Occupation laws for some time-nonhomogeneous Markov chains

Zach Dietz\textsuperscript{1} and Sunder Sethuraman\textsuperscript{2}

August 5, 2018

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 some different, perhaps unexpected, occupation behaviors depending on parameters.

Although it is shown, on the one hand, that the position at time $n$ converges to a point-mixture for all $\zeta > 0$, on the other hand, the average occupation vector up to time $n$, when variously $0 < \zeta < 1$, $\zeta > 1$ or $\zeta = 1$, is seen to converge to a constant, a point-mixture, or a distribution $\mu_G$ 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 $\zeta > 1$.

In particular, when $G$ is appropriately chosen, intriguingly, $\mu_G$ is a Dirichlet distribution, reminiscent of results in Pólya urns.

Research supported in part by nsa-lh982300510041 and NSF-DMS-0504193

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

Abbreviated title: Occupation laws for nonhomogeneous Markov chains.

AMS (2000) subject classifications: Primary 60J10; secondary 60F10.

\textsuperscript{1} Department of Mathematics, Tulane University, 6823 St. Charles Ave., New Orleans, LA 70118; zdietz@math.tulane.edu.

\textsuperscript{2} Department of Mathematics, Iowa State University, 396 Carver Hall, Ames, IA 50011; sethuram@iastate.edu.

1 Introduction and Results

In this article, we study laws of large numbers (LLN) 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, LLN occupation behaviors emerge depending on parameters. A specific case, as in Example 1.1, was first introduced in Gantert \cite{8} in connection with analysis of certain simulated annealing LLN 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. One can see the probability of being in state 1 tends to $1/2$ regardless of the initial distribution. One may ask, however, how the average location, or frequency, of state 1 behaves asymptotically. For this example, it was shown in \cite{8} and Ex. 7.1.1. \cite{28}, perhaps surprisingly, that any LLN limit could not be a constant, or even converge in probability, without further identification. However, a quick consequence of our results is that the average occupation limit of state 1 converges weakly 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 $i \rightarrow j \neq i$ and $i \rightarrow i$ 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 probabilities tend to a distribution which depends on $G$, $\zeta$, and initial probability $\pi$ when $\zeta > 1$, but does not depend on $\zeta$ and $\pi$ when $\zeta \leq 1$ (Theorem 1.1), the results on the average occupation vector limit separate roughly into three cases depending on whether $0 < \zeta < 1$, $\zeta = 1$, or $\zeta > 1$.

When $0 < \zeta < 1$, following \cite{8}, the average occupation is seen to converge to a constant 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 average occupation converges to a mixture of point masses (Theorem 1.2).

Our main results are when $\zeta = 1$. In this case, we show the average occupation converges to a non-atomic distribution $\mu_G$, 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 transitions into a state $j$ are constant, $\mu_G$ 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 is that when $0 < \zeta < 1$ the chance of switching is strong and sufficient mixing leads to constant limits, but when $\zeta > 1$ there is little movement giving point-mixture limits. The 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 $c/n^\zeta$, the Beta($c, c$) limit when $\zeta = 1$ interpolates, as $c$ varies on $(0, \infty)$, between the point-mass at 1/2, the frequency limit of state 1 when $0 < \zeta < 1$, and the
fair mixture of point-masses 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 LLN’s 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 case “$\max \zeta(i, j) < 1$” when the LLN limit is a constant [8], Ch. 7 [28], [9]. See also Ch. 1 [16], [19], [20]; and texts [6], [14], [15] for more on nonhomogeneous Markov chains. In this light, the non-degenerate limits $\mu_G$ found here seem to be novel objects. In terms of simulated annealing, these limits suggest a more complicated LLN picture at the “critical” cooling schedule when $\zeta(i, j) = 1$ for some pairs $i, j$ in the state space.

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], [10], as it hints at an even larger role for Dirichlet measures in related but different “reinforcement”-type models (see [17], [23], [22], and references therein, for more on urn and reinforcement schemes). In this context, the set of “spreading” limits $\mu_G$ 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 $n$th trial. For such trials much is known about the spacings between successes, and connections to GEM random allocation models and Poisson-Dirichlet measures [27], [1], [2], [3], [24], [25].

We also mention, in a different, neighbor setting, some interesting but distinct LLN’s have been shown for arrays of time-\textit{homogeneous} Markov sequences where the transition matrix $P_n$ for the $n$th row converges to a limit matrix $P$ [7], [11], Section 5.3 [15]; see also [21] 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 $\Sigma$ 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 $1 \leq i, j \leq m$ distinct, and $M(i, i) < 0$ for $1 \leq i \leq m$.

To avoid technicalities, e.g. with reducibility, we work with the following matrices, $G \subseteq \mathbb{R}^{m \times m}$: $G$ is a generator matrix with nonzero entries if $G(i, j) > 0$ for $1 \leq i, j \leq m$ distinct, and $G(i, i) < 0$ for $1 \leq i \leq m$.

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

\[ P_{G, \pi}(X_{n+1} = j | X_0, X_1, \ldots, X_n, X_n = i) = P_{G, \pi}^n(i, j) \]

for all $i, j \in \Sigma$ and $n \geq 0$. Our convention then is that $P_{G, \pi}^n$ controls “transitions” between times $n$ and $n + 1$. Let also $E_{G, \pi}$ be expectation with respect to $P_{G, \pi}$. More generally, $E_\mu$ denotes expectation with respect to measure $\mu$.

