

# THE CLASS OF TENABLE ZERO-BALANCED PÓLYA URN SCHEMES: CHARACTERIZATION AND GAUSSIAN PHASES

SANAA KHOLFI\* AND

HOSAM M. MAHMOUD,\*\* *The George Washington University*

## Abstract

We study a class of tenable, irreducible, nondegenerate zero-balanced Pólya urn schemes. We give a full characterization of the class by sufficient and necessary conditions. Only forms with a certain cyclic structure in their replacement matrix are admissible. The scheme has a steady state into proportions governed by the principal (left) eigenvector of the average replacement matrix. We study the gradual change for any such urn containing  $n \rightarrow \infty$  balls from the initial condition to the steady state. We look at the status of an urn starting with an asymptotically positive proportion of each color after  $j_n$  draws. We identify three phases of  $j_n$ : the growing sublinear, the linear, and the superlinear. In the growing sublinear phase the number of balls of different colors has an asymptotic joint multivariate normal distribution, with mean and covariance structure that are influenced by the initial conditions. In the linear phase a different multivariate normal distribution kicks in, in which the influence of the initial conditions is attenuated. The steady state is not a good approximation until a certain superlinear amount of time has elapsed. We give interpretations for how the results in different phases conjoin at the ‘seam lines’. In fact, these Gaussian phases are all manifestations of one master theorem. The results are obtained via multivariate martingale theory. We conclude with some illustrating examples.

*Keywords:* Urn; Pólya urn; combinatorial probability; phase transition; multivariate martingale

2010 Mathematics Subject Classification: Primary 60C05; 60F05; 05A05  
Secondary 60G42

## 1. Introduction

A  $k$ -color Pólya urn is an urn containing balls of up to  $k$  different colors and some rules of evolution in discrete time units. At each stage a ball is drawn uniformly at random from among all the balls in the urn and its color is noted. If a ball of color  $r$ ,  $r = 1, \dots, k$ , is drawn at a stage, it is returned to the urn together with  $A_{r,s}$  (a random variable, possibly with support including negative numbers) balls of color  $s$ ,  $s = 1, \dots, k$ . It is customary to represent these dynamics by the square replacement matrix

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,k} \\ A_{2,1} & A_{2,2} & \dots & A_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k,1} & A_{k,2} & \dots & A_{k,k} \end{pmatrix}.$$

Received 11 August 2011; revision received 12 April 2012.

\* Postal address: Department of Statistics, The George Washington University, Washington, DC 20052, USA.

\*\* Email address: hosam@gwu.edu

The replacement matrix together with the initial conditions (the number of balls of each color initially present) will be referred to as the *Pólya urn scheme*, or simply as the ‘scheme’.

Much research in the classic era was devoted to growing schemes, with deterministic replacement matrix, where the sum of the rows of the replacement matrix is a constant called the *balance factor*; see, for example, [4], [7], [8], [11], [12], [32], and [33]. In more recent years, the tradition of investigating balanced urns continues on more sophisticated types of deterministic urn schemes; see, for example, [2], [9], [10], [13]–[15], and [19]. Smythe [34] assumed random matrix entries, with average balance. For growing urns, [1], [5], [17], and [34] give the broadest theories to date. Surveys are provided in [23] and [28]. For textbook discussions, see [20] and [29].

We wish to deal with the class of zero-balance tenable urns. In these urn schemes the balance factor is 0, with the row sum equal to 0 on each row of the replacement matrix. Thus, in these urns the number of balls never changes, it is only the proportion of balls of different colors that does. We will refer to the number of balls in a zero-balanced urn by  $n$ . The zero-balance urn might have negative entries in the replacement matrix. A myriad of applications, particularly in the area of random trees, dictated that attention be paid to matrices with negative entries. To name a few, we refer the reader to [18], [24]–[27], [30], and [31].

Let  $X_{j,r}^{(n)}$  be the number of balls in the urn of color  $r$  in a zero-balanced urn scheme, for  $r = 1, \dots, k$ , after  $j$  draws, and set

$$X_j^{(n)} := \begin{pmatrix} X_{j,1}^{(n)} \\ X_{j,2}^{(n)} \\ \vdots \\ X_{j,k}^{(n)} \end{pmatrix}.$$

So, a zero-balanced scheme is a pair  $(A, X_0^{(n)})$  consisting of a replacement matrix and a vector of initial values.

An instance of the class we investigate is the classic Ehrenfest scheme [8]. A class of urn schemes is *tenable* if we can perpetuate the drawing following any stochastic path whatsoever. Tenability is a proper combination of a certain matrix structure and initial conditions. For instance, suppose that we have an urn of balls of two colors, say white and blue, with replacement matrix

$$A = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}.$$

This matrix, together with an even initial number of white balls and an even initial number of blue balls, forms a tenable scheme. The same matrix with other initial conditions may become an untenable urn scheme. For instance, if we start with one white and three blue balls, the urn gets ‘stuck’ if the first draw produces a white ball.

When  $n$  is large, we describe the initial conditions in terms of proportions, i.e.  $X_0^{(n)} = \alpha_n n + o(n)$ , where the vector of proportions is

$$\alpha_n = \begin{pmatrix} \alpha_1(n) \\ \alpha_2(n) \\ \vdots \\ \alpha_k(n) \end{pmatrix},$$

with each  $\alpha_r(n) \in [0, 1]$  for  $r = 1, \dots, k$  and

$$\sum_{r=1}^k \alpha_r(n) = 1.$$

In this paper  $\mathbf{o}(g(n))$  and  $\mathbf{O}(g(n))$  are matrices or vectors, where each component is  $o(g(n))$  and  $O(g(n))$ , respectively, in the usual sense. Toward convergence of a suitably normalized version of  $\mathbf{X}_{j_n}^{(n)}$ , we will only consider cases where we have componentwise convergence, i.e.

$$\alpha_n \rightarrow \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix},$$

with each  $\alpha_r \in [0, 1]$  for  $r = 1, \dots, k$  and  $\sum_{r=1}^k \alpha_r = 1$ . Our interest is in the characterization of the class of zero-balanced tenable urns, and in the short- and long-term behaviors. We will see that there is a steady state with a multinomial distribution, which for large urns ( $n \rightarrow \infty$ ) can be approximated by a multivariate normal distribution. We also explore the transition of Gaussian phases along the way to that multivariate normal state.

### 2. Matrix notation

Many of the results are of multivariate nature, and can be compactly written in matrix notation. Matrices and vectors are denoted using bold characters. As we deal with  $k$ -color schemes, all matrices are  $k \times k$  and all vectors have  $k$  components; we write them unindexed by  $k$ . The transpose of a vector or a matrix  $\mathbf{W}$  is denoted by  $\overline{\mathbf{W}}$ . We denote by  $\mathbf{A}_j$  the realization of  $\mathbf{A}$  after  $j$  draws, and by  $\lambda_1, \lambda_2, \dots, \lambda_k$  the eigenvalues of its expectation  $E[\mathbf{A}]$ . The eigenvalues are ordered according to their decreasing real parts, that is,  $\text{Re } \lambda_1 \geq \text{Re } \lambda_2 \geq \dots \geq \text{Re } \lambda_k$ . An eigenvalue with largest real part will be called the *principal eigenvalue*, and the corresponding left eigenvector will be called the *principal eigenvector*. Of the continuum of scales that may be used for eigenvectors, we will use the scales that render the sum of the components of the principal eigenvector equal to 1. For a vector or a matrix  $\mathbf{W}$ , we will often need to extract the  $d$ th row. For this, we introduce the operator  $\text{Row}(\mathbf{W}, d)$ , which we succinctly write as  $\mathbf{W}_{[d]}$ . We will also use  $\mathbf{0}$  to refer to a vector of zeros or a null matrix, and  $\mathbf{I}$  for the identity matrix. The matrix  $\mathbf{G}(d)$  is the  $k \times k$  diagonal matrix that has all entries equal to 0, except for the  $(d, d)$ th entry which is equal to 1. To simplify the notation throughout, we use the special matrices  $\mathbf{M} = \mathbf{I} + (1/n) E[\overline{\mathbf{A}}]$  and  $\mathbf{B} = (1/n) E[\mathbf{A}]$ . For real numbers  $y_1, \dots, y_k$ , the notation  $\text{diag}(y_1, \dots, y_k)$  stands for a diagonal matrix of all zeros, except for the diagonal elements, with the  $r$ th diagonal element equal to  $y_r$ . A matrix  $\mathbf{W}$  with distinct eigenvalues can be diagonalized in terms of its modal matrix and a diagonal matrix of eigenvalues. (The modal matrix of  $\mathbf{W}$  is a matrix in which each column is an eigenvector of  $\mathbf{W}$ .) A matrix  $\mathbf{W}$  with distinct eigenvalues can be written as  $\mathbf{H} \text{diag}(\lambda_1, \dots, \lambda_k) \mathbf{H}^{-1}$ , where  $\mathbf{H}$  is its modal matrix.

### 3. The results

Tenability is the property of being able to draw balls and execute the rules forever, no matter which stochastic path is traced. Irreducibility means that all the colors are essential and feed into each other. Nondegeneracy means that no two colors are tied together to be basically one

combined color. These verbose definitions will be made formally precise in the sequel. We will denote the class of tenable, irreducible, nondegenerate zero-balanced  $k$ -color Pólya urn schemes by  $\mathcal{C}_k$ . We will first characterize the class  $\mathcal{C}_k$ . It will turn out that the nontrivial forms have matrices that are a certain positive multiple of dependent Bernoulli random variables.

**Theorem 1.** *An urn scheme  $(A, X_0^{(n)})$  is in the class  $\mathcal{C}_k$  if and only if*

$$A = c\tilde{A} = c \begin{pmatrix} -B_{1,1} & B_{1,2} & \dots & B_{1,k} \\ B_{2,1} & -B_{2,2} & \dots & B_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ B_{k,1} & B_{k,2} & \dots & -B_{k,k} \end{pmatrix} \tag{1}$$

for an integer  $c \geq 1$ , and the scheme satisfies the following conditions.

- (a) *The entries of row  $r$  of  $\tilde{A}$  are Bernoulli random variables, where  $B_{r,r}$  is a  $\text{Ber}(p_r)$  random variable for some  $p_r \in (0, 1]$ .*
- (b) *The off-diagonal random variables of row  $r$  are Bernoulli random variables with conditional distribution that has mass 1 at 0 if  $B_{r,r} = 0$ , and a conditional distribution on the  $(k - 1)$ th unit simplex if  $B_{r,r} = 1$ .*
- (c) *There is a positive probability of having a cyclic structure in the form of a cyclic permutation  $\{s_1, \dots, s_k\}$  of  $\{1, \dots, k\}$  such that  $B_{1,s_1} = 1, B_{s_1,s_2} = 1, \dots, B_{s_{k-1},s_k} = 1$ , and  $B_{s_k,1} = 1$ .*
- (d) *Every component of  $X_0^{(n)}$  is a multiple of  $c$ .*

Each admissible replacement matrix in Theorem 1 has an average with a certain eigenvalue structure that has a unique principal eigenvalue 0 (as will be shown). Let  $\mathbf{v} = (v_1, v_2, \dots, v_k)$  be the corresponding (principal) left eigenvector.

**Theorem 2.** *Let  $X_j^{(n)}$  be the vector of the number of balls of different colors in a Pólya urn scheme  $(A, X_0^{(n)})$  in the class  $\mathcal{C}_k$  with  $n \geq 1$  (fixed) balls. Then  $X_j^{(n)}$  is a Markov chain with a stationary multinomial distribution on  $n/c$  trials and probabilities  $v_1, \dots, v_k$ .*

When the number of balls  $n$  is very large, the multinomial distribution of Theorem 2 becomes asymptotically multivariate normal. We will further identify rates of convergence. Let  $j_n$  be the number of draws. There are three phases of  $j_n$ .

- (a) *The growing sublinear phase, when  $j_n \rightarrow \infty$  and  $j_n = o(n)$ .*
- (b) *The linear phase, when  $j_n \sim \beta_n n$  for some  $\beta_n > 0$  of a magnitude bounded from above and below.*
- (c) *The superlinear phase, when  $n = o(j_n)$ .*

At the very low end of the sublinear phase, when  $j_n = O(1)$  as  $n \rightarrow \infty$ , there is not much change in the content of the urn, only a finite perturbation on the initial composition appears, which is negligible. Changes begin to happen when  $j_n$  grows to  $\infty$ . In what follows, we denote by  $\mathcal{N}_k(\mathbf{0}, \Sigma)$  the normally distributed  $k$ -component random vector with mean vector  $\mathbf{0}$  (of  $k$  components) and  $k \times k$  covariance matrix  $\Sigma$ , and by ‘ $\xrightarrow{D}$ ’ convergence in distribution. We develop multivariate central limit theorems in the phases, when we do not have an initially dominant proper subset of colors, that is, when there is no proper subset of colors with total

initial ball counts adding asymptotically to  $n$ . In other words, we assume that  $0 < \alpha_r < 1$  for  $r = 1, \dots, k$ . In this investigation the multivariate normal distributions refer to what some books call singular multivariate normal distributions, where  $\Sigma$  is a singular matrix, but a number of linear combinations together define a proper multivariate normal distribution of lower dimension. In the absence of an initially dominant proper subset of colors, a proper multivariate central limit theorem holds among the counts of any  $k - 1$  colors, and the count of the remaining color is, of course, normally distributed as it is the total number of balls  $n$  minus the sum of the  $k - 1$  ball counts.

