{i=1}^{\alpha_j-1} 1_s(\omega_i)$, $|(1/n)\sum_{i=1}^{\alpha_j-1} 1_s(\omega_i) - p_s| \le \delta$, and $p_s - \delta > 0$, the path must visit state s before time α_j, e.g. $V_i = s$ for some $1 \le i \le j$.

Then, from the deduction just after (6.4), we have

$$n(\check{p}-\delta) \ \le\ n\max_{\substack{V_i\ne V_{j+1}\\ 1\le i\le j}}(p_{V_i}-\delta) \ \le\ \max_{\substack{V_i\ne V_{j+1}\\ 1\le i\le j}}\overline{n}^i_{\tau_i} \ \le\ \overline{n}^j_{\tau_j} \ =\ \alpha_j - 1$$

giving the first statement.

For the second statement, note that $-\underline{n}^j_{\tau_j} + \sum_{i=1}^{\tau_j-1}(\overline{n}^j_i - \underline{n}^j_i + 1)$ (with convention the sum vanishes when $\tau_j = 1$) is independent of paths in $A_j(\mathbf{U},\mathbf{V})$ being some combination of $\{U_i : 1 \le i \le j-1\}$. Hence, with $k = j$ in (6.4), we observe $\alpha_j = \overline{n}^j_{\tau_j} + 1$ takes on at most $\lfloor 2n\delta + 1\rfloor$ distinct values. The result now follows as paths in $A_j(\mathbf{U},\mathbf{V})$ are in $1:1$ correspondence with last switch times α_j. □

Acknowledgement. We thank M. Balázs and J. Pitman for helpful communications, and the referees for their constructive comments.

References

[1] Arratia, R. (1998) On the central role of scale invariant Poisson processes on $(0,\infty)$. Microsurveys in discrete probability (Princeton, NJ, 1997), 21-41, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., **41**, Amer. Math. Soc., Providence, RI MR1630407

[2] Arratia, R., Barbour, A. D., Tavaré, S. (1999) On Poisson-Dirichlet limits for random decomposable combinatorial structures. *Combin. Probab. Comput.* **8** 193–208. MR1702562

[3] Arratia, R., Barbour, A. D., Tavaré, S. (2003) *Logarithmic Combinatorial Structures: A Probabilistic Approach* EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich. MR2032426

[4] Athreya, K.B. (1969) On a characteristic property of Pólya's urn. *Studia Sci. Math. Hungar.* **4** 31-35. MR0247643

[5] Brémaud, P. (1999) *Markov chains. Gibbs fields, Monte Carlo simulation, and Queues.* Texts in Applied Mathematics **31**. Springer-Verlag, New York MR1689633

[6] Dobrushin, R. (1953) Limit theorems for Markov chains with two states. (Russian) *Izv. Adad. Nauk SSSR* **17:4** 291-330. MR0058150

[7] Gantert, N. (1990) Laws of large numbers for the annealing algorithm. *Stochastic Process. Appl.* **35** 309-313. MR1067115

[8] Gouet, R. (1997) Strong convergence of proportions in a multicolor Pólya urn. *J. Appl. Probab.* **34** 426-435. MR1447347

[9] Hanen, A. (1963) Théorèmes limites pour une suite de chaînes de Markov. *Ann. Inst. H. Poincaré* **18** 197-301. MR0168017

[10] Hannig, J., Chong, E.K.P., Kulkarni, S.R. (2006) Relative frequencies of generalized simulated annealing. *Math. Oper. Res.* **31** 199-216. MR2205528

[11] Hartfiel, D. J.(1995) Dense sets of diagonalizable matrices. *Proc. Amer. Math. Soc.* **123** 1669–1672. MR1264813

[12] Horn, R. A., Johnson, C. R. (1990) *Matrix Analysis.* Corrected reprint of the 1985 original. Cambridge University Press, Cambridge. MR1084815

[13] Isaacson, D.L., and Madsen, R.W. (1976) *Markov Chains. Theory and Applications.* John Wiley and Sons, New York. MR0407991

[14] Iosifescu, M. (1980) *Finite Markov Processes and Their Applications.* John Wiley and Sons, New York. MR0587116

[15] Iosifescu, M., and Theodorescu, R. (1969) *Random Processes and Learning.* Springer-Verlag, Berlin. MR0293704

[16] Kotz, S., Balakrishnan, N. (1997) Advances in urn models during the past two decades. In *Advances in Combinatorial Methods and Applications to Probability and Statistics,* 203–257, Stat. Ind. Technol., Birkhäuser Boston, Boston, MA MR1456736

[17] Kotz, S., Balakrishnan, N., Johnson, N.L. (2000) *Continuous Multivariate Distributions* John Wiley and Sons, New York.

[18] Liu, Wen, Liu, Guoxin (1995) A class of strong laws for functionals of countable nonhomogeneous Markov chains. *Stat. and Prob. Letters* **22** 87-96. MR1327733

[19] Miclo, L. (1998) Sur les temps d'occupations des processus de Markov finis inhomogènes à basse température. (French) *Stochastics Stochastics Rep.* **63** 65-137. MR1639780

[20] Del Moral, P., Miclo, L. (2006) Self-interacting Markov chains. *Stoch. Anal. Appl.* **24** 615-660. MR2220075

[21] Pemantle, R. (2007) A survey of random processes with reinforcement. *Probability Surveys* **4** 1-79. MR2282181

[22] Pitman, J. (1996) Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory; Papers in honor of David Blackwell, Ed. T.S. Ferguson et al.* Institute of Mathematical Statistics Lecture Notes-Monograph Series **30** 245-267. MR1481784

[23] Pitman, J. (2006) *Combinatorial stochastic processes.* Lecture Notes in Mathematics, **1875** Springer-Verlag, Berlin. MR2245368

[24] Sethuraman, J. (1994) A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

[25] Vervaat, W. (1972) *Success Epochs in Bernoulli Trials.* Math. Centre Tracts, Amsterdam MR0328989

[26] Wen, Liu, Weiguo, Yang (1996) An extension of Shannon-McMillan theorem and some limit properties for nonhomogeneous Markov chains. *Stoch. Proc. Appl.* **61** 129-145. MR1378852

[27] Winkler, G. (2003) *Image Analysis, Random Fields and Dynamic Monte Carlo Methods.* Second edition. Springer-Verlag, New York. MR1316400