Define the occupation statistic $Z_n = \langle Z_1, \ldots, Z_m \rangle$ for $n \geq 1$ where

\[ Z_{i, n} = \frac{1}{n} \sum_{k=1}^n 1_i(X_k) \]

for $1 \leq i \leq m$. Then, $Z_n$ is an element of the $m - 1$-dimensional simplex,

\[ \Delta_m = \left\{ \mathbf{x} : \sum_{i=1}^m x_i = 1, 0 \leq x_i \leq 1 \text{ for } 1 \leq i \leq m \right\} \]

The first result is on convergence of the position of the process. For $G \in G$, let $\nu_G$ be the stationary distribution 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 $0$. 

\[ \nu_G = \left\{ \mathbf{G} \in \mathbb{R}^{m \times m} : \mathbf{G} \text{ is a generator matrix with nonzero entries} \right\}, \]
Theorem 1.1 For $G \in \mathbb{G}$, $\zeta > 0$, and initial distribution $\pi$, under $P^{G,\zeta}_\pi$,

$$X_n \xrightarrow{d} \nu_{G,\pi,\zeta}$$

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

Remark 1.1 For $\zeta > 1$, with only finitely many moves, the convergence is a.s., and $\nu_{G,\pi,\zeta}$ is explicit when $G = V_G D_G V^{-1}_G$ is diagonalizable with $D_G$ diagonal and $D_G(i,i) = \lambda_i^G$, the $i$th eigenvalue of $G$, for $1 \leq i \leq m$. By calculation, $\nu_{G,\pi,\zeta} = \pi t \prod_{n \geq 1} P^{G,\zeta}_n = \pi V_G D'^{-1}_G$ with $D'$ diagonal and $D'(i,i) = \prod_{n \geq n_0(G,\zeta)+1} (1 + \lambda_i^G/n^\zeta)$.

We now consider the cases $\zeta \neq 1$ with respect to average occupation limits. Let $i$ be the basis vector $i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \Delta_m$ with a 1 in the $i$th component and $\delta_i$ be the point mass at $i$ for $1 \leq i \leq m$.

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

$$Z_n \xrightarrow{d} \nu_G$$

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

However, when $\zeta > 1$, under $P^{G,\zeta}_\pi$,

$$Z_n \xrightarrow{d} \sum_{i=1}^m \nu_{G,\pi,\zeta}(i) \delta_i.$$

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 $A = \{a_i : 1 \leq i \leq \bar{\gamma}\} = \{1, \ldots, 1, 2, \ldots, 2, \ldots, m, \ldots, m\}$. Let $S(\gamma_1, \ldots, \gamma_m)$ be the $\bar{\gamma}!$ permutations of $A$, although there are only $(\gamma_1, \gamma_2, \ldots, \gamma_m)$ 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 $-1 \leq 1 + \text{Re}(\lambda_i^G)/k \leq 1$, yielding the non-positivity), and so the resolvent $(xI - G)^{-1}$ is well defined for $x \geq 1$.

Theorem 1.3 For $\zeta = 1$, $G \in \mathbb{G}$, and initial distribution $\pi$, we have under $P^{G,\zeta}_\pi$ that

$$Z_n \xrightarrow{d} \mu_G$$
where $\mu_G$ is a measure on the simplex $\Delta_m$ characterized by its moments: For $1 \leq i \leq m$,

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

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

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

**Remark 1.3** However, as in Ex. 1.1 and [8], when $\zeta = 1$ as above, $Z_n$ cannot converge in probability (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 [16] and (2.3), but the limit distribution $\mu_G$ is not a point-mass 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 $\mu_G$ 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 $\bar{\theta} = \sum_{i=1}^m \theta_i$. It is clear $\Theta \in \mathbb{S}$. Recall identification of the Dirichlet distribution by its density and moments; see [18], [26] for more on these distributions. Namely, the Dirichlet distribution on the simplex $\Delta_m$ with parameters $\theta_1, \ldots, \theta_m$ (abbreviated as $\text{Dir}(\theta_1, \ldots, \theta_m)$) has density

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

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

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

(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 $\mu_\Theta = \text{Dir}(\theta_1, \ldots, \theta_m)$.

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

6
We now characterize the measures \( \{ \mu_G : G \in \mathcal{G} \} \) as “spreading” measures different from the limits when \( 0 < \zeta < 1 \) and \( \zeta > 1 \).

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

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

\[
G_{\text{left}} = \begin{bmatrix} -3 & 1 & 2 \\ 2 & -3 & 1 \\ 1 & 2 & -3 \end{bmatrix}, \quad \text{and} \quad G_{\text{right}} = \begin{bmatrix} -0.4 & 0.2 & 0.2 \\ 0.3 & -0.6 & 0.3 \\ 0.5 & 0.5 & -1 \end{bmatrix}.
\]

To help visualize plots, \( \Delta_3 \) is mapped to the plane by linear transformation \( f(x) = x_1 f((1, 0, 0)) + x_2 f((0, 1, 0)) + x_3 f((0, 0, 1)) \) where \( f((1, 0, 0)) = (\sqrt{2}, 0), f((0, 1, 0)) = (0, 0) \) and \( f(0, 0, 1) = \sqrt{2}(1/2, \sqrt{3}/2) \). 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 $\Sigma$, define the “contraction coefficient”

$$
c(P) = \max_{x,y} \frac{1}{2} \sum_z \left| P(x, z) - P(y, z) \right|
$$

$$
= 1 - \min_{x,y} \sum_z \min \left\{ P(x, z), P(y, z) \right\}
$$

(2.1)

The following is, for instance, Theorem 4.5.1 [28].

**Proposition 2.1** Let $X_n$ be a time-nonhomogeneous Markov chain on $\Sigma$ connected by transition matrices $\{P_n\}$ with corresponding stationary distributions $\{\nu_n\}$. Suppose

$$
\prod_{n=1}^{\infty} c(P_n) = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} \|\nu_n - \nu_{n+1}\|_{\text{Var}} < \infty.
$$

(2.2)

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

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

The following is stated in Section 2 [8] as a consequence of results (1.2.22) and Theorem 1.2.23 in [16].

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

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

in the following senses:

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

(ii) a.s. when $\sum_{n \geq n_0} 2^{-n}(1 - c_{2^n})^{-2} < \infty$.

**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 $\sum_{n \geq 1} \mathbb{P}_{\pi}^G(X_n \neq X_{n+1}) \leq C \sum_{n \geq 1} n^{-\zeta} < \infty$. Hence there is a time of last movement $N < \infty$ a.s. Then, $\lim X_n = X_N$ a.s., and, for $k \in \Sigma$, the limit distribution $\nu_{G,\pi,\zeta}$ is defined and given by $\mathbb{P}_{\pi}^G(X_N = k) = \nu_{G,\pi,\zeta}(k)$.