**Theorem 3.** *Let  $X_j^{(n)}$  be the vector of the number of balls of different colors in an urn scheme  $(A, X_0^{(n)})$  in  $\mathcal{C}_k$  with  $n$  balls. The mean and covariance are given by*

$$\begin{aligned} E[X_j^{(n)}] &= M^j X_0^{(n)}, \\ \text{cov}[X_j^{(n)}] &= \sum_{i=1}^j \binom{j}{i} M^{j-i} X_0^{(n)} \bar{X}_0^{(n)} \frac{1}{n^i} (E[A])^i \\ &\quad + \frac{1}{n} \sum_{i=1}^j \sum_{r=0}^{i-1} \sum_{d=1}^k \binom{i-1}{r} M^{i-1-r} M_{[d]}^{(j-i)} X_0^{(n)} E[A_{[d]} \bar{A}_{[d]}] \frac{1}{n^r} (E[A])^r \\ &\quad - M^j X_0^{(n)} \bar{X}_0^{(n)} \bar{M}^j. \end{aligned}$$

Furthermore, after  $j_n$  draws from the urn with  $n$ ,  $n \rightarrow \infty$ , balls, where  $j_n \rightarrow \infty$ , in a sublinear (when there is no initially dominant proper subset of colors), linear, or superlinear fashion,

$$\frac{X_{j_n}^{(n)} - (I + E[\bar{A}]/n)^{j_n} X_0^{(n)}}{\xi_n} \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \Sigma_{\text{phase}}),$$

where  $\xi_n = \sqrt{j_n}$  in the sublinear phase,  $\xi_n = \sqrt{n}$  in the linear and superlinear phases, and  $\Sigma_{\text{phase}}$  is a constant covariance matrix depending on the phase.

**Remark.** In the presence of an initially dominant proper subset of colors, either a proper subset of  $k - 2$  or less colors have a proper multivariate central limit theorem under the scaling  $\sqrt{j_n}$  in the sublinear phase, or in critical cases a different scale factor is needed for all  $k$  different colors to follow Gaussian laws, starting at ranges differing from those found in this paper. We give results on this topic in a companion paper [22].

Theorem 3 has the following manifestations in various phases. When  $j_n$  grows sublinearly to  $\infty$ , the initial conditions persist, and the asymptotic normal result in this case contains the vector  $\alpha$ . The Gaussian law in Theorem 3 takes the form

$$\frac{X_{j_n}^{(n)} - (I + E[\bar{A}]/n)^{j_n} X_0^{(n)}}{\sqrt{j_n}} \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \Sigma_{\text{sub}});$$

we specify  $\Sigma_{\text{sub}}$  in the sequel (the covariance matrix  $\Sigma_{\text{sub}}$  involves  $\alpha$ ).

Normality continues to hold in the linear and superlinear phases. However, in each phase we obtain a different normal distribution. The mean and scale factors are essentially different. In the linear phase a different normal distribution (with the usual  $\sqrt{n}$  scaling of central limit theorems) is in effect, and the parameters of the distribution depend on both  $\alpha$  and the coefficient of linearity.

Deep in the linear phase (with large linearity coefficient), when  $j_n \sim \beta_n n$ , the Gaussian law in Theorem 3 takes the form

$$\frac{\mathbf{X}_{j_n}^{(n)} - \boldsymbol{\mu}_n n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}_{\text{lin}});$$

we specify  $\boldsymbol{\mu}_n$  and  $\boldsymbol{\Sigma}_{\text{lin}}$  in the sequel. The vector  $\boldsymbol{\mu}_n$  contains a fixed component and an exponentially decaying part.

After a very long period of time, as in the superlinear phase, when  $n = o(j_n)$ , the distribution approaches the multinomial (asymptotically normal) steady state. Deep in the superlinear phase the Gaussian law in Theorem 3 takes the form

$$\frac{\mathbf{X}_{j_n}^{(n)} - \bar{\mathbf{v}} n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}_{\text{sup}});$$

we specify  $\boldsymbol{\Sigma}_{\text{sup}}$  in the sequel (the covariance matrix  $\boldsymbol{\Sigma}_{\text{sup}}$  does not involve  $\boldsymbol{\alpha}$ ).

Note how the effect of any initial conditions is eventually obliterated. Though we have one master theorem for all the phases, certain aspects of the proof are different in each phase. Basically, a different approximation technique is used in each of the three phases.

#### 4. Organization

The rest of this paper has the following organization. In Section 5 we set up a basic stochastic recurrence. In Section 6 we characterize the class  $\mathcal{C}_k$  by necessary and sufficient conditions (proof of Theorem 1). Subsections 6.1 and 6.2 are devoted to defining the notions of irreducibility and nondegeneracy, respectively. Sufficient conditions are discussed in Subsection 6.3; necessary conditions are discussed in Subsection 6.4. In Section 7 we derive the steady-state multinomial distribution (proof of Theorem 2). In Section 8 we present the mean and variance of  $\mathbf{X}_j^{(n)}$ . In Section 9 we derive the underlying martingale. In Section 10 we discuss the three phases, the growing sublinear, the linear, and the superlinear, with a subsection devoted to each phase. In each subsection we prove the master theorem in one phase. We conclude in Section 11 with some examples.

#### 5. Basic stochastic recurrence

In a tenable, zero-balanced scheme a stochastic path of  $j$  steps corresponds to a history of nonhalting replacements up to  $j - 1$  steps, a determination of the color of the  $j$ th drawn ball, and an independent realization of  $\mathbf{A}$ . Let  $\mathbf{1}_{j,r}^{(n)}$  be an indicator of the event that a ball of color  $r$  is sampled in the  $j$ th draw. (The indicators  $\mathbf{1}_{j,1}^{(n)}, \mathbf{1}_{j,2}^{(n)}, \dots, \mathbf{1}_{j,k}^{(n)}$  are mutually exclusive: only one can be 1 while all others are 0.) Each indicator, when it realizes the value 1, takes the urn along a different stochastic path. By tenability,  $\mathbf{X}_{j-1}^{(n)}$  is a vector with nonnegative components. Let  $\mathbf{A}_j = [A_{r,s}^{(j)}]$  be a realization of  $\mathbf{A}$  in the  $j$ th step. For each color  $r$ , we have

$$X_{j,r}^{(n)} = X_{j-1,r}^{(n)} + A_{1,r}^{(j)} \mathbf{1}_{j,1}^{(n)} + A_{2,r}^{(j)} \mathbf{1}_{j,2}^{(n)} + \dots + A_{k,r}^{(j)} \mathbf{1}_{j,k}^{(n)}. \tag{2}$$

When we write these components (for  $r = 1, \dots, k$ ) in matrix form, we obtain the stochastic recursive formula

$$\mathbf{X}_j^{(n)} = \mathbf{X}_{j-1}^{(n)} + \bar{\mathbf{A}}_j \mathbf{J}_j^{(n)}, \tag{3}$$

where  $\bar{\mathbf{J}}_j^{(n)} = (\mathbf{1}_{j,1}^{(n)}, \mathbf{1}_{j,2}^{(n)}, \dots, \mathbf{1}_{j,k}^{(n)})$ .

## 6. Characterization

In this section we prove Theorem 1. It is helpful to get rid of redundant situations. We go over the concepts of irreducibility and nondegeneracy first. At the core of these concepts is the idea that every color is essential in some nonartificial way.

### 6.1. Irreducibility

Color  $r$  is said to be *essential* if there is probability 1 of finding balls of that color after some number of draws from an urn for every monochromatic starting condition void of that color (that is, whenever all the balls at the start are of one color, which is not  $r$ ). A scheme is *irreducible* if all its colors are essential. If a scheme is not irreducible, it is said to be *reducible*. For example, the Pólya–Eggenberger scheme is irreducible. An example of a reducible scheme is

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix};$$

we cannot ever find balls of color 3 in the urn if we start with five balls of color 2. Color 3, and in fact all the colors, are not essential.

**Lemma 1.** *In the replacement matrix  $A$  of a tenable, irreducible, zero-balanced scheme, entry  $A_{i,i}$  for  $i = 1, \dots, k$  is nonnegative, and cannot be identically 0.*

*Proof.* Assume that the diagonal element  $A_{r,r}$  of  $A$  has a positive realization. The zero-balance condition necessitates that some entry on the same row, say  $A_{r,s}$ , realizes a negative value. The irreducibility condition ensures that, sooner or later, a ball of color  $r$  appears (with probability 1), and subsequent draws of color- $r$  balls will deplete color- $s$  balls, while increasing color- $r$  balls. So, we can follow one such stochastic path, then realize a positive value of  $A_{r,r}$  and a negative value for  $A_{r,s}$ , with the scheme coming to a halt after one extra draw of a color- $r$  ball, contradicting its tenability. Entry  $A_{r,r}$  is nonpositive.

Next, assume that  $A_{r,r} \equiv 0$ , and that we start with a monochromatic urn, in which all the balls are of color  $r$ . The entire  $r$ th row must be 0, otherwise, by the zero balance condition we have positive and negative entries. Then we can deplete some other color by repeatedly drawing color- $r$  balls (while keeping the number of color- $r$  balls unchanged), and drive the process to a halt, contradicting tenability. However, if the entire  $r$ th row is 0, no other color will ever appear, contradicting irreducibility.

### 6.2. Nondegeneracy

When two colors, say  $r$  and  $r'$ , are tied together (that is, when the number of balls of one color is always the same multiple of the number of balls of the other color), we consider the situation a degeneracy. More precisely, we say that a tenable, zero-balanced urn scheme is degenerate if, for a positive constant  $m \geq 1$ ,  $X_{j,r}^{(n)} = mX_{j,r'}^{(n)}$  for all  $j \geq 0$ . In this case, we can reduce the dimensionality of the problem from  $k$  to  $k - 1$  by combining the two colors. That is, we can study a  $(k - 1)$ th scheme in which colors  $r$  and  $r'$  are replaced with a new color. Without loss of generality, we can assume that the two colors tied together are the colors  $k - 1$  (crimson) and  $k$  (pink). We construct a new equivalent urn scheme in which the two colors are replaced by a new color (simply très chic red). The new scheme has a  $(k - 1) \times (k - 1)$  replacement matrix, in which simply très chic red corresponds to a new  $(k - 1)$ th row. All the entries of the new replacement matrix remain the same as those in the old  $1, \dots, k - 2$  columns. In the new

$(k - 1)$ th column all the entries are replaced with the new entries  $A'_{r,k-1} = A_{r,k-1} + A_{r,k}$  for  $r = 1, \dots, k - 1$ . In the starting vector we also combine the  $(k - 1)$ th and  $k$ th components into one component at position  $k - 1$ , given by  $\tilde{X}_{0,k-1}^{(n)} = X_{0,k-1}^{(n)} + X_{0,k}^{(n)}$ . Let  $\tilde{X}_{j,k-1}^{(n)}$  be the number of balls of the simply très chic red color after  $j$  draws. We can develop all the results for the simply très chic red color, then recover the behavior of the crimson and pink colors from their proportionality relations:  $X_{j,k-1}^{(n)} = m\tilde{X}_{j,k-1}^{(n)}/(m + 1)$  and  $X_{j,k}^{(n)} = \tilde{X}_{j,k-1}^{(n)}/(m + 1)$ . Subsequently, if we analyze the class  $\mathcal{C}_k$ , we can still say a definite word about a tenable, irreducible, degenerate zero-balanced  $k$ -color urn scheme. We illustrate this with an example in Subsection 11.4.

**Lemma 2.** *A tenable, irreducible, zero-balanced scheme is degenerate if and only if there is one column of its replacement matrix that is proportional (in all its components) to another.*

*Proof.* Suppose that the scheme is degenerate. There are two distinct colors, say  $r$  and  $s$  (with  $r \neq s$ ), that are tied together—for some integer  $m \geq 1$  and all  $j \geq 0$ ,  $X_{j,r}^{(n)} = mX_{j,s}^{(n)}$ . Drawing a ball of any color keeps this proportionality. Thus, if we draw a ball of color  $i$  for any  $i = 1, \dots, k$ , we obtain

$$X_{j,r}^{(n)} = X_{j-1,r}^{(n)} + A_{i,r}^{(j)}, \quad X_{j,s}^{(n)} = X_{j-1,s}^{(n)} + A_{i,s}^{(j)}.$$

But then  $X_{j-1,r}^{(n)} = mX_{j-1,s}^{(n)}$  and  $X_{j,r}^{(n)} = mX_{j,s}^{(n)}$ . Solving these equations for  $A_{i,r}^{(j)}$  and  $A_{i,s}^{(j)}$ , we find (for any  $i$ ) that  $A_{i,r}^{(j)} = mA_{i,s}^{(j)}$ ; column  $s$  of the replacement matrix is proportional (in all its components) to column  $r$ .

The converse is obviously true (note that if two colors are tied and one disappears, the other must also disappear).

**6.3. Sufficient conditions for the class  $\mathcal{C}_k$**

We will show that a Pólya urn scheme is in the class  $\mathcal{C}_k$  if its replacement matrix has the form (1), provided that the components of the initial vector  $X_0^{(n)}$  are nonnegative multiples of  $c$  (which we call the *initial divisibility condition*). By virtue of its construction, the form (1) is zero balanced and, by Lemma 2, it is also nondegenerate.

**Lemma 3.** *Under the form (1) and the initial divisibility condition, each component of  $X_j^{(n)}$  is a nonnegative multiple of  $c$  (that is, the scheme is tenable), with at least one being a positive multiple of  $c$ .*

*Proof.* We prove the lemma by induction on  $j$ . The initial divisibility condition (Theorem 1(d)) guarantees this for  $j = 0$ , providing a basis for induction. Assume that the hypothesis is true for  $X_{j-1}^{(n)}$ . Suppose that  $\mathbf{1}_{j,s}^{(n)} = 1$  (with all other indicators equal to 0), and write (2) as

$$X_{j,r}^{(n)} = X_{j-1,r}^{(n)} + A_{s,r}^{(j)}.$$

By hypothesis,  $X_{j-1,r}^{(n)}$  is a nonnegative multiple of  $c$ , and, according to the form of the replacement matrix,  $A_{s,r}^{(j)}$  is a nonnegative multiple of  $c$  for  $r \neq s$ . Hence,  $X_{j,r}^{(n)}$  is a multiple of  $c$  for  $r \neq s$ . As for the component  $X_{j,s}^{(n)}$ , we have

$$X_{j,s}^{(n)} = X_{j-1,s}^{(n)} - cB_{s,s}.$$

By induction,  $X_{j-1,s}^{(n)}$  is a nonnegative multiple of  $c$ . This multiple cannot be 0 as the  $s$ th indicator is 1, which ensures that there are balls of color  $s$  at the  $j$ th draw; thus,  $X_{j-1,s}^{(n)}$  is a

positive multiple of  $c$ , and  $X_{j,s}^{(n)}$  is also a multiple of  $c$ , regardless of whether the realization of the Bernoulli random variable  $B_{s,s}$  is 0 or 1. The induction is complete. It is evident that not all these multiples are 0 simultaneously; at least one must be positive, as we have a total of  $n$  balls.

**Lemma 4.** *A scheme with replacement matrix of the form (1) and starting values satisfying the initial divisibility condition is irreducible.*

*Proof.* A scheme in the form (1) has a cyclic structure. Let  $\mathcal{E}_{j,s_i}$  be the event that, starting at some  $j$ th draw, we successively go through the permutation of all colors starting with a  $j$ th draw of color  $s_i$ . That is,  $\mathcal{E}_{j,s_i}$  is the event that  $\mathbf{1}_{j,s_i}^{(n)} = 1, \mathbf{1}_{j+1,s_{i+1}}^{(n)} = 1, \dots, \mathbf{1}_{j+k-1,s_{i+k-1}}^{(n)} = 1$  (interpret the index of  $s$  to wrap around into 1 if it exceeds  $k$ ), and  $p = P(\mathcal{E}_{j,s_i}) > 0$ . (By Lemma 3, drawing forever is ensured, and  $\mathcal{E}_{j,s_i}$  is well defined.)

The sequence  $\{\mathcal{E}_{jk+1,s_i}\}_{j=0}^\infty$  comprises independent events, and

$$\sum_{j=0}^\infty P(\mathcal{E}_{jk+1,s_i}) = \sum_{j=0}^\infty p = \infty.$$

By the Borel–Cantelli lemma,  $P(\mathcal{E}_{jk+1,s_i}$  infinitely often) = 1, and, consequently,  $P(\mathcal{E}_{j,s_i}$  infinitely often) = 1. So, any color has probability 1 of appearing infinitely often, if we start monochromatically with balls of color  $s_i$ . This being true for  $s_i, i = 1, 2, \dots, k$ , shows that all colors are essential; an urn scheme with replacement matrix of the form (1) and starting values satisfying the initial divisibility is irreducible.

The zero-balance nature and nondegeneracy of (1) and Lemmas 3–4 assert that this choice of replacement matrix with the initial divisibility condition are sufficient conditions for a scheme to be in the class  $\mathcal{C}_k$ . This proves the sufficiency part of Theorem 1.

**6.4. Necessary conditions for the class  $\mathcal{C}_k$**

We show here that the form (1) and the initial divisibility condition are necessary for a Pólya urn scheme to qualify in  $\mathcal{C}_k$ . Throughout this section, we assume that the scheme  $(A, X_0^{(n)})$  is in  $\mathcal{C}_k$ . It was shown in Lemma 1 that the diagonal entries of  $A$  are nonpositive. It remains to determine the ranges of these entries.

**Lemma 5.** *For some positive integer  $c_r$  and some  $p_r \in (0, 1]$ ,*

$$A_{r,r} = -c_r \text{Ber}(p_r).$$

*Proof.* We have established in Lemma 1 that, for a scheme to be tenable, irreducible, and zero-balanced, a diagonal random variable such as  $A_{r,r}$  must have a nonpositive support (and cannot be identically 0). Thus,

$$A_{r,r} \in \{-n, -(n - 1), \dots, -2, -1, 0\}.$$

Suppose (toward a contradiction) that realizations of  $A_{r,r}$  can assume two negative values,  $-y$  and  $-z$ , with  $y < z$ . There is a stochastic path that brings the process to a halt. We describe this path. If  $X_{0,r}^{(n)} > 0$ , we can draw balls of color  $r$ ; otherwise, there are stochastic paths (with probability 1) on which, sooner or later, there will be balls of color  $r$  by the irreducibility condition. Let  $j$  be the smallest index on such a path, such that  $X_{j,r}^{(n)} > 0$ . Clearly,  $X_{j,r}^{(n)}$  must be a multiple of both  $y$  and  $z$ . Now  $\lfloor X_{j,r}^{(n)} / y \rfloor - 1$  draws of color- $r$  balls, each time having  $A_{r,r}$  realize the value  $-y$ , leave a remainder of  $y$  balls of color  $r$ . One more draw of a color- $r$  ball, for which  $A_{r,r}$  realizes the value  $-z$ , will halt the process, as there are only  $y < z$  balls of

color  $r$  to take out. In view of this contradiction,  $A_{r,r}$  can realize at most one negative value or 0 (but cannot be identically 0). That is, for some positive integer  $c_r$  and some  $p_r \in (0, 1]$ ,  $A_{r,r}$  falls in  $\{-c_r, 0\}$ .

Henceforth, we assume that  $A_{r,r} = -c_r \text{Ber}(p_r)$  for  $p_r \in (0, 1]$ .

**Corollary 1.** *The off-diagonal entries of  $A$  are nonnegative random variables.*

*Proof.* As the scheme is nondegenerate, colors  $r$  and  $s$  are not tied together and do not have to disappear together. It is possible for the urn to have balls of color  $r$  but not  $s$ . Starting from such a state, if  $A_{r,s}$  can realize a negative value, the drawing halts, contradicting tenability.

**Lemma 6.** *The positive values  $c_1, c_2, \dots, c_k$  are all the same.*

*Proof.* Without loss of generality, assume that

$$c_1 \leq c_2 \leq \dots \leq c_k.$$

Suppose that, for some  $1 \leq q_1, q_2, \dots, q_v \leq k$ ,  $c_1 = \dots = c_{q_1} < c_{q_1+1} = c_{q_1+2} = \dots = c_{q_1+q_2} < c_{q_1+q_2+1} = \dots < c_{q_1} + \dots + c_{q_{v-1}+1} = c_{q_1} + \dots + c_{q_{v-1}+2} = \dots = c_k$ . We argue that the submatrix in rows 1 to  $q_1 + \dots + q_{v-1}$  and columns  $q_1 + \dots + q_{v-1} + 1$  to  $k$  is a block of zeros. Let  $A_{r,s}$  be a variable from that submatrix. By tenability, it must be a multiple of  $c_k$ . However, in view of the zero-balance condition and the sign structure determined by Lemma 1 and Corollary 1,  $A_{r,s}$  should not exceed  $c_r < c_k$ . The only possible such multiple is 0. But then the scheme is not irreducible (as can be seen from starting void of the colors  $q_1 + \dots + q_{v-1} + 1, \dots, k$ ). This is a contradiction.

**Corollary 2.** *The off-diagonal entries of  $A$  are nonnegative random variables that are multiples of  $c$ .*

**Corollary 3.** *The initial value  $X_{0,r}^{(n)}$  is a nonnegative multiple of  $c$ .*

**Corollary 4.** *The entries of row  $r$  are random variables that are of the form  $cB_{r,1}, cB_{r,2}, \dots, cB_{r,r-1}, -cB_{r,r}, cB_{r,r+1}, \dots, cB_{r,k}$ , where the diagonal element is  $-c\text{Ber}(p_r)$  for some  $p_r \in (0, 1]$ , and the off-diagonal random variables are  $c$ -times Bernoulli random variables with conditional distribution that has mass 1 at 0 if  $B_{r,r} = 0$ , and a conditional distribution on the  $(k - 1)$ th unit simplex if  $B_{r,r} = 1$ .*

*Proof.* If  $B_{r,r}$  realizes the value 0, all other (nonnegative) random variables on the same row must be 0, by the zero-balance condition and the sign structure determined by Lemma 1 and Corollary 1. If, alternatively,  $B_{r,r}$  realizes the value 1, by the zero-balance condition, the (nonnegative integer-valued) off-diagonal random variables on the same row must add up to 1. One and only one assumes the value 1, the rest are 0.

Finally, we argue the cyclic structure.

**Corollary 5.** *For any given realization of  $A$ , there is a cyclic permutation  $\{s_1, \dots, s_k\}$  of  $\{1, \dots, k\}$  such that  $B_{1,s_1} = 1, B_{s_1,s_2} = 1, \dots, B_{s_{k-1},s_k} = 1$ , and  $B_{s_k,1} = 1$ .*

*Proof.* Suppose that, for every possible realization of  $A$ , there is no such cyclic permutation. So, any cycle has length shorter than  $k$ . In other words, for any given sequence of  $k$  realizations of  $A$ , there is only a cyclic permutation of  $\{s_1, \dots, s_\ell\}$  such that  $B_{s_1,s_2} = 1, B_{s_2,s_3} = 1, \dots, B_{s_{\ell-1},s_\ell} = 1$ , and  $B_{s_\ell,s_1} = 1$  for some  $\ell < k$ . Starting with only balls of colors  $s_1, \dots, s_\ell$ , the other  $k - \ell$  colors never appear, contradicting irreducibility.

Lemmas 5–6 and their corollaries demonstrate that, indeed, the replacement matrices of the form (1) in addition to the initial divisibility conditions are the only admissible schemes for the class  $\mathcal{C}_k$ .

### 7. The Markov chain and its stationary distribution

For a  $k$ -color scheme in the class  $\mathcal{C}_k$ , the vector  $X_j^{(n)}$  alternates in value between vectors of the form  $\bar{x} = c(x_1, x_2, \dots, x_k)$ , where the components of  $x$  are nonnegative solutions in nonnegative integers to the equation

$$x_1 + \dots + x_k = \frac{n}{c} =: N;$$

we call such a solution a *feasible partition*. Given any particular feasible partition of colors in the urn, the change of composition after one drawing depends only on the color of the ball sampled from the given feasible partition and the realization of  $A$  at that step. The process  $X_j^{(n)}$  is therefore a Markov chain on  $\binom{N+k-1}{k-1}$  states, each of which corresponds to one feasible partition. The irreducibility aspect of the scheme makes at least  $k$  states of the Markov chain nontransient.

Let  $p_{r,s}$  be the probability that  $A_{s,r} = 1$  and  $A_{s,s} = -1$ . The diagonal elements of the Markov transition matrix are given by

$$P(\bar{X}_j^{(n)} = (x_1, \dots, x_k) \mid \bar{X}_{j-1}^{(n)} = (x_1, \dots, x_k)) = \sum_{r=1}^k \frac{x_r}{N} (1 - p_r),$$

and the off-diagonal elements are given by

$$\begin{aligned} P(\bar{X}_j^{(n)} = (x_1, \dots, x_k) \mid \bar{X}_{j-1}^{(n)} = (x_1, x_2, \dots, x_{r-1}, x_r - 1, x_{r+1}, \dots, \\ x_{s-1}, x_s + 1, x_{s+1}, \dots, x_k)) \\ = \frac{x_s + 1}{N} p_{r,s}. \end{aligned}$$

Let  $\pi(x_1, \dots, x_k)$  be the stationary probability that the process is in state  $(x_1, \dots, x_k)$ , and let  $\pi$  be the  $\binom{N+k-1}{k-1}$ -component vector of stationary probabilities. We solve the equation  $\pi = \pi Q$ , where  $Q$  is the Markov transition matrix as follows.