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

$$
\prod_{n \geq 1} c(P_n^{G,\zeta}) = \prod_{n \geq n_0(G, \zeta)} \left( 1 - \frac{C_G}{n^\zeta} \right) = 0.
$$

(2.3)
Since for \( n > n(G, \zeta) \), \( \nu^i_G P^G_{n, \zeta} = \nu^i_G (I - G/n^\zeta) = \nu^i_G \), the second condition of Proposition 2.1 is trivially satisfied, and hence the result follows.

**Proof of Theorem 1.3.** 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. \( \lim Z_n = \sum_{k=1}^m 1_{X_N = k} \) concentrates on basis vectors \( \{k\} \). Hence, as defined in proof of Theorem 1.1, \( P^G_{n, \zeta}(X_N = k) = \nu^i_{G, \pi, \zeta}(k) \), and the result follows.

When \( 0 < \zeta < 1 \), we apply Proposition 2.2 and follow the method in [8]. First, as in the proof of Theorem 1.1, (2.2) holds, and \( c(P^G_{n, \zeta}) = 1 - c(G/n^\zeta) \) for a constant \( C_G > 0 \) and all \( n \geq n_0(G, \zeta) \). Then, \( c_n = \max_{n_0(G, \zeta) \leq i \leq n} c(P^G_{i, \zeta}) = 1 - C_G/n^\zeta < 1 \). Now, \( n(1 - c_n) = C_G n^{1 - \zeta} \to \infty \) to give the probability convergence in part (i). For a.s. convergence in part (ii) when \( 0 < \zeta < 1/2 \), note

\[
\sum_n 2^n (1 - c_n)^2 = \sum_n \frac{1}{2^n (C_G/(2^n)^\zeta)^2} = \sum_n \frac{1}{C_G^2 (2^{1-2\zeta})^n} < \infty.
\]

**3 Proof of Theorem 1.3.**

In this section, as \( \zeta = 1 \) is fixed, we suppress notational dependence on \( \zeta \). Also, as \( Z_n \) takes values on the compact set \( \Delta_m \), 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 \( G \in \mathcal{G}, 1 \leq k \leq m \), and initial distribution \( \pi \),

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

**Proof.** From Theorem 1.1 and Cesaro convergence,

\[
\lim_n \mathbb{E}_{\pi}^G (Z_{k,n}) = \lim_n \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\pi}^G (1_k(X_i)) = \lim_n \frac{1}{n} \sum_{i=1}^n \mathbb{P}_{\pi}^G (X_i = k) = \nu_G(k). \]

We now turn to the joint moment limits in several steps, and will assume in the following that \( \gamma_1, \ldots, \gamma_m \geq 0 \) with \( \bar{\gamma} \geq 2 \). The first step is an “ordering of terms.”

**Lemma 3.2** For \( G \in \mathcal{G} \), and initial distribution \( \pi \), we have

\[
\lim_{n \to \infty} \left| \mathbb{E}_{\pi}^G \left( Z_{1,n}^{\gamma_1} \cdots Z_{m,n}^{\gamma_m} \right) - \sum_{\sigma \in S(\gamma_1, \ldots, \gamma_m)} \frac{1}{n^{\bar{\gamma}}} \sum_{i_1=1}^{n-\bar{\gamma}+1} \sum_{i_2>1} \cdots \sum_{i_{\tilde{\bar{\gamma}}}>i_{\tilde{\bar{\gamma}}-1}} \mathbb{E}_{\pi}^G \left( \prod_{l=1}^{\tilde{\bar{\gamma}}} 1_{\sigma_l(X_{i_l})} \right) \right| = 0.
\]
Lemma 3.3 For \( G \in G \) and initial distribution \( \pi \), we have

\[
\lim_{n \to \infty} \left| \sum_{\sigma \in S(\gamma_1, \ldots, \gamma_m)} \frac{1}{n^\gamma} \sum_{i_1=1}^{n-\gamma+1} \cdots \sum_{i_{\bar{\gamma}-1}=1} \sum_{i_\bar{\gamma}}^{n} \mathbb{E}_\pi^G \left( \prod_{l=1}^{\bar{\gamma}} \sigma_l(X_{i_l}) \right) \right| = 0.
\]

Proof. As \( \mathbb{P}_G^G(X_j = t | X_i = s) = P_{i+1,j}^G(s, t) \) for \( 1 \leq i < j \) and \( s, t \in \Sigma \), we have

\[
\sum_{\sigma \in S(\gamma_1, \ldots, \gamma_m)} \frac{1}{n^\gamma} \sum_{i_1=1}^{n-\gamma+1} \cdots \sum_{i_{\bar{\gamma}-1}=1} \sum_{i_\bar{\gamma}}^{n} \mathbb{E}_\pi^G \left( \prod_{l=1}^{\bar{\gamma}} \sigma_l(X_{i_l}) \right) = \sum_{\sigma \in S(\gamma_1, \ldots, \gamma_m)} \frac{1}{n^\gamma} \sum_{i_1=1}^{n-\gamma+1} \cdots \sum_{i_{\bar{\gamma}-1}=1} \sum_{i_\bar{\gamma}}^{n} \mathbb{P}_\pi^G(X_{i_1} = \sigma_1) \prod_{l=1}^{\bar{\gamma}-1} P_{i+1,i_{l+1}}^G(s_l, s_{l+1}).
\]
which differs from the second expression in (3.1) by at most
\[
\sum_{\sigma \in \mathcal{G}(\gamma_1, \ldots, \gamma_m)} \frac{1}{n} \sum_{i_1=1}^{n-\gamma+1} \left| P^G_{i_1}(X_{i_1} = \sigma_1) - \nu_G(\sigma_1) \right|,
\]
which vanishes by Theorem 1.1.