**Lemma 7.** *The stationary distribution is Multinomial( $N; v_1, \dots, v_k$ ), where  $(v_1, v_2, \dots, v_k) = v$  is the left row eigenvector corresponding to the eigenvalue 0 of  $A$ .*

*Proof.* Writing the relation  $\pi = \pi Q$  component by component we obtain the  $\binom{N+k-1}{k-1}$  equations

$$\begin{aligned} \pi(x_1, \dots, x_k) = \pi(x_1, \dots, x_k) \sum_{r=1}^k \frac{x_r}{N} (1 - p_r) \\ + \sum_{s=1}^k \frac{x_s + 1}{N} \sum_{r \neq s} p_{r,s} \pi(x_1, x_2, \dots, x_{r-1}, x_r - 1, x_{r+1}, \dots, \\ x_{s-1}, x_s + 1, x_{s+1}, \dots, x_k) \end{aligned} \tag{4}$$

(one equation for every feasible partition). Let  $\mathbf{a} = \overline{(a_1, a_2, \dots, a_k)}$  be a vector of positive numbers (probabilities) in  $(0, 1]$  such that  $\sum_{j=1}^k a_k = 1$ . If we substitute the multinomial probabilities

$$\pi(x_1, \dots, x_k) = \binom{N}{x_1, x_2, \dots, x_k} a_1^{x_1} a_2^{x_2} \dots a_k^{x_k}$$

on the right-hand side of (4), we obtain

$$\begin{aligned} & \binom{N}{x_1, x_2, \dots, x_k} a_1^{x_1} a_2^{x_2} \dots a_k^{x_k} \sum_{r=1}^k \frac{x_r}{N} (1 - p_r) \\ & + \sum_{s=1}^k \frac{x_s + 1}{N} \sum_{r \neq s} p_{r,s} a_1^{x_1} \dots a_{r-1}^{x_{r-1}} a_r^{x_r - 1} a_{r+1}^{x_{r+1}} \dots a_{s-1}^{x_{s-1}} a_s^{x_s + 1} a_{s+1}^{x_{s+1}} \dots a_k^{x_k} \\ & \quad \times \binom{N}{x_1, \dots, x_{r-1}, x_r - 1, x_{r+1}, \dots, x_{s-1}, x_s + 1, x_{s+1}, \dots, x_k} \\ & = \binom{N}{x_1, x_2, \dots, x_k} a_1^{x_1} a_2^{x_2} \dots a_k^{x_k} \left( \sum_{r=1}^k \frac{x_r}{N} (1 - p_r) + \sum_{s=1}^k \frac{x_s + 1}{N} \sum_{r \neq s} p_{r,s} a_s \frac{x_r}{a_r (x_s + 1)} \right). \end{aligned}$$

Indeed,  $\pi(x_1, \dots, x_k) = \binom{N}{x_1, x_2, \dots, x_k} a_1^{x_1} a_2^{x_2} \dots a_k^{x_k}$  is a solution to the system of equations (4), if the numbers  $a_1, \dots, a_k$  satisfy the equations

$$N = \sum_{r=1}^k x_r (1 - p_r) + \sum_{s=1}^k \sum_{r \neq s} p_{r,s} a_s \frac{x_r}{a_r}$$

for every feasible partition of  $\mathbf{x}$ . There are  $\binom{N+k-1}{k-1}$  such partitions. So, as equations in the  $k$  variables  $a_1, \dots, a_k$ , this is an over-determined system. We also need to satisfy  $\sum_{j=1}^k a_j = 1$ . We can solve any independent  $k$  of these combined equations and verify that such a solution is unique, and consistent with the rest of the equations in the over-determined system. Let us solve the  $k$  equations where only one  $x_r$  is  $N$  and all the remaining shares in the partition are 0. This gives us the linear system of equations

$$(a_1, a_2, \dots, a_k) \begin{pmatrix} -p_1 & p_{2,1} & \dots & p_{k,1} \\ p_{1,2} & -p_2 & \dots & p_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1,k} & p_{2,k} & \dots & -p_k \end{pmatrix} = (0, 0, \dots, 0).$$

A solution to this equation is an eigenvector of  $E[\overline{\mathbf{A}}]$  corresponding to the eigenvalue 0. The constraint  $\sum_{j=1}^k a_j = 1$  further determines the scale as already chosen.

### 8. Exact and asymptotic mean and variance

In this section we obtain explicit expressions for the mean and the covariance matrix of  $\mathbf{X}_j^{(n)}$ , and describe the asymptotic behavior of this random vector after  $j = j_n$  draws, for functions  $j_n$  of various growth rates. In what follows,  $\mathcal{F}_j$  is the sigma-field generated by the first  $j$  draws.

**Lemma 8.** *We have*

$$\begin{aligned}
 E[X_j^{(n)}] &= M^j X_0^{(n)}, \\
 \text{cov}[X_j^{(n)}] &= \sum_{i=1}^j \binom{j}{i} M^{j-i} X_0^{(n)} \bar{X}_0^{(n)} \frac{1}{n^i} (E[A])^i \\
 &\quad + \frac{1}{n} \sum_{i=1}^j \sum_{r=0}^{i-1} \sum_{d=1}^k \binom{i-1}{r} M^{i-1-r} M_{[d]}^{(j-i)} X_0^{(n)} E[A_{[d]} \bar{A}_{[d]}] \frac{1}{n^r} (E[A])^r \\
 &\quad - M^j X_0^{(n)} \bar{X}_0^{(n)} \bar{M}^j.
 \end{aligned}$$

*Proof.* The expectation of (3) is

$$\begin{aligned}
 E[X_j^{(n)}] &= E[X_{j-1}^{(n)}] + E[\bar{A} J_j^{(n)}] \\
 &= E[X_{j-1}^{(n)}] + E[\bar{A}] E[E[J_j^{(n)} \mid \mathcal{F}_{j-1}]] \\
 &= E[X_{j-1}^{(n)}] + E[\bar{A}] E\left[\frac{1}{n} X_{j-1}^{(n)}\right] \\
 &= M^j X_0^{(n)}.
 \end{aligned}$$

A recurrence for the covariance follows from the expectation of the product of (3) and its transpose:

$$\text{cov}[X_j^{(n)}] = E[X_j^{(n)} \bar{X}_j^{(n)}] - E[X_j^{(n)}] E[\bar{X}_j^{(n)}] = E[X_j^{(n)} \bar{X}_j^{(n)}] - M^j X_0^{(n)} \bar{X}_0^{(n)} \bar{M}^j. \tag{5}$$

We need to only evaluate the first term  $E[X_j^{(n)} \bar{X}_j^{(n)}]$ . From (3) we have

$$E[X_j^{(n)} \bar{X}_j^{(n)}] = E[X_{j-1}^{(n)} \bar{X}_{j-1}^{(n)}] + E[\bar{A} J_j^{(n)} \bar{X}_{j-1}^{(n)}] + E[X_{j-1}^{(n)} \bar{J}_j^{(n)} A] + E[\bar{A} J_j^{(n)} \bar{J}_j^{(n)} A]. \tag{6}$$

We now consider the last three components on the right-hand of (6). We have

$$E[\bar{A} J_j^{(n)} \bar{X}_{j-1}^{(n)}] = \overline{E[X_{j-1}^{(n)} \bar{J}_j^{(n)} A]} = \frac{1}{n} E[\bar{A}] E[X_{j-1}^{(n)} \bar{X}_{j-1}^{(n)}], \tag{7}$$

the transpose of which is

$$E[X_{j-1}^{(n)} \bar{J}_j^{(n)} A] = \frac{1}{n} E[X_{j-1}^{(n)} \bar{X}_{j-1}^{(n)}] E[A]. \tag{8}$$

We also have

$$\begin{aligned}
 E[\bar{A} J_j^{(n)} \bar{J}_j^{(n)} A] &= E[\bar{A} \text{diag}(\mathbf{1}_{j,1}^{(n)}, \dots, \mathbf{1}_{j,k}^{(n)}) A] \\
 &= \sum_{d=1}^k E[\mathbf{1}_{j,d}^{(n)}] E[\bar{A} G(d) A] \\
 &= \sum_{d=1}^k E[\mathbf{1}_{j,d}^{(n)}] E[\bar{A}_{[d]} A_{[d]}] \\
 &= \frac{1}{n} \sum_{d=1}^k E[\bar{A}_{[d]} A_{[d]}] M_{[d]}^{(j-1)} X_0^{(n)}.
 \end{aligned} \tag{9}$$

We substitute (7), (8), and (9) into (6) to obtain

$$\begin{aligned}
 E[X_j^{(n)} \bar{X}_j^{(n)}] &= \mathbf{M} E[X_{j-1}^{(n)} \bar{X}_{j-1}^{(n)}] + \frac{1}{n} E[X_{j-1}^{(n)} \bar{X}_{j-1}^{(n)}] E[\mathbf{A}] \\
 &\quad + \frac{1}{n} \sum_{d=1}^k E[\bar{\mathbf{A}}_{[d]} \mathbf{A}_{[d]}] \mathbf{M}_{[d]}^{(j-1)} \mathbf{X}_0^{(n)}.
 \end{aligned}
 \tag{10}$$

To simplify the notation of the above equation, we introduce

$$\mathbf{S}_j = E[X_j^{(n)} \bar{X}_j^{(n)}], \quad \mathbf{C}_j = \frac{1}{n} \sum_{d=1}^k E[\bar{\mathbf{A}}_{[d]} \mathbf{A}_{[d]}] \mathbf{M}_{[d]}^{(j-1)} \mathbf{X}_0^{(n)},$$

so that (10) takes the recursive form  $\mathbf{S}_j = \mathbf{M} \mathbf{S}_{j-1} + \mathbf{S}_{j-1} \mathbf{B} + \mathbf{C}_j$ . It can be easily shown by induction that the solution is

$$\mathbf{S}_j = \sum_{i=0}^j \binom{j}{i} \mathbf{M}^{j-i} \mathbf{S}_0 \mathbf{B}^i + \sum_{i=1}^j \sum_{r=0}^{i-1} \binom{i-1}{r} \mathbf{M}^{i-1-r} \mathbf{C}_j \mathbf{B}^r.$$

Hence, we have

$$\begin{aligned}
 E[X_j^{(n)} \bar{X}_j^{(n)}] &= \sum_{i=0}^j \binom{j}{i} \mathbf{M}^{j-i} \mathbf{X}_0^{(n)} \bar{\mathbf{X}}_0^{(n)} \frac{1}{n^i} (E[\mathbf{A}])^i \\
 &\quad + \frac{1}{n} \sum_{i=1}^j \sum_{r=0}^{i-1} \sum_{d=1}^k \binom{i-1}{r} \mathbf{M}^{i-1-r} E[\bar{\mathbf{A}}_{[d]} \mathbf{A}_{[d]}] \mathbf{M}_{[d]}^{(j-i)} \mathbf{X}_0^{(n)} \frac{1}{n^r} (E[\mathbf{A}])^r.
 \end{aligned}$$

Substituting this into (5), the covariance follows in the form stated.

The following corollaries provide the mean and covariance asymptotics for the growing sublinear, linear, and superlinear phases.

**Corollary 6.** *We have*

$$E[X_{j_n}^{(n)}] = \begin{cases} \boldsymbol{\alpha} n + \mathbf{o}(n) & \text{for } j_n = o(n), \\ \mathbf{H} \text{diag}(e^{\beta_n \lambda_1}, \dots, e^{\beta_n \lambda_k}) \mathbf{H}^{-1} \boldsymbol{\alpha} n + \mathbf{o}(n) & \text{for } j_n \sim \beta_n n, \\ \bar{\mathbf{v}} n + \mathbf{o}(n) & \text{for } n = o(j_n). \end{cases}$$

*Proof.* From Lemma 8 we have

$$\begin{aligned}
 E[X_{j_n}^{(n)}] &= \left( \mathbf{I} + \frac{1}{n} E[\bar{\mathbf{A}}] \right)^{j_n} \mathbf{X}_0^{(n)} \\
 &= \mathbf{H} \text{diag} \left( \left( 1 + \frac{1}{n} \lambda_1 \right)^{j_n}, \dots, \left( 1 + \frac{1}{n} \lambda_k \right)^{j_n} \right) \mathbf{H}^{-1} \mathbf{X}_0^{(n)}.
 \end{aligned}
 \tag{11}$$

We have

$$\left( 1 + \frac{1}{n} \lambda_r \right)^{j_n} = \begin{cases} 1 + O\left(\frac{j_n}{n}\right) & \text{in the sublinear phase,} \\ e^{\beta_n \lambda_r} + o(1) & \text{in the linear phase, when } j_n \sim \beta_n n. \end{cases}$$

Thus, in the sublinear phase (11) asymptotically becomes

$$E[X_{j_n}^{(n)}] = \left(1 + O\left(\frac{j_n}{n}\right)\right) \mathbf{H} \text{diag}(1, \dots, 1) \mathbf{H}^{-1} \mathbf{X}_0^{(n)} = \left(1 + O\left(\frac{j_n}{n}\right)\right) \mathbf{X}_0^{(n)} = \alpha n + o(n).$$

The argument for the linear phase is similar.

Note that if there are positive eigenvalues, one or more of the expressions  $(1 + \lambda_r/n)^{j_n}$  would grow exponentially in the superlinear phase, and the sum of the components of  $E[X_{j_n}^{(n)}]$  (i.e. the total number of balls) would exceed  $n$ . This contradiction shows that the eigenvalues are all nonpositive. As  $E[\mathbf{A}]$  is singular, 0 is an eigenvalue, so it is the principal eigenvalue. The nondegeneracy of  $\mathcal{C}_k$  renders all the eigenvalues distinct; the real parts of the eigenvalues form a decreasing sequence. Whence, in the superlinear phase

$$\left(1 + \frac{1}{n} \lambda_i\right)^{j_n} \rightarrow \begin{cases} 1 & \text{if } i = 1, \\ 0 & \text{if } 2 \leq i \leq k. \end{cases}$$

In the superlinear phase (11) asymptotically becomes

$$E[X_{j_n}^{(n)}] = \mathbf{H} \mathbf{G}(1) \mathbf{H}^{-1} \mathbf{X}_0^{(n)} + o(n).$$

Let  $\mathbf{H}^{-1} = [h_{r,s}]_{1 \leq r,s \leq k}$ . Multiplying out, we obtain

$$E[X_{j_n}^{(n)}] = \bar{v} \sum_{r=1}^k \alpha_r h_{1,r} n + o(n).$$

According to the chosen scale for the principal eigenvector,  $h_{1,r} = 1$  for  $r = 1, \dots, k$ , and we have  $\sum_{r=1}^k \alpha_r = 1$ , so

$$E[X_{j_n}^{(n)}] = \bar{v} n + o(n).$$

**Corollary 7.** (a) *For the growing sublinear phase, when  $j_n \rightarrow \infty$  and  $j_n = o(n)$ ,*

$$\text{cov}[X_{j_n}^{(n)}] = j_n \left( \sum_{d=1}^k \alpha_d E[\bar{\mathbf{A}}_{[d]} \mathbf{A}_{[d]}] - E[\bar{\mathbf{A}}] \alpha \bar{\alpha} E[\mathbf{A}] \right) + o(j_n).$$

(b) *For the linear phase, when  $j_n \sim \beta_n n$  for some  $\beta_n > 0$  of a magnitude bounded from above and below,*

$$\begin{aligned} \text{cov}[X_{j_n}^{(n)}] = & n \left( - \sum_{d=2}^k \lambda_d e^{\beta_n \lambda_d} \mathbf{H} \mathbf{G}(d) \mathbf{H}^{-1} \alpha \bar{\alpha} \bar{\mathbf{H}}^{-1} \text{diag}(0, \dots, \beta_n \lambda_k e^{\beta_n \lambda_k}) \bar{\mathbf{H}} \right. \\ & + \sum_{r=2}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1} (\mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1})_{[d]} \alpha E[\bar{\mathbf{A}}_{[d]} \mathbf{A}_{[d]}] \bar{\mathbf{H}}^{-1} e^{\beta_n \lambda_r} \\ & \quad \times \text{diag}\left(\beta_n, \dots, \frac{1 - e^{\beta_n \lambda_k}}{\lambda_1 - \lambda_k}\right) \bar{\mathbf{H}} \\ & + \sum_{r=1}^k \sum_{s=2, s \neq r}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(s) \mathbf{H}^{-1} (\mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1})_{[d]} \alpha E[\bar{\mathbf{A}}_{[d]} \mathbf{A}_{[d]}] \bar{\mathbf{H}}^{-1} \\ & \quad \left. \times \text{diag}\left(\frac{e^{\beta_n \lambda_r} - e^{\beta_n (\lambda_s + \lambda_1)}}{\lambda_r - \lambda_s - \lambda_1}, \dots, \frac{e^{\beta_n \lambda_r} - e^{\beta_n (\lambda_s + \lambda_k)}}{\lambda_r - \lambda_s - \lambda_k}\right) \bar{\mathbf{H}} \right) \\ & + o(n). \end{aligned}$$

(c) For the superlinear phase, when  $n = o(j_n)$ ,

$$\text{cov}[X_{j_n}^{(n)}] = -n \sum_{s=2}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(s) \mathbf{H}^{-1} v_d \mathbb{E}[\bar{\mathbf{A}}_{[d]} \mathbf{A}_{[d]}] \bar{\mathbf{H}}^{-1} \text{diag}\left(\frac{1}{\lambda_s + \lambda_1}, \dots, \frac{1}{\lambda_s + \lambda_k}\right) \bar{\mathbf{H}} + o(n).$$

*Proof.* The proof follows similarly to the arguments based on asymptotic expansions used in the proof of Corollary 6, and is thus omitted.

Note that the mean and variance in the three phases coincide at the phase boundary. The linear phase with  $\beta_n = 0$  gives the result in the growing sublinear phase, and the linear phase with  $\beta_n = \infty$  gives the result in the superlinear phase.

### 9. The underlying multivariate martingale

We use a multivariate martingale approach to derive the master central limit theorem. Other approaches may also work here, such as analytic techniques based on moment generating functions and the method of moments tailored to random combinatorial structures [6]. These methods may tend to be more computationally intensive than this presentation.

Conditioning recurrence (3) on the content of the sigma-field  $\mathcal{F}_{j-1}$ , we obtain

$$\mathbb{E}[X_j^{(n)} \mid \mathcal{F}_{j-1}] = \mathbb{E}[(X_{j-1}^{(n)} + \bar{\mathbf{A}}_j \mathbf{J}_j^{(n)}) \mid \mathcal{F}_{j-1}] = X_{j-1}^{(n)} + \frac{1}{n} \mathbb{E}[\bar{\mathbf{A}}_j] X_{j-1}^{(n)} = \mathbf{M} X_{j-1}^{(n)}.$$

So, it is evident that

$$\tilde{\mathbf{Y}}_j^{(n)} = \mathbf{M}^{j_n} \mathbf{Y}_j^{(n)} - \mathbf{M}^{j_n} \mathbf{Y}_0^{(n)}$$

is a centered martingale (with mean  $\mathbf{0}$ ). For any nonzero scalar sequence  $\xi_n$ ,  $\xi_n^{-1} \tilde{\mathbf{Y}}_j^{(n)}$  is a martingale difference sequence. It is sufficient for our purpose to check the conditional Lindeberg condition and the conditional variance condition (see [16, p. 58]). The former condition requires that, for all  $\varepsilon > 0$ ,

$$U_n := \sum_{i=1}^{j_n} \mathbb{E} \left[ \left\| \frac{1}{\xi_n} \nabla \tilde{\mathbf{Y}}_i^{(n)} \right\|^2 \mathbf{1}_{\{\|\nabla \tilde{\mathbf{Y}}_i^{(n)} / \xi_n\| > \varepsilon\}} \mid \mathcal{F}_{i-1} \right] \xrightarrow{p} 0,$$

where the indicator  $\mathbf{1}_{\mathcal{E}}$  is the function (of a sample space) that assumes the value 1 if  $\mathcal{E}$  occurs, and the value 0 otherwise, ' $\xrightarrow{p}$ ' denotes convergence in probability, and  $\|\cdot\|$  is the usual norm of a matrix. The matrix norm is defined as the root of the sum of the squares of its matrix components, or, equivalently, the root of the sum of the squares of its eigenvalues.

A  $\Sigma$ -conditional variance condition requires that

$$V_n := \sum_{i=1}^{j_n} \text{cov} \left[ \frac{1}{\xi_n} \nabla \tilde{\mathbf{Y}}_i^{(n)} \mid \mathcal{F}_{i-1} \right] \xrightarrow{p} \Sigma$$

for a covariance matrix  $\Sigma$ , where the convergence takes place component by component.

When both conditions hold, the sum  $\sum_{i=1}^{j_n} \xi_n^{-1} \nabla \tilde{\mathbf{Y}}_i^{(n)} = \xi_n^{-1} (\mathbf{M}^{j_n} \mathbf{Y}_{j_n}^{(n)} - \mathbf{M}^{j_n} \mathbf{Y}_0^{(n)}) = \xi_n^{-1} (X_{j_n}^{(n)} - \mathbb{E}[X_{j_n}^{(n)}])$  converges to the multinormally distributed random vector  $\mathcal{N}_k(0, \Sigma_{\text{phase}})$ .

In the sublinear phase we take  $\xi_n = \sqrt{j_n}$ , and in the linear and superlinear phases we take  $\xi_n = \sqrt{n}$ . To check the conditional Lindeberg condition, the following uniform bound is helpful in all the phases.

**Lemma 9.** *It holds that*

$$\|\nabla \tilde{Y}_j^{(n)}\| \leq 2ck^2.$$

*Proof.* In view of (3) we have

$$\begin{aligned} \|\nabla \tilde{Y}_j^{(n)}\| &= \|\mathbf{M}^{j_n} \mathbf{Y}_j^{(n)} - \mathbf{M}^{j_n} \mathbf{Y}_{j-1}^{(n)}\| \\ &\leq \|\mathbf{M}^{j_n-j}\| \|\mathbf{X}_j^{(n)} - \mathbf{M} \mathbf{X}_{j-1}^{(n)}\| \\ &= \|\mathbf{M}^{j_n-j}\| \|\mathbf{X}_{j-1}^{(n)} + \bar{\mathbf{A}} \mathbf{J}_j^{(n)} - \mathbf{M} \mathbf{X}_{j-1}^{(n)}\| \\ &\leq \|\mathbf{M}^{j_n-j}\| \left( \|\bar{\mathbf{A}} \mathbf{J}_j^{(n)}\| + \left\| \frac{1}{n} \mathbb{E}[\bar{\mathbf{A}}] \mathbf{X}_{j-1}^{(n)} \right\| \right) \\ &\leq \|\mathbf{M}^{j_n-j}\| \|\bar{\mathbf{A}}\| \|\mathbf{J}_j^{(n)}\| + \frac{1}{n} \|\mathbf{M}^{j_n-j}\| \|\mathbb{E}[\bar{\mathbf{A}}]\| \|\mathbf{X}_{j-1}^{(n)}\|. \end{aligned} \tag{12}$$

By the definition of the matrix norm we go further with

$$\|\mathbf{M}^{j_n-j}\| \leq \sqrt{k}, \quad \|\bar{\mathbf{A}}\| \leq ck, \quad \|\mathbb{E}[\bar{\mathbf{A}}]\| \leq ck, \quad \|\mathbf{J}_j^{(n)}\| = 1, \quad \|\mathbf{X}_{j-1}^{(n)}\| \leq n\sqrt{k}.$$

The lemma follows from these inequalities and (12).

**Lemma 10.** *It holds that*

$$U_n := \sum_{i=1}^{j_n} \mathbb{E} \left[ \left\| \frac{1}{\xi_n} \nabla \tilde{Y}_i^{(n)} \right\|^2 \mathbf{1}_{\{\|\nabla \tilde{Y}_i^{(n)}\|/\xi_n > \varepsilon\}} \middle| \mathcal{F}_{i-1} \right] \xrightarrow{P} 0.$$

*Proof.* Recall that in the sublinear phase  $\xi_n = \sqrt{j_n}$ , and in the linear and superlinear phases  $\xi_n = \sqrt{n}$ . Therefore, in all the growing phases  $\xi_n$  grows with  $n$ . For any given  $\varepsilon > 0$ , the uniform bound in Lemma 9 asserts that the sets  $\{\|\xi_n^{-1} \nabla \tilde{Y}_i^{(n)}\| > \varepsilon\}$  are empty for all  $j_n$  greater than some positive integer  $n_0(\varepsilon)$ . For  $j_n \geq n_0(\varepsilon)$ , we have

$$U_n = \sum_{i=1}^{n_0(\varepsilon)} \mathbb{E} \left[ \left\| \frac{1}{\xi_n} \nabla \tilde{Y}_i^{(n)} \right\|^2 \mathbf{1}_{\{\|\nabla \tilde{Y}_i^{(n)}\|/\xi_n > \varepsilon\}} \middle| \mathcal{F}_{i-1} \right].$$

According to Lemma 9, in all the growing phases we obtain

$$U_n \leq \sum_{i=1}^{n_0(\varepsilon)} \mathbb{E} \left[ \frac{4c^2k^4}{\xi_n^2} \middle| \mathcal{F}_{i-1} \right] \leq \frac{4c^2k^4 n_0(\varepsilon)}{\xi_n^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This completes the proof.