We now focus on a useful class of diagonalizable matrices
\[
\mathcal{G}^* = \left\{ G \in \mathbb{R}^m \times \mathbb{R}^m : \operatorname{Re}(\lambda^G_l) < 1 \text{ for } 1 \leq l \leq m, \text{ and } G \text{ is diagonalizable} \right\}
\]
where \{\lambda^G_l\} are the eigenvalues of \(G\). As \(\operatorname{Re}(\lambda^G_l) \leq 0\) for \(1 \leq l \leq m\) when \(G \in \mathcal{G}\), certainly all diagonalizable \(G \in \mathcal{G}\) belong to \(\mathcal{G}^*\). The relevance of this class, in the subsequent arguments, is that for \(G \in \mathcal{G}^*\) the resolvent \((xI - G)^{-1}\) exists for \(x \geq 1\).

For \(G \in \mathcal{G}^*\), let \(V_G\) be the matrix of eigenvectors and \(D_G\) be a diagonal matrix with corresponding eigenvalue entries \(D_G(i, i) = \lambda^G_i\) so that \(G = V_G D_G V_G^{-1}\). Define also for \(1 \leq s, t, k \leq m\),
\[
g(k; s, t) = V_G(s, k) V_G^{-1}(k, t).
\]
We also denote for \(a_1, \ldots, a_m \in \mathbb{C}\), the diagonal matrix \(\operatorname{Diag}(a_i)\) with \(i\)th diagonal entry \(a_i\) for \(1 \leq i \leq m\). We also extend the definitions of \(P^G_{n, i, j}\) and \(P^G_{i, j}\) to \(G \in \mathcal{G}^*\) 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 \(G \in \mathcal{G}^*, s, t \in \Sigma, \) and \(C \leq i \leq j\) where \(C = C(G)\) is a large enough constant,
\[
P^G_{i, j}(s, t) = \sum_{k=1}^{m} v(k; i, j) g(k; s, t) \left( \frac{j}{i-1} \right)^{\lambda^G_k} ;
\]
moreover, \(v(k; i, j) \to 1\) as \(i \to \infty\) uniformly over \(k\) and \(j\).

**Proof.** Straightforwardly,
\[
P^G_{i, j} = V_G \prod_{k=i}^{j} \left( I + \frac{1}{k} D_G \right) V_G^{-1} = V_G \operatorname{Diag} \left( \prod_{k=i}^{j} \left( 1 + \frac{\lambda^G_k}{k} \right) \right) V_G^{-1}.
\]
To expand further, we note for \(z \in \mathbb{C}\) such that \(|z - 1| < 1\), we have
\[
\log(z) = (z - 1) + (z - 1)^2 \sum_{n=0}^{\infty} \frac{1}{n+2} (z - 1)^n.
\]
and estimate
\[
\left| \sum_{n=0}^{\infty} \frac{1}{n+2} (z - 1)^n \right| \leq \sum_{n=0}^{\infty} |z - 1|^n = \left( 1 - |z - 1| \right)^{-1}.
\]
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

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

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

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

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

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

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

$$P_{i,j}^G = V_G \text{Diag} \left(\nu(\cdot;i,j) \left(\frac{j}{i-1}\right)^{\lambda_s^G}\right) V_G^{-1}$$

$$= \left[\sum_{k=1}^{m} \nu(k;i,j) g(k;s,t) \left(\frac{j}{i-1}\right)^{\lambda_k^G}\right]_{s,t \in \Sigma}$$

completing the proof.

□

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

$$T^G_{x,y}(s,t) = \sum_{k=1}^{m} g(k;s,t) \left(\frac{x}{y}\right)^{-\lambda_k^G}.$$
Lemma 3.5 For $G \in \mathcal{G}$,

$$
\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \sum_{\sigma \in \mathcal{S}(\gamma_1, \ldots, \gamma_m)} \frac{\nu_G(\sigma)}{n^\gamma} \sum_{i_1=1}^{[n\epsilon]} \sum_{i_2>i_1}^{n-\gamma+2} \ldots \sum_{i_\gamma>i_{\gamma-1}}^{n} \prod_{l=1}^{\gamma-1} P_{i_l+1, i_l+1}^G(\sigma_l, \sigma_{l+1}) = 0.
$$

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

$$
0 \leq \lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n^\gamma} \sum_{i_1=1}^{[n\epsilon]} \sum_{i_2>i_1}^{n-\gamma+2} \ldots \sum_{i_\gamma>i_{\gamma-1}}^{n} \prod_{l=1}^{\gamma-1} P_{i_l+1, i_l+1}^G(\sigma_l, \sigma_{l+1})
\leq \lim_{\epsilon \downarrow 0} \frac{1}{n^\gamma} (n\epsilon)^{\gamma-1} = 0.
$$

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

$$
\lim_{n \to \infty} \frac{1}{n^\gamma} \sum_{i_1=[n\epsilon]+1}^{n-\gamma+1} \sum_{i_2>i_1}^{n-\gamma+2} \ldots \sum_{i_\gamma>i_{\gamma-1}}^{n} \prod_{l=1}^{\gamma-1} P_{i_l+1, i_l+1}^G(\sigma_l, \sigma_{l+1})
= \int_{\epsilon \leq x_1 \leq x_2 \leq \ldots \leq x_\gamma \leq 1} \prod_{l=1}^{\gamma-1} T_{x_l, x_l+1}^G(\sigma_l, \sigma_{l+1}) \, dx_1 dx_2 \ldots dx_\gamma.
$$

Proof. From Lemma 3.5 as $\nu(s; i, j) \to 1$ as $i \to \infty$ 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

$$
\lim_{n} \frac{1}{n^\gamma} \sum_{i_1=[n\epsilon]+1}^{n-\gamma+1} \sum_{i_2>i_1}^{n-\gamma+2} \ldots \sum_{i_\gamma>i_{\gamma-1}}^{n} \prod_{l=1}^{\gamma-1} P_{i_l+1, i_l+1}^G(\sigma_l, \sigma_{l+1})
= \int_{\epsilon \leq x_1 \leq x_2 \leq \ldots \leq x_\gamma \leq 1} \prod_{l=1}^{\gamma-1} T_{x_l, x_l+1}^G(\sigma_l, \sigma_{l+1}) \, dx_1 dx_2 \ldots dx_\gamma.
$$

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

$$
\lim_{\epsilon \downarrow 0} \int_{\epsilon \leq x_1 \leq x_2 \leq \ldots \leq x_\gamma \leq 1} \prod_{l=1}^{\gamma-1} T_{x_l, x_l+1}^G(\sigma_l, \sigma_{l+1}) \, dx_1 dx_2 \ldots dx_\gamma
= \int_0^1 \int_0^{x_\gamma} \ldots \int_0^{x_2} T_{x_{\gamma-1}, x_{\gamma}}^G(\sigma_{\gamma-1}, \sigma_{\gamma}) \ldots T_{x_1, x_2}^G(\sigma_1, \sigma_2) \, dx_1 dx_2 \ldots dx_\gamma.
$$
Proof. Let
\[ f_\epsilon = 1_{\{\epsilon \leq x_1 \leq x_2 \leq \cdots \leq x_\gamma \leq 1\}} \prod_{l=1}^{\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l, \sigma_{l+1}). \]

Then,
\[ \lim_{\epsilon \to 0} f_\epsilon = 1_{\{0 < x_1 \leq x_2 \leq \cdots \leq x_\gamma \leq 1\}} \prod_{l=1}^{\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l, \sigma_{l+1}), \]
and \( f_\epsilon \) is uniformly bounded over \( \epsilon \) as
\[ |f_\epsilon| \leq \tilde{f} = 1_{\{0 < x_1 \leq x_2 \leq \cdots \leq x_\gamma \leq 1\}} \prod_{l=1}^{\gamma-1} \sum_{k=1}^m |g(k; \sigma_l, \sigma_{l+1})| \left( \frac{x_l}{x_{l+1}} \right)^{-\text{Re}(\lambda_k^G)}. \]
The right-hand bound is integrable: Indeed, by Tonelli’s Lemma and induction, we have
\[
\int \int_0^{x_\gamma} \cdots \int_0^{x_\gamma} \prod_{l=1}^{\gamma-1} \sum_{k=1}^m |g(k; \sigma_l, \sigma_{l+1})| \left( \frac{x_l}{x_{l+1}} \right)^{-\text{Re}(\lambda_k^G)} \; dx_1 \cdots dx_\gamma
= \frac{1}{\gamma} \prod_{l=1}^{\gamma-1} \left( \sum_{k=1}^m \frac{|g(k; \sigma_l, \sigma_{l+1})|}{l - \text{Re}(\lambda_k^G)} \right).
\]

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

Lemma 3.8 For \( G \in \mathbb{G}^* \) and \( \sigma \in \mathbb{S}(\gamma_1, \ldots, \gamma_m) \),
\[
\int_0^{x_\gamma} \cdots \int_0^{x_\gamma} \prod_{l=1}^{\gamma-1} T^G_{x_l,x_{l+1}}(\sigma_l, \sigma_{l+1}) \; dx_1 \cdots dx_\gamma = \frac{1}{\gamma} \prod_{l=1}^{\gamma-1} \left( I - G \right)^{-1}(\sigma_l, \sigma_{l+1}).
\]

Proof. By induction, the integral equals
\[
\int_0^{x_\gamma} \cdots \int_0^{x_\gamma} \prod_{l=1}^{\gamma-1} T^G_{x_{\gamma-l},x_{\gamma}(\sigma_{\gamma-1}, \sigma_{\gamma})} \cdots T^G_{x_1,x_2}(\sigma_1, \sigma_2) \; dx_1 \cdots dx_\gamma
= \frac{1}{\gamma} \prod_{l=1}^{\gamma-1} \left( \sum_{k=1}^m \frac{g(k; \sigma_l, \sigma_{l+1})}{l - \lambda_k^G} \right).
\]
However, for \( x \geq 1 \), we have
\[
\left( xI - G \right)^{-1}(s,t) = V_G \left( xI - D_G \right)^{-1} V^{-1}_G(s,t) = \sum_{k=1}^m \frac{g(k; s,t)}{x - \lambda_k^G}
\]
to finish the identification.

At this point, by straightforwardly combining the previous lemmas, we have proved Theorem 1.2 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 $G \in \mathbb{G}$,

$$
\lim_{n \to \infty} \sum_{\sigma \in S(\gamma_1, \ldots, \gamma_m)} \frac{\nu_G(\sigma_1)}{n^{\gamma}} \sum_{i_1=1}^{n-\gamma+1} \sum_{i_2 > i_1} \cdots \sum_{i_l > i_{l-1}} \prod_{l=1}^{\gamma_l-1} p_{i_l+1, i_l+1}^G(\sigma_l, \sigma_{l+1})
$$

$$
= \frac{1}{\gamma} \sum_{\sigma \in S(\gamma_1, \ldots, \gamma_m)} \nu_G(\sigma_1) \prod_{l=1}^{\gamma_l-1} (I - G)^{-1}(\sigma_l, \sigma_{l+1}). \quad (3.2)
$$

Proof. For an $m \times m$ matrix $A$, let $G[A] = G + A$. Let $\| \cdot \|_M$ be the matrix norm $\|A\|_M = \max \{|A(s,t)| : 1 \leq s, t \leq m\}$. Now, for small $\epsilon > 0$, choose matrices $A_1$ and $A_2$ with non-negative entries so that $\|A_1\|_M, \|A_2\|_M < \epsilon, I + G[-A_1]/l, I + G[A_2]/l$ 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 $\| \cdot \|_M$ (cf. Appendix D [13]), and (2) diagonalizable real matrices are dense (cf. Theorem 1 [12]).