To check the conditional variance condition, we need to determine  $\text{cov}[\nabla \tilde{Y}_i^{(n)} \mid \mathcal{F}_{i-1}]$ , see the definition of  $\mathbf{V}_n$  in (9), which is given by

$$\begin{aligned} \text{cov}[\nabla \tilde{Y}_i^{(n)} \mid \mathcal{F}_{i-1}] &= \text{cov}[\mathbf{M}^{j_n} \mathbf{Y}_i^{(n)} - \mathbf{M}^{j_n} \mathbf{Y}_{i-1}^{(n)} \mid \mathcal{F}_{i-1}] \\ &= \text{cov}[\mathbf{M}^{j_n} \mathbf{Y}_i^{(n)} \mid \mathcal{F}_{i-1}] \\ &= \text{cov}[\mathbf{M}^{j_n-i} \mathbf{X}_i^{(n)} \mid \mathcal{F}_{i-1}] \\ &= \mathbf{M}^{j_n-i} \text{cov}[\mathbf{X}_i^{(n)} \mid \mathcal{F}_{i-1}] \bar{\mathbf{M}}^{j_n-i} \\ &= \mathbf{M}^{j_n-i} \text{cov}[\bar{\mathbf{A}} \mathbf{J}_i^{(n)} \mid \mathcal{F}_{i-1}] \bar{\mathbf{M}}^{j_n-i}. \end{aligned} \tag{13}$$

We have

$$\begin{aligned} \text{cov}[\bar{A}J_i^{(n)} \mid \mathcal{F}_{i-1}] &= E[\bar{A}J_i^{(n)}\bar{J}_i^{(n)}\mathbf{A} \mid \mathcal{F}_{i-1}] - E[\bar{A}J_i^{(n)} \mid \mathcal{F}_{i-1}]E[\bar{J}_i^{(n)}\mathbf{A} \mid \mathcal{F}_{i-1}] \\ &= \sum_{d=1}^k E[\mathbf{1}_{i,d}^{(n)} \mid \mathcal{F}_{i-1}]E[\bar{A}\mathbf{G}(d)\mathbf{A}] - \frac{1}{n^2}E[\bar{A}]X_{i-1}^{(n)}\bar{X}_{i-1}^{(n)}E[\mathbf{A}] \\ &= \frac{1}{n}\sum_{d=1}^k X_{i-1,d}^{(n)}E[\bar{A}_{[d]}\mathbf{A}_{[d]}] - \frac{1}{n^2}E[\bar{A}]X_{i-1}^{(n)}\bar{X}_{i-1}^{(n)}E[\mathbf{A}] \\ &= \frac{1}{n}\sum_{d=1}^k E[\bar{A}_{[d]}\mathbf{A}_{[d]}]I_{[d]}X_{i-1}^{(n)} - \frac{1}{n^2}E[\bar{A}]X_{i-1}^{(n)}\bar{X}_{i-1}^{(n)}E[\mathbf{A}]. \end{aligned}$$

From (13) and the above equation, we obtain

$$\begin{aligned} \text{cov}[\nabla\tilde{Y}_i^{(n)} \mid \mathcal{F}_{i-1}] &= \frac{1}{n}M^{j_n-i}\sum_{d=1}^k E[\bar{A}_{[d]}\mathbf{A}_{[d]}]I_{[d]}X_{i-1}^{(n)}\bar{M}^{j_n-i} \\ &\quad - \frac{1}{n^2}M^{j_n-i}E[\bar{A}]X_{i-1}^{(n)}\bar{X}_{i-1}^{(n)}E[\mathbf{A}]\bar{M}^{j_n-i}. \end{aligned}$$

Substituting this into (9) we obtain the explicit formula

$$\begin{aligned} V_n &= \frac{1}{\xi_n^2}\sum_{i=1}^{j_n}\left(\frac{1}{n}M^{j_n-i}\sum_{d=1}^k E[\bar{A}_{[d]}\mathbf{A}_{[d]}]I_{[d]}X_{i-1}^{(n)}\bar{M}^{j_n-i} \right. \\ &\quad \left. - \frac{1}{n^2}M^{j_n-i}E[\bar{A}]X_{i-1}^{(n)}\bar{X}_{i-1}^{(n)}E[\mathbf{A}]\bar{M}^{j_n-i}\right). \end{aligned} \tag{14}$$

### 10. The inherent Gaussian phases

Suppose that the drawing process continues indefinitely. We will see that, as the balls continue to be drawn from a Pólya scheme in the class  $\mathcal{C}_k$ , the urn passes through different phases.

#### 10.1. The growing sublinear phase

Let  $j_n$  be in the growing sublinear phase ( $j_n$  grows to  $\infty$  and  $j_n = o(n)$ ). In this phase we have, for  $r = 1, \dots, k$  (as  $n \rightarrow \infty$ ), the bounds

$$\alpha_r n + o(n) - cj_n \leq X_{j_n,r}^{(n)} \leq \alpha_r n + o(n) + cj_n.$$

In other words, we have

$$X_{j_n}^{(n)} = \alpha n + o(n). \tag{15}$$

*Proof of Theorem 3 in the sublinear phase.* In Lemma 10 the conditional Lindeberg condition was verified throughout the growing sublinear phase. We completed the proof by showing that  $V_n$  converges to a constant covariance matrix in probability.

In (14) replace  $X_{i-1}^{(n)}$  by the asymptotic equivalent in (15) to obtain

$$V_n = \frac{1}{j_n}\sum_{i=1}^{j_n}\left(M^{j_n-i}\sum_{d=1}^k E[\bar{A}_{[d]}\mathbf{A}_{[d]}]\alpha_d\bar{M}^{j_n-i} - M^{j_n-i}E[\bar{A}]\alpha\bar{\alpha}E[\mathbf{A}]\bar{M}^{j_n-i}\right) + o(1).$$

In this phase, for all  $0 \leq i \leq j_n$ , we have  $M^{j_n-i} = I + O(j_n/n)$ . It follows that

$$V_n \rightarrow \Sigma_{\text{sub}} := \sum_{d=1}^k \alpha_d E[\bar{A}_{[d]} A_{[d]}] - E[\bar{A}] \alpha \bar{\alpha} E[A].$$

We can show that in the sublinear phase the coefficient of  $j_n$  in the covariance is not  $\mathbf{0}$  if there is no initially dominant proper subset of colors. In fact, in this case  $\Sigma_{\text{sub}}$  is a covariance matrix of rank  $k - 1$ . The extremal nuance with a proper subset of dominant color requires a much more detailed asymptotic analysis [22].

The  $\Sigma_{\text{sub}}$ -conditional variance condition has also been verified in the growing sublinear phase. With both conditions checked, the martingale central limit theorem gives

$$\sum_{i=1}^{j_n} \frac{1}{\sqrt{j_n}} \nabla \tilde{Y}_i^{(n)} = \frac{X_{j_n}^{(n)} - E[X_{j_n}^{(n)}]}{\sqrt{j_n}} \xrightarrow{D} \mathcal{N}(0, \Sigma_{\text{sub}}).$$

After reorganization, this becomes the statement of the theorem.

**10.2. The linear phase**

In the linear phase  $j_n \sim \beta_n n$  for some  $\beta_n > 0$  of a magnitude uniformly bounded from above and below, that is, for two positive constants,  $S_1$  and  $S_2$ , and all  $n$ ,

$$S_1 \leq \beta_n \leq S_2.$$

In Lemma 10 the conditional Lindeberg condition was verified throughout the linear phase. The proof will be complete if we show that  $V_n$  converges to a constant covariance matrix in probability. The calculation involves summing random variables. To carry out these calculations, we approximate  $X_j^{(n)}$ . The simple bound  $X_{j_n,r}^{(n)} \leq n$  always holds. As we sum from 1 to  $j_n \sim \beta_n n$ , we can use this bound throughout the sublinear phase and for a small portion of the linear phase. As we advance in the linear phase, we need a sharper bound for  $V_n$  to converge. More precisely, for some small  $\varepsilon$ ,  $0 < \varepsilon < S_1 \leq \beta_n$ , we break up the sum from 1 to  $j_n$  into a sum from 1 to  $j'_n = \lfloor \varepsilon n \rfloor$  (in which we apply the bound  $X_{j_n,r}^{(n)} \leq n$ ), and a sum from  $j'_n + 1$  to  $n$  (in which we apply a sharper bound, valid for the linear phase).

**Lemma 11.** *In the linear phase*

$$X_{j_n}^{(n)} = \mu_n n + o_{\mathcal{L}_1}(n),$$

where

$$\mu_n = H \text{diag}(e^{\beta_n \lambda_1}, \dots, e^{\beta_n \lambda_k}) H^{-1} \alpha.$$

*Proof.* Let  $\mu_n(r)$  be the  $r$ th component of  $\mu_n$ . From the asymptotics of the mean and covariance in the linear phase, as given in Corollaries 6 and 7, for large  $n$ , we have

$$E[(X_{j_n,r}^{(n)} - \mu_n(r)n)^2] = \text{var}[X_{j_n,r}^{(n)}] + (E[X_{j_n,r}^{(n)}] - \mu_n(r)n)^2 = O(n) + o(n^2) = o(n^2).$$

So, by Jensen’s inequality,

$$E[|X_{j_n,r}^{(n)} - \mu_n(r)n|] \leq \sqrt{E[(X_{j_n,r}^{(n)} - \mu_n(r)n)^2]} = o(n),$$

implying that

$$X_{j_n}^{(n)} = \mu_n n + o_{\mathcal{L}_1}(n).$$

The  $\mathbf{o}_{\mathcal{L}_1}(g(n))$  term represents a vector or a matrix of random variables with each component being  $o(g(n))$  in the usual  $\mathcal{L}_1$  sense.

*Proof of Theorem 3 in the linear phase.* We refer to (14). As discussed in the introductory paragraph of this subsection, for some small  $\varepsilon$ ,  $0 < \varepsilon < S_1 \leq \beta_n$ , we break up  $\mathbf{V}_n$  at  $j'_n = \lfloor \varepsilon n \rfloor$ , to obtain

$$\mathbf{V}_n = \frac{1}{\xi_n^2} \left( \sum_{i=1}^{j'_n} + \sum_{i=j'_n+1}^{j_n} \right).$$

Recall that in this phase  $\xi_n = \sqrt{n}$ . In the first sum, replace  $X_{i-1}^{(n)}$  by the obvious bound  $\mathbf{O}(n)$ , and in the second sum replace  $X_{i-1}^{(n)}$  by the asymptotic equivalent in (11) to obtain

$$\begin{aligned} \mathbf{V}_n = \mathbf{O}(\varepsilon) + \frac{1}{n} \sum_{i=1}^{j_n} & \left( M^{j_n-i} \sum_{d=1}^k \mathbb{E}[\bar{A}_{[d]} A_{[d]}] I_{[d]} \boldsymbol{\mu}_n \bar{M}^{j_n-i} \right. \\ & \left. - M^{j_n-i} \mathbb{E}[\bar{A}] \boldsymbol{\mu}_n \bar{\boldsymbol{\mu}}_n \mathbb{E}[A] \bar{M}^{j_n-i} \right) + \mathbf{o}_{\mathcal{L}_1}(1). \end{aligned}$$

The  $\mathbf{O}(\varepsilon)$  term vanishes when we take the limit as  $\varepsilon \rightarrow 0$ . With the help of Corollary 7, we recognize what is left to be

$$\begin{aligned} \mathbf{V}_n = & - \sum_{d=2}^k \lambda_d e^{\beta_n \lambda_d} \mathbf{H} \mathbf{G}(d) \mathbf{H}^{-1} \boldsymbol{\alpha} \bar{\boldsymbol{\alpha}} \bar{\mathbf{H}}^{-1} \text{diag}(0, \dots, \beta_n \lambda_k e^{\beta_n \lambda_k}) \bar{\mathbf{H}} \\ & + \sum_{r=2}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1} (\mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1})_{[d]} \boldsymbol{\alpha} \mathbb{E}[\bar{A}_{[d]} A_{[d]}] \bar{\mathbf{H}}^{-1} e^{\beta_n \lambda_r} \\ & \quad \times \text{diag} \left( \beta_n, \dots, \frac{1 - e^{\beta_n \lambda_k}}{-\lambda_k} \right) \bar{\mathbf{H}} \\ & + \sum_{r=1}^k \sum_{s=2, s \neq r}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(s) \mathbf{H}^{-1} (\mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1})_{[d]} \boldsymbol{\alpha} \mathbb{E}[\bar{A}_{[d]} A_{[d]}] \\ & \quad \times \bar{\mathbf{H}}^{-1} \text{diag} \left( \frac{e^{\beta_n \lambda_r} - e^{\beta_n \lambda_s}}{\lambda_r - \lambda_s}, \dots, \frac{e^{\beta_n \lambda_r} - e^{\beta_n (\lambda_s + \lambda_k)}}{\lambda_r - \lambda_s - \lambda_k} \right) \bar{\mathbf{H}} + \mathbf{o}_{\mathcal{L}_1}(1). \end{aligned}$$

We have

$$I_{[d]} \boldsymbol{\mu}_n = I_{[d]} \mathbf{H} \text{diag}(e^{\beta_n \lambda_1}, \dots, e^{\beta_n \lambda_k}) \mathbf{H}^{-1} \boldsymbol{\alpha} = \sum_{r=1}^k e^{\beta_n \lambda_r} (\mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1})_{[d]} \boldsymbol{\alpha}.$$