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

$$
P_{i,j}^{G[-A_1]}(s,t) \leq P_{i,j}^G(s,t) \leq P_{i,j}^{G[A_2]}(s,t).
$$

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

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

and

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

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

$$
\prod_{l=1}^{\gamma_l-1} (I - G[-A_1])^{-1}(\sigma_l, \sigma_{l+1}), \quad \prod_{l=1}^{\gamma_l-1} (I - G[A_2])^{-1}(\sigma_l, \sigma_{l+1}) \rightarrow \prod_{l=1}^{\gamma_l-1} (I - G)^{-1}(\sigma_l, \sigma_{l+1})
$$

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

\[ \square \]

4 Proof of Theorem [14]

The proof follows by evaluating the moment expressions in Theorem [12] when $G = \Theta$ as those corresponding to the Dirichlet distribution with parameters $\theta_1, \ldots, \theta_m$ [11].
Lemma 4.1 The stationary distribution $\nu_\Theta$ 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,

$$\left( I - \Theta \right)^{-1} = \frac{1}{l(l + \theta)} F_{l+1}.$$

Proof. The form of $\nu_\Theta$ follows by inspection. For the second statement, write $F_{l+1} = I + \hat{\Theta}$ where the matrix $\hat{\Theta}$ has $i$th column equal to $\theta_i(1, \ldots, 1)^t$. Then, also $\Theta = \Theta - \hat{\Theta}I$.

As $(1, \ldots, 1)^t$ is an eigenvector of $\Theta$ with eigenvalue 0, we see $(I - \Theta)(I + \Theta) = (I^2 + \theta I)$ finishing the proof. \hfill $\square$

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

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

$$E_{\mu_\Theta}\left[ \prod_{i=1}^{m} x_i^{\gamma_i} \right] = \sum_{\sigma \in \mathcal{S}(\gamma_1, \ldots, \gamma_m)} \nu_\Theta(\sigma_1) \frac{1}{\bar{\gamma}} \prod_{l=1}^{\bar{\gamma}-1} \left( I - \Theta \right)^{-1} (\sigma_l, \sigma_{l+1})$$

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

We now evaluate the last expression of Lemma 4.2 by first specifying of the value of $\sigma_{\bar{\gamma}}$. Recall, by convention $\theta_1 \cdots (\theta_l + \gamma_l - 1) = 1$ when $\gamma_l = 0$ for $1 \leq l \leq m$.

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

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

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 $\theta_1 F_2(i, k) = \theta_1 \theta_2$. 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

$$\sum_{\sigma \in \mathcal{S}(\gamma_1, \ldots, \gamma_m)} \theta_{\sigma_1} \prod_{l=2}^{\bar{\gamma}} F_l(\sigma_l, \sigma_{l-1}) = \gamma_1 (\theta_1 + \bar{\gamma} - 1) \sum_{\sigma \in \mathcal{S}(\gamma_1 - 1, \ldots, \gamma_m)} \theta_{\sigma_1} \prod_{l=2}^{\bar{\gamma}-1} F_l(\sigma_l, \sigma_{l-1})$$

$$+ \sum_{j=2}^{m} \gamma_1 \theta_1 \sum_{\sigma \in \mathcal{S}(\gamma_1 - 1, \ldots, \gamma_m)} \theta_{\sigma_1} \prod_{l=2}^{\bar{\gamma}-1} F_l(\sigma_l, \sigma_{l-1}).$$
We now use induction to evaluate the right-side above as

\[
\theta_1 \cdots (\theta_1 + \gamma_1 - 2) \prod_{i=2}^{m} \theta_i \cdots (\theta_i + \gamma_i - 1)
\times \left\{ \gamma_1(\theta_1 + \bar{\gamma} - 1)(\gamma_1 - 1)(\bar{\gamma} - 2)! + \sum_{j=2}^{m} \gamma_j \theta_1 \gamma_j (\bar{\gamma} - 2)\right\}
= \theta_1 \cdots (\theta_1 + \gamma_1 - 2) \prod_{i=2}^{m} \theta_i \cdots (\theta_i + \gamma_i - 1)
\times \left\{ \gamma_1(\theta_1 + \bar{\gamma} - 1)(\gamma_1 - 1)(\bar{\gamma} - 2)! + \gamma_1 \theta_1 (\bar{\gamma} - 1)\right\}
= \gamma_1(\bar{\gamma} - 1)! \prod_{i=1}^{m} \theta_i \cdots (\theta_i + \gamma_i - 1).
\]

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

**Lemma 4.4** When \( \bar{\gamma} \geq 2 \),

\[
\sum_{\sigma = \{1, \ldots, m\}} \frac{\theta_{\sigma_1} \prod_{i=2}^{\bar{\gamma}} F_i(\sigma_{i-1}, \sigma_i)}{\gamma! \prod_{l=0}^{\gamma-1}(\theta + l)} = \frac{\prod_{i=1}^{m} \theta_i \cdots (\theta_i + \gamma_i - 1)}{\prod_{l=0}^{\gamma-1}(\theta + l)}.
\]

**Proof.**

\[
\sum_{\sigma = \{1, \ldots, m\}} \frac{\theta_{\sigma_1} \prod_{i=2}^{\bar{\gamma}} F_i(\sigma_{i-1}, \sigma_i)}{\gamma! \prod_{l=0}^{\gamma-1}(\theta + l)} = \sum_{k=1}^{m} \sum_{\sigma = \{1, \ldots, m\}} \frac{\theta_{\sigma_1} \prod_{i=2}^{\bar{\gamma}} F_i(\sigma_{i-1}, \sigma_i)}{\gamma! \prod_{l=0}^{\gamma-1}(\theta + l)}
= \sum_{k=1}^{m} \gamma_k (\bar{\gamma} - 1)! \prod_{l=1}^{m} \theta_i \cdots (\theta_i + \gamma_i - 1) \prod_{l=0}^{\gamma-1}(\theta + l)
= \prod_{l=0}^{\gamma-1}(\theta + l).
\]

\section{5 Proof of Theorem 1.5 (1)}

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

$$\mathbb{P}_n^G \left( \mathbf{Z}_n \in B(p, \epsilon) \right) > C(p, \epsilon) > 0.$$ 

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

$$\{\mathbf{Z}_n \in B(p, \epsilon) \} \supset \bigcup_{0 \leq k_1, \ldots, k_m \leq \lfloor n \delta \rfloor} \left\{ \mathbf{X}^{[\bar{p}_j] - k_1} = \bar{i} \right\} \cap \left( \bigcap_{j=2}^m \left\{ \mathbf{X}^{[\bar{p}_j] - k_j} = \bar{j} \right\} \right)$$

where $k_n = \sum_{l=1}^a k_l$, and $\bar{i}$ 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 $[n \bar{p}_{j-1}] - k_{j-1} + 1$ to $[n \bar{p}_j] - k_j$ for $1 \leq j \leq m$ where we take $1 - k_0 = [n \delta]$.