Hence,

$$\begin{aligned} \mathbf{V}'_n := & \sum_{i=1}^{j_n} M^{j_n-i} \sum_{d=1}^k \mathbb{E}[\bar{A}_{[d]} A_{[d]}] I_{[d]} \boldsymbol{\mu}_n \bar{M}^{j_n-i} \\ = & \sum_{r=1}^k \sum_{s=1}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(s) \mathbf{H}^{-1} e^{\beta_n \lambda_r} (\mathbf{H} \mathbf{G}(r) \mathbf{H}^{-1})_{[d]} \boldsymbol{\alpha} \mathbb{E}[\bar{A}_{[d]} A_{[d]}] \bar{\mathbf{H}}^{-1} \\ & \times \sum_{i=0}^{j_n-1} \text{diag} \left( \left( 1 + \frac{\lambda_s}{n} \right)^i, \dots, \left( 1 + \frac{\lambda_s}{n} \right)^i \left( 1 + \frac{\lambda_k}{n} \right)^i \right) \bar{\mathbf{H}} \end{aligned}$$

$$= \sum_{h=2}^k \sum_{r=1}^k \sum_{s=1}^k \sum_{d=1}^k \mathbf{H}\mathbf{G}(s)\mathbf{H}^{-1}e^{\beta_n\lambda_r}(\mathbf{H}\mathbf{G}(r)\mathbf{H}^{-1})_{[d]}\boldsymbol{\alpha} \mathbb{E}[\bar{\mathbf{A}}_{[d]}\mathbf{A}_{[d]}] \times \bar{\mathbf{H}}^{-1}\mathbf{G}(h)\bar{\mathbf{H}} \frac{1 - (1 + \lambda_s/n)^{j_n}(1 + \lambda_h/n)^{j_n}}{-\lambda_s - \lambda_h - \lambda_s\lambda_h/n}.$$

Similarly, we have

$$\begin{aligned} \mathbf{V}_n'' &:= \sum_{i=1}^{j_n} \mathbf{M}^{j_n-i} \mathbb{E}[\bar{\mathbf{A}}]\boldsymbol{\mu}_n\bar{\boldsymbol{\mu}}_n \mathbb{E}[\mathbf{A}]\bar{\mathbf{M}}^{j_n-i} \\ &= \sum_{h=2}^k \sum_{s=1}^k \mathbf{H}\mathbf{G}(s)\mathbf{H}^{-1} \mathbb{E}[\bar{\mathbf{A}}]\boldsymbol{\mu}_n\bar{\boldsymbol{\mu}}_n \mathbb{E}[\mathbf{A}]\bar{\mathbf{H}}^{-1}\mathbf{G}(h)\bar{\mathbf{H}} \\ &\quad \times \frac{1 - (1 + \lambda_s/n)^{j_n}(1 + \lambda_h/n)^{j_n}}{-\lambda_s - \lambda_h - \lambda_s\lambda_h/n}. \end{aligned}$$

Putting the two terms  $\mathbf{V}_n'$  and  $\mathbf{V}_n''$  together, we obtain

$$\begin{aligned} \mathbf{V}_n &= \sum_{h=1}^k \sum_{r=1}^k \sum_{s=1}^k \sum_{d=1}^k \mathbf{H}\mathbf{G}(s)\mathbf{H}^{-1}e^{\beta_n\lambda_r}(\mathbf{H}\mathbf{G}(r)\mathbf{H}^{-1})_{[d]}\boldsymbol{\alpha} \mathbb{E}[\bar{\mathbf{A}}_{[d]}\mathbf{A}_{[d]}] \\ &\quad \times \bar{\mathbf{H}}^{-1}\mathbf{G}(h)\bar{\mathbf{H}} \frac{1 - e^{\beta_n(\lambda_s+\lambda_h)}}{-\lambda_s - \lambda_h} \\ &\quad - \sum_{h=1}^k \sum_{s=1}^k \mathbf{H}\mathbf{G}(s)\mathbf{H}^{-1} \mathbb{E}[\bar{\mathbf{A}}]\boldsymbol{\mu}_n\bar{\boldsymbol{\mu}}_n \mathbb{E}[\mathbf{A}]\bar{\mathbf{H}}^{-1}\mathbf{G}(h)\bar{\mathbf{H}} \frac{1 - e^{\beta_n(\lambda_s+\lambda_h)}}{-\lambda_s - \lambda_h} + o_{\mathcal{L}_1}(1). \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathbb{E}[\bar{\mathbf{A}}]\boldsymbol{\mu}_n\bar{\boldsymbol{\mu}}_n \mathbb{E}[\mathbf{A}] &= \mathbf{H}\text{diag}(\lambda_1e^{\beta_n\lambda_1}, \dots, \lambda_ke^{\beta_n\lambda_k})\mathbf{H}^{-1}\boldsymbol{\alpha}\bar{\boldsymbol{\alpha}}\bar{\mathbf{H}}^{-1}\text{diag}(\lambda_1e^{\beta_n\lambda_1}, \dots, \lambda_ke^{\beta_n\lambda_k})\bar{\mathbf{H}} \\ &= \sum_{d=1}^k \mathbf{H}\mathbf{G}(d)\mathbf{H}^{-1}\boldsymbol{\alpha}\bar{\boldsymbol{\alpha}}\bar{\mathbf{H}}^{-1}\lambda_d e^{\beta_n\lambda_d}\text{diag}(\lambda_1e^{\beta_n\lambda_1}, \dots, \lambda_ke^{\beta_n\lambda_k})\bar{\mathbf{H}}. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbf{V}_n &\rightarrow \boldsymbol{\Sigma}_{\text{lin}} \\ &:= \sum_{h=2}^k \sum_{s=2}^k \sum_{r=1}^k \sum_{d=1}^k \mathbf{H}\mathbf{G}(s)\mathbf{H}^{-1}e^{\beta_n\lambda_r}(\mathbf{H}\mathbf{G}(r)\mathbf{H}^{-1})_{[d]}\boldsymbol{\alpha} \mathbb{E}[\bar{\mathbf{A}}_{[d]}\mathbf{A}_{[d]}] \bar{\mathbf{H}}^{-1}\mathbf{G}(h)\bar{\mathbf{H}} \\ &\quad \times \frac{1 - e^{\beta_n(\lambda_s+\lambda_h)}}{-\lambda_s - \lambda_h} \\ &\quad - \sum_{h=2}^k \sum_{s=2}^k \mathbf{H}\mathbf{G}(s)\mathbf{H}^{-1}\boldsymbol{\alpha}\bar{\boldsymbol{\alpha}}\bar{\mathbf{H}}^{-1}\lambda_s\lambda_h e^{\beta_n(\lambda_s+\lambda_h)}\mathbf{G}(h). \end{aligned}$$

The  $\boldsymbol{\Sigma}_{\text{lin}}$ -conditional variance condition has been verified in the growing linear phase. With both conditions checked, the martingale central limit theorem gives

$$\sum_{i=1}^{j_n} \frac{1}{\sqrt{n}} \nabla \tilde{\mathbf{Y}}_i^{(n)} = \frac{\mathbf{X}_{j_n}^{(n)} - \mathbb{E}[\mathbf{X}_{j_n}^{(n)}]}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, \boldsymbol{\Sigma}_{\text{lin}}),$$

which is the statement of the theorem.

### 10.3. The superlinear phase

Suppose that the drawing continues for a superlinearly long period of time, namely,  $n = o(j_n)$ . As in the linear phase, for the sum in the conditional variance condition, we apply the bound  $X_{j,r}^{(n)} \leq n$  until the superlinear phase. After that we apply sharper estimates. More precisely, to asymptotically handle the sums in the conditional Lindeberg condition (going from 1 to  $j_n$ ), we break up the sums in  $V_n$  into sums going from 1 to  $j'_n$ , which is any superlinear function of order less than  $j_n$  (giving ignorable contribution), and sums starting at  $j'_n + 1$  and ending at  $j_n$  (most of the contribution occurs near  $j_n$ ). We can take  $j'_n = \lfloor j_n / \ln(j_n/n) \rfloor$ . Since the proof structure is very similar to that in the linear phase, the argument is only sketched. We approximate  $X_{j_n}^{(n)}$  by its average:

$$X_{j_n}^{(n)} = \bar{v}n + o_{\mathcal{L}_1}(n).$$

This statement is equivalent to Lemma 11, and the proof is essentially the same.

*Proof of Theorem 3 in the superlinear phase.* In Lemma 10 the conditional Lindeberg condition was verified throughout the superlinear phase. The proof will be complete if we show that  $V_n$  converges to a constant covariance matrix in probability.

Breaking up the sum as mentioned above, and keeping the essential contribution (as  $\varepsilon \rightarrow 0$ ), we write

$$V_n = \frac{1}{n} \sum_{i=1}^{j_n} M^{j_n-i} \sum_{d=1}^k E[\bar{A}_{[d]} A_{[d]}] v_d \bar{M}^{j_n-i} + o_{\mathcal{L}_1}(1).$$

That is,

$$\begin{aligned} V_n &= \frac{1}{n} \sum_{s=2}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(s) \mathbf{H}^{-1} v_d E[\bar{A}_{[d]} A_{[d]}] \bar{\mathbf{H}}^{-1} \\ &\quad \times \sum_{i=0}^{j_n-1} \text{diag} \left( \left( 1 + \frac{\lambda_s}{n} \right)^i, \dots, \left( 1 + \frac{\lambda_s}{n} \right)^i \left( 1 + \frac{\lambda_k}{n} \right)^i \right) \bar{\mathbf{H}} + o_P(1) \\ &= \sum_{s=2}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(s) \mathbf{H}^{-1} v_d E[\bar{A}_{[d]} A_{[d]}] \bar{\mathbf{H}}^{-1} \\ &\quad \times \text{diag} \left( \frac{1 - (1 + \lambda_s/n)^{j_n}}{-\lambda_s - \lambda_1 - \lambda_s \lambda_1/n}, \dots, \frac{1 - (1 + \lambda_s/n)^{j_n} (1 + \lambda_k/n)^{j_n}}{-\lambda_s - \lambda_k - \lambda_s \lambda_k/n} \right) \bar{\mathbf{H}} \\ &\quad + o_{\mathcal{L}_1}(1). \end{aligned}$$

Recalling that the nonprincipal eigenvalues have negative real parts for all  $r = 2, \dots, k$ , we have  $(1 + \lambda_r/n)^{j_n} \rightarrow 0$ . It follows that

$$\begin{aligned} V_n &\rightarrow \Sigma_{\text{sup}} \\ &:= - \sum_{s=2}^k \sum_{d=1}^k \mathbf{H} \mathbf{G}(s) \mathbf{H}^{-1} v_d E[\bar{A}_{[d]} A_{[d]}] \bar{\mathbf{H}}^{-1} \text{diag} \left( \frac{1}{\lambda_1 + \lambda_s}, \dots, \frac{1}{\lambda_s + \lambda_k} \right) \bar{\mathbf{H}}. \end{aligned}$$

The  $\Sigma_{\text{sup}}$ -conditional variance condition has been verified in the superlinear phase. With both conditions checked, the martingale central limit theorem gives

$$\sum_{i=1}^{j_n} \frac{1}{\sqrt{n}} \nabla \tilde{Y}_i^{(n)} = \frac{X_{j_n}^{(n)} - E[X_{j_n}^{(n)}]}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, \Sigma_{\text{sup}}).$$

### 11. Examples

In this section we give a few examples, including the classical Ehrenfest urn, schemes with both deterministic and random replacement matrices, and degenerate schemes. The examples elicit the asymptotic Gaussian distributions.

#### 11.1. The Ehrenfest scheme

The Ehrenfest urn was first proposed as a model for the diffusion of nonreacting gases [8]. It is governed by the replacement matrix

$$A = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Avogadro’s number, asserting that the number of particles of gas in a container of a common size is very large, gives a real-world justification for an approximation with  $n \rightarrow \infty$ .

Assume that the two colors are white and blue, and that their numbers after  $j$  draws are respectively  $W_j$  and  $B_j$ . The case is dichromatic, and it is adequate to study only the number of balls of one color. Starting out with  $X_0^{(n)} = (\lfloor \frac{1}{3}n \rfloor, \lceil \frac{2}{3}n \rceil)$ , in the sublinear phase the Gaussian law in Theorem 3 takes the form

$$\frac{1}{\sqrt{j_n}} \left( W_{j_n} - n \left( \frac{1}{2} - \frac{1}{6} \left( \frac{n-2}{n} \right)^{j_n} \right) \right) \xrightarrow{D} \mathcal{N} \left( 0, \frac{8}{9} \right).$$

So, after  $j_n = \lfloor n^{3/4} \rfloor$  draws, for instance, we have

$$\frac{1}{n^{3/8}} \left( W_{\lfloor n^{3/4} \rfloor} - \frac{1}{3}n + \frac{1}{3}\sqrt{n} - \frac{1}{3}n^{3/4} \right) \xrightarrow{D} \mathcal{N} \left( 0, \frac{8}{9} \right).$$

A typical instance of the linear phase is when  $j_n = 2n$ , in which case the Gaussian law in Theorem 3 takes the form

$$\frac{1}{\sqrt{n}} \left( W_{2n} - \left( \frac{1}{2} - \frac{1}{6}e^{-4} \right) n \right) \xrightarrow{D} \mathcal{N} \left( 0, \frac{1}{4} - \frac{17}{36}e^{-8} \right).$$

According to Theorem 2, the stationary distribution of this diffusion process is Multinomial( $n, \frac{1}{2}, \frac{1}{2}$ ). This is a classic result [21]. Deep into the superlinear phase, the diffusion is nearly complete. For instance, after  $j_n = n^2$  draws, Theorem 3 takes the symmetric form

$$\frac{1}{\sqrt{n}} \left( W_{n^2} - \frac{1}{2}n \right) \xrightarrow{D} \mathcal{N} \left( 0, \frac{1}{4} \right),$$

which is the usual approximation of the stationary Multinomial( $n, \frac{1}{2}, \frac{1}{2}$ ) distribution. For a detailed discussion of the phases of the Ehrenfest urn, see [3].