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

$$\mathbb{P}_n^G \left( X^{[\bar{p}_j] - k_j} = \bar{i} | X^{[\bar{p}_i] - k_i - 1} = \bar{j} \right) \geq \prod_{j=1}^n 1 - \frac{C_1}{j} \geq \frac{\delta C_1}{2}.$$ 

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

$$\mathbb{P}_n^G \left( X^{[\bar{p}_i] - k_i - 1} = i | X^{[\bar{p}_i] - k_i - 1} = i - 1 \right) \geq \frac{C_2}{[n \bar{p}_i - 1] - k_i - i + 1}.$$ 

Hence, for all large $n$, as $\mathbb{P}_n^G(X^{[n \delta]} = 1) \geq \nu_G(1)/2$ (Theorem 1.1),

$$\mathbb{P}_n^G \left( \mathbf{Z}_n \in B(p, \epsilon) \right) \geq \sum_{0 \leq k_1, \ldots, k_m \leq \lfloor n \delta \rfloor} \mathbb{P}_n^G \left( \mathbf{X}^{[n \delta]} = \bar{i} \right) \cap \left( \bigcap_{j=2}^m \left\{ \mathbf{X}^{[\bar{p}_j] - k_j} = \bar{j} \right\} \right)$$

$$\geq \left[ \frac{\delta C_1}{2} \right]^m \sum_{0 \leq k_1, \ldots, k_m \leq \lfloor n \delta \rfloor} \frac{\nu_G(1)}{2} \prod_{j=2}^m \frac{C_2}{[n \bar{p}_{j-1}] - k_{j-1} + 1}$$

$$\geq \left[ \frac{\delta C_1}{2} \right]^m \sum_{0 \leq k_1, \ldots, k_m \leq \lfloor n \delta \rfloor} \frac{\nu_G(1)}{2} \prod_{j=2}^m \frac{C_2}{[n \bar{p}_{j-1}] - k_{j-1} + 1}$$

$$\geq \frac{\nu_G(1)}{4} \left[ \frac{C_2 \delta C_1}{2} \right]^m \prod_{j=2}^m \log \left( \frac{[n \bar{p}_{j-1}]}{[n \bar{p}_{j-1}] - [n \delta]} \right)$$

$$\geq \frac{\nu_G(1)}{8} \left[ \frac{C_2 \delta C_1}{2} \right]^m \prod_{j=2}^m \log \left( \frac{\bar{p}_{j-1} - \beta}{\bar{p}_{j-1} - \beta} \right).$$
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 $\Delta_m$, $1, \ldots, m$, 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 $k$,

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

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 $\mu_G(|l\rangle) \leq E_{\mu_G}[(x_l)^k] \to 0$ as $k \to \infty$, the point $\mathbf{l}$ cannot be an atom of the limit distribution. $\square$

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

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

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

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

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

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

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 $\{\omega^n : (1/n)\sum_{i=1}^n(1_1(\omega_i), \ldots, 1_m(\omega_i)) \in B(p, \delta)\} \cap T(j)$ into non-empty sets $A_j(U, V)$ where $U = \langle U_1, \ldots, U_{j-1} \rangle$ and $V = \langle V_1, \ldots, V_{j+1} \rangle$ denote possible switch times (up to the $j - 1$st switch time) and visit locations respectively:

$$A_j(U, V) = \left\{\omega^n : \omega^n \in T(j), \frac{1}{n}\sum_{i=1}^n(1_1(\omega_i), \ldots, 1_m(\omega_i)) \in B(p, \delta), \alpha_i = U_i, \theta_k = V_k \text{ for } 1 \leq i \leq j - 1, 1 \leq k \leq j + 1\right\}.$$ 

In this decomposition, paths in $A_j(U, V)$ are in 1:1 correspondence with $j$th switch times $\alpha_j$—the only feature allowed to vary.
Now, for each set $A_j(U, V)$, we define a path $\eta(j, U, V) = (\eta_1, \ldots, \eta_n)$ where the last $j$th 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 $\eta(j, U, V)$ 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 $(U_1, \ldots, U_{j-1})$ and $(V_1, \ldots, V_j)$. In particular, recalling $X_1^n = (X_1, \ldots, X_n)$ denotes the coordinate sequence up to time $n$, we have

$$\sum_{U, V} \mathbb{P}_{\pi}^G(X_1^n = \eta(j, U, V)) \leq m \mathbb{P}_{\pi}^G(X_1^n \in T(j - 1)) \quad (6.1)$$

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

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

**Lemma 6.1** On $A_j(U, V)$, we have $[n(p - \delta) + 1] \leq \alpha_j$. Also, $|A_j(U, V)| \leq [2n\delta + 1]$.

A consequence of these bounds on the position and cardinality of $\alpha_j$’s associated to a fixed set $A_j(U, V)$, is that

$$\sum^{'} \frac{1}{U_j} \leq \sum_{k=n(p - \delta) + 1}^{[n(p + \delta) + 2]} \frac{1}{k} \leq \log \left( \frac{\hat{p} + \delta + 3/n}{\hat{p} - \delta} \right) \quad (6.2)$$

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

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

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

$$\mathbb{P}_{\pi}^G(X_1^n = \omega^n) \leq \frac{\hat{G}(\hat{p}/2)^{-2\hat{G}}}{U_j} \mathbb{P}_{\pi}^G(X_1^n = \eta(j, U, V)). \quad (6.3)$$