#### 11.2. A three-color scheme with deterministic replacements

Consider the scheme

$$A = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix}, \quad X_0^{(n)} = \begin{pmatrix} n - 2\lfloor \frac{2}{5}n \rfloor \\ \lfloor \frac{2}{5}n \rfloor \\ \lfloor \frac{2}{5}n \rfloor \end{pmatrix}.$$

The principal row eigenvector is  $\frac{1}{3}(1, 1, 1)$ . Therefore, the stationary distribution is Multinomial( $n; \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ ); cf. Theorem 2. For very large  $n$  and an excessive number of draws, we have a multivariate normal distribution that approximates this multinomial. Specifically, according to Theorem 3, in the sublinear phase

$$\frac{1}{\sqrt{j_n}} \left( \mathbf{X}_{j_n}^{(n)} - \frac{1}{n j_n} \begin{pmatrix} n-1 & 1 & 0 \\ 0 & n-1 & 1 \\ 1 & 0 & n-1 \end{pmatrix}^{j_n} \begin{pmatrix} n-2 \lfloor \frac{2}{5}n \rfloor \\ \lfloor \frac{2}{5}n \rfloor \\ \lfloor \frac{2}{5}n \rfloor \end{pmatrix} \right) \xrightarrow{D} \mathcal{N}_3 \left( \mathbf{0}, \frac{1}{25} \begin{pmatrix} 14 & -4 & -10 \\ -4 & 14 & -10 \\ -10 & -10 & 20 \end{pmatrix} \right),$$

whereas in the linear phase, where  $\beta_n = 2$ , we have the Gaussian law

$$\frac{\mathbf{X}_{2n}^{(n)} - \boldsymbol{\mu}_{\text{lin}} n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}_3(\mathbf{0}, \boldsymbol{\Sigma}_{\text{lin}}),$$

where

$$\boldsymbol{\mu}_{\text{lin}} \approx \begin{pmatrix} 0.32 \\ 0.33 \\ 0.34 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\text{lin}} \approx \begin{pmatrix} 0.223 & -0.110 & -0.112 \\ -0.110 & 0.219 & -0.109 \\ -0.112 & -0.109 & 0.222 \end{pmatrix}.$$

In the superlinear phase we have the Gaussian law

$$\frac{1}{\sqrt{n}} \left( \mathbf{X}_{j_n}^{(n)} - \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} n \right) \xrightarrow{D} \mathcal{N}_3 \left( \mathbf{0}, \frac{1}{9} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \right).$$

### 11.3. A three-color scheme with random replacements

Let  $B$  be  $\text{Ber}(\frac{2}{3})$ , and let  $B'$  be an independent  $\text{Ber}(\frac{1}{2})$ . Consider the scheme

$$\mathbf{A} = 2 \begin{pmatrix} -B & BB' & B(1-B') \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix}, \quad \mathbf{X}_0^{(n)} = \begin{pmatrix} 18n \\ 66n \\ 0 \end{pmatrix}.$$

This replacement matrix has average

$$\mathbb{E}[\mathbf{A}] = 2 \begin{pmatrix} -\frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix},$$

with principal left eigenvector  $(\frac{1}{2}, \frac{1}{6}, \frac{1}{3})$ . Therefore, the stationary distribution is Multinomial( $42n; \frac{1}{2}, \frac{1}{6}, \frac{1}{3}$ ); cf. Theorem 2. For very large  $n$  and an excessive number of draws, we have a multivariate normal distribution that approximates this multinomial. Specifically, according to Theorem 3, in the sublinear phase

$$\frac{1}{\sqrt{j_n}} \left( \mathbf{X}_{j_n}^{(84n)} - \frac{1}{(3n)j_n} \begin{pmatrix} 3n-4 & 2 & 2 \\ 0 & 3n-2 & 2 \\ 2 & 0 & 3n-2 \end{pmatrix}^{j_n} \begin{pmatrix} 18n \\ 66n \\ 0 \end{pmatrix} \right) \xrightarrow{D} \mathcal{N}_3 \left( \mathbf{0}, \frac{1}{49} \begin{pmatrix} 24 & -34 & -10 \\ -34 & 68 & -34 \\ -10 & -34 & 24 \end{pmatrix} \right),$$

whereas in the linear phase, where  $\beta_n = 2$ , we have the Gaussian law

$$\frac{X_{j_n}^{(42n)} - \mu_{\text{lin}} n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}_3(0, \Sigma_{\text{lin}}),$$

where

$$\mu_{\text{lin}} \approx \begin{pmatrix} 41.06 \\ 14.01 \\ 28.38 \end{pmatrix} \quad \text{and} \quad \Sigma_{\text{lin}} \approx \begin{pmatrix} 41.99 & -13.87 & -28.12 \\ -13.87 & 23.26 & -9.38 \\ -28.12 & -9.38 & 37.50 \end{pmatrix}.$$

In the superlinear phase we have the Gaussian law

$$\frac{1}{\sqrt{n}} \left( X_{j_n}^{(84n)} - \begin{pmatrix} 42 \\ 14 \\ 28 \end{pmatrix} n \right) \xrightarrow{D} \mathcal{N}_3 \left( \mathbf{0}, \frac{1}{3} \begin{pmatrix} 126 & -42 & -84 \\ -42 & 70 & -28 \\ -84 & -28 & 112 \end{pmatrix} \right).$$

**11.4. An example of a degenerate scheme**

Consider the following tenable, irreducible, zero-balanced degenerate urn scheme:

$$A = \begin{pmatrix} -6 & 2 & 4 \\ 6 & -2 & -4 \\ 6 & -2 & -4 \end{pmatrix}, \quad X_0^{(18n)} = \begin{pmatrix} 12n \\ 2n \\ 4n \end{pmatrix}.$$

Colors 2 and 3 are tied together—at any point in time the number of balls of color 3 is twice the number of balls of color 2.

To comprehend the behavior of this urn, we can simply study the following tenable, irreducible, nondegenerate zero-balanced urn scheme:

$$A = \begin{pmatrix} -6 & 6 \\ 6 & -6 \end{pmatrix}, \quad X_0^{(18n)} = \begin{pmatrix} 12n \\ 6n \end{pmatrix}.$$

Many of the results derived for the class  $\mathcal{C}_k$  apply in an almost direct way. For example, with  $X_{n^2}^{(n)}$  being the two-component partition vector associated with the dichromatic scheme, the reduced scheme has the central limit behavior

$$\frac{1}{\sqrt{n}} \left( X_{n^2}^{(n)} - \begin{pmatrix} 9 \\ 9 \end{pmatrix} n \right) \xrightarrow{D} \mathcal{N}_2 \left( 0, \begin{pmatrix} 9 & -9 \\ -9 & 9 \end{pmatrix} \right),$$

which is a classic result for Ehrenfest urn models. As balls of colors 2 and 3 in the trivariate scheme occur in proportion 1:2, we can then assert that

$$\frac{1}{\sqrt{n}} \left( X_{n^2}^{(n)} - \begin{pmatrix} 9 \\ 3 \\ 6 \end{pmatrix} n \right) \xrightarrow{D} \mathcal{N}_3 \left( 0, \begin{pmatrix} 9 & -3 & -6 \\ -3 & 1 & 2 \\ -6 & 2 & 4 \end{pmatrix} \right).$$

**Acknowledgements**

The first author thanks Columbian College of The George Washington University for travel money that enabled her to present a preliminary draft of this paper at the Eighth Workshop on Analytic Algorithmics and Combinatorics, San Francisco, USA, 2011. The second author was funded by the Institute for Integrating Statistics in Decision Sciences at The George

Washington University to attend the same workshop. He also wishes to acknowledge the support of Professors Rebecca Doerge and Mark Daniel Ward for their hospitality and support during his sabbatical visit to the Department of Statistics at Purdue University. The authors are grateful to an anonymous referee for numerous insightful comments that helped improve the presentation.

## References

- [1] ATHREYA, K. B. AND KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39**, 1801–1817.
- [2] BAGCHI, A. AND PAL, A. K. (1985). Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Meth.* **6**, 394–405.
- [3] BALAJI, S., MAHMOUD, H. AND ZHANG, T. (2010). Phases in the diffusion of gases via the Ehrenfest urn model. *J. Appl. Prob.* **47**, 841–855.
- [4] BERNSTEIN, S. (1940). Sur un problème du schéma des urnes à composition variable. *C. R. (Doklady) Acad. Sci. URSS* **28**, 5–7.
- [5] CHAUVIN, B., POYANNE, N. AND SAHNOUN, R. (2011). Limit distributions for large Pólya urns. *Ann. Appl. Prob.* **21**, 1–32.
- [6] CHERN, H.-H. AND HWANG, H.-K. (2001). Phase changes in random  $m$ -ary search trees and generalized quicksort. *Random Structures Algorithms* **19**, 316–358.
- [7] EGGENBERGER, F. AND PÓLYA, G. (1923). Über die statistik verketteter vorgänge. *Z. Angew. Math. Mech.* **3**, 279–289.
- [8] EHRENFEST, P. AND EHRENFEST, T. (1907). Über zwei bekannte einwände gegen das Boltzmannsche H-theorem. *Phys. Z.* **8**, 311–314.
- [9] FLAJOLET, P., DUMAS, P. AND PUYHAUBERT, V. (2006). Some exactly solvable models of urn process theory. In *4th Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities* (Discrete Math. Theoret. Comput. Sci. Proc. **AG**), Assoc. Discrete Math. Theoret. Comput. Sci. Nancy, pp. 59–118.
- [10] FLAJOLET, P., GABARRÓ, J. AND PEKARI, H. (2005). Analytic urns. *Ann. Prob.* **33**, 1200–1233.
- [11] FREEDMAN, D. A. (1965). Bernard Friedman’s urn. *Ann. Math. Statist.* **36**, 956–970.
- [12] FRIEDMAN, B. (1949). A simple urn model. *Commun. Pure Appl. Math.* **2**, 59–70.
- [13] GOUET, R. (1989). A martingale approach to strong convergence in a generalized Pólya-Eggenberger urn model. *Statist. Prob. Lett.* **8**, 225–228.
- [14] GOUET, R. (1993). Martingale functional central limit theorems for a generalized Pólya urn. *Ann. Prob.* **21**, 1624–1639.
- [15] GOUET, R. (1997). Strong convergence of proportions in a multicolor Pólya urn. *J. Appl. Prob.* **34**, 426–435.
- [16] HALL, P. AND HEYDE, C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.
- [17] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110**, 177–245.
- [18] JANSON, S. (2005). Asymptotic degree distribution in random recursive trees. *Random Structures Algorithms* **26**, 69–83.
- [19] JANSON, S. (2005). Limit theorems for triangular urn schemes. *Prob. Theory Relat. Fields* **134**, 417–452.
- [20] JOHNSON, N. AND KOTZ, S. (1977). *Urn models and Their Application*. John Wiley, New York.
- [21] KARLIN, S. AND MCGREGOR, J. (1965). Ehrenfest urn models. *J. Appl. Prob.* **2**, 352–376.
- [22] KHOLFI, S. AND MAHMOUD, H. M. (2012). The class of tenable zero-balanced Pólya urns with an initially dominant subset of colors. *Statist. Prob. Lett.* **82**, 49–57.
- [23] KOTZ, S. AND BALAKRISHNAN, N. (1997). Advances in urn models during the past two decades. In *Advances in Combinatorial Methods and Applications to Probability and Statistics*. Birkhäuser, Boston, MA, pp. 203–257.
- [24] MAHMOUD, H. M. (1998). On rotations in fringe-balanced binary trees. *Inform. Process. Lett.* **65**, 41–46.
- [25] MAHMOUD, H. M. AND SMYTHE, R. T. (1991). On the distribution of leaves in rooted subtrees of recursive trees. *Ann. Appl. Prob.* **1**, 406–418.
- [26] MAHMOUD, H. M. AND SMYTHE, R. T. (1992). Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms* **3**, 255–266.
- [27] MAHMOUD, H. M. AND SMYTHE, R. T. (1995). Probabilistic analysis of bucket recursive trees. *Theoret. Comput. Sci.* **144**, 221–249.
- [28] MAHMOUD, H. (2003). Pólya urn models and connections to random trees: a review. *J. Iranian Statist. Soc.* **2**, 53–114.
- [29] MAHMOUD, H. M. (2008). *Pólya Urn Models*. Chapman & Hall/CRC, Boca Raton, FL.

- [30] MAHMOUD, H. M., SMYTHE, R. T. AND SZYMAŃSKI, J. (1993). On the structure of plane-oriented recursive trees and their branches. *Random Structures Algorithms* **4**, 151–176.
- [31] MCKENZIE, A. AND STEELE, M. (2000). Distributions of cherries for two models of trees. *Math. Biosci.* **164**, 81–92.
- [32] PÓLYA, G. (1931). Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* **1**, 117–161.
- [33] SAVKEVICH, V. (1940). Sur le schéma des urnes à composition variable. *C. R. (Doklady) Acad. Sci. URSS* **28**, 8–12.
- [34] SMYTHE, R. T. (1996). Central limit theorems for urn models. *Stoch. Process. Appl.* **65**, 115–137.