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

$$\frac{\mathbb{P}_{\pi}^G(X_1^n = \omega^n)}{\mathbb{P}_{\pi}^G(X_1^n = \eta(j, U, V))} = \frac{G(V_j, V_{j+1})}{U_j(1 + G(V_j, V_j)/U_j)} \prod_{l=U_{j+1}}^n \left( \frac{1 + G(V_{j+1}, V_{j+1})/l}{1 + G(V_j, V_j)/l} \right).$$

20
Now bounding $G(V_j, V_{j+1}) \leq \hat{G}, 1 + G(V_{j+1}, V_{j+1})/l \leq 1, 1 + G(V_j, V_j)/l \geq 1 - \hat{G}/l,$ and noting $U_j \geq n(\bar{p} - \delta) + 1$ (by Lemma [6.11], $-\ln(1 - x) \leq 2x$ for $x > 0$ small, and $\delta < \tilde{p}/2$, give for large $n$,

$$\frac{G(V_j, V_{j+1})}{1 + G(V_j, V_j)/U_j} \prod_{i=U_{j+1}}^{n} \left(1 + \frac{G(V_{j+1}, V_{j+1})/l}{1 + G(V_j, V_j)/l}\right) \leq \hat{G} \left(\frac{n}{n(\bar{p} - \delta)}\right)^{2\hat{G}} \leq \hat{G}(\tilde{p}/2)^{-2\hat{G}}. \quad \Box$$

**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 $\sum_j \mathbb{P}_\pi^G(X^n \in T(j-1)) \leq 1$, we have for all large $n$,

$$\mathbb{P}_\pi^G\left(Z_n \in B(p, \delta)\right) = \sum_{j=1}^{n-1} \mathbb{P}_\pi^G\left(Z_n \in B(p, \delta), X^n \in T(j)\right) = \sum_{j=1}^{n-1} \sum_{U, V} \mathbb{P}_\pi^G\left(A_j(U, V)\right) \leq \sum_{j=1}^{n-1} \sum_{U, V} \frac{C(G, p)}{U_j} \mathbb{P}_\pi^G\left(X^n = \eta(j, U, V)\right) \leq C(G, p) \log\left(\frac{\tilde{p} + 2\delta}{\bar{p} - \delta}\right) \sum_{j=1}^{n-1} \mathbb{P}_\pi^G\left(X^n \in T(j-1)\right) \leq mC(G, p) \log\left(\frac{\tilde{p} + 2\delta}{\bar{p} - \delta}\right).
$$

The proposition follows by taking limit on $n$, and weak convergence. $\Box$

**Proof of Lemma 6.4.** For a path $\omega^n \in A_j(U, V)$ and $1 \leq k \leq j + 1$, let $\tau_k$ be the number of visits to state $V_k$ (some $\tau_k$‘s may be the same if $V_k$ is repeated). For $1 \leq i \leq \tau_k$, let $\underline{w}_i^k$ and $\overline{w}_i^k$ be the start and end of the ith visit to $V_k$. Certainly, $\sum_{i=1}^{\tau_k} 1_{V_k}(\omega_i) = \sum_{i=1}^{\tau_k} (\overline{w}_i^k - \underline{w}_i^k + 1)$. Moreover, as $(1/n) \sum_{i=1}^{\tau_k} 1_{V_k}(\omega_i) \in B(p, \delta)$, we have $|(1/n) \sum_{i=1}^{\tau_k} 1_{V_k}(\omega_i) - n\nu_k| \leq \delta$, and so

$$n(p\nu_k - \delta) \leq \sum_{i=1}^{\tau_k} (\overline{w}_i^k - \underline{w}_i^k + 1) \leq n(p\nu_k + \delta). \quad (6.4)$$

Hence, as the disjoint sojourns $\{[\underline{w}_i^k, \overline{w}_i^k] : 1 \leq i \leq \tau_k\}$ occur between times 1 and $\overline{w}_{\tau_k}^k$, their total sum length is less than $\overline{w}_{\tau_k}^k$, and we deduce $n(p\nu_k - \delta) \leq \overline{w}_{\tau_k}^k$. Now, for $p \in \Delta_m \setminus \{1, \ldots, m\}$, at least one of the $\{p\nu_i : V_i \neq V_{j+1}, 1 \leq i \leq j\}$ is positive: Indeed, there are two coordinates of $p$, say $p_s$ and $p_t$, which are positive. Say
$V_{j+1} \neq s$; then, as $(1/n) \sum_{i=1}^{n} 1_s(\omega_i) = (1/n) \sum_{i=1}^{n} 1_{s}(\omega_i)$, $|(1/n) \sum_{i=1}^{n} 1_s(\omega_i) - p_s| \leq \delta$, and $p_s - \delta > 0$, the path must visit state $s$ before time $\alpha_j$, e.g. $V_i = s$ for some $1 \leq i \leq j$.

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

$$n(\bar{p} - \delta) \leq n \max_{V_i \neq V_{j+1}} \frac{p_{V_i} - \delta}{1 \leq i \leq j} \leq \max_{V_i \neq V_{j+1}} \frac{\bar{\pi}_{\tau_i}}{1 \leq i \leq j} \leq \bar{\pi}_{\tau_j} = \alpha_j - 1$$

giving the first statement.

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

Acknowledgement. We thank M. Balázs and J. Pitman for helpful communications.

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

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

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

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

[5] Baumgärtel, H. (1985) Analytic Perturbation Theory for Matrices and Operators. Birkhäuser, Boston.

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

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

[8] Gantert, N. (1990) Laws of large numbers for the annealing algorithm. Stochastic Process. Appl. 35 309-313.
[9] Gidas, B. (1985) Nonstationary Markov chains and convergence of the annealing algorithm. *J. Stat. Phys.* **39** 73-131.

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

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

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

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

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

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

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

[17] 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

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

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

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

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

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

[23] Pemantle, R. (2005) Random processes with reinforcement. Pre-print available at [http://www.math.upenn.edu/~pemantle/papers/Papers.html](http://www.math.upenn.edu/~pemantle/papers/Papers.html)

[24] 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.
[25] Pitman, J. (2005) *Combinatorial Stochastic Processes*. Manuscript available at http://bibserver.berkeley.edu/csp/april05/bookcsp.pdf

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

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